The Nature and Significance of Differential Keratin Gene Expression

ELAINE FUCHS, ISRAEL HANUKOGLU, DOUGLAS MARCHUK, MICHAEL P. GRACE, AND KWAN HEE KIM

The Department of Molecular Genetics and Cell Biology
The University of Chicago
Chicago, Illinois 60637

Keratin filaments are 8 nm in diameter, and together with the 6 nm actin microfilaments and the 23 nm microtubules, they constitute the cytoskeletal architecture of most vertebrate epithelial cells. The filaments are assembled from two classes of subunits, type I and type II. The type I keratins are acidic and generally small in size (40–57 kD). The type II keratins are more basic and larger in size (53–67 kD). Each class is encoded by a small multigene family, the members of which are differentially expressed in different epithelial cells at different stages of differentiation and development, and upon malignant transformation. At any given time, an epithelial cell typically expresses two to five keratin polypeptides, including at least one member of each type. Prior to the discovery of the two distinct types of keratin subunits and their coordinate expression, it was shown that no single keratin protein of either type by itself was competent for filament assembly. Collectively, these results suggest that the expression of both types may be essential for subsequent filament formation.

The details of how the keratin subunits assemble remain to be elucidated. Until recently, most of what we knew about the structure of the keratin filament came from the early X-ray diffraction studies of Crick and Pauling back in the early 1950s. The 8 nm keratin filament was known to be made up of an as yet undetermined number of smaller, closely packed filaments, called protofilaments. Each protofilament has a diameter of about 2 nm and is itself made up of several \( \alpha \)-helical polypeptide strands intertwined in a coiled-coil configuration.

The early X-ray studies could not decipher whether two or three helical polypeptide subunits were interwoven to give the coiled-coil substructure of the protofilament. Although physicochemical studies suggested a three-stranded coiled-coil model.

\( \alpha \) E. F. is a Searle Scholar, the recipient of a National Science Foundation Presidential Young Investigator Award, and a National Institutes of Health Career Development Award. K. H. K. is a Postdoctoral Fellow funded by the National Cancer Institute. M. P. G. is a Postdoctoral Fellow funded by the Damon Runyon-Walter Winchell Cancer Fund. D. M. is a National Institutes of Health Predoctoral Trainee. This work was supported by a grant from the National Institutes of Health.

436
building using helical portions of a wool keratin showed that the sequence is compatible with a two-stranded, but not a three-stranded coiled coil. Additional evidence from chemical cross-linking analyses also favors a protofilament consisting of a pair of fibrous polypeptides interwoven in a coiled-coil structure. More recent electron microscopy studies suggest that several higher order coiled coiling of protofilaments may take place during the assembly of the 8 nm filament. Although the controversies over the substructure of the keratin filament have not yet been unequivocally resolved, it seems likely that a type I and a type II keratins are both involved in the formation of the basic subunit structure of the protofilament.

Four major questions must now be considered: (1) What are the details of the structural interactions between a pair of type I and type II keratin subunits that lead to the formation of the protofilament? (2) What is the nature of the differences between the individual members of the type I or type II family and how do these differences influence the resulting structure or function of the 8 nm filaments expressed in different epithelial tissues? (3) What is the relation between the keratins and other IF subunits? (4) What is the complexity of the genes encoding the different keratins and how is their expression regulated? In this paper, we discuss work done in our own laboratory and the laboratory of others where techniques in molecular biology have been utilized to begin to unravel the answers to these questions.

THE AMINO ACID SEQUENCES FOR A TYPE I AND TYPE II KERATIN EXPRESSED IN HUMAN EPIDERMAL CELLS AND THEIR PREDICTED SECONDARY STRUCTURES

To explore whether a type I and type II keratin might interact to form the coiled coil of the protofilament, it is necessary to determine sequences of a pair of type I and type II keratins that are co-expressed in a particular epithelial tissue. We have obtained the predicted amino acid sequences for the 50 kD type I and the 56 kD type II keratin co-expressed in cultured human epidermal cells. These sequences were determined in large part from cloning and sequencing near full-length cDNAs complementary to their mRNAs. We have recently completed the amino acid sequence of the type I keratin by isolating and sequencing the gene encoding the human 50 kD keratin.

A comparison of the predicted amino acid sequences for the two keratins reveals that they are only very distantly related to one another with < 30% homology (Figure 1, A). The strongest sequence homology resides in residues 200-350. A glycine-serine-rich sequence in the amino terminal region of the type I 50 kD keratin (residues 1-50) shows weak homology with a similar sequence in the carboxy terminal region of the type II 56 kD keratin (residues 380-440).

A remarkable feature of the two keratins is that even though they have very different sequences, their predicted secondary structures are largely indistinguishable (Figure 2). Computer-assisted secondary structures are nearly equal in length for the two keratins, even though the type II keratin is always substantially larger than the co-expressed type I keratin. The difference in size of the keratins seems to
FIGURE 1. Matrix plot of homology between (A) a human type I (50 kD) and a human type II (56 kD) epidermal cytoskeletal keratin; (B) the human type I and a wool microfibrillar type I keratin; (C) chicken desmin and the human type I keratin; and (D) chicken desmin and hamster vimentin. The numbers on the axes correspond to the residue numbers of the respective sequences. The sequences were obtained from the following sources: human type I, human type II, wool type I, chicken desmin, and hamster vimentin. The positions of the four predicted helical domains (see Figures 2 and 3) are marked by bars (I-IV) along the residue axes. Note that the sequences of the amino terminal end and the first predicted helical region (I) of the human type II keratin are not yet known, although their structures are expected to be similar to those of the human type I keratin. The matrix plot analysis was carried out by a modified version of Pustell and Kafatos (1982) program. The parameters used in the analysis: range = 20, distance weight factor = 1, compression factor = 4, cutoff point = 34%. Thus, the position of each number on the plots indicates that within a 20 residue long segment to the right of the marked position, at least 35% of the residues are identical in both sequences. The numbers indicate the percentile range of identical residues in the 20 residue stretch, 3 = 35-39%, 4 = 40-49%, . . . 9 = 90-99%.
FIGURE 2. Comparison of the predicted secondary structures of the 56 kD type II and 50 kD type I keratins that are coordinately expressed in cultured human epidermal cells. The predicted amino acid sequences of the two keratins and their predicted secondary structures were determined previously. The values on the abscissa indicate amino acid residue number. The sequences have been aligned according to their predicted secondary structures. The Chou and Fasman and Garnier, Osguthorpe and Robson methods used in our analyses make predictions for four conformational states: α-helix, β-sheet, β-turn, and random coil. For each keratin, central α-helical domains are represented by the thick bars. The thinner internal bars mark the positions of the β-turns, or helix breakers, that are predicted by the Chou and Fasman method for all IF protein sequences. The bars flanking the helical region represent the non-helical termini. Largely different sequence homologies are indicated by different shades of gray.

reside solely in their nonhelical amino and carboxy termini, which are highly variable in length. These general structural features of the pair of human epidermal keratins are also very similar to those of the pair of partially sequenced wool keratins, and they are likely to represent structural characteristics common to all keratins.

The central α-helical domains of the keratin subunits are likely to be involved in the coiled coil of the 2 nm protofilament. The sequence characteristic of a periodic interchain interaction, i.e., a heptad repeat of a b c d e f g, where a and d are hydrophobic residues, has been identified in the helical portions of all sequenced keratins. In addition, an interesting periodicity in positively and negatively charged residues has been observed for different keratin subunits, indicating that electrostatic interactions may also play an important role in stabilizing the keratin structure.

The overall structural features of the type I and type II keratins support the conservation and coordinate expression data that implicate these two types of sequences in forming the structural backbone of the keratin filament. These data do not necessarily rule out the possibility that two members of a single keratin type could form a keratin filament. However, coupled with the finding that no single keratin can self-assemble into a filament and with recent assembly data that suggest that filaments from particular mixtures of type I and type II keratin subunits form with different stabilities, it seems likely that the two subunits are intimately involved in the coiled coil of the protofilament. This apparent requirement for two subunits in filament formation seems to set the keratins apart from other intermediate filament subunits, which have structural features similar to the keratins, but which can assemble as homopolymers.
THE SEQUENCE VARIATIONS AMONG DIFFERENT MEMBERS OF A SINGLE KERATIN TYPE AND THEIR FUNCTIONAL SIGNIFICANCE

Although the type I and type II keratin subunits are only distantly evolutionarily related, hybridization studies indicate that the individual members within a particular class are quite similar to each other. The nature of these more subtle differences is important, since it is the members of each class that are differentially expressed in different tissues. Sequence analyses of all the different type I and type II keratins should resolve whether the multiplicity of keratin sequences is simply an evolutionary phenomenon, or whether the use of different keratin subunits in filament assembly influences the properties, structure, or interactions of the resulting filaments.

To date sequence data for the keratins are limited. In addition to the sequences of the two human keratins that we have obtained, there is complete sequence data for a type I keratin of mouse and partial sequence data for a type I and a type II keratin of wool. Nevertheless, from a comparison of these few sequences, an intriguing relation is emerging. For any two keratins of the same type, the $\alpha$-helical domains of the polypeptides are quite highly conserved (FIGURE 1B and FIGURE 3). This is in direct contrast to a comparison of two keratins of different types, where these domains are largely distinct (FIGURE 1A and FIGURE 2). The major differences in the sequences of different subunits of the same type reside almost exclusively in the nonhelical termini of the keratins. These differences are most pronounced when the keratins of wool are compared with the keratins of epidermis (FIGURE 1, B). Whereas the end segments of the epidermal keratins contain a strange inexact
repeat consisting of glycine and serine residues interspersed with phenylalanine, tyrosine, and leucine, the wool keratins have termini that are conspicuously rich in cysteine (TABLE 1). Such differences are quite remarkable when it is considered that the central helical domains of the two type I or type II keratins are highly homologous. Thus, the keratins of a particular class, either type I or type II, seem to have constant helical domains and variable nonhelical termini.

The functional significance of the multiplicity of keratin sequences and their differential expression still remains largely undetermined. However, a closer look at the nonhelical termini and the role that these termini play in filament assembly has shed some light on this issue. Elegant physicochemical studies by Skerrow, Matoltsy, and Matoltzy,35 and also by Steinert and colleagues36 have revealed that the removal of these ends by limited proteolytic digestion render the subunits incompetent for filament assembly. The subunits intertwine to form the 2 nm coiled coil of the protofilament, but no protofibrillar interactions take place to yield the 8 nm diameter of the filament. The end-to-end interactions of the subunits are also disrupted, leaving rods that are 0.05 μm rather than 40 μm in length. Thus, the use of subunits with different termini in filament assembly might not necessarily change the structure of the resulting 8 nm filaments, but it is likely to give rise to filaments with quite different properties. The microfilbrillar keratin filaments are insoluble and possibly rigid, with numerous disulfide bridges linking the termini of the subunits to one another, whereas the epidermal cytoskeletal keratin filament is probably more flexible with glycine-serine-rich termini and relatively few cysteine residues.37,38,40 Although the variations among the individual keratin subunits of each class are not likely to be more pronounced than those already seen between the keratins of wool and epidermis, it is expected that smaller variations in the sequences of other epithelial keratins will give rise to subtle variations in the properties of the corresponding filaments, e.g. their solubility, their flexibility, their tensile strength, and their ability to interact with the cell or nuclear membrane or other cytoplasmic components.

A COMPARISON OF TYPE I AND TYPE II KERATIN WITH OTHER IF SUBUNITS

Sequence analysis of different IF subunits has revealed that all IF proteins are evolutionarily related.13,14,16,40-44,46,55-58,61,62 A summary of the IF sequence data suggests that if all IF proteins originated from a single gene, then two distant duplications must have led to the formation of a third IF gene in addition to the primordial type I and type II keratin genes.64 Each of these genes must have undergone further, more recent, duplication to give rise to three related multigene subfamilies. The third gene subfamily is made up of the genes encoding the neurofilament proteins, glial filament proteins, desmin, and vimentin, all of which share 60-70% homology with one another.65 This IF multigene subfamily shows 25-40% homology with both the type I and type II keratin subfamilies. FIGURE 1 illustrates the low but substantial homology between desmin and the 50 kD human type I keratin (1C) and the much higher homology between desmin and vimentin (1D).

For all IF subunits, the predicted secondary structures are strikingly similar.16,40,42,48,62,65 In comparing any two members of the same subfamily, the central α-helical domains share marked homology, whereas the variation in sequence resides primarily at the nonhelical termini (TABLE 1).
Table 1. The Amino Acid Compositions of the Different Structural Domains of Epidermal (Cytoskeletal) and Wool (Microfibrillar) Keratins and Intermediate Filament Proteins Desmin and Vimentin

<table>
<thead>
<tr>
<th>Amino Terminus</th>
<th>Central Region</th>
<th>Carboxy Terminus</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>WT-I MT-I HT-I WT-II CDES HVIM</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>3.7 0.0 5.2 5.5 6.1 4.0</td>
<td>4.8 6.3 8.0 10.0</td>
</tr>
<tr>
<td>Arg(+)</td>
<td>5.6 2.8 3.5 5.5 11.2 13.0</td>
<td>9.0 7.0 7.7 7.1</td>
</tr>
<tr>
<td>Asn</td>
<td>11.1 0.7 0.9 2.0 5.0</td>
<td>7.1 6.7 6.1 5.9</td>
</tr>
<tr>
<td>Asp(−)</td>
<td>1.9 0.7 0.9 (7.3) 2.0 3.0</td>
<td>4.8 4.8 6.4 8.2</td>
</tr>
<tr>
<td>Cys</td>
<td>14.8 2.1 2.6 6.4 0.0 0.0</td>
<td>2.6 0.6 0.6 0.7</td>
</tr>
<tr>
<td>Gin</td>
<td>0.0 0.0 0.7 0.9 6.1 2.0</td>
<td>8.7 8.6 6.4 7.1</td>
</tr>
<tr>
<td>Glu(−)</td>
<td>1.9 0.0 0.9 (12.8) 2.0 1.0</td>
<td>15.4 13.3 14.1 11.9</td>
</tr>
<tr>
<td>Gly</td>
<td>9.3 49.6 35.0 9.2 9.2 7.0</td>
<td>1.0 2.9 2.9 3.0</td>
</tr>
<tr>
<td>His</td>
<td>0.0 0.0 0.0 0.0 0.0 0.0</td>
<td>0.6 0.5 0.6 1.1</td>
</tr>
<tr>
<td>Ile</td>
<td>1.9 0.0 2.6 4.6 0.0 0.0</td>
<td>4.8 4.8 4.5 4.8</td>
</tr>
<tr>
<td>Leu</td>
<td>5.6 3.5 6.0 11.9 6.1 7.0</td>
<td>14.7 14.6 12.8 12.3</td>
</tr>
<tr>
<td>Lys(+)</td>
<td>0.0 0.0 0.7 1.7 5.5 0.0 0.0</td>
<td>3.5 3.5 3.9 2.6</td>
</tr>
<tr>
<td>Met</td>
<td>0.0 0.0 0.0 1.7 0.9 1.0 1.0</td>
<td>0.3 1.9 3.8 2.6</td>
</tr>
<tr>
<td>Phe</td>
<td>7.4 9.9 6.0 4.6 6.1 3.0</td>
<td>1.0 0.6 1.3 1.5</td>
</tr>
<tr>
<td>Pro</td>
<td>9.3 0.0 0.0 0.9 5.1 4.1 4.0</td>
<td>1.0 1.0 1.0 0.4</td>
</tr>
<tr>
<td>Ser</td>
<td>18.5 22.7 21.6 7.3 19.4 25.0</td>
<td>7.4 7.6 4.5 4.1</td>
</tr>
<tr>
<td>Thr</td>
<td>3.7 0.7 3.5 4.6 12.7 11.0</td>
<td>4.5 5.4 4.5 3.7</td>
</tr>
<tr>
<td>Trp</td>
<td>1.9 0.0 0.0 0.0 0.0 0.0 0.0 0.0</td>
<td>0.3 0.3 0.3 0.3</td>
</tr>
<tr>
<td>Tyr</td>
<td>0.0 3.5 3.5 3.7 0.5 5.0</td>
<td>2.9 3.2 3.2 3.3</td>
</tr>
<tr>
<td>Val</td>
<td>3.7 2.1 2.6 4.6 7.1 6.0</td>
<td>6.4 4.8 4.8 4.5</td>
</tr>
</tbody>
</table>

The values are presented as percentages and are based on the sequences from the following sources: (1) WT-I: Type I microfibrillar keratin from sheep wool(11,14,43,62), (2) MT-I: Type I cytoskeletal keratin from mouse epidermis ($M_r = 59K$)(4), (3) HT-I: Type I cytoskeletal keratin from human epidermal cells ($M_r = 51K$)(42,62), (4) HT-II: Type II cytoskeletal keratin from human epidermal cells ($M_r = 56K$)(16), (5) WT-II: Type II microfibrillar keratin from sheep wool(11,14,43,62), (6) CDES: Desmin from chicken muscle ($M_r = 53K$)(62), (7) HVIM: Vimentin from hamster ($M_r = 53K$)(4).

The "Central Region" represents the segment extending from the first amino acid of Helical Domain I to the last amino acid of Helical Domain IV as previously described(4).

* The composition is based on incomplete sequence for this region.
THE REGULATION OF THE DIFFERENTIAL EXPRESSION OF HUMAN KERATINS IN DIFFERENT TISSUES AND IN MALIGNANT NEOPLASMS DERIVED FROM NORMAL EPITHELIA

It is interesting that the nonhelical termini of the two keratins expressed in human epidermal cells are rich in glycine and serine residues, whereas the two keratins produced by wool are rich in cysteine and serine residues. This suggests that the expression of different pairs of type I and type II keratins may be coordinately regulated. An investigation of the keratins expressed in different tissues suggests that indeed this may be the case. It has been well documented that the expression of most keratins arises from the appearance of newly synthesized mRNAs in these cells. Whether the coordinate expression of all of these proteins extends to the level of the genome has not yet been determined, although there are clearly multiple genes (≥10) for each of the two keratin subfamilies. We have begun to isolate and characterize the multiple human genes encoding the keratins and have shown that they are different, albeit related. In at least some cases, the genes encoding the keratins are differentially expressed. However, we have not ruled out the possibility of additional regulation at the post-transcriptional level.

To further probe the regulation of keratin gene expression, we have chosen differentiating human epidermis as a model system. Previously, we had found that as a human basal epidermal cell is triggered to terminally differentiate, it changes the pattern of keratins that it makes. RNAs encoding keratins of 46, 50, 56, and 58 kD are made by the basal cells, and additional mRNAs encoding keratins of 56.5 and 67 kD are produced by the differentiating keratinocytes (FIGURE 4). The 46 and 50 kD keratins are acidic and belong to the type I class, whereas the 56 and 58 kD keratins are basic and belong to the type II class. We have recently shown that the newly synthesized 56.5 kD keratin is an acidic type I keratin and the 67 kD keratin is a basic type II keratin. We have confirmed this by two-dimensional gel electrophoresis, by immunocross-reactivity and by Northern blot analysis. The shift towards the synthesis of type I and type II keratins of unusually large size occurs in all vertebrate epidermis and seems to occur without an imbalance in the normal ratio of type I to type II keratins within each cell.

We now know that the shift to the synthesis of large keratins is regulated in a negative fashion by the fat-soluble isoprenoid, vitamin A. When human epidermal cells are cultured in vitro in the absence of vitamin A, an induction in the synthesis of these large keratins and their mRNAs is observed. Thus, whether directly or indirectly, the vitamin appears to regulate coordinately the expression of at least some pairs of keratins. In addition to the influence of vitamin A on epithelial growth and differentiation, a number of other environmental and extracellular roles have been shown to play a role in gene regulation in these cells. These include dermal-epidermal interactions, CAMP effectors, steroid hormones, and growth factors.

That extracellular factors play a prominent role in keratin gene expression can be illustrated by investigating the altered pattern of keratins produced under different conditions in human esophageal tissue (FIGURE 5). Normal esophageal tissue produces three major keratins, 58, 56, and 52 kD (lane 1). When these cells are cultured in vitro, they change their pattern of keratins and bands of 56, 50, and 46 kD are seen (lane 2). This pattern is indistinguishable from that produced by the cells upon malignant transformation (lane 3). However, in a liver metastasis of the esophageal carcinoma, the pattern of keratins was found to change again, with new keratins of 56, 46, and 40 kD being synthesized (lane 4).
FIGURE 4. Synthesis of human keratins from mRNA of differentiating epidermis and of basal epidermal cells. Poly(A)+ RNAs from whole thigh skin and from cultured basal epidermal cells were purified and translated in vitro in the presence of [35S]methionine. The synthesized products were immunoprecipitated and resolved by SDS-polyacrylamide gel electrophoresis. Lanes 1 and 3: [35S]methionine-labeled extracts from basal cells and epidermis, respectively. Lanes 2 and 4: Antikeratin immunoprecipitates of translated products of the mRNAs extracted from basal cells and epidermis, respectively.
We have shown that at least for cells that have been cultured from squamous cell carcinomas of the tongue and epidermis, the altered pattern of keratins produced on malignant transformation of these cells is transient and is influenced profoundly by the levels of vitamin A in the culture medium. This opens the possibility that altered patterns of keratin synthesis in tumors stem from either a change in the sensitivity of the malignant tissue to the vitamin, or alternatively, from a change in the environment of the tumor cells as a consequence of disruption of the normal organization of the tissue.

The exact nature of the roles played by these extracellular factors in regulating the expression of keratin genes must await an investigation of the structure of the keratin genes and their chromosomal organization. Further sequencing will also be necessary to begin to elucidate the functional significance of the differential expression of these genes. Nonetheless, it is clear that many of the mysteries of the most variable cytoskeletal component in higher eukaryotic organisms have begun to unravel. The field of keratin filaments appears to be at the threshold of elucidating the functional significance of the multiplicity of subunits that can all assemble into the abundant 8 nm filaments that form the structural web within an epithelial cell.

ACKNOWLEDGMENT

We thank Ms. Claire Doyle for her expeditious typing of this manuscript.
REFERENCES


36. **Geisler, N. & K. Weber. 1982. The amino acid sequence of chicken muscle desmin provides a common structural model for intermediate proteins including the wool a-keratins. EMBO J. 1: 1649-1656.**


44. **Steinert, P. M., R. H. Rice, D. R. Roop, B. L. Trus & A. C. Steven. 1983. Complete


