Cell, Vol. 33, 915–924, July 1983, Copyright © 1983 by MIT	0092-8674/83/070915-10 $02.00/0

# The cDNA Sequence of a Type II Cytoskeletal Keratin Reveals Constant and Variable Structural Domains among Keratins

**Israel Hanukoglu and Elaine Fuchs**
Department of Biochemistry
The University of Chicago
Chicago, Illinois 60637

## Summary

**We present the cDNA and amino acid sequences of a cytoskeletal keratin from human epidermis ($M_r$ = 56K) that belongs to one of the two classes of keratins (Type I and Type II) present in all vertebrates. In these two types of keratins the central ~300 residue long regions share ~30% homology both with one another and with the sequences of other IF proteins. Within this region, all IF proteins are predicted to contain four helical domains demarcated from one another by three regions of $\beta$-turns. The amino and carboxy termini of the Type II keratin are very different from those of microfibrillar keratins and other nonkeratin IF proteins. However, they contain unusual glycine-rich tandem repeats similar to the amino terminus of the Type I keratin. Thus the size heterogeneity among keratins appears to be a result of differences in the length of the terminal ends rather than the structurally conserved central region.**

## Introduction

The cytoskeleton of most mammalian cells includes a prominent network of 8–10 nm filaments called intermediate filaments (IF) (Lazarides, 1982). In the past ten years a number of biochemical, immunological, and physicochemical studies have indicated that the structure of IF is highly similar to that of microfibrils which form the backbone of such epidermal appendages as hair and wool (Crewther and Harrap, 1967; Fraser, MacRae, and Rogers, 1972; Skerrow, Matoltsy, and Matoltsy, 1973; Jones, 1975; Fraser, MacRae, and Suzuki, 1976; Steinert, 1978; Steinert, Idler, and Goldman, 1980; Weber, Osborn, and Franke 1980; Geisler, Kaufmann, and Weber, 1982). Thus the models that were developed earlier for the structure of microfibrils have been more recently extended to IF. The microfibrils are composed of about 10 protofibrils and each protofibril represents a polymer of two or three polypeptide chains (for review see Fraser, MacRae and Rogers, 1972). These proteins that form the subunits of protofibrils contain long regions of mainly $\alpha$-helical conformation in the center, and a staggered conformation of unknown structure at the amino and carboxy terminal ends. Since the early studies of Pauling and Corey (1953) and Crick (1953) it has been thought that the helical regions of the protofibrillar subunits intertwine around one another to form a coiled-coil rod, and that these rods are linked end-to-end to form a rope-like filament that constitutes the protofibril.

Until recently, the complete amino acid sequence of any IF or microfibrillar protein was not known. Therefore, many important aspects of the structure of IF and of microfibrils, and the relationships among the various proteins that form these filaments, remained unclear. In different tissues and at different stages of development there are at least 20–30 distinct polypeptides ($M_r$ 40–70 K) capable of forming IF (Lazarides, 1982). Recently, amino acid sequencing of desmin, vimentin, glial filament protein, and neurofilament protein fragments indicated that these four classes of IF proteins are highly homologous (70%–80%) (Geisler and Weber, 1981; 1982; Geisler, Plessmann, and Weber, 1982; Hong and Davison, 1981). The amino acid sequence of a 50 K human epidermal keratin, predicted from a cDNA sequence, also revealed a significant yet low homology with these four IF proteins (~30%) (Hanukoglu and Fuchs, 1982). Furthermore, the sequences of all of these IF proteins showed significant homology with the partial sequences of wool microfibrillar keratins (Hanukoglu and Fuchs, 1982; Geisler and Weber, 1982; Dowling, Parry, and Sparrow, 1983) thus finally directly confirming earlier indications that IF and microfibrillar keratins are related.

Within the family of IF proteins, the keratins represent the largest and the most diverse class (Moll et al., 1982). In contrast to the other types of IF proteins, the cytoskeletal keratins can be further subgrouped into at least two distinct classes on the basis of the homologies of their corresponding mRNAs to two different cloned human epidermal keratin cDNAs (Fuchs et al., 1981). At least one member of each of these two classes of keratins is present in all epithelial cells (Kim, Rheinwald, and Fuchs, 1983), and genomic DNA sequences complementary to each of the two classes are observed in all vertebrates (Fuchs et al. 1981; Fuchs and Marchuk, 1983). Recently, studies on cloned cDNAs for mouse cytoskeletal keratins have similarly demonstrated that there is more than one class of keratin mRNAs (Roop et al., 1983).

The partial sequences of two wool keratin fragments have also revealed that there are at least two distinct sequences of microfibrillar keratins (Type I and Type II) that share with each other only about 30% homology (Crewther et al., 1978; Gough et al., 1978). We previously observed that the human cytoskeletal keratin sequence of one class shares 60% homology with the Type I wool microfibrillar keratin sequence and 30% with the Type II sequence. Thus we predicted that the other class of epidermal keratins would be more homologous to the Type II wool keratin sequence than to the Type I sequence. In this paper, we report the cDNA and the predicted amino acid sequence of a 56 K keratin that represents a member of the second class of keratin mRNAs. This sequence confirms our earlier predictions, hence we now name these two classes of keratins as Type I and Type II, on the basis of their homologies to the wool microfibrillar keratin fragments. In addition, we present comparisons of the sequence and the predicted structure of this Type II keratin with those of the Type I keratin and other IF proteins. Our results reveal the molecular basis of the variability in the sizes of the two classes of keratins and indicate that despite the variations in their sequences, all IF proteins contain a central region with four helical domains demarcated by three conserved sites of helix interruptions. The nonhelical ends of both Type I and Type II cytoskeletal keratins contain short tandem repeats of an unusual sequence which distinguishes them not only from other IF proteins but also from the microfibrillar Type I and Type II keratins.

## Results

### The cDNAs Sequenced and the DNA Sequencing Strategy

The cDNAs of Type I and Type II keratin mRNAs present in cultured human epidermal cells were cloned previously using pBR322 as vector and E. coli χ1776 as host (Fuchs et al., 1981). We have already reported the DNA sequence of two cloned cDNAs that represented copies of a Type I mRNA for a 50 K cytoskeletal keratin (Hanukoglu and Fuchs, 1982). For the present study we selected two different cloned cDNAs, KA-1 and KA-13, that belong to the other class (Type II). As determined by positive hybridization–selection of the corresponding mRNA, the insert of KA-1 codes for a 56 K cytoskeletal keratin (Kim, Rheinwald, and Fuchs, 1983). KA-1 was selected because it represented the cDNA insert nearest in size to the complete mRNA from which it was derived (Fuchs et al., 1981). Preliminary restriction maps indicated that KA-13 is shorter than KA-1 but otherwise contains the same sites as KA-1, suggesting that it represents a copy of the same mRNA as KA-1. Thus KA-13 was also selected for sequencing to ensure the fidelity of the sequence of KA-1. The strategies used in determining the DNA sequence of the cDNA inserts, KA-1 and KA-13, are shown in Figure 1.

### Coding and Noncoding Regions of the cDNAs

The complete sequence of the KA-1 keratin cDNA insert is shown in Figure 2. The size of the insert is 1685 nucleotides. The DNA sequence of the shorter insert, KA-13, is identical to the corresponding sequenced region of KA-1 with the exception of one nucleotide which is indicated in the legend. The cDNA insert KA-1 includes a complete copy of the 3′ end of the mRNA, as it has a 32 nucleotide long poly(A) stretch that most likely represents a copy of a portion of the 3′ end poly(A) tail of the mRNA. In addition, the putative polyadenylation signal sequence AAUAAA (Fitzgerald and Shenk, 1981) appears 21 nucleotides from the first A of the poly(A) region.

The open reading frame shown in Figure 2 extends for 1071 nucleotides and ends with the stop codon TAA. The other two possible reading frames are interrupted by 11 and 15 stop codons (either TAG or TGA) up to nucleotide 1071 and contain an additional 9 and 12 stop codons (TAG, TGA, or TAA) beyond this point. It is interesting that the sequence TAA does not exist within the coding segment of the mRNA in any reading frame. This may suggest that during evolution TAA has been screened out of the coding region. Overall, these results clearly indicate that the reading frame depicted in Figure 2 is the only frame of this mRNA that codes for a protein. The unusually long (545 nucleotides) 3′-end untranslated region predicted by this sequence is consistent with previous estimates of the size of the noncoding regions of the mRNA for 56 K epidermal keratin (Fuchs and Green, 1979). The cDNA does not contain a complete copy of the 5′ end of the mRNA because the size of the coding region sequence is not sufficient to code for the full length of the 56 K keratin. The estimated size of the mRNA for the 56 K keratin is 2150 nucleotides (Fuchs and Green, 1979), indicating that the KA-1 keratin cDNA represents about 80% of the total length of the mRNA.

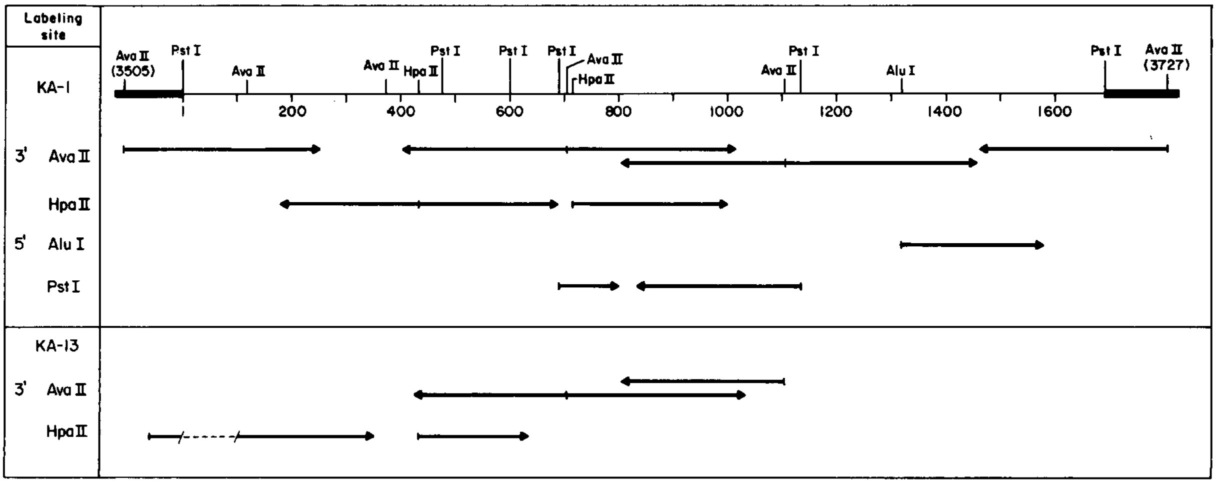

Figure 1. The DNA Sequencing Strategies for the Human Epidermal 56 K Type II Keratin cDNA Inserts KA-1 and KA-13

The KA-1 cDNA insert (thin lined) flanked by pBR322 sequences (heavy lined) is shown at the top. The nucleotide numbers within the KA-1 insert are in the 5′ to 3′ direction of the mRNA strand and the positions of all recognition sites for each enzyme, except Alu I, are indicated. The $^{32}P$ labeling site for each series of restriction fragments is shown at the left, and the direction and extent of DNA sequence determination are indicated by the arrows.

(G)$^{10}$

```
                                     10                                      20                                      30
Gln Asn Leu Glu Pro Leu Phe Glu Gln Tyr Ile Asn Asn Leu Arg Arg Gln Leu Asp Ser Ile Val Gly Glu Arg Gly Arg Leu Asp Ser
CAG AAC CTG GAG CCG TTG TTC GAG CAG TAC ATC AAC AAC CTC AGG AGG CAG CTG GAC AGC ATT GTC GGG GAA CGG GGC CGC CTG GAC TCA
                                     30                                      60                                      90

                                     40                                      50                                      60
Glu Leu Arg Gly Met Gln Asp Leu Val Glu Asp Phe Lys Asn Lys Tyr Glu Asp Glu Ile Asn Lys Arg Thr Ala Ala Glu Asn Glu Phe
GAG CTC AGA GGC ATG CAG GAC CTG GTG GAG GAC TTC AAG AAC AAA TAT GAG GAT GAA ATC AAC AAG CGC ACA GCA GCA GAG AAT GAA TTT
                                    120                                     150                                     180

                                     70                                      80                                      90
Val Thr Leu Lys Lys Asp Val Asp Ala Ala Tyr Met Asn Lys Val Glu Leu Gln Ala Lys Ala Asp Thr Leu Thr Asp Glu Ile Asn Phe
GTG ACT CTG AAG AAG GAT GTG GAT GCT GCC TAC ATG AAC AAG GTT GAA CTG CAA GCC AAG GCA GAC ACT CTC ACA GAC GAG ATC AAC TTC
                                    210                                     240                                     270

                                    100                                     110                                     120
Leu Arg Ala Leu Tyr Asp Ala Glu Leu Ser Gln Met Gln Thr His Ile Ser Asp Thr Ser Val Val Leu Ser Met Asp Asn Asn Arg Asn
CTG AGA GCC TTG TAT GAT GCA GAG CTG TCC CAG ATG CAG ACC CAC ATC TCA GAC ACA TCT GTG GTG CTG TCC ATG GAC AAC AAC CGC AAC
                                    300                                     330                                     360

                                    130                                     140                                     150
Leu Asp Leu Asp Ser Ile Ile Ala Glu Val Lys Ala Gln Tyr Glu Glu Ile Ala Gln Arg Ser Arg Ala Glu Ala Glu Ser Trp Tyr Gln
CTG GAC CTG GAC AGC ATC ATC GCT GAG GTC AAG GCC CAA TAT GAG GAG ATT GCT CAG AGA AGC CGG GCT GAG GCT GAG TCC TGG TAC CAG
                                    390                                     420                                     450

                                    160                                     170                                     180
Thr Lys Tyr Glu Glu Leu Gln Val Thr Ala Gly Arg His Gly Asp Asp Leu Arg Asn Thr Lys Gln Glu Ile Ala Glu Ile Asn Arg Met
ACC AAG TAC GAG GAG CTG CAG GTC ACA GCA GGC AGA CAT GGG GAC GAC CTG CGC AAC ACC AAG CAG GAG ATT GCT GAG ATC AAC CGC ATG
                                    480                                     510                                     540

                                    190                                     200                                     210
Ile Gln Arg Leu Arg Ser Glu Ser Asp His Val Lys Lys Gln Cys Ala Asn Leu Gln Ala Ala Ile Ala Asp Ala Glu Gln Arg Gly Glu
ATC CAG AGG CTG AGA TCT GAG AGC GAC CAC GTC AAG AAG CAG TGC GCC AAC CTG CAG GCC GCC ATT GCT GAT GCT GAG CAG CGT GGG GAG
                                    570                                     600                                     630

                                    220                                     230                                     240
Met Ala Leu Lys Asp Ala Lys Asn Lys Leu Glu Gly Leu Glu Asp Ala Leu Gln Lys Ala Lys Gln Asp Leu Ala Arg Leu Leu Lys Glu
ATG GCC CTC AAG GAT GCC AAG AAC AAG CTG GAA GGG CTG GAG GAT GCC CTG CAG AAG GCC AAG CAG GAC CTG GCC CGG CTG CTG AAG GAG
                                    660                                     690                                     720

                                    250                                     260                                     270
Tyr Gln Glu Leu Met Asn Val Lys Leu Ala Leu Asp Val Glu Ile Ala Thr Tyr Arg Lys Leu Leu Glu Gly Glu Glu Cys Arg Leu Asn
TAC CAG GAG CTG ATG AAT GTC AAG CTG GCC CTG GAC GTG GAG ATC GCC ACC TAC CGC AAG CTG CTG GAG GGT GAG GAG TGC AGG CTG AAT
                                    750                                     780                                     810

                                    280                                     290                                     300
Gly Glu Gly Val Gly Gln Val Asn Ile Ser Val Val Gln Ser Thr Val Ser Ser Gly Tyr Gly Gly Ala Ser Gly Val Gly Ser Gly Leu
GGC GAA GGC GTT GGA CAA GTC AAC ATC TCT GTG GTG CAG TCC ACC GTC TCC AGT GGC TAT GGC GGT GCC AGT GGT GTC GGC AGT GGC TTA
                                    840                                     870                                     900

                                    310                                     320                                     330
Gly Leu Gly Gly Gly Ser Ser Tyr Ser Tyr Gly Ser Gly Leu Gly Val Gly Gly Gly Phe Ser Ser Ser Ser Gly Arg Ala Ile Gly Gly
GGC CTG GGT GGA GGA AGC AGC TAC TCC TAT GGC AGT GGT CTT GGC GTT GGA GGT GGC TTC AGT TCC AGC AGT GGC AGA GCC ATT GGG GGT
                                    930                                     960                                     990

                                    340                                     350
Gly Leu Ser Ser Val Gly Gly Gly Ser Ser Thr Ile Lys Tyr Thr Thr Thr Ser Ser Ser Ser Arg Lys Ser Tyr Lys His  .
GGC CTC AGC TCT GTT GGA GGC GGC AGT TCC ACC ATC AAG TAC ACC ACC ACC TCC TCC TCC AGC AGG AAG AGC TAT AAG CAC TAA AGTGCGT
                                   1020                                    1050


CTGCTAGCTCTCGGTCCCACAGTCCTCAGGCCCCTCTCTGGCTGCAGAGCCCTCTCCTCAGGTTGCCTGTCCTCTCCTGGCCTCCAGTCTCCCCTGCTGTCCCAGGTAGAGCTGGGGAT
          1094                1114                1134                1154                1174                1194

GAATGCTTAGTGCCCTCACTTCTTCTCTCTCTCTCTATACCATCTGAGCACCCATTGCTCACCATCAGATCAACCTCTGATTTTACATCATGATGTAATCACCACTGGAGCTTCACTGT
           1214                1234                1254                1274                1294                1314

TACTAAATTATTAATTTCTTGCCTCCAGTGTTCTATCTCTGAGGCTGAGCATTATAAGAAAATGACCTCTGCTCCTTTTCATTGCAGAAAATTGCCAGGGGCTTATTTCAGAACAACTT
            1334                1354                1374                1394                1414                1434

CCACTTACTTTCCACTGGCTCTCAAACTCTCTAACTTATAAGTGTTGTGAACCCCCACCCAGGCAGTATCCATGAAAGCACAAGTGACTAGTCCTATGATGTACAAAGCCTGTATCTCT
             1454                1474                1494                1514                1534                1554
```

GTGATGATTTCTGTGCTCTTCACTCTTTGCAATTGCTAAATAAAGCAGATTTATAATAC(A)$^{32}$(C)$^{27}$

```
       1574                1594                1614
```

Figure 2. The DNA Sequence of the KA-1 cDNA Insert, and the Predicted Amino Acid Sequence of the Human Epidermal 56 K Type II Keratin

The sequence is shown in the 5′ to 3′ direction of the mRNA strand. The numbers above the amino acids mark the position of the amino acids, and those below mark the position of the nucleotides (N-terminus of the predicted amino acid sequence = 1). The cluster of Gs at the 5′ end and the cluster of Cs at the 3′ end represent the enzymatically-tailed regions of the plasmid and the ds-cDNA used for cloning (Fuchs et al., 1981). The stop codon of the reading frame shown here is marked with a dot. The underlined nucleotides represent the putative polyadenylation signal sequence. The nucleotide at position 208 is a C instead of a T in the KA-13 cDNA.

To provide further evidence that the open reading frame indeed codes for a 56 K keratin, we isolated the 56 K keratin from cultured human epidermal cells and determined its amino acid composition. The agreement between the results of this actual amino acid analysis and the amino acid composition predicted from the cDNA is significant (Table 1). However, the percentage of Gly and Phe are lower in the predicted sequence. As we discuss in the next section, we expect that the missing amino terminal portion of the cDNA sequence codes for a protein segment that is similar to the amino terminal portion of the Type I 50 K keratin. This region in the 50 K keratin contains an unusual sequence that consists of triplets of glycines separated by hydrophobic residues, including phenylalanine (Hanukoglu and Fuchs, 1982). Thus with this assumption, the amino acid composition of the predicted portion of the 56 K keratin sequence corresponds very well to the actual amino acid analysis results (Table 1).

In the predicted sequence of the 56 K keratin, the codon frequencies of many amino acids are not random (tabulated results are not shown). In many (but not all) animal genes, codons ending in G or C are preferred (Wain-Hobson et al., 1980). The magnitude of this bias was shown to be especially pronounced for the 50 K keratin cDNA sequence (Hanukoglu and Fuchs, 1982), and our results here indicate that the codon bias for the 56 K keratin cDNA sequence is also markedly conspicuous. For example, in the 56 K keratin sequence, 28 codons for Leu end in G, only five end in C, and one ends in A or T; similarly, 22 codons for Lys end in G and only one ends in A. In the absence of bias in some genes (Hendy et al., 1981), these findings suggest that codon bias of related genes may be similar.

Table 1. Comparison of the Amino Acid Composition of the Type I 50 K and Type II 56 K Human Epidermal Keratins

| | 50 K[a] | | 56 K | |
|---|---|---|---|---|
| Amino Acid | AA Analysis | cDNA | AA[b] Analysis | cDNA[c] |
| Ala | 7.7 | 6.4 | 6.8 | 8.1 |
| Arg | 6.1 | 6.8 | 5.0 | 5.9 |
| Asn | – | 4.9 | – | 5.0 |
| Asp | – | 5.9 | – | 6.2 |
| Asn+Asp | 9.6 | 10.8 | 9.3 | 11.2 |
| Cys | ND | 0.5 | ND | 0.6 |
| Gln | – | 5.9 | – | 5.9 |
| Glu | – | 11.0 | – | 9.2 |
| Gln+Glu | 15.6 | 16.9 | 13.7 | 15.1 |
| Gly | 11.8 | 9.3 | 15.8 | 9.5 |
| His | 1.0 | 1.0 | 1.1 | 1.1 |
| Ile | 3.8 | 3.7 | 4.4 | 4.5 |
| Leu | 10.6 | 11.2 | 10.0 | 10.4 |
| Lys | 4.8 | 5.4 | 5.2 | 6.4 |
| Met | 2.4 | 3.2 | 1.1 | 2.0 |
| Phe | 3.2 | 2.4 | 4.0 | 1.4 |
| Pro | 1.3 | 0.7 | 1.5 | 0.3 |
| Ser | 9.7 | 8.3 | 10.2 | 9.5 |
| Thr | 4.8 | 4.6 | 4.2 | 4.2 |
| Trp | – | 0.5 | – | 0.3 |
| Tyr | 3.0 | 2.9 | 2.9 | 3.9 |
| Val | 4.9 | 5.4 | 4.7 | 5.6 |
| Residues Sequenced | | 410 | | 357 |

Values are presented as percentages.

[a] From Hanukoglu and Fuchs, 1982.

[b] Determined as described in Experimental Procedures.

[c] Calculated from the predicted protein sequence in Figure 2.

## Sequence Homologies and Differences between Type I and Type II Cytoskeletal Keratins

At the bottom of Figure 3, we list the sequences of the human Type II and Type I keratins and provide a comparison of their sequences (see legend). As shown, the amino acid sequences of the 56 K (Type II) and 50 K (Type I) cytoskeletal keratins could be aligned to reveal a significant but low (27%) homology over the entire central region (approximately 300 residues long) of the two proteins. To achieve optimal homology between these two sequences within this region, it was necessary to assume only a single deletion at position 256 in the 50 K keratin sequence (Figure 3). With this alignment of the two protein sequences, the corresponding coding DNA sequences showed 47% homology in this central region. When the homology between the two DNA sequences was analyzed by the method of Brutlag et al. (1982), statistically significant homologies were observed for the following segments of KA-1 and KB-2 respectively: 751–797 with 1033–1079, 760–789 with 757–786, and 943–977 with 13–47 (the numbers refer to the DNA sequences presented in Figure 2 for KA-1, and in Hanukoglu and Fuchs, 1982, for KB-2).

The ready alignment of the 50 K and 56 K keratin sequences immediately revealed that the size differences between Type I and Type II keratins result from differing lengths of both the amino and the carboxy terminal regions of the two proteins rather than from any significant insertions or deletions in their central regions (Figure 3). Remarkably, the carboxy terminal region of the 56 K keratin contains an unusual sequence that is highly similar to the glycine and serine-rich amino terminal portion of the 50 K keratin (Figure 3; Hanukoglu and Fuchs, 1982). Recently, Steinert and Roop (J. Cell. Biol. 95:228a, 1982) reported that both the amino and carboxy terminal ends of two mouse epidermal keratins (59 K and 67 K) contain tandem repeats rich in glycine and serine, which are probably similar to those found in our two human keratin sequences. Thus on the basis of the mouse keratin sequences and the fact that the percentage of glycine in our predicted partial sequence is lower than that observed by actual amino acid analysis of the protein (Table 1), we expect that the missing sequence for the amino terminal portion of the 56 K sequence will be analogous to that of the 50 K keratin in containing these unusual repeats.

## Primary and Secondary Structural Homologies among Type I and Type II Cytoskeletal and Microfibrillar Keratins

We previously showed that a segment of the human epidermal 50 K keratin shares 59% and 27% homology, respectively, with two Type I and Type II microfibrillar wool

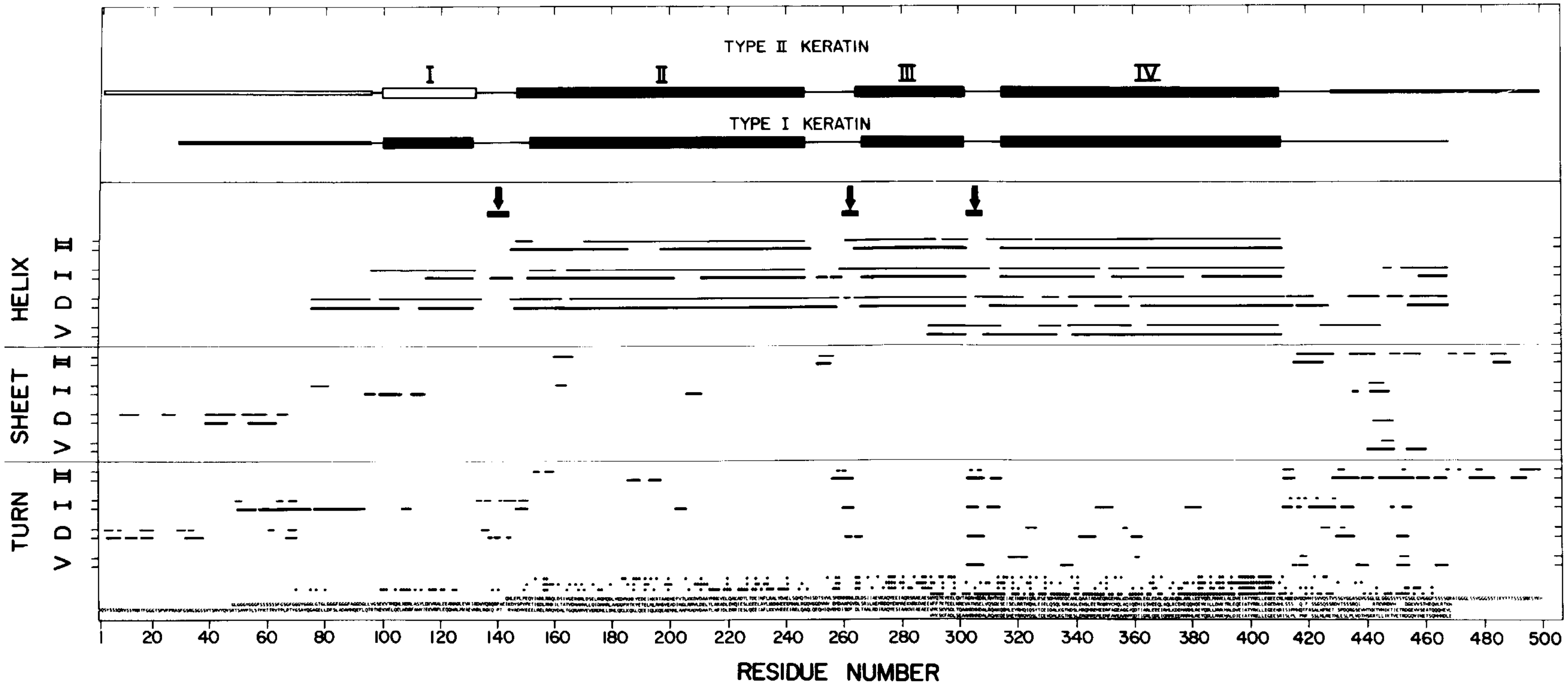

Figure 3 Comparison of the Amino Acid Sequences and Predicted Secondary Structures of 56 K Type II (II) and 50 K Type I (I) Human Epidermal Keratins, Chicken Desmin (D), and Porcine Vimentin (V)

The chicken desmin and porcine vimentin sequences are from Geisler and Weber (1981; 1982). The known amino acid sequences of the four proteins are given at the bottom of the figure and their order is the same as listed in the ordinate. The four rows of dots on top of the sequence mark the positions of homology for the following comparisons: (top to bottom) (1) I-II, (2) II-D, (3) I-D, and (4) D-V. The gaps in the sequences indicate gaps introduced by us in order to align the sequences for optimal homology. The values on the abscissa start with the first amino terminal residue of D. The sequence of D is complete whereas those for I, II, and V represent >90%, 70%, and 40% of the complete sequences of the respective proteins. The missing portions of these sequences are at their amino terminal ends. Thus for all proteins shown, the sequences extend to the last carboxy terminal residue. The Chou and Fasman (1978; 1979) and Garnier, Osguthorpe, and Robson (1978) methods used in our analyses make predictions for four conformational states: $\alpha$-helix, $\beta$-sheet, $\beta$-turn, and random coil. The predictions for the first three states are shown in the central portion of the figure. Contiguous lines mark the prediction for the corresponding region of the sequence indicated on the left. In each case, the thin lines represent the predictions of the Garnier et al. method and the thick lines the prediction of the Chou and Fasman method. The regions of the sequence which do not have any prediction indicated are predicted to be in random coil conformation. The arrows mark the positions of the $\beta$-turns that are predicted by the Chou and Fasman method for all IF protein sequences. For the Chou and Fasman method, the cutoff point for turns was taken as $1.25 \times 10^{-4}$ (Chou and Fasman, 1979). For the Garnier et al. method, the helical potential was assigned a decision constant of $-75$.

The top two lines marked as Type I Keratin and Type II Keratin indicate a summary model for the major structural domains of these proteins. The thickest bars represent the helical domains, and the bars of intermediate thickness represent the unusual glycine-rich sequence. The open bars mark the segment of the 56 K Type II keratin for which the sequence is not available (see text).

```
EK1-WK1:     ●  ●● ●●● ●  ●●● ●●   ●      ●●●● ●● ●●●●●●● ●  ●●  ●● ●●● ●●● ●●●●●●    ●●   ●●●●●●●  ●●●●●●●●    ●
EK1-WK2:            ●● ●●     ●  ●                         ●●● ●  ●●   ●     ●    ●    ●   ●●●   ● ●  ●   ●    ●●●    ●
EK1     :  IKDYSPYFKTIEDLRNKILTATVDNANVLLQIDNARLAADDFRTKYETELNLRMSVEADINGLRRVLDELTLARADLEMQIESLKEELAYLKKNHEEEMNAL
WK1     :  CPNYQSYFRTIEELQQKILCAKSENSRLVIEIDNAKLASDDFRTKYESERSLRQLVESDINSLRRILDELTLCKSNLEAEVESLKEELLCLKKNHEEEADSL
WK1-WK2:            ●● ●     ●●     ●●  ●                  ●●● ●  ●●   ●     ●    ●    ●   ● ●● ●●● ●● ●  ●   ●    ●●●    ●
WK2     :  SNLEPLFSGYIETLRREAECAEADSGRLSSELNSLQEVLEGYKKKYEEEIALRATAENEFVALKKDVDCAYLRKSDLEANVEALIQETDFLRRLYEEEIRVL
EK2     :  QNLEPLFEQYINNLRRQLDSIVGERGRLDSELRGMQDLVEDFKNKYEDEINKRTAAENEFVTLKKDVDAAYMNKVELQAKADTLTDEINFLRALYDAELSQM
EK2-WK1:            ●  ●         ●  ●●  ●         ●●  ●●● ●   ●   ●     ●    ●     ●  ● ●    ●  ●   ●      ●
EK2-WK2:   ●●●●●●  ●●  ●●●          ●●● ●●●   ●   ●  ● ●●● ●●  ●  ●●●●●● ●●●●●● ●●  ●  ● ●    ●  ●  ●●● ●●  ●
```

Figure 4. Amino Acid Sequence Homologies between the 50 K (EK1) and 56 K (EK2) Human Epidermal Keratin and Type I (WK1) and Type II (WK2) Fragments of Wool Microfibrillar Keratins

The wool keratin sequences are from Gough, Inglis, and Crewther (1978). The sequences of the human epidermal keratins represent the helical domain II and are located at position 143–244 in Figure 3. The dots mark the positions of homology between the two sequences indicated at the left side of each line. The percentage of homologous positions for the five comparisons are: (1) EK1-WK1: 59%, (2) EK1-WK2: 26%, (3) WK1-WK2: 31%, (4) EK2-WK1: 22%, and (5) EK2-WK2: 50%. In this region the homology between EK1 and EK2 is 18%.

keratin fragments (Hanukoglu and Fuchs, 1982), which in turn share 30% homology with each other (Crewther, Inglis, and McKern, 1978; Gough, Inglis, and Crewther, 1978). As predicted, the comparison in Figure 4 reveals that the 56 K cytoskeletal keratin is more homologous to Type II than to Type I wool microfibrillar keratin fragments (51% vs. 23%).

Empirical estimation of the helix content of the microfibrillar Type I and Type II keratin fragments revealed that these polypeptides are highly helical (Crewther and Dowling, 1971): The helix content of the Type I microfibrillar keratin fragment was found to be 85%, with the Type II showing somewhat less helicity. For both of the epidermal cytoskeletal keratins, the region homologous to these fragments is also predicted to be highly helical (Position 136–244 in Figure 3). Thus within the regions compared here, the amino acid differences between these proteins are still compatible with the same secondary structural conformation even though the homology between microfibrillar and cytoskeletal keratins is at most 60%.

Additional sequence information for other fragments within the helical central region of the microfibrillar keratins is also available (Crewther, Dowling, and Inglis, 1980; Dowling, Parry, and Sparrow, 1983) and an examination of these sequences indicates that the above conclusions can be extended to these other regions as well. It is important to note, however, that the amino and the carboxy terminal ends of the microfibrillar keratins do not contain the unusual tandem repeats of glycine triplets that we observed in the cytoskeletal ones. Instead, the end regions of the microfibrillar keratins are very different in their own right, being rich in cysteine, serine, and proline.

## Primary and Secondary Structural Homologies and Differences among Intermediate Filament Proteins

At present, the most substantial sequence information available is for four IF proteins: a Type II human cytoskeletal keratin (this paper); a Type I human cytoskeletal keratin (Hanukoglu and Fuchs, 1982); chicken desmin (Geisler and Weber, 1982); and porcine vimentin (Geisler and Weber, 1981; Geisler, Plessman, and Weber, 1982). In Figure 3 we present a comparison of the sequences and predicted secondarv structures of these four proteins.

The results in Figure 3 clearly extend to the Type II cytoskeletal keratins two of our earlier observations on the Type I keratins: One, the sequences of cytoskeletal keratins share overall low homology with other IF proteins; two, despite major sequence divergence of the keratins as compared to other IF proteins, they maintain a sequence that is compatible with the formation of long helical regions which appear to be common to all IF proteins.

In Figure 3 we juxtaposed the results of the secondary structure prediction analyses in order to determine what structural features are predicted to be common to all IF proteins. In a study validating the use of statistical analysis to predict the positions of β-turns in proteins, Chou and Fasman (1979a, 1979b) noted that despite divergence of primary structures of related proteins, the positions of β-turns are highly conserved. Using their procedure, we showed that very few β-turns are predicted within the 300 residue segment in the center of the 50 K keratin (Hanukoglu and Fuchs, 1982). At that time, in the absence of sequence information for the amino terminal half of desmin and vimentin, we could not determine which β-turns are conserved in all IF proteins. However, with the availability of three IF protein sequences that extend over at least 70% of the complete length of the protein, we are now able to observe a striking pattern of positions wherein β-turns are conserved (see positions marked with arrows in Figure 3). It is noteworthy that the prediction of conserved turns at these positions is not associated with conserved sequences (Figure 3). Within or adjacent to the first two of these β-turn positions, some of the sequences include a proline residue which is strongly disruptive to the propagation of helical conformation. However, the third conserved β-turn position, which separates helical regions III and IV, does not include a proline (Figure 3).

The most conspicuous feature of the predicted secondary structures of the four IF proteins is the long domains of helical conformation (Figure 3). When the three conserved β-turns noted above are taken as the demarcation points, these helical domains clearly appear in four major clusters in all IF proteins shown in Figure 3, and they are marked as I, II, III, and IV in the models for the secondary structure of Type I and Type II keratins. These four domains reveal an interesting tandemly repeated structure for these proteins, with the length of the domains being approximately 30–50, 95, 35, and 95 residues respectively. Regions II, III, and IV appear to be nearly constant in size in all IF proteins, whereas domain I may be of varying length, at least for desmin and Type I (50 K) keratin for which sequence information is available.

The size variability among different IF proteins is a result of differences in the length of the carboxy and amino terminal regions rather than insertions or deletions in the central region of the molecules. The homologous positions among all four sequences are concentrated within the predicted helical regions. The two most conserved regions include nearly the entire length of Region III and the carboxy terminal segment of Region IV (Figure 3). The homology between either Type I or Type II cytoskeletal keratins and the other IF proteins cannot be ascribed to the similar secondary structural conformations of the proteins, because this homology is reduced to nonsignificant levels (3%-6%) when the alignment of the sequences is shifted by one or more amino acid residues in either direction. As noted above for the Type I and Type II keratins, and previously for Type I keratin and other IF proteins, the sequence homology is especially low in the carboxy terminal regions and the optimal alignment of the sequences in the carboxy terminal regions requires addition of gaps in the Type I 50 K keratin sequence (Hanukoglu and Fuchs, 1982). Nonetheless, it appears that in this carboxy terminal region the Type I keratin is more related to desmin and vimentin than it is to the Type II keratins.

## Discussion

### What Are the Sequence and Structural Relations between Type I and Type II Cytoskeletal Keratins and Other IF Proteins?

Within the family of intermediate filament proteins, keratins ($M_r$ = 40–70 K) constitute the largest group, with at least 10–20 members in most mammals. As recent studies indicate, cytoskeletal keratins of all vertebrates can be grouped into at least two classes on the basis of their sequence homology deduced from cloned cDNA–mRNA hybridizations. These studies further reveal that the $M_r$ range of the keratins in the two classes are largely nonoverlapping and that they are 40–55 K and 56–70 K respectively, for the Type I and Type II keratins.

The amino acid sequence of a 50 K human epidermal keratin, which we predicted previously from a cDNA sequence, provided direct sequence evidence that a cytoskeletal keratin shares 60% homology with a Type I and 30% homology with a Type II wool microfibrillar keratin fragment (Hanukoglu and Fuchs, 1982; Crewther et al., 1978; Gough et al., 1978). In this paper we present the predicted amino acid sequence of a 56 K keratin that represents a member of the second class of cytoskeletal keratins. This sequence shows greater homology with Type II than with Type I wool microfibrillar keratin fragments. Thus we have named these two classes of cytoskeletal keratins Type I and Type II in accordance with the wool keratin nomenclature.

A comparison of the sequences and predicted secondary structures of these two cytoskeletal keratins both with one another and with other available IF protein sequences reveal certain important similarities and differences among these proteins: (1) Within the central approximately 300 residue long portion of the IF proteins (position 100–410, Figure 3), Type I and Type II keratins share a low but significant (30%) sequence homology both with each other and with the other nonkeratin IF proteins (desmin, vimentin, and neurofilament protein share with each other >70% homology; Geisler and Weber, 1982). For optimal homology, the alignment of all IF protein sequences in this region requires the addition of no or only a few (1–5) gaps. (2) Within this central region, there are four richly $\alpha$-helical domains which are demarcated from one another by three regions for which $\beta$-turns are predicted with a high degree of probability in all IF sequences (see Results). The first two of these $\beta$-turn regions contain proline(s) in some, but not all IF sequences. (3) The four helical domains (marked as I, II, III, and IV in Figure 3) are predicted to be nearly constant in size in all IF proteins and they are approximately 30-50, 95, 35, and 95 residues long respectively. (4) Although amino acid sequence homology is higher within the predicted helical domains, it is especially prominent in domain III and in the carboxy terminal end of domain IV. (5) Beyond the central conserved region, the carboxy terminal end of the 56 K Type II keratin shows no significant similarity to the carboxy terminal end of Type I keratin or to other IF proteins. In addition, it is about 30 residues longer than those of other IF proteins. Although there is significant homology among the carboxy terminal regions of the 50 K Type I keratin, desmin, and vimentin, this is less than that within the helical domains. (6) The amino (but not carboxy) terminal end of the Type I 50 K keratin contains a tandemly repeating pattern of Gly–Gly–Gly–X, (where X is an aromatic or hydrophobic amino acid) separated by stretches of serines. Not only does the carboxy terminal end of the Type II 56 K keratin contain a similar sequence, but its amino terminal end is expected to be rich in glycine (see Results).

At present, we do not know the sequences of other IF proteins. However, recently Steinert and Roop (J. Cell Biol., 95:228a, 1982) reported that they have determined the sequences of the 59 K and 67 K mouse epidermal keratins, that the sequences of these keratins are compatible with the formation of two long helical regions, and that both contain glycine-rich tandemly repeated sequences in their carboxy and amino terminal ends. Thus from their observations it appears that the sequence and structural characteristics outlined above for the cytoskeletal keratins holds true for IF keratins from different species.

The Type II cytoskeletal keratins generally show isoelectric points that are more basic than those of Type I keratins (6.5-7.5 vs. 4.5-5.5) (for a review see Moll et al., 1982). The sequence and amino acid composition differences between these two types of keratins (Table 1) do not readily explain this isoelectric point difference. Nonetheless, it is possible that this may be a result of small but important sequence or conformational differences, or alternatively, posttranslational modifications of these two groups of proteins.

### What Is the Role of the Constant Regions among IF Proteins?

Previously, many physicochemical and biochemical studies have indicated that the proteins that form the IF or the microfibrillar filaments contain two helical domains the sizes of which have been estimated to be 100–150 residues (or 15–23 nm assuming 1.5 nm axial rise per amino acid residue in an $\alpha$-helix) (Crewther and Harrap, 1967; Skerrow, Matoltsy, and Matoltsy, 1973; Fraser, MacRae, and Suzuki, 1976; Steinert, 1978; Steinert, Idler, and Goldman, 1980; Geisler, Kaufmann, and Weber, 1982). Despite the many differences noted among the various IF proteins, the ultrastructures of the filaments formed from these proteins are highly similar and display an axial periodicity of 21 nm (Kallman and Wessells, 1967; Henderson, Geisler, and Weber, 1982; Milam and Erickson, 1982). As already noted by these authors, this periodicity may reflect the unit length of the helical domains within each IF protein.

The results of our analyses on the sequences of the two human cytoskeletal keratins are, in general, consistent with these previous observations. However, our analyses, based primarily on sequence data, present a more precise model for the structure of these two proteins, one that is, in some important respects, different from those of some of the earlier studies. The models shown in Figure 3 depict four centrally located helical domains which are demarcated from one another by three conserved $\beta$-turn regions. The analyses we carried out reflect predictions for single subunits. At present we do not know whether the domains are completely helical in their native IF conformation and whether the $\beta$-turn(s) predicted in between these domains reverses the direction of or causes a loop in the polypeptide chain at these points. Studies on the microfibrillar keratins have already indicated that the helicity of major domains may be enhanced by the interactions of the proteins with one another in their polymerized states (Crewther and Dowling, 1971).

The original observations of Pauling and Corey (1953) and Crick (1953) suggested that the helical domains of protein subunits intertwine around one another to form the protofibrils that constitute the IF or the microfibrils. More recent studies indicate that the assembly of coiled-coil protofibrillar structure of IF may be dependent on hydrophobic and ionic interactions among the helical domains of the protein subunits (Fraser, MacRae, and Suzuki, 1976; Parry et al., 1977; Elleman, Crewther, and Touw, 1978; McLachlan and Stewart, 1982; Steinert et al., 1982; Dowling, Parry, and Sparrow, 1983). As an extension of studies on tropomyosin structure, it has been noted that there is a periodicity in the occurrence of nonpolar residues in the sequence of wool microfibrillar keratin fragments known to constitute a major helical domain of these proteins (Fraser, MacRae, and Suzuki, 1976; Elleman, Crewther, and Touw, 1978). Similar periodicities also occur in Type I and Type II cytoskeletal keratins (Hanukoglu and Fuchs, 1982; and this work) and desmin (Geisler and Weber, 1982). We observe that our predicted helical domains I, II, and IV correspond well with these periodicities. However, within a region of sequence that corresponds to our domain III there are irregularities in this periodicity in both microfibrillar keratins and IF proteins (see Figures 6, 2, and 1 in, respectively, Geisler, Kaufmann, and Weber, 1982; Geisler and Weber, 1982; and Dowling, Parry, and Sparrow, 1983). This does not necessarily eliminate the possibility of a coiled-coil helical structure in this region, because such a structure may be stabilized by other types of molecular interactions. In fact, using Fourier analysis, evidence has also been provided for periodicities in the distribution of acidic and basic residues in microfibrillar keratins and IF proteins indicating that electrostatic interactions may be important in IF assembly (Parry et al., 1977; McLachlan and Stewart, 1982; Dowling, Parry, and Sparrow, 1983). Consistent with these observations, an examination of our keratin sequences reveals that both hydrophobic and charged residues are clearly conserved. The general concordance of our secondary structural prediction analyses with these significant periodicities increases our confidence in the model indicated in Figure 3 for the cytoskeletal keratins and the involvement of the four predicted helical domains in coiled-coil formation in protofibrils.

### What Is the Role of the Variable Regions among IF?

In the past ten years, biochemical studies have indicated that in IF proteins the central region that is richly helical is flanked by nonhelical regions and that these may be involved in end-to-end linkage of IF proteins (Skerrow, Matoltsy, and Matoltsy, 1973; Steinert, 1978; Geisler, Kaufmann, and Weber, 1982). Indeed, analyses of the sequence of cytoskeletal keratins (Figure 3; Hanukoglu and Fuchs, 1982), and desmin (Geisler, Kaufmann, and Weber, 1982) show little or no helical conformation at the amino and carboxy terminal regions of these proteins. However, the sequence information further reveals that, unlike the central helical domains, there is no sequence or structure that is common to all IF proteins at their terminal ends. In this region, the most unusual sequence is displayed by the cytoskeletal keratins which contain the tandem repeats of glycine noted above. Although the secondary structure prediction methods we used indicate that these regions contain mostly $\beta$-turns or $\beta$-sheet (Figure 3), the presence of these unusual repeats in this region may reduce the validity of the predictions for this region (Chou and Fasman, 1974). Thus the structure of these regions and their packing within the IF remain to be elucidated and in the future, true structural information may disclose novel features at the ends of these proteins.

Unlike actin filaments or microtubules, in vitro assembly of IF from purified protein subunits can be achieved with minimal requirements by a combination of purified IF proteins (Lee and Baden, 1976; Steinert, Idler, and Zimmerman, 1976; Steinert et al., 1982). However, the conditions and efficiency of the assembly of IF varies depending on the stoichiometry and the type of protein(s) that is copolymerized. For example, some IF proteins, e.g., desmin and vimentin, can form IF by themselves, whereas the formation of filaments from keratins seems to be dependent upon the presence of at least two distinct keratin polypeptides (Lee and Baden, 1976; Steinert et al., 1982). If, as previously suggested, the terminal regions are involved in the assembly of IF proteins, then the variability in these regions may explain the specificity observed in the copolymerization of these proteins.

At present we wonder whether the two distinct types of keratins that we have characterized are the two necessary building blocks for keratin filaments. The observations that both of these keratins are present in all tissues that have keratin filaments and that both are evolutionarily conserved (Fuchs et al., 1981; Kim et al., 1983; Fuchs and Marchuk, 1983) are consistent with this possibility. This will be further tested by specific in vitro studies combining different keratin subunits. The sequence and the predicted structure of the two types of keratins do not provide an answer to this question but will ultimately be necessary in order to resolve the tertiary and supersecondary structures of these proteins. Nonetheless, the presence of unusual tandem repeats at the amino terminal of Type I and at both the amino and carboxy termini of Type II cytoskeletal keratins may be indicative of some interactions between these two types of keratins.

We have presented the sequence of only one Type II keratin, but in some mammals there are at least 10 different proteins that belong to this Type. Most likely, variations in the sequences of these proteins are important in determining the structure of the filaments and the interactions of these proteins with other cellular molecules. The recent availability of cDNA clones and in vitro procedures for the formation of modified genes and the expression of their protein products provides some tools with which the function of the variable and constant regions in IF proteins can be further dissected.

## Experimental Procedures

All materials and methods used for this study are similar to those we previously used in the sequencing and analysis of the 50 K Type I keratin cDNA sequence (Hanukoglu and Fuchs, 1982). In brief, the DNA sequence analysis was carried out according to the procedure of Maxam and Gilbert (1980). The purification and amino acid analysis of the 56 K keratin was carried out as previously described (Hanukoglu and Fuchs, 1982). The prediction of the secondary structures of the different IF proteins according to the method of Garnier, Osguthorpe, and Robson (1978) was done using a computer program we obtained from B. Robson (University of Manchester). For the prediction of the secondary structures of these proteins according to the method of Chou and Fasman (1978; 1979a), we used a program written by one of us.

## Acknowledgments

We are grateful to Naoko Tanese and Joan Harper for their very capable technical assistance. In addition we gratefully acknowledge that we used the Stanford University Molgen Project resources for DNA sequence analysis.

I. H. is the recipient of a National Research Service Award #1F32-CA07145. E. F. is a Searle Scholar and the recipient of a Research Career Development Award #1K04-AM00997. This work was supported by National Institutes of Health grant 1R01-AM27883.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received March 30, 1983

## References

Brutlag, D. L., Clayton, J., Friedland, P., and Kedes, L. H. (1982). SEQ: a nucleotide sequence analysis and recombination system. Nucl Acids Res. *10,* 279–294.

Chou, P. Y., and Fasman, G. D. (1974). Prediction of protein conformation. Biochemistry *13,* 222–245.

Chou, P. Y., and Fasman, G. D. (1978). Prediction of the secondary structure of proteins from their amino acid sequence. Adv. Enzymol. *47,* 45–148.

Chou, P. Y., and Fasman, G. D. (1979a). Prediction of β-turns. Biophys. J. *26,* 367–383.

Chou, P. Y., and Fasman, G. D. (1979b). Conservation of chain reversal regions in proteins. Biophys. J. *26,* 385–399.

Crewther, W. G., and Dowling, L. M. (1971). The preparation and properties of large peptides from the helical regions of the low-sulfur proteins of wool. Appl. Polym. Symp. *18,* 1–20.

Crewther, W. G., and Harrap, B. S. (1967). The preparation and properties a helix-rich fraction obtained by partial proteolysis of low sulfur S-carboxymethylkerateine from wool. J. Biol. Chem. *242,* 4310–4319.

Crewther, W. G., Inglis, A. S., and McKern, N. M. (1978). Amino acid sequences of α-helical segments from S-carboxymethylkerateine-A. Biochem. J. *173,* 365–371.

Crewther, W. G., Dowling, L. M., and Inglis, A. S. (1980). Amino acid sequence data from a microfibrillar protein of α-keratin. In The Structure and Chemical Reactions of Keratins. Vol. *2,* pp. 79–91.

Crick, F. H. C. (1953). The packing of α-helices: simple coiled-coils. Acta Cryst. *6,* 689–697.

Culbertson, V. B., and Freedberg, I. M. (1977). Isolation and characterization of the α-helical proteins from new born rat. Biochim. Biophys. Acta *490,* 178–191.

Dowling, L. M., Parry, D. A. D., and Sparrow, L. G. (1983). Structural homology between hard α-keratin and the intermediate filament proteins desmin and vimentin. Biosci. Rep. *3,* 73–78.

Elleman, T. C., Crewther, W. G., and Touw, J. V. D. (1978). Amino acid sequences of α-helical segments from S-carboxymethylkerateine-A: statistical analysis. Biochem. J. *173,* 387–391.

Fitzgerald, M., and Shenk, T. (1981). The sequence 5′-AAUAAA-3′ forms part of the recognition site for polyadenylation of late SV40 mRNAs. Cell *24,* 251–260.

Fraser, R. D. B., MacRae, T. P., and Rogers, G. E. (1972). Keratins: their composition, structure and biosynthesis. C. C. Thomas, Springfield, USA.

Fraser, R. D. B., MacRae, T. P., and Suzuki, E. (1976). Structure of the α-keratin microfibril. J. Mol. Biol. *108,* 435–452.

Fuchs, E., and Green, H. (1979). Multiple keratins of cultured human epidermal cells are translated from different mRNA molecules. Cell *17,* 573–582.

Fuchs, E. V., Coppock, S. M., Green, H., and Cleveland, D. W. (1981) Two distinct classes of keratin genes and their evolutionary significance. Cell *27,* 75–84.

Fuchs, E., and Marchuk, D. (1983). Type I and Type II keratins have evolved from lower eucaryotes to form the epidermal intermediate filaments in mammalian skin. Proc. Nat. Acad. Sci. USA, in press.

Garnier, J., Osguthorpe, D. J., and Robson, B. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. J. Mol. Biol. *120,* 97–120.

Geisler, N., and Weber, K. (1981). Comparison of the proteins of two immunologically distinct intermediate-sized filaments by amino acid sequence analysis: desmin and vimentin. Proc. Nat. Acad. Sci. USA *78,* 4120–4123.

Geisler, N., Kaufmann, E., and Weber, K. (1982). Proteinchemical characterization of three structurally distinct domains along the protofilament unit of desmin 10 nm filaments. Cell *30,* 277–286.

Geisler, N., Plessmann, U., and Weber, K. (1982). Related amino acid sequences in neurofilaments and non-neuronal intermediate filaments. Nature *296,* 448–450.

Geisler, N., and Weber, K. (1982). The amino acid sequence of chicken muscle desmin provides a common structural model for intermediate filament proteins. EMBO J. *1,* 1649–1656.

Gough, K. H., Inglis, A. S., and Crewther, W. G. (1978). Amino acids sequences of $\alpha$-helical segments from S-carboxymethylkerateine-A. Biochem. J. *173,* 373–385.

Hanukoglu, I., and 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.

Henderson, D., Geisler, N., and Weber, K. (1982). A periodic ultrastructure in intermediate filaments. J. Mol. Biol. *155,* 173–176.

Hendy, G. N., Kronenberg, H. M., Potts, J. T., and Rich, A. (1981). Nucleotide sequence of cloned cDNAs encoding human preproparathyroid hormone. Proc. Nat. Acad. Sci. USA *78,* 7365–7369.

Hong, B.-S., and Davison, P. F. (1981). Isolation and characterization of a soluble, immunoactive peptide of glial fibrillary acidic protein. Biochim. Biophys. Acta *670,* 139–145.

Jones, L. N. (1975). The isolation and characterization of $\alpha$-keratin microfibrils. Biochim. Biophys. Acta. *412,* 91–98.

Kallman, F., and Wessells, N. K. (1967). Periodic repeat units of epithelial cell tonofilaments. J. Cell Biol. *32,* 227–231.

Kim, K.-H., Rheinwald, J. G., and Fuchs, E. (1983). Tissue specificity of epithelial keratins: differential expression of mRNAs from two multigene families. Mol. Cell. Biol. *3,* in press.

Lazarides, E. (1982). Intermediate filaments: a chemically heterogeneous, developmentally regulated class of proteins. Ann. Rev. Biochem. *51,* 219–250.

Lee, L. D., and Baden, H. P. (1976). Organisation of the polypeptide chains in mammalian keratin. Nature *264,* 377–379.

McLachlan, A. D. (1978). Coiled coil formation and sequence regularities in the helical regions of $\alpha$-keratin. J. Mol. Biol. *124,* 297–304.

McLachlan, A. D., and Stewart, M. (1982). Periodic charge distribution in the intermediate filament proteins desmin and vimentin. J. Mol. Biol. *162,* 693–698.

Milam, L., and Erickson, H. P. (1982). Visualization of a 21-nm axial periodicity in shadowed keratin filaments and neurofilaments. J. Cell Biol. *94,* 592–596.

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

Parry, D. A. D., Crewther, W. G., Fraser, R. D., and MacRae, T. P. (1977). Structure of $\alpha$-keratin: structural implications of the amino acid sequence of the Type I and Type II chain segments. J. Mol. Biol. *113,* 449–454.

Pauling, L., and Corey, R. B. (1953). Compound helical configurations of polypeptide chains: structure of proteins of the $\alpha$-keratin type. Nature *171,* 59–61.

Roop, D. R., Hawley-Nelson, P., Cheng, C. K., and Yuspa, S. H. (1983). Keratin gene expression in mouse epidermis and cultured epidermal cells. Proc. Nat. Acad. Sci. USA *80,* 716–720.

Skerrow, D., Matoltsy, G., and Matoltsy, M. (1973). Isolation and characterization of the helical regions of epidermal prekeratin. J. Biol. Chem. *248,* 4820–4826.

Steinert, P. M., and Idler, W. W. (1975). The polypeptide composition of bovine epidermal $\alpha$-keratin. Biochem. J. *151,* 603–614.

Steinert, P. M., Idler, W. W., and Zimmerman, S. B. (1976). Self-assembly of bovine epidermal keratin filaments in vitro. J. Mol. Biol. *108,* 547–567.

Steinert, P. M. (1978). Structure of the three-chain unit of the bovine epidermal keratin filament. J. Mol. Biol. *123,* 49–70.

Steinert, P. M., Idler, W. W., and Goldman, R. D. (1980). Intermediate filaments of baby hamster kidney (BHK-21) cells and bovine epidermal keratinocytes have similar ultrastructures and subunit domain structures. Proc. Nat. Acad. Sci. USA *77,* 4534–4538.

Steinert, P., Idler, W., Aynardi-Whitman, M., Zackroff, R., and Goldman, R. D. (1982). Heterogeneity of intermediate filaments assembled in vitro. Cold Spring Harbor Symp. Quant. Biol. *46,* 465–474.

Sun, T.-T., Shih, C., and Green, H. (1979). Keratin cytoskeletons in epithelial cells of internal organs. Proc. Nat. Acad. Sci. USA *76,* 2813–2817.

Wain-Hobson, S., Nussinov, R., Brown, R. J., and Sussman, J. L. (1981). Preferential codon usage in genes. Gene *13,* 355–364.

Weber, K., Osborn, M., and Franke, W. W. (1980). Antibodies against merokeratin from sheep wool decorate cytokeratin filaments in non-keratinizing epithelial cells. Eur. J. Cell Biol. *23,* 110–114.

## Note Added in Proof

The complete amino acid sequence of a 59 K epidermal keratin sequence has been recently reported (Steinert, P. M., Rice, R. H., Roop, D. R., Trus, B. L., and Steven, A. C. Nature *302,* 794–800, 1983). Although the size of this mouse keratin (59 K) is larger than most Type I keratins ($M_r$ range 40–55 K), its amino acid sequence unambiguously identifies it as a Type I rather than a Type II keratin. Steinert et al. also present a predicted structure for this keratin wherein the general location of the helical domains are highly consistent with our previous analyses (Hanukoglu and Fuchs, 1982).