*Proc. Natl. Acad. Sci. USA*
Vol. 81, pp. 5709–5713, September 1984
Biochemistry

# 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

(cDNA clones/amino acid sequences/protein structure predictions)

PETER M. STEINERT*, DAVID A. D. PARRY†, ESTHER L. RACOOSIN*, WILLIAM W. IDLER*, ALASDAIR C. STEVEN‡, BENES L. TRUS§, AND DENNIS R. ROOP¶

*Dermatology Branch, Building 10, Room 12N238, National Cancer Institute; ‡Laboratory of Physical Biology, National Institute of Arthritis, Diabetes, and Digestive and Kidney Diseases; §Computer Systems Laboratory, Division of Computer Research and Technology; ¶Laboratory of Cellular Carcinogenesis and Tumor Promotion, National Cancer Institute, National Institutes of Health, Bethesda, MD 20205; and †Department of Chemistry, Biochemistry and Biophysics, Massey University, Palmerston North, New Zealand

*Communicated by W. A. Hagins, June 11, 1984*

ABSTRACT **We present the complete nucleotide and deduced amino acid sequences of a mouse epidermal keratin subunit of 60,000 Da. The keratin possesses a central α-helical domain of four tracts (termed 1A, 1B, 2A, and 2B) that can form coiled-coils, interspersed by short linker sequences, and has non-α-helical terminal domains. This pattern of secondary structure is emerging as common to all intermediate filament subunits. The α-helical sequences conform to the type II class of keratins. Accordingly, this is the first type II keratin for which complete sequence information is available, and thus it facilitates elucidation of the fundamental distinctions between type I and type II keratins. It has been observed that type I keratins are acidic and type II keratins are neutral–basic in charge. We suggest that the basis for this empirical correlation between type and charge resides in the respective net charges of the 1A and 2B tracts. Calculations on interchain interactions between charged residues in the α-helical domains indicate that this keratin prefers to participate in dimers according to an in-register parallel arrangement. The terminal domains of this keratin possess characteristic glycine-rich sequences, and the carboxyl-terminal domain is highly homologous to that of a human epidermal keratin of 56,000 Da. According to the hypothesis that end-domains are located on the periphery of keratin filaments, we conclude that the corresponding mouse and human keratins are closely related, both structurally and functionally.**

Keratins constitute the intermediate filaments (IF) of epithelial cells. Based on detailed one- and two-dimensional gel electrophoretic analysis of various human epithelial tissues, a total of ≈19 different subunits have been identified, including those of specialized epithelia such as the epidermis (1), but when the keratin IF subunits of other epidermal derivatives, such as hair, are included, the total number may approach 30. Their molecular sizes vary within the range of 40–70 kDa. Like all other IF subunits, keratins contain extended sequences that favor the formation of α-helices, bounded on each end by sequences that are not α-helical (2–7). Interestingly, when resolved by two-dimensional gel electrophoresis, the keratin IF subunits fall into distinct acidic (pI, 4.5–5.5) and neutral–basic (pI, 6.5–7.5) groups. These analyses also show that any one epithelial cell-type expresses only a few of these subunits, and usually a coordinated pair or group containing at least one acidic and one neutral–basic subunit (1, 8). Two other recent types of experiments support this idea of the coordinated expression of pairs of keratins in epithelial cells. In the first, two monoclonal antibodies termed AE1 and AE3 have been described that, respectively, recognize acidic and neutral–basic keratins (8, 9). Second, recombinant cDNA probes to one acidic and one basic human epidermal keratin have been used to demonstrate the presence of these two classes of keratins throughout vertebrates (10). These observations are all consistent with earlier findings that at least two different keratin subunits are required for keratin IF assembly (11–13).

Amino acid sequence analyses of keratin subunits have also revealed the likelihood of two distinct keratin classes. Peptides derived from the α-helical regions of two wool keratins were termed type I and type II, because their sequences were similar but not identical (14). Analyses of the α-helical regions identified from sequencing the cDNA clones for the human acidic 50 kDa (4) and basic 56 kDa (5) keratins were similar to each other, but were respectively more similar to the wool type I and type II keratins (5). Thus, the idea has arisen that type I keratins are acidic and type II keratins are neutral–basic (5, 10).

The structural and functional significance of the apparently coincident expression in epithelial cells of representatives of these two types of keratins remains unclear. For example, what, if any, is the role of charges on the non-α-helical terminal domains of the subunits in specifying the type of helical sequence? Are there specific regions on the α-helical sequences that might correlate with the helical type and subunit charge differences? In addition, how do similar or dissimilar keratin subunits combine sterically to form the proposed two-chain coiled-coil building blocks of keratin IF? To explore these and related questions, we have determined the complete amino acid sequence of a mouse epidermal keratin of 60 kDa. We have carried out quantitative comparative analyses of this sequence with those available for other keratins. We show that its helical sequences are very closely related to the wool type II helical sequence, and that there is indeed a close correlation between the basic charge of this subunit and the type and charge of its helical sequences.

## EXPERIMENTAL

**Source of cDNA Clones.** Normal newborn mouse epidermis expresses about six different keratins. Cells committed to terminal differentiation express two major keratins, an acidic 59 kDa and a basic 67 kDa keratin (8, 15). Basal cells, which are normally only a minor component of the total cell population, and cells that can be propagated in cell culture, synthesize four keratin subunits, two neutral–basic keratins of 60 and 59 kDa, and two acidic keratins of 55 and 50 kDa

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

Abbreviation: IF, intermediate filaments.

0.2 kbp

Smal I, Sac I, Pvu II, Nci I, Taq I, Pvu II, Sac I, Nci I, Pst I, Hha I, Hae III, Pst I, Ava II, Nci I, Nco I, Ava II, Pst I, Dde I

*pK* 3.27
*pK* 3.68
*pK* 13.27
*pK* 5.23

FIG. 1. DNA sequencing strategy. Top line represents the full-length mRNA [2005 base pairs long, minus the poly(A) tract], with the restriction endonuclease sites used in sequencing. Below are shown the regions encoded by the four clones identified by hybrid selection analyses. The arrows delineate the beginning and end of the coding region.

(unpublished data). cDNA libraries have been prepared from poly(A) mRNA of terminally differentiating cells (15) and of basal cells grown in culture (unpublished data). From this library, four clones were identified by hybrid selection analyses to encode the 60 kDa keratin, by using procedures described (15).

**DNA Sequence Analysis.** The inserts of the four clones used here were isolated as their *Pst* I pieces. Each *Pst* I piece, or a smaller fragment derived from it, was labeled with [$^{32}$P]phosphate at either its 5′ end (16) or its 3′ end with cordycepin (17), and then sequenced by use of Maxam–Gilbert (16) procedures, except that 5% (vol/vol) formic acid was substituted for 1 M piperidine formate in the adenine + guanine reaction.

**Computer Analyses of Amino Acid Sequences.** Fourier transform analyses were carried out as described (6, 18–20). Sequence homologies were determined using the methods described (20, 21).

```
                                                  -52                                         -1
                                                  5'-GTCTGCCCTGCCGTTTCTCTACTTCCCAGCCTTCTCATCTCCAGGAACCATG

1                                                                                                                                                      50
S  T  K  T  T  I  K  S  Q  T  S  H  R  G  Y  S  A  S  S  A  R  V  L  G  L  N  R  S  G  F  S  S  V  S  V  C  R  S  R  G  S  R  G  S  S  A  M  C  G  G
TCTACCAAAACCACCATCAAAAGTCAAACCAGCCACCGTGGCTACAGTGCCAGCTCAGCCAGAGTGCTTGGACTCAACCGCTCGGGCTTCAGCAGTGTGTCCGTGTGCCGCTCCCGGGGCAGCGGTGGCTCCAGTGCAATGTGTGGAGGA
                                                                              100
                                                                                                                                                      100
A  G  F  G  S  R  S  L  Y  G  V  G  S  S  K  R  I  S  I  G  G  G  S  C  G  I  G  G  G  Y  G  S  R  F  G  G  S  F  G  I  G  G  G  A  G  S  G  F  G  F
GCTGGCTTTGGCAGCAGGAGCCTCTATGGTGTGGGGAGCTCCAAGAGGATCTCCATCGGAGGGGGCAGCTGTGGCATTGGAGGAGGCTATGGCAGCCGATTCGGAGGAAGCTTCGGCATTGGTGGTGGAGCTGGTAGTGGCTTTGGCTTC
                                      200                                                                                                             300
                                                                                                                                                      150
G  G  G  A  G  F  G  G  G  Y  G  G  A  G  F  P  V  C  P  L  G  G  I  Q  E  V  T  I  N  Q  S  L  L  T  P  L  N  L  Q  I  D  P  T  I  Q  R  V  R  T  E
GGTGGTGGAGCTGGCTTTGGTGGTGGCTATGGGGGAGCTGGCTTCCCGGTGTGCCCACTTGGAGGCATCCAAGAGGTCACCATCAACCAGAGCCTCCTCACACCCCTGAACCTGCAAATTGACCCCACCATCCAGCGGGTCAGGACTGAG
                                                                              400
                                                                                                                                                      200
E  R  E  Q  I  K  T  L  N  N  K  F  A  S  F  I  D  K  V  R  F  M  E  Q  Q  N  K  V  M  D  T  K  W  A  L  L  Q  E  Q  D  T  K  T  V  R  Q  N  M  E  P
GAGAGGGAGCAGATCAAGACCCTCAATAACAAGTTTGCCTCCTTCATCGACAAGGTGCGGTTCATGGAGCAGCAGAACAAGGTCATGGACACCAAGTGGGCCCTGCTGCAAGAGCAGGACACCAAGACCGTGAGGCAGAACATGGAGCCC
                                      500                                                                                                             600
                                                                                                                                                      250
M  F  E  Q  Y  I  S  N  L  R  R  Q  L  D  S  I  I  G  E  R  G  R  M  N  S  E  L  R  N  M  Q  E  L  V  E  E  L  R  N  K  Y  E  D  E  I  N  K  R  T  D
ATGTTTGAGCAGTACATCAGCAACCTCCGCAGACAGCTGGACAGCATCATTGGAGAGAGGGGTCGCATGAACTCAGAGCTGAGGAACATGCAGGAACTCGTAGAAGAACTACGGAACAAATATGAAGATGAAATCAACAAGCGCACAGAC
                                                                              700
                                                                                                                                                      300
A  E  N  E  F  V  T  L  K  K  D  V  D  A  A  Y  M  N  K  V  E  L  Q  A  K  A  D  S  L  T  D  D  I  N  F  L  R  A  L  Y  E  A  E  L  S  Q  M  Q  T  H
GCAGAGAATGAATTCGTGACCCTGAAGAAGGATGTAGATGCTGCCTACATGAACAAAGTTGAACTGCAAGCCAAGGCAGACAGTCTAACAGATGATATCAACTTCTTGAGAGCTCTCTATGAAGCAGAACTGTCTCAGATGCAAACTCAC
                                      800                                                                                                             900
                                                                                                                                                      350
I  S  D  T  S  V  V  L  S  M  V  N  N  R  S  L  V  L  D  S  I  I  A  E  V  K  A  Q  F  E  V  I  A  Q  R  S  R  A  E  A  E  S  L  Y  Q  T  K  Y  E  E
ATCTCAGACACATCTGTGGTCCTCTCCATGGTCAACAACCGTAGCCTCGTCCTAGACAGCATCATCGCTGAGGTCAAGGCCCAGTTTGAGGTCATAGCTCAGAGAAGTCGGGCTGAAGCTGAGTCATTGTACCAGACTAAATATGAGGAG
                                                                              1000
                                                                                                                                                      400
L  Q  V  T  A  G  R  H  G  D  D  L  R  N  T  K  Q  E  I  A  E  I  N  R  M  I  Q  R  L  R  S  E  I  D  H  V  K  K  Q  C  A  N  L  Q  A  A  I  A  D  A
CTGCAGGTCACAGCTGGCAGACATGGGGACGACCTGCGCAACACCAAGCAGGAGATTGCTGAGATCAACCGCATGATCCAGAGGCTGAGATCTGAGATCGACCACGTTAAGAAGCAGTGTGCCAACCTGCAAGCTGCTATTGCTGATGCT
                                      1100                                                                                                            1200
                                                                                                                                                      450
E  Q  R  G  E  M  A  L  K  D  A  R  G  K  L  E  G  L  E  D  A  L  Q  K  A  K  Q  D  M  A  M  L  L  K  E  Y  H  E  L  M  N  V  K  L  A  L  D  V  E  I
GAGCAACGTGGGGAGATGGCCCTGAAGGATGCAAGGGGCAAGCTGGAAGGGCTGGAGGATGCCCTGCAGAAGGCCAAACAGGACATGGCCATGCTGCTGAAGGAGTACCATGAACTCATGAATGTCAAGCTGGCCCTTGATGTGGAAATT
                                                                              1300
                                                                                                                                                      500
A  T  Y  R  K  L  L  E  G  E  E  C  R  L  N  G  E  G  V  G  P  V  N  I  S  V  V  Q  S  T  V  S  S  G  Y  G  S  A  G  G  A  S  S  S  L  G  M  G  G  G
GCCACCTACAGGAAGCTGCTGGAAGGAGAGGAGTGCAGGTTGAATGGTGAAGGTGTTGGACCAGTCAACATCTCTGTGGTGCAGTCCACCGTGTCCAGCGGCTATGGCAGTGCCGGGGGTGCCAGCAGCAGCTTAGGCATGGGTGGAGGC
                                      1400                                                                                                            1500
                                                                                                                                                      550
S  S  Y  S  T  S  S  S  H  G  L  G  G  G  F  S  A  G  S  G  R  A  I  G  G  G  L  S  S  S  G  G  L  S  S  S  T  I  K  Y  T  T  T  S  S  S  K  K  S  Y
AGCAGCTACTCCTATAGCAGCAGCCATGGCCTTGGAGGTGGCTTCAGTGCTGGCAGTGGCAGAGCCATCGGAGGTGGCCTCAGCTCTTCTGGTGGCCTCAGCTCTTCTACCATCAAATACACCACCACCTCCTCCAGCAAGAAGAGCTAC
                                                                              1600

R  Q  .
AGGCAGTGAATTCTGTCACCAAGAGCTTGTCTCTGGTCCCAGATGTCATGGCTGCAGCTGAACCACATGCTTTGGTTCCCGGAAGGGAACGAATCCCAACCTCTGGCCTCCCCATGGCTCAGTTCTACATTTGTGTTGCACGTCAGCACC
                                      1700                                                                                                            1800

TATACATGTTCTTTGGTGACCCAGACCCCAAAATGTTGCAGAATGTAGACCTCCAAGACGAAACCCCAAACCCTACCCAGAATACCCACCTAAATTCTGTCATGGTTCTGACTTCCTCCAGAGTCTGTAAAATAAAATGCCCCCACAACA
                                                                              1900

AAC(A18).
```

FIG. 2. Complete nucleotide and deduced amino acid sequence. Nucleotides are numbered from the first base following the likely initiating codon. The poly(dC)·poly(dG) tracts derived in cloning are not included. The likely initiating codon and polyadenylylation signal sequence are underlined. The deduced amino acid sequence is given above the nucleotide sequence, using the single-letter code, and is numbered from the serine residue following the initiating codon, because *N*-acetylserine was previously identified as the amino-terminal residue of the $K_2$ group of proteins (12).

## RESULTS AND DISCUSSION

**DNA Sequence and Deduced Amino Acid Sequence of cDNA Clones.** Preliminary restriction enzyme maps (Fig. 1) revealed that the four clones contain overlapping sequences that together encode the entire mRNA for the 60-kDa keratin. The complete DNA sequence was determined (Fig. 2). This sequence includes the complete 3′ end of the mRNA, owing to the presence of an 18-nucleotide-long poly(A) tract and a polyadenylylation signal sequence of A-A-T-A-A-A (22) that appears 22 nucleotides before the poly(A) region. The sequence reveals a putative initiating codon near the 5′ end that corresponds to the only open frame.

The deduced amino acid sequence shown in Fig. 2 (using the single-letter amino acid code) contains a total of 552 amino acids, giving a calculated molecular size of 59,486 Da. While the phosphate content of this protein is unknown, the larger basal epidermal cell $K_2$ keratins, of which this is one (12), contain ≈5 mol of *o*-phosphoserine per mol (23). Thus, the actual molecular size of the protein is 60,000 Da, as estimated by $NaDodSO_4$ gel electrophoresis. The protein has a net charge of +7 (excluding phosphate).

**Structural Organization of the Subunit.** Preliminary secondary structural analyses by the methods of Chou and Fasman (24) and Garnier *et al.* (25) reveal that this keratin IF subunit, like others so far examined (20), contains a central domain ≈315 residues long (residues 152–464) that favors the formation of four extended tracts of α-helix, separated by short regions of less regular structure. This central domain is bounded on the amino- and carboxyl-terminal ends by regions that are likely to adopt non-α-helical conformations. Analysis by the Fourier transform method (18) confirms such an arrangement, and in addition, detects prominent nonrandom repeats in several types of amino acids in certain subsequences (Table 1), which provide more detailed information on the secondary and supersecondary structure of this protein.

**Structural Order of the Central α-Helical Domain.** The Fourier transform periodicities of 2.3 (7/3) and 3.5 (7/2) in apolar amino acid residues (Table 1) are harmonics of the 7-residue (heptad) repeat, which is characteristic of α-type proteins, that can adopt a coiled-coil conformation. Residues with apolar side chains are usually located in the *a* and *d* heptad positions and form a closely packed hydrophobic interface between the chains of a coiled-coil, whereas polar and charged residues are usually located in the other 5 heptad positions and project radially from the coiled-coil where they are involved in interchain and inter-coiled-coil interactions (see below) (18, 19). By identifying those regions of the sequence that exhibit these periodicities, it is possible to predict which residues participate in the coiled-coils. Fig. 3 shows the arrangement of four coiled-coil tracts (1A, 1B, 2A, and 2B) separated by non-coiled-coil linkers L1, L12, and L2. The specification of the amino acids that enter and leave the coiled-coils in this way substantiates the concept that the sizes and relative locations of the coiled-coil and non-coiled-coil sequences have been highly conserved among different IF subunits (4–7). The distributions of the

Table 1. Fourier transform analysis of most significant residue repeats

| Subsequence and residue type | Periodicity | Intensity, $-\log_{10}$ |
|---|---|---|
| Central α-helical domain | | |
| 199–299 (tract 1B) | | |
| Apolar (AVMILFY) | 3.51 | 4.51 |
| | 2.32 | 2.05 |
| Basic (HKR) | 9.57 | 3.31 |
| Acidic (DE) | 9.62 | 2.24 |
| 344–464 (tract 2B) | | |
| Apolar | 3.52 | 5.32 |
| | 2.22 | 1.78 |
| Basic | 9.77 | 2.42 |
| Acidic | 9.85 | 2.66 |
| Amino-terminal domain | | |
| 1–114 | | |
| Apolar | 2.04 | 3.04 |
| | 2.60 | 2.38 |
| Glycine | 2.07 | 1.83 |
| Carboxyl-terminal domain | | |
| 480–552 | | |
| Apolar | 2.03 | 2.26 |
| Glycine | 2.00 | 2.50 |

Preliminary analyses of the complete sequence gave strong indications of these repeats, but in more refined analyses, the strongest signals were associated with the specific sequence regions tabulated. The probability that the observed repeats are not random statistical fluctuations is related to the calculated intensity by $P = 10^{-I}$.

acidic and basic residues exhibit repeats of 9.5–9.6 in tract 1B, which is statistically significantly different from the repeat of ≈9.8 in tract 2B (20). This protein also displays a stutter in the progression of heptads near the middle of tract 2B. This feature presumably introduces a kink in the regularity of this coiled-coil and represents a structural feature that has been strictly conserved among all IF subunits (6).

A detailed analysis of the relatedness of the coiled-coil α-helical sequences of this keratin to the available sequences of other keratins is shown in Table 2. This procedure affords a quantitative estimate of the probability that any resemblance between two sequences under comparison did not arise by chance. This provides a more quantitative view of the relatedness of the sequences than the simple empirical comparisons done earlier (4–6). It is clear from Table 2 that the 60-kDa keratin is unequivocally more closely related to the other type II keratins, wool 7c and human 56 kDa, than to type I keratins. The lower degrees of homology observed between the type I and type II groups of Table 2 may reflect, in part, the occurrence in both of the heptad repeats and the similar charge distributions. Based on our hybrid selection analyses using human poly(A) mRNA and the present mouse cDNAs, the 60-kDa mouse keratin is the equivalent of the human 58-kDa keratin (unpublished data). Thus, this keratin is of the neutral–basic charge group and corresponds to protein 5 of the human cytokeratin catalog (1).

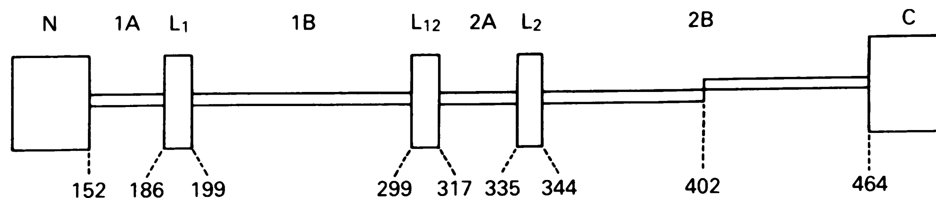


FIG. 3. Amino acid sequence arrangement and predicted secondary structure. N and C are, respectively, the amino- and carboxyl-terminal non-α-helical end domains of the subunit. The rods represent the four coiled-coil tracts (1A, 1B, 2A, and 2B) separated at the non-coiled-coil linkers L1, L12, and L2, by the residue numbers shown. The discontinuity on tract 2B indicates the position of the heptad polarity reversal.

Table 2. Homologies of coiled-coil tracts of 60-kDa keratin with similar tracts of other keratin subunits

| Protein (type) | Coiled-coil tract | Probability of unrelatedness to 60-kDa keratin, $-\log_{10}$ | Ref. |
|---|---|---|---|
| Mouse, 59-kDa (I) | 1A | 3.5 | 6 |
| Human, 50-kDa (I) | 1A | 2.7 | 4 |
| Wool, 8c (I) | 1A | 2.9 | 19 |
| Mouse, 59-kDa (I) | 1B | 2.5 | |
| Human, 50-kDa (I) | 1B | 3.1 | |
| Human, 56-kDa (II) | 1B | >38 | 5 |
| Wool, 8c (I) | 1B | 5.6 | |
| Wool, 7c (II) | 1B | 17 | 26 |
| Mouse, 59-kDa (I) | 2B | 3.2 | |
| Human, 50-kDa (I) | 2B | 8.8 | |
| Human, 56-kDa (II) | 2B | >38 | |
| Wool, 8c (I) | 2B | 6.5 | |
| Wool, 7c (II) | 2B | >38 | |

**Structural Order of the Terminal Non-α-Helical Domains.** At least three structural forms are evident in each domain. Immediately adjacent to the central α-helical domain are sequences (residues 116–151 and 465–483) that favor the formation of several β-turns; these portions are, therefore, likely to adopt a globular conformation. Adjacent to these are sequences enriched in glycines and hydrophobic residues (approximately residues 47–115 and 484–500) with a pronounced periodicity of ≈2 residues (Table 1) and that are reminiscent of the glycine-rich sections previously reported for the mouse type I 59-kDa keratin (6). The frequent occurrence of consecutive glycines presumably contributes to a highly convoluted and flexible structure. Unlike the 59-kDa keratin, however, the 60-kDa keratin lacks the defined tandem peptide repeats in these glycine-rich sequences. The third distinct region of these domains, which encompasses the 40–50 amino acids at either end of the subunit, is enriched in serine residues and has a strongly basic charge. Overall, the type I 59-kDa and type II 60-kDa keratins exhibit considerable inter-type homology in the amino-terminal but not in the carboxyl-terminal domains (the probability of the former occurring independently is only $2.2 \times 10^{-32}$).

The entire carboxyl-terminal domain of the subunit is identical in length and is highly homologous ($P = 2.3 \times 10^{-18}$) to the type II human 56-kDa keratin (5). The apparent difference in mass between the mouse 60-kDa and the human 56-kDa keratins presumably resides in the size of their amino-terminal domains. Previous studies have suggested that much of the end-domains of the subunits occupy peripheral positions on the keratin IF (6), from which it follows that the structural features of these domains prescribe the function of the keratin IF in cells. Since the mouse 60-kDa and the human 56-kDa keratins represent the type II keratins expressed in basal cells of their respective species, we conclude that their functions must be very similar or complementary. It remains to be seen whether the sequences of the two type I keratins expressed in basal cells are also similar.

Table 4. Interchain ionic interactions of subunits in a two-chain coiled-coil as a function of heptad stagger

| Heptad stagger | 60(II) 60(II) | 60(II) 56(II) | 60(II) 59(I) | 60(II) 50(I) | 56(II) 50(I) | 59(I) 59(I) |
|---|---|---|---|---|---|---|
| | Parallel subunits | | | | | |
| −2 | 3 | 2 | 1 | 2 | 1 | 2 |
| −1 | 7 | 4 | 5 | 8 | 6 | 6 |
| 0 | 14 | 11 | 13 | 11 | 8 | 10 |
| +1 | 7 | 3 | 8 | 4 | 1 | 6 |
| +2 | 3 | 2 | 0 | 5 | 4 | 2 |
| | Antiparallel subunits | | | | | |
| −2 | 9 | 7 | 5 | 4 | 2 | 1 |
| −1 | −14 | −14 | −7 | −9 | −10 | −7 |
| 0 | 0 | 2 | −3 | −2 | 1 | 3 |
| +1 | −10 | −7 | −9 | −10 | −7 | −8 |
| +2 | −10 | −6 | −9 | −9 | −5 | −12 |

Column headers give the subunit pair (type). Numbers of ionic interactions that result when the two subunits are moved along each other in full heptad (7 residue) steps are listed. Interactions of unlike charges (attraction) contribute +1, while like charges (repulsion) contribute −1 to the total scores.

**Distribution of Ionic Charges Along Keratin Subunits.** Table 3 lists the numbers of charged amino acid residues in each domain of the 6 keratins for which sequence information is currently available. Overall, type I keratins are acidic and type II keratins are basic, in concurrence with the two-dimensional gel electrophoretic data (1, 8). The amino- and carboxyl-terminal non-α-helical domains are all strongly basic, as seems to be the case for other IF subunits (27, 28). In type I keratins, the 4 coiled-coil tracts are all acidic. In type II keratins, tracts 1B and 2A are also acidic, but interestingly, tract 2B is neutral and tract 1A is basic. The tract 2B contains 8–10 more basic residues than the other tracts. Thus, even though the terminal domains are generally strongly basic, we predict that the net basic charge of type II keratins is largely due to these extra basic charges in tract 2B. Moreover, we have found that the last 30 amino acids of tract 2B (E-Y-Q-E-L-L-N-V-K-L-A-L-D-E-I-A-T-Y-R-E-L-L-E-G-E-E-C-R-L) of net charge −2 have been highly conserved between type I and type II keratins, and indeed all IF subunits so far sequenced. Thus, the remaining sequences of tract 2B of type II keratins are in fact basic in charge in comparison to type I keratins. These observations offer an explanation for the correlation of helical type and net subunit charge.

**Interactions Between Subunits to Form Two-Chain Coiled-Coils.** Several studies involving cross-linking reagents (27, 29), recombinations of isolated helical sequences (30), and analyses of available sequence (21) and x-ray diffraction (31) data, have suggested that the coiled-coil unit of keratin and other IF consists of two chains rather than three chains as

Table 3. Distribution of ionic charges along domains and subdomains of different keratin IF subunits

| Protein (type) | Amino-terminal domain | Central α-helical domain: 1A | L1 | 1B | L12 | 2A | L2 | 2B | Carboxyl-terminal domain |
|---|---|---|---|---|---|---|---|---|---|
| Wool, 8c (I) | NK | 8-7 | 1-0 | 21-14 | 3-1 | 4-2 | 3-2 | 21-13 | NK |
| Mouse, 59-kDa (I) | 1-5 | 7-7 | 3-5 | 20-12 | 2-0 | 3-2 | 3-2 | 20-12 | 1-8 |
| Human, 50-kDa (I) | NK | 7-6 | 2-3 | 21-14 | 3-0 | 5-4 | 3-2 | 23-15 | 4-9 |
| Wool, 7c (II) | NK | NK | 1-1 | 25-15 | 3-2 | 5-2 | 2-1 | 21-21 | NK |
| Mouse, 60-kDa (II) | 4-14 | 4-7 | 2-2 | 22-14 | 1-2 | 3-2 | 2-1 | 23-23 | 1-6 |
| Human, 56-kDa (II) | NK | NK | NK | 22-14 | 2-1 | 5-2 | 2-1 | 23-23 | 1-6 |

In each pair of numbers, the first is the number of acidic residues (D + E) and the second is the number of basic residues (H + K + R). NK, not known due to unavailable or incomplete sequence information.

previously suggested (2, 3). The availability of complete $\alpha$-helical domain sequences permits predictions on how two subunits may interact to form the coiled-coil (32, 33). Such models are based on calculations of the numbers of ionic interactions that are possible in the coiled-coil tracts when two subunits are brought together in a parallel or antiparallel manner and staggered by various amounts. In Table 4, it is seen that the in-register arrangement of parallel subunits always results in the largest numbers of favorable ionic interactions. Analysis by this method, however, suggests that two type I or two type II subunits interact with each other as effectively as combinations of one type I and one type II subunit. Therefore, the question of the subunit composition of the two-chain coiled-coil of epidermal keratin IF remains open. Even though *in vivo* expression data (refs. 1, 8, 9, and 15; unpublished data) and *in vitro* IF reassembly data (11–13, 34) imply that type I and type II keratins are required for epidermal keratin IF assembly, it is not yet clear at which level of IF structure this requirement is imposed. More detailed biochemical and structural analyses will be necessary to resolve this question.

1. Moll, R., Franke, W. W., Schiller, D. L., Geiger, B. & Krepler, R. (1982) *Cell* **31**, 11–24.
2. Steinert, P. M. (1978) *J. Mol. Biol.* **123**, 49–70.
3. Steinert, P. M., Idler, W. W. & Goldman, R. D. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 4534–4538.
4. Hanukoglu, I. & Fuchs, E. V. (1982) *Cell* **31**, 243–252.
5. Hanukoglu, I. & Fuchs, E. V. (1983) *Cell* **33**, 915–924.
6. Steinert, P. M., Rice, R. H., Roop, D. R., Trus, B. L. & Steven, A. C. (1983) *Nature (London)* **302**, 794–800.
7. Steinert, P. M., Jones, J. C. R. & Goldman, R. D. (1984) *J. Cell Biol.*, in press.
8. Eichner, R., Bonitz, P. & Sun, T.-T. (1984) *J. Cell Biol.* **98**, 1388–1396.
9. Sun, T.-T., Eichner, R., Nelson, W. G., Tseng, S. C. G. & Weiss, R. A. (1983) *J. Invest. Dermatol.* **81**, 109s–115s.
10. Fuchs, E. V. & Marchuk, D. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 4321–4325.
11. Steinert, P. M., Idler, W. W. & Zimmerman, S. B. (1976) *J. Mol. Biol.* **108**, 547–567.
12. Steinert, P. M., Idler, W. W., Poirier, M. G., Katoh, Y., Stoner, G. D. & Yuspa, S. H. (1979) *Biochim. Biophys. Acta* **577**, 11–21.
13. Steinert, P. M., Idler, W. W., Anyardi-Whitman, M. & Goldman, R. D. (1982) *Cold Spring Harbor Symp. Quant. Biol.* **46**, 465–474.
14. Gough, K. H., Inglis, A. S. & Crewther, W. G. (1978) *Biochem. J.* **173**, 373–385.
15. Roop, D. R., Hawley-Nelson, P., Cheng, C. K. & Yuspa, S. H. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 716–720.
16. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
17. Tu, C.-P. D. & Cohen, S. N. (1980) *Gene* **10**, 177–183.
18. Parry, D. A. D., Crewther, W. G., Fraser, R. D. B. & MacRae, T. P. (1977) *J. Mol. Biol.* **113**, 449–454.
19. Dowling, L. M., Parry, D. A. D. & Sparrow, L. A. (1983) *Biosci. Rep.* **3**, 73–78.
20. Crewther, W. G., Dowling, L. M., Steinert, P. M. & Parry, D. A. D. (1983) *Int. J. Biol. Macromol.* **5**, 267–274.
21. Gibbs, A. & MacIntyre, J. (1970) *Eur. J. Biochem.* **16**, 1–11.
22. Fitzgerald, M. & Shenk, T. (1981) *Cell* **24**, 251–260.
23. Steinert, P. M., Wantz, M. L. & Idler, W. W. (1982) *Biochemistry* **21**, 177–183.
24. Chou, F. Y. & Fasman, G. D. (1974) *Biochemistry* **13**, 211–245.
25. Garnier, J., Osguthorpe, D. J. & Robson, B. (1978) *J. Mol. Biol.* **120**, 97–120.
26. Sparrow, L. G. & Inglis, A. S. (1980) *Proceedings of the Sixth International Wool Textile Research Conference*, pp. 237–246.
27. Geisler, N. & Weber, K. (1982) *EMBO J.* **1**, 1649–1656.
28. Quax, W., Egberts, W. V., Hendricks, W., Quax-Jeuken, Y. & Bloemendahl, H. (1983) *Cell* **35**, 215–223.
29. Pang, Y.-Y. S., Robson, R. M., Hartzer, M. K. & Stromer, M. H. (1983) *J. Cell Biol.* **97**, 226.
30. Gruen, L. C. & Woods, E. F. (1983) *Biochem. J.* **209**, 587–598.
31. Fraser, R. D. B. & MacRae, T. P. (1983) *Biosci. Rep.* **3**, 517–525.
32. Trus, B. L. & Piez, K. A. (1976) *J. Mol. Biol.* **108**, 705–732.
33. Parry, D. A. D. (1977) *Polymer* **18**, 1091–1094.
34. Franke, W. W., Schiller, D. L., Hatzfeld, M. & Winter, S. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 7113–7117.