Nonepidermal members of the keratin multigene family: cDNA sequences and in situ localization of the mRNAs

Bernhard Knapp, Martin Rentrop, Jürgen Schweizer and Hermelita Winter

Institute of Experimental Pathology, German Cancer Research Center, Im Neuenheimer Feld 280, 6900 Heidelberg, FRG

Received 22 October 1985; Revised and Accepted 13 December 1985

ABSTRACT

A keratin set which consists of a type I 47kd and a type II 57kd protein occurs as a major constituent of the keratin patterns of various internal stratified epithelia of the mouse. We have isolated specific cDNA clones of the two complementary keratin subunits from a cDNA library constructed with polyA⁺ RNA of mouse tongue epithelium and present the complete nucleotide and deduced amino acid sequences of the 57kd protein and about 75% of the corresponding data of the 47kd protein.

The comparison of the sequence data with those of known epidermal keratin mRNAs coding for the two types of keratin proteins reveals a fundamentally identical and type-specific organization of the mRNAs into both highly conserved and variable domains. In order to avoid cross-reactions with other members of the keratin multigene family, appropriately tailored ³⁵S-labeled cDNA probes comprising the low and non-homologous 3' coding and noncoding domains of the mRNAs were used for in situ hybridization to tissue sections. The localization and distribution of the corresponding transcripts indicates a strongly compartmentalized keratin expression in mouse tongue epithelium.

INTRODUCTION

Keratins, a multigene family of water-insoluble proteins are the constituent proteins of intermediate filaments (IF) present in almost all epithelial cells. The complexity of the keratin polypeptide patterns increases from simple to stratified non-keratinizing tissues and is most complex in keratinizing terminally differentiating epithelia (1-5). There is strong evidence that at least two complementary, oppositely charged keratin proteins are required for keratin IF assembly in vivo (6,7).

Recently, sequence data from both cDNA and genomic clones of epidermal keratin proteins have extended our knowledge on the structural organization of this class of IF proteins (8-20). There is now general agreement that both acidic type I proteins and basic to neutral type II keratin proteins possess a common structure in which a highly conserved central α -helical domain is flanked by rather variable non α -helical carboxy and amino termi-

nal regions. At the nucleotide level, the mRNAs of both types of keratins within one species reveal a high degree of divergence within their 3' and (as far as determined) 5' noncoding regions (12).

These data stem rather exclusively from epidermal keratin subunits of different species. In the present paper we report on the isolation and characterization of cDNA clones of nonepidermal members of both keratin subfamilies. The complementary type I 47kd and type II 57kd proteins are co-expressed in internal, stratified mouse epithelia and represent the main keratin subunits of palate, cheek (21), esophagus, forestomach (22) and tongue (23). Comparisons with presently known sequence data provided evidence of generally identical structural properties of keratin proteins in epithelia displaying different forms of cell differentiation and enabled us to prepare highly specific cDNA probes for the study of details of epithelial differentiation by means of in situ hybridization.

METHODS

Cloning of cDNAs and screening procedures

PolyA⁺RNA from heat-separated mouse tongue epithelium was isolated as described (24). Double-stranded cDNA > 600 bp was prepared from 10 µg of polyA⁺RNA (25), blunt-end ligated into pUC8 plasmid which was used to transform *E. coli* OM 214 cells (26). The library was screened for keratin cDNA clones (27) with ³²P-labeled enriched mRNAs (26,28) coding either for the 57kd or the 47kd keratin protein. Clones strongly hybridizing with one of the enriched keratin mRNAs were selected and screened by positive hybrid selection (26). The clones pkt57-1 and pkt47-1 containing the largest inserts were selected for further investigation.

All experiments were carried out according to the guide-lines for recombinant DNA published by the "Bundesminister für Forschung und Technologie" of the Federal Republic of Germany.

RNA blot analysis

Ten µg polyA⁺RNA from different tissues were electrophoresed in 1.4% agarose gels containing 15 mM methylmercury hydroxide (28), transferred to nitrocellulose filters (29), and hybridized overnight at 42°C (50% formamide) with the ³²P-labeled inserts (30) of pkt57-1 and pkt47-1. Blots were washed at 68°C in 0.1xSSC/0.1% SDS and autoradiographed.

DNA sequence analysis

Restriction site mapping was carried out according to Smith and Birnstiel (31). Suitable restriction fragments were subcloned into the M13 mp10 and

mp11 vectors (32), and sequencing was performed by the dideoxy chain terminator method of Sanger et al. (33). Both strands of the two clones were completely sequenced.

The translation of the two cDNA clone sequences into amino acid sequences, the nucleotide and protein sequence comparisons, and molecular weight determinations were performed using the programs of Osterburg et al. (34). For the prediction of the secondary structure of the two keratin proteins both, the probabilistic methods of Chou and Fasman (35,36) and Garnier et al. (37) as well as the heptade convention analysis (10,38,39) were used.

In situ hybridization

In situ hybridization to tissue sections was performed according to the method described by Lawrence and Singer (40) with minor modifications. Briefly, 5 μ m frozen sections of mouse tongue were hybridized with ^{35}S -labeled cDNA probes (specific activities $3\text{--}4 \times 10^8 \text{cpm}/\mu\text{g}$). As a control for the specificity of hybridization tissue sections were treated with RNase (100 μg RNase A and 100 units RNase T1/ml) prior to hybridization. After hybridization slides were washed and dried by passage through ethanol-series. Subsequently slides were dipped in Kodak NTB2 emulsion and exposed for 4–21 days at 4°C. Slides were developed in Kodak D19 and sections were stained with hematoxylin-eosin. Details of the in situ hybridization technique will be published elsewhere.

RESULTS AND DISCUSSION

Isolation and identification of cDNA clones of the 57kd and 47 kd keratin proteins

A cDNA library was constructed with polyA⁺RNA from mouse tongue epithelium and screened with enriched mRNA probes coding for the 57kd and 47kd proteins. These probes were obtained by size fractionation of polyA⁺RNA from tongue epithelium using methylmercury hydroxide-agarose gels. Strongly positive recombinants identified with these mRNA probes by means of colony hybridization were further screened by positive hybrid selection (results not shown). These screening procedures yielded the plasmids pkt57-1 and pkt47-1 with inserts of about 2200 bp and 1300 bp, respectively.

Size of mRNAs of the 57kd and 47kd keratin proteins and tissue-specific expression

The size of the mRNAs coding for the two keratin proteins was determined by Northern blot analysis. Methylmercury hydroxide-agarose gel fractionated polyA⁺RNA from mouse tongue epithelium was hybridized under stringent con-

Fig. 1. Identification in different tissues of mRNAs coding for the 57kd and 47kd keratin proteins. Ten μ g of polyA⁺ RNA from mouse tongue epithelium (slots a), forestomach epithelium (slots b), footpad epidermis (slots c) and liver (slots d) were separated by gel electrophoresis on 1.4% agarose gels containing 15 mM methylmercury hydroxide. The RNA was blotted onto nitrocellulose paper and the blots were hybridized with ³²P-labeled insert DNAs of plasmid pkt57-1 and pkt47-1, washed at 68°C with 0.1xSSC/0.1% SDS and autoradiographed. The 16S/23S and 18S/28S rRNAs (indicated in kilobases) were run on the same gel as size markers.

ditions with ³²P-labeled pkt57-1 and pkt47-1 inserts (Fig. 1a). Each probe hybridized with a single mRNA species containing 2250 ± 100 and 1750 ± 50 bases, respectively. As expected (22), mRNAs corresponding to tongue mRNAs could be identified when polyA⁺ RNA from mouse forestomach epithelium was subjected to blot analysis (Fig. 1b). In contrast under the highly stringent hybridization and wash conditions, epidermal and liver mRNAs do not show cross reaction with pkt57-1 and pkt47-1 inserts (Fig. 1 c,d).

cDNA sequences of pkt57-1 and pkt47-1

The nucleotide sequence of the pkt57-1 insert, which represents the full length transcript of the mRNA is shown in Figure 2. The size of this single insert is 2205 nucleotides. It includes the complete 3' end and extends far (117 nucleotides) into the non-coding 5' end of the mRNA. Thirty five nucleotides upstream from the ATG initiation codon appears a putative ribosome binding consensus sequence CTTCTG which is frequently encountered in eukaryotic mRNAs (41,42) and has also been described in the mRNA coding for the human 50kd keratin (14). The non-coding 3' end of the sequence comprises 492 nucleotides and contains a polyadenylation signal AATAAA (43)

```

1 CCT GCC ATA GGC GGC TTT GCA GGA CCC AGC CAC GCT CAT GGA GAT ACT CTT GTG ACC CCC GAG GTT GTA AAG

75 ACT TCT GTT CTA AGC TCG TTG CAG CTG CCT CCT CTC TCA GGC ATG ATC GGC AGA CAG TCA AGT GTC GGC GGA GGC
Met Ile Ala Arg Gln Ser Ser Val Arg Gly Ala 11

150 TCC CGG GGC TTT AGC AGT GGC TCA GCT ATT GCT GGT GGT GTC AAG GGA GTG GGC TTT AGC TCA GGC TCC ATG TCT
Ser Pro Gly Phe Thr Ser Gly Ser Ala Ile Ala Gly Gly Val Lys Arg Val Ala Phe Ser Ser Gly Ser Met Ser 36

225 GGC GGT GCA GGG GGC TGC TCT TCT GGG GGC TTC GGC AGC AGA AGT CTT TAC AAC CTC GGG GGT CAC AAG AGT ATC
Gly Gly Ala Gly Arg Cys Ser Ser Gly Gly Phe Gly Ser Arg Ser Leu Tyr Asn Leu Gly Gly His Lys Ser Ile 61

300 TCC ATG AGC GTG GCT GGG TCC TCC CAA GGC GGC GGC TAT GGG GGT GGC GGA GGC TTC GGC GTT GGA GGA TAC GGT
Ser Met Ser Val Ala Gly Ser Cys Gln Gly Gly Gly Tyr Gly Gly Ala Gly Gly Phe Gly Val Gly Gly Tyr Gly 86

375 GCT GGG TTT GGT GGC GGT TTT GGC GGT GGC TTT GGA GGC TCC TTC AAC GGT GGA GGA GGT CCT GGC TTC CCA
Ala Gly Phe Gly Ala Gly Gly Phe Gly Gly Phe Gly Gly Ser Phe Asn Gly Arg Gly Gly Pro Gly Phe Pro 111

450 GTC TGT OCT GCT GGA GGT ATT CAG GAA GTC ACC ATC AAC CAG AGC CTG CTG ACA CCT CTT CAG GTG GAG ATT GAC
Val Cys Pro Ala Gly Gly Ile Gln Glu Val Thr Ile Asn Gln Ser Leu Leu Thr Pro Leu Gln Val Leu Ile Asp 136

525 CCA GAG ATC CAG AAA ATC GGC AGC GGC GAG GGT GAG CAG ATC AAG ACC CTC AAC AAC AAA TTC GGG TCC TTC ATC
Pro Glu Ile Gln Lys Ile Arg Thr Ala Glu Arg Glu Gln Ile Lys Thr Leu Asn Asn Lys Phe Ala Ser Phe Ile 161

600 GAT AAG GTG GGA TTC TTA GAG CAA CAG AAC AAG GTC CTG GAG ACC AAA TGG AAC CTC CTC CAG CAG CAG ACA ACC
Asp Lys Val Arg Phe Leu Glu Gln Asn Lys Val Leu Glu Thr Lys Trp Asn Leu Leu Gln Gln Thr Thr 186

675 ACC ACA TCT CCC AAA AGC CTG GAT OCT TTC TTT GAG ACT TAC ATC AAC GGC TTG AGG AAG AAC CTG GAC ACT TTG
Thr Thr Ser Pro Lys Ser Leu Asp Pro Phe Phe Glu Thr Tyr Ile Asn Ala Leu Arg Lys Asn Leu Asp Thr Leu 211

750 AGC AAT GAC AAA GGT CCC CTA CAG TCA GAG CTA AAG ATG ATG CAG GAC AGT GTG GAG GAC TTC AAG ACC AAG TAT
Ser Asn Asp Lys Gly Arg Leu Gln Ser Glu Leu Lys Met Met Gln Asp Ser Val Gln Asn Lys Phe Ala Ser Phe Ile 236

825 GAA GAG GAG ATC AAC AAG OCT ACA GCT GCA GAG AAT GAC TTC GTG CTC CTC AAG AAA GAT GTG GAT GCT GGC TAC
Glu Glu Glu Ile Asn Lys Arg Thr Ala Ala Glu Asn Asp Phe Val Val Leu Lys Lys Asp Val Asp Ala Ala Tyr 261

900 ATG ATC AAG GTG GAG TTA GAG GGC AAG ATG GAA AGC CTT AAG GAT GAG ATC AAC TTC AAG AGA GTC CTC TAT GAA
Met Ile Lys Val Glu Leu Glu Ala Lys Met Glu Ser Leu Lys Asp Glu Ile Asn Leu Thr Arg Val Leu Thr Asp 286

975 GGG GAG CTG GGC CAG ATG CAG ACA CAT GTC TCA GAC ACA TCT GTA GTG CTG TCC ATG GAC AAC AAC GGC AAC CTG
Ala Glu Leu Ala Gln Met Gln Thr His Val Ser Asp Thr Ser Val Val Leu Ser Met Asp Asn Asn Arg Asn Leu 311

1050 GAG CTG GAT GGC ATC ATC GCA GAG GTC GGG GGC CAG TAT GAG GAC ATT GCT AGG AAG AOC AAG GCT GAG GTT GAA
Asp Leu Asp Gly Ile Ile Ala Glu Val Arg Ala Gln Tyr Glu Asp Ile Ala Arg Lys Ser Lys Ala Glu Val Glu 336

1125 TCC TGG TAC CAG ATC AAG GTC CAG CAG CTC CAG ATG TCA GCT GAC CAA CAT GGA GAT AGC CTG AAG ACC ACC AAG
Ser Trp Tyr Gln Ile Lys Val Gln Gln Leu Gln Met Ser Ala Asp Gln His Gly Asp Ser Leu Lys Thr Thr Lys 361

1200 AAT GAG ATC TCA GAA CTC AAC AGG ATG ATC CAG AGG CTG GGC GCA GAG ATT GAG AAC ATC AAG AAG CAG AGC CAG
Asn Glu Ile Ser Glu Leu Asn Arg Met Ile Gln Arg Leu Arg Ala Glu Ile Glu Asn Ile Lys Lys Gln Ser Gln 386

1275 ACT CCG CAG GCA TCT GTG GCT GAT GCA GAG CAG CCG GGA GAG CTG GGC CTC AAA GAT GGC TAC AGC AAA GGC GCA
Thr Pro Gln Ala Ser Val Ala Asp Ala Glu Gln Arg Gly Glu Leu Ala Leu Lys Asp Ala Tyr Ser Lys Arg Ala 411

1350 GAA CTG GAG ACT GCA CTG CAG AAG GGC AAG GAG GAC CTG GGC GGC CTG CTG AAG GAC TAC CAG GCT CTC ATG AAC
Glu Leu Glu Thr Ala Leu Gln Lys Ala Lys Glu Asp Leu Ala Arg Leu Leu Arg Asp Tyr Gln Ala Leu Met Asn 436

1425 GTC AAG CTG GGC CTG GAT GTG GAG ATT GGC ACC TAC AGG AAG CTG CTG GAG GGC GAG GAG TCC AGG ATG TCT GGA
Val Lys Leu Ala Leu Asp Val Glu Ile Ala Thr Tyr Arg Lys Leu Leu Glu Gly Glu Glu Cys Arg Met Ser Gly 461

1500 GAA TCC AAG AGT GCT GTG AOC ATC TCG GTA GTT GGC GGC AGC CAG CAT TGG CCG AGT GGC CTT GGC CTG GGT AGC
Glu Cys Lys Ser Ala Val Ser Ile Ser Val Val Gly Gly Ser Gln His Trp Arg Ser Gly Leu Gly Leu Gly Ser 486

1575 GGT TTT TGC TCC GGT TCT GGT TCT GGA AGT GGC TTT GGA TTT GGT GGT GGC ATT TAT GGC GGT TCT GGT TCC AAG
Gly Phe Cys Ser Gly Ser Gly Ser Gly Phe Gly Phe Gly Gly Ile Tyr Gly Gly Ser Gly Ser Gly Ser Lys 511

1650 ATC ACC TCT TCC GGC ACC ATC ACC AAG AGG TCC CCA GGA TAG GCG AGA GCG GAT CCT GAT GGC CCT GAG CCC GGC
Ile Thr Ser Ser Ala Thr Ile Thr Lys Arg Ser Pro Arg ***

1725 AAG GTC AGC GGT TCC TGA TAT CTC CTG TGC GAT CTT CAC TTC ACC TCC AGT CTT OCT GGT GCT CAT CTT GTT ACT

1800 CTT CCC OCT CTG OCT TGG TCC CTC AGT TCC AAG AGC ATC TTC TTG GGC ATG GTG TGA OCT CTG TTC TAT TGG AAC

1875 CTG GTA GCT TCT TCT CAA CTT GGG TCA AGT GGC CCC CCA CCC CCT GCA TGG GAG GGG TGT CTG ATT TCA CCC CAA

1950 TGG ACA GAG GGG ATG AAG AGC AGA AAC TCC TTC TCC AGA TGT ACA CTG TCA CTC TCA GCT CCT CTC GTC ACA CCG

2025 GCT TCC ACC TGA TCC TGG AGA TCT ATG TTC TAC AAT GAT GCT CAT TCT CCA TCA CCT GGC AGC AAC GGG CTC TGC

2100 GAG CTG AAC AAG GAC TGC TTG GTG TCC GGT GCA ACC CTC CTC CTC TCT CTA CTC ATC CCC AAT AAA ATG CTT TTG

2175 TCA TTC AAA AAA AAA AAA AAA AAA AAA

```

Fig. 2. Complete nucleotide sequence of the pkt57-1 insert and deduced amino acid sequence of the 57kd keratin protein of mouse tongue epithelium. The nucleotide sequence is shown in the 5' to 3' direction of the mRNA strand. Nucleotides are numbered from the first base of the cDNA insert at the left side and amino acids are numbered at the right side. The stop codon is marked by three asterisks. Arrows denote the deduced demarcation of the central α -helical domain. H1, V1, N and H2, V2, C indicate the proposed threefold subdivision of the non- α -helical end domains (19). Both the putative polyadenylation signal and the putative ribosome binding consensus sequence are underlined.

```

1 GTG AAG ATT CGT GAC TGG CAT CTG AAA CAG AGC CCA GCT AGT CCA GAG CGG GAC TAC AGT GCT TAT TAT AAG ACC 25
  Val Lys Ile Arg Asp Trp His Leu Lys Gln Ser Pro Ala Ser Pro Glu Arg Asp Tyr Ser Ala Tyr Tyr Lys Thr

75 ATC GAG GAG CTC CGG ATC AAG ATT CTG GAA GGC ACC ACG GAC AAC AAC CGG ATC ATC CTG GAG ATT GAC AAT GCC 50
  Ile Glu Glu Leu Arg Ile Lys Ile Leu Glu Ala Thr Thr Asp Asn Asn Arg Ile Ile Leu Glu Ile Asp Asn Ala

150 AGG CTG GCT GCA GAT GAC TTC AGG CTC AAG TAC GAA AAT GAG CTG ACC TTG CGC CAG AGC GTG GAG GCC GAC ATC 75
  Arg Leu Ala Ala Asp Asp Phe Arg Leu Lys Tyr Glu Asn Glu Leu Thr Leu Arg Gln Ser Val Glu Ala Asp Ile

225 AAC GGC CTG CGC AGG GTG CTG GAT GAG CTG ACT CTG GCT AAG ACT GAC CTG GAA ATG CAG ATT GAG AGC CTG AAT 100
  Asn Gly Leu Arg Arg Val Leu Asp Glu Leu Thr Leu Ala Lys Thr Asp Leu Glu Met Gln Ile Glu Ser Leu Asn

300 GAA GAA CTG GCT TAC CTG AAG AAG AAC CAC GAA GAG GAG ATG AAG GAA TTC AGC AAC CAG GTA GTA GGC CAG GTC 125
  Glu Glu Leu Ala Tyr Leu Lys Lys Asn His Glu Glu Glu Met Lys Glu Phe Ser Asn Gln Val Val Gly Gln Val

375 AAT GTG GAG ATG GAC GCC ACC CCT GGC ATT GAT CTG ACC CGT GTG CTG GCT GAG ATG AGG GAG CAG TAT GAA GCC 150
  Asn Val Glu Met Asp Ala Thr Pro Gly Ile Asp Leu Thr Arg Val Leu Ala Glu Met Arg Glu Gln Tyr Glu Ala

450 CTG GCA GAG AAG AAT CGG AGG GAT GCT GAG GAA TGG TTC CAG ACC AAG AGT GCA GAG CTG AAC AAG GAA GTA TOC 175
  Leu Ala Glu Lys Asn Arg Arg Asp Ala Glu Glu Trp Phe Gln Thr Lys Ser Ala Glu Leu Asn Lys Glu Val Ser

525 TCC AAC GCT GAA ATG ATC CAG ACC AGC AAG GAG ATC ACA GAA CTC AGG CGC ACA CTC CAG GGC CTG GAG ATT 200
  Ser Asn Ala Glu Met Ile Gln Thr Ser Lys Thr Glu Ile Thr Glu Leu Arg Arg Thr Leu Gln Gly Leu Glu Ile

600 GAG CTG CAG TCC CAG CTG AGC ATG AAA GCC GGC CTG GAA AGT ACG TTG GCA GAG ACA GAG TGC CGC TAC GCC CTG 225
  Glu Leu Gln Ser Gln Leu Ser Met Lys Ala Gly Leu Glu Ser Thr Leu Ala Glu Thr Glu Cys Arg Tyr Ala Leu

675 CAG CTG CAA CAG ATC CAG GGG CTC ATC AGC AGC ATC GAG GCT CAG CTG AGC GAG CTC CGA AGT GAG ATG GAG TGC 250
  Gln Leu Gln Gln Ile Gln Gly Leu Ile Ser Ser Ile Glu Ala Gln Leu Ser Glu Leu Arg Ser Glu Met Glu Cys

750 CAG AAC CAG GAA TAC AAG ATG CTG TTG GAC ATC AAA ACA AGG CTG GAA CAG GAG ATC GCC ACC TAC CGC AGC CTG 275
  Gln Asn Gln Glu Tyr Lys Met Leu Leu Asp Ile Lys Thr Arg Leu Glu Gln Glu Ile Ala Thr Tyr Arg Ser Leu

825 CTA GAG GGC CAG GAT GCT AAG ATG ACC GGC TTC AAC TCC GGA GGA AAT AAC ACT ACT ACC TCC AAC GGG TCC CCC 300
  Leu Glu Gly Gln Asp Ala Lys Met Thr Gly Phe Asn Ser Gly Gly Asn Asn Thr Thr Thr Ser Asn Gly Ser Pro

900 TCC TCC AAT TCC GGT CGC CCA GAT TTC CGA AAG TAT TAA CTC AGC CTG GCG TCC CTT CCC TGG CCT TCA GCA GCC
  Ser Ser Asn Ser Gly Arg Pro Asp Phe Arg Lys Tyr ***

975 ACA GGA GGA GAG CTG ATG TCG CCT GTA GGA TGA GAG GAC TGC AGG ATG TGA TGC CCA GTC CCA TTG TCC CCC CAG

1050 GAC CCA TGG CTC TGA CAA CCC CTG AGG TGG GCC CTT CCC CAT CTC TTC TGA TTG GAA TCT TTC TTC TCC CTA CCC

1125 CAT TCT CTC TAT TCT TTG GCT TCT CAG TGT CTC TAC AAC AAA GAA TAG ATG GTT TCA ATA AAG TTC TCT GAT TTC

1200 CTG GAA AAA AAA AAA AAA AAA AAA AAA AAA AAA AAA AAA AAA AAA AAA AAA AAA AAA AAA AAA AAA

1275 AAA Val AAA AAA AA

```

Fig. 3. Nucleotide sequence of the pkt47-1 insert and deduced amino acid sequence of the 47kd keratin protein of mouse tongue epithelium. The stop codon, the end of the α -helical domain and the putative polyadenylation signal are indicated as described in legend to Figure 4.

fifteen nucleotides upstream from the polyA tail (A_{24}). The polyadenylation signal is followed by a ATGC-sequence occurring at the same position in the mRNAs of the type II mouse 60kd (11, see also Fig.5) and bovine 69kd keratin proteins (13). The central 1572 nucleotides represent the coding region for the entire length of the 57kd keratin protein.

The nucleotide sequence of the pkt47-1 insert is shown in Figure 3. The size of this insert is 1289 nucleotides, thus comprising approximately 75% of the calculated length of the mRNA. The reading frame reveals a stop codon TAA which divides the sequence into a coding region (936 nucleotides) and a 3' non-coding region (268 nucleotides) containing a putative polyadenylation signal AATAAA followed 17 nucleotides downstream by a 85 nucleotide long polyA tract.

These sequence informations supplement the list of currently available

Fig. 4. Structural organization of the 57kd (upper part) and the 47kd (lower part) proteins. N and C indicate the non- α -helical amino and carboxy terminal regions. 1A, 1B, 2A and 2B represent the coiled coil regions and L1, L12 and L2 the non-coiled coil linker regions of the α -helical domains. [Adoption of the nomenclature used by Steinert et al. (11)]. The numbers denote (from left to right) the entering positions of the individual structural segments. ND = not determined.

sequence data of epidermal keratins by one representative nonepidermal member of the type I and type II keratin multigene family.

Structural organization of the 57kd and 47kd keratin proteins

The 57kd protein comprises 523 amino acids if the initial methionine residue is excluded (Fig. 2). The calculated molecular size of 56,429 Da (without the unknown phosphate content) is in good agreement with the size estimated by gel electrophoresis (21). The amino acid sequence of the 47kd protein (Fig. 3) could be deduced over a length of 312 amino acids and comprises the carboxy terminal part and almost the complete α -helical region.

Secondary structure analysis (10,35-39) of both nonepidermal proteins revealed a high conformity with the structural organization elucidated so far for epidermal keratin proteins. This is especially relevant for the typical subregions of the central α -helical domain into four coiled coil regions and three short linker regions which cannot adopt coiled coils (19;Fig.4). In addition both the non α -helical N and C termini of the type II 57kd protein can be divided into three distinct subdomains according to the natures of their amino acid sequences (19). These comprise highly conserved H subdomains on either side of the rod domain, glycine and serine rich V subdomains of variable size and sequence and highly variable, usually basic N and C termini (Fig.2). In contrast these structural particularities cannot be detected in the C terminus of the type I 47kd protein (Fig.3). The entire region encompasses only 29 amino acid residues. It is therefore the shortest type I protein carboxy terminal part determined so far and confirms the prediction (19) that especially small type I keratin proteins may lack the substructuration of their non α -helical end domains.

Fig. 5. Comparison of the nucleotide sequences of the mRNAs coding for the nonepidermal 57kd keratin protein and the mouse epidermal 60kd keratin protein (11). Large arrows indicate the positions of the region coding for the α -helical domains, small arrows denote the ends of conserved regions coding for the non α -helical H subdomains (compare Fig. 2). Initiation and stop codons and the putative polyadenylation signals are in boxes. Asterisks indicate conserved nucleotides.

Sequence comparisons of nonepidermal and epidermal keratin mRNAs

mRNAs which code for epidermal type I and type II keratin proteins of mouse and man (8-11, 14-19) were compared with the mRNAs of the nonepider-

mal 47kd and the 57kd keratin proteins. In each case, nucleotide sequences with the highest degree of sequence identity correspond to those coding for the α -helical domains of the proteins. The calculated values are in the order of magnitude of 70-80% and therefore resemble those determined for epidermal keratin mRNAs coding for the same type of proteins (12). Figure 5 which shows the alignment of the 57kd protein mRNA and the mRNA of the mouse epidermal type II 60kd protein (11) is representative for these multiple comparisons. In this case the sequence identity of the α -helix-coding region is 75%. In addition the alignment reveals highly homologous regions of constant lengths immediately upstream (108 nucleotides; sequence identity 70%) and downstream (39 nucleotides; sequence identity 53%) of the α -helix-coding region. These sequences are common to all known epidermal type II keratin mRNAs (11,16,19) and code for the non α -helical H subdomains of this keratin protein subfamily (Fig. 2). They are absent from both type I epidermal (10,17-19) and the type I nonepidermal keratin mRNA (Fig. 3). In contrast both the lengths and the sequences of the remaining 5' and 3' regions, encompassing coding as well as noncoding sequences of the mRNAs are highly variable, unless functionally related epidermal keratin mRNAs of different species are compared (12). However, Figure 5 shows that the 5' end regions of type II keratin mRNAs are slightly more conserved than the 3' end regions. Occasionally they contain rather extended homologous nucleotide sequences i.e. positions 254-278 of the 57kd protein mRNA which code for a characteristic type II protein octade (11, 16, 19). As a rule, this is not the case for the 3'end regions which at most exhibit rare stretches of 4-6 conserved nucleotides and therefore seem to be most specific for any given keratin mRNA.

Besides the demonstration of fundamentally identically organized keratin mRNAs in epithelia showing different forms of cell differentiation, these comparisons are of paramount importance in view of the generation of specific cDNA probes for in situ hybridization experiments. cDNA probes containing sequences complementary to the 3' non-homologous end regions of a keratin mRNA should exclude the danger of crossreactivity and therefore be ideally suited to investigate the specific cellular location of the corresponding mRNA.

In situ hybridization

Figure 6 illustrates the results of the in situ hybridization to sections of adult mouse tongue using a ^{35}S -labeled 553bp HhaI-HindII fragment derived from the 3' region of the pkt57-1 insert (nucleotides 1553-2106,

Fig. 6. In situ hybridization of the 553 bp ^{35}S -labeled Hha I - Hind II fragment of the pkt57-1 insert. The probe was hybridized to a sagittal section of the ventral side (a) and to a sagittal (b), frontal (c) and tangential (d) section of the dorsal side of adult mouse tongue. Similar results were obtained with a specific probe of the 47kd insert. Control sections treated with RNase prior to hybridization showed only background noise. Competition experiments with increasing amounts of unlabelled specific cDNA fragments abolished the label, whereas unrelated DNA was ineffective. BC, basal cells; SC, stratum corneum; CT, connective tissue; IP, inter-papillary region; FP, filiform papillae; DP, dermal papillae (H&E; x200).

i.e. containing parts of the low homologous terminal coding region and the noncoding region without the polyA tract; Fig. 2). Figure 6a shows a sagittal section of the ventral side of the tongue which is covered by a conti-

nuous, multilayered epithelium (44). It can easily be seen that the label is restricted to the epithelial cells of this tissue with negligible background density of silver grains over non-epithelial cells. It should be emphasized that treatment of the sections with RNase prior to hybridization did not lead to any detectable label in the tissue. Whereas most basal cells show a level of hybridization that is not significantly above background noise, the entire suprabasal compartment of living cells is heavily labeled. The highest density of silver grains is found over the spinous cell layers. The label gradually decreases in the granular cell layers and is absent from the dead stratum corneum. Unlike the ventral epithelium, the dorsal epithelium of the tongue shows a rather complex morphology in that numerous papillae - mainly of the filiform type - are distributed in parallel rows within an interpapillary epithelium (44,45). In situ hybridization with the same specific probe reveals that the corresponding mRNA is abundantly present in the suprabasal interpapillary epithelium, however, again absent from its basal cell layer as well as from the entire filiform papillae (Fig. 6b-d). As observed for the morphologically related ventral epithelium, there is a clear cut decrease of the suprabasal interpapillary label from the spinous to the granular cell layers (best visible in Fig. 6d). Whether the progressive death of cells at the transition from the granular to the cornified layer correlates with a simultaneous decrease of keratin gene expression or if there exists a temporal difference between cell death and cessation of keratin expression remains to be investigated. Modulations of epithelial keratin mRNA expression by certain drugs (46,47) and the localization of these alterations at the cellular level by in situ hybridization with specific cDNA probes may be a promising approach to study elements which control the differential expression of this multigene family.

ACKNOWLEDGMENTS

We thank H.P. Schmitt and A. Alonso for helpful advice and discussion and C. Frees for skillful technical assistance. We are indebted to E. Amtmann for facilities with methylmercury hydroxide-agarose gel electrophoresis. This work was supported by the Deutsche Forschungsgemeinschaft (Wi 489/2-1).

REFERENCES

1. Moll, R., Franke, W.W., Schiller, D.L., Geiger, B. and Krepler, R. (1982) *Cell* 31, 11-24.
 2. Tseng, S.C.G., Jarvinen, M.J., Nelson, W.G., Huang, J.W., Woodcock-Mitchell, J. and Sun, T.-T. (1982) *Cell* 30, 361-372.
-

3. Wu, Y.J., Parker, L.M., Binder, N.E., Beckett, M.A., Sinard, J.H., Griffiths, C. T. and Rheinwald, J.G. (1982) *Cell* 31, 693-703.
4. Sun, T.-T., Eichner, R., Nelson, W.G., Tseng, S.C.G., Weiss, R.A., Jarvinen, W. and Woodcock-Mitchell, J. (1983) *J. Invest. Dermatol.* 81, 109s-115s.
5. Schweizer, J., Kinjo, M., Fürstenberger, G. and Winter, H. (1984) *Cell* 37, 159-170.
6. Franke, W.W., Schiller, D.L., Hatzfeld, M. and Winter, S. (1983) *Proc. Natl. Acad. Sci. USA* 80, 7113-7117.
7. Steinert, P.M., Jones, J.C.R. and Goldman, R.D. (1984) *J. Cell Biol.* 99, 22s-27s.
8. Hanukoglu, I. and Fuchs, E. (1982) *Cell* 31, 243-252.
9. Hanukoglu, I. and Fuchs, E. (1983) *Cell* 33, 915-924.
10. Steinert, P.W., Rice, R.H., Roop, D.R., Trus, B.L. and Steven, A.C. (1983) *Nature* 302, 794-800.
11. Steinert, P.M., Parry, D.A.D., Racoosin, E.L., Idler, W.W., Steven, A.C., Trus, B.L. and Roop, D.R. (1984) *Proc. Natl. Acad. Sci. USA* 81, 5709-5713.
12. Jorcano, J.L., Rieger, M., Franz, J.K., Schiller, D.L., Moll, R. and Franke, W.W. (1984) *J. Mol. Biol.* 179, 257-281.
13. Jorcano, J.L., Franz, J.K. and Franke, W.W. (1984) *Differentiation* 28, 155-163.
14. Marchuk, D., McCrohon, S. and Fuchs, E. (1984) *Cell* 39, 491-498.
15. Johnson, L.D., Idler, W.W., Zhou, X., Roop, D.R. and Steinert, P.M. (1985) *Proc. Natl. Acad. Sci. USA* 82, 1896-1900.
16. Tyner, A.L., Eichman, M.J. and Fuchs, E. (1985) *Proc. Natl. Acad. Sci. USA* 82, 4683-4687.
17. Marchuk, D., McCrohon, S. and Fuchs, E. (1985) *Proc. Natl. Acad. Sci. USA* 82, 1609-1613.
18. Krieg, T.M., Schafer, M.P., Cheng, C.K., Filpula, D., Flaherty, P., Steinert, P.M. and Roop, D.R. (1985) *J. Biol. Chem.* 260, 5867-5870.
19. Steinert, P.M., Parry, D.A.D., Idler, W.W., Johnson, L.D., Steven, A.C. and Roop, D.R. (1985) *J. Biol. Chem.* 260, 7142-7149.
20. Rieger, M., Jorcano, J.L. and Franke, W.W. (1985) *EMBO J.* 4, 2261-2267.
21. Schweizer, J., Winter, H., Hill, M.W. and Mackenzie, I.C. (1984) *Differentiation* 26, 144-153.
22. Schweizer, J. and Winter, H. (1983) *J. Biol. Chem.* 258, 13268-13272.
23. Franke, W.W., Schmid, E., Schiller, D.L., Winter, S., Jarasch, D.E., Moll, R., Denk, H., Jackson, B.W. and Ilmensee, K. (1982) *Cold Spring Harbor Symposium on Quantitative Biology*, Vol. XLVI, 431-453.
24. Schweizer, J. and Goerttler, K. (1980) *Eur. J. Biochem.* 112, 243-249.
25. Wickens, M.P., Buell, G.N. and Schimke, R.T. (1978) *J. Biol. Chem.* 253, 2483-2495.
26. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) *Molecular cloning. A Laboratory Manual*. Cold Spring Harbor Laboratory.
27. Grunstein, M. and Hogness, D.S. (1975) *Proc. Natl. Acad. Sci. USA* 72, 3961-3965.
28. Bailey, J.M. and Davidson, N. (1975) *Anal. Biochem.* 70, 75-85.
29. Thomas, P. (1980) *Proc. Natl. Acad. Sci. USA* 77, 5201-5205.
30. Rigby, P.W.J., Dieckmann, M., Rhodes, C. and Berg, P. (1977) *J. Mol. Biol.* 113, 237-251.
31. Smith, H.O. and Birnstiel, M.L. (1976) *Nucleic Acids Res.* 3, 2387-2398.
32. Messing, J. (1983) *Meth. Enzymol.* 101, 20-89.
33. Sanger, P., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* 74, 5463-5467.
34. Osterburg, G., Glatting, K.H. and Sommer, R. (1982) *Nucleic Acids Res.* 10, 207-216.

35. Chou, P.Y. and Fasman, G.D. (1978) *J. Biol. Chem.* 252, 2127-2133.
36. Chou, P.Y. and Fasman, G.D. (1979) *Biophys. J.* 26, 367-383.
37. Garnier, J., Osguthorpe, D.J. and Robson, B. (1978) *J. Mol. Biol.* 120, 97-120.
38. Parry, D.A.D., Crewther, W.G., Fraser, R.D.B. and McRae, T.P. (1977) *J. Mol. Biol.* 133, 449-454.
39. Crewther, W.G., Inglis, A.S. and McKern, N.M. (1978) *Biochem. J.* 173, 365-371.
40. Lawrence, J.B. and Singer, R.H. (1985) *Nucleic Acids Res.* 13, 1777-1799.
41. Baralle, F.E. and Brownlee, G.G. (1978) *Nature* 274, 84-87.
42. Hagenbuechle, O., Santer, M., Argetsinger-Steitz, J. and Mans, R.J. (1978) *Cell* 13, 551-563.
43. Fitzgerald, M. and Shenk, T. (1981) *Cell* 24, 251-260.
44. Cameron, I.L. (1966) *J. Exp. Zool.* 163, 271-284.
45. Potten, C.S., Schofield, R. and Lajtha, L.G. (1979) *Biochim. Biophys. Acta* 560, 281-299.
46. Schweizer, J., and Winter, H. (1982) *Cancer Res.* 42, 1517-1529.
47. Fuchs, E., and Green, H. (1981) *Cell* 25, 617-625.