Proc. Natl. Acad. Sci. USA
Vol. 87, pp. 9333–9337, December 1990
Biochemistry

Primary structure of keratinocyte transglutaminase

(molecular cloning/evolution/cross-linked envelopes)

M. A. PHILLIPS*, B. E. STEWART*, Q. QIN*, R. CHAKRAVARTY†, E. E. FLOYD‡, A. M. JETTEN‡, AND R. H. RICE*§

*Department of Environmental Toxicology, University of California, Davis, CA 95616; †Department of Internal Medicine, University of Cincinnati Medical Center, Cincinnati, OH 45267; and ‡Laboratory of Pulmonary Pathobiology, National Institute of Environmental Health Sciences, Research Triangle Park, NC 27713

Communicated by Howard Green, September 6, 1990

ABSTRACT **The nucleotide and deduced amino acid sequences of the coding regions of human and rat keratinocyte transglutaminases (protein-glutamine: amine γ-glutamyltransferase; EC 2.3.2.13) have been determined. These yield proteins of ≈90 kDa that are 92% identical, indicative of the conservation of important structural features. Alignments of amino acid sequences show substantial similarity among the keratinocyte transglutaminase, human clotting factor XIII catalytic subunit, guinea pig liver tissue transglutaminase, and the human erythrocyte band-4.2 protein. The keratinocyte enzyme is most similar to factor XIII, whereas the band-4.2 protein is most similar to the tissue transglutaminase. A salient feature of the keratinocyte transglutaminase is its 105-residue extension beyond the N terminus of the tissue transglutaminase. This extension and the unrelated activation peptide of factor XIII (a 37-residue extension) appear to be added for specialized functions after divergence of the tissue transglutaminase from their common lineage.**

During terminal differentiation, keratinocytes of the epidermis and other stratified squamous epithelia synthesize an envelope consisting of cross-linked protein beneath the plasma membrane (1). Localization of the envelope appears due to the presence of a membrane-bound transglutaminase (2, 3) and several of its substrate proteins at the cell periphery (4). The enzyme is anchored in the membrane by acylated fatty acid (5) and is activated by flux of Ca^{2+} into the cytoplasm when cellular membranes lose their integrity during the final maturation stage (6). The biochemical events resulting in mature envelopes have been difficult to follow due to the intractable nature of the highly cross-linked product. In view of the many proteins and amines in keratinocytes serving as transglutaminase substrates (4, 7), further study of the enzyme structure may help in analysis of this process. In addition to acylation, for example, phosphorylation of the membrane anchorage region has been seen, which could alter the interaction of the enzyme with potential substrate proteins (8).

The blood clotting factor XIII catalytic subunit (9–11) and tissue transglutaminase (12) have recently been cloned and sequenced. These enzymes are distantly related to each other but display significant similarity in certain regions, especially around the active site. An origin of the latter region in common with thiol proteases has been proposed (9). The more recent demonstration of striking similarity between the active site and a corresponding region in the erythrocyte band-4.2 protein (13, 14), however, indicates that closer relatives of transglutaminases exist. A cDNA clone for the keratinocyte-specific enzyme of the rabbit was originally identified by using an oligonucleotide probe directed toward the active site and was partially sequenced (15). Using that clone as probe, we have now cloned and determined the complete primary structure of this third type of transglutaminase for the human and rat.¶ Although many types of transglutaminase have been reported (16), little is known of their interrelations. Study of more members of this family may offer insight into their origin and relationship to other protein families as well. The present results reveal a salient structural feature of the keratinocyte transglutaminase (protein-glutamine: amine γ-glutamyltransferase; EC 2.3.2.13) and the evolutionary relationship of this enzyme to other family members.

MATERIALS AND METHODS

Materials. A cDNA library prepared in λgt11 by using poly(A)$^+$ RNA from cultured human epidermal cells and primed with oligo(dT)-cellulose (catalog no. HL1045b) was obtained from Clontech. In addition, λgt11 libraries prepared by using random primers [including oligo(dT)] were provided by N. Riedel (Boston University School of Medicine, Boston, MA). For this and other purposes, the RNA used was prepared by CsCl step-gradient centrifugation (19) from cultured human epidermal (17) or rat esophageal (18) keratinocytes that were dissolved in 6 M guanidine thiocyanate.

Cloning and Sequencing. With the use of a *Sac* I fragment [2.2 kilobases (kb)] of the pTG-7 cDNA probe previously obtained for the rabbit enzyme (15), positive λgt11 clones were selected from the Clontech human cDNA library described above. Few (<0.1%) of the plaques in the unscreened library were reactive with B.C1 monoclonal antibody (3) by immunoblotting on nitrocellulose filters according to standard methods (20). After two screenings with the rabbit cDNA probe, however, 4 of 16 positive clones were immunoreactive (≈1 of 6 clones were expected to be positive due to reading frame and directional requirements). Inserts from the 16 positive clones were excised and sized, and the two longest sequences (2.3 and 1.7 kb, the former displaying immunoreactivity) were inserted in M13. These inserts were characterized by partial dideoxynucleotide chain-termination sequencing (21) with a Sequenase kit (United States Biochemical), and then appropriate fragments were prepared by Bal-31 exonuclease digestion (22) and sequenced. A probe (0.7 kb) was prepared from the 5′ end of the longest clone by digestion with *Eco*RI and *Sma* I and used to select positive clones from a cDNA library prepared with random primers. As above, Bal-31-deleted fragments were prepared and sequenced from the clones selected. Positive clones from a random-primed rat keratinocyte cDNA library were selected by using a 2.3-kb human cDNA probe, and these clones were sequenced similarly.

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: nt, nucleotide(s).

§To whom reprint requests should be addressed.

¶The sequences reported in this paper have been deposited in the GenBank data base (accession nos. M55183 and M57263).

Primer Extension. The 32-base oligonucleotide 5′-CGT GGT AGG GGG CTG CAA GGG GTT GCC ACC CC-3′ was ^{32}P-labeled at the 5′ end with T4 polynucleotide kinase. Total human keratinocyte RNA was primed with this oligonucleotide, and cDNA was synthesized by using avian myeloblastosis virus reverse transcriptase according to standard procedures. The resulting product was then sized on a sequencing gel by electrophoresis in parallel with a sequencing ladder. Two experiments with different enzyme/substrate ratios gave equivalent results.

Northern (RNA) Blotting. Samples of poly(A)$^+$ (2 μg) or total RNA (25 μg) were electrophoresed in agarose gels (1.0–1.3%) containing formaldehyde (23) and transferred to Biotrans nylon membrane (ICN). Blots were hybridized at 65°C with 2.3-kb human or 2.2-kb rat cDNA probes or the 32-base oligonucleotide primer and washed at this temperature. The last wash with the oligonucleotide probe was 2× SSC (1× SSC is 0.15 M sodium chloride and 0.015 M sodium citrate, pH 7)/0.1% SDS and with cDNA probes was 0.5× SSC/0.1% SDS.

RESULTS AND DISCUSSION

The nucleotide sequences obtained for the rat and human keratinocyte transglutaminases were each represented in at least two independent clones that were sequenced in opposite directions. The subclones analyzed are presented in Fig. 1.

The transglutaminase cDNA clones selected in these experiments correspond to the keratinocyte enzyme because they match well the partial nucleotide sequence of a clone previously selected from cultured rabbit tracheal epithelial cells (15), and the proteins expressed by some of them in bacteria were recognized by the B.C1 monoclonal antibody specific for the keratinocyte enzyme (3). Most of the transglutaminase in cultures of human and rat keratinocytes is the keratinocyte-specific enzyme, although these cells can express small amounts of the tissue transglutaminase as well (3, 24). Only one class of message was observed in RNA blots, with an estimated size of 2.9–3.0 kb in parallel with commercial RNA size markers (Fig. 2). Thus, the mRNA is smaller than the tissue transglutaminase mRNA (≈4 kb), as shown previously (15). Although factor XIII has been reported in epidermis (25), a similarly soluble and thrombin-activated form of transglutaminase appears not to be synthesized in cultured epidermal cells (26).

Alignment of the compiled human and rat sequences (Fig. 3) gives open reading frames of ≈2.5 kb for each species. The nucleotide similarity in the coding region is high (84% identity), and the deduced amino acid sequences are 92% identical. In the human sequence, the 3′ end of the cDNA extends 176 nucleotides beyond a translation stop codon sequence, includes an AATAAA polyadenylylation signal, and terminates in a poly(A) tail. This sequence is much shorter than the 3′-untranslated regions (1.5 kb) of factor XIII (11) and the tissue transglutaminase (12), accounting for the smaller size of the keratinocyte message. The untranslated region in the cloned rat sequence is nearly as long but does not extend to a polyadenylylation signal or poly(A) tail. At the 5′ end of the rat sequence, an initiation ATG codon for protein synthesis is evident 113 bases from the beginning of the cDNA. This codon is preceded by two in-phase termination signals (TAA and TGA) 3 and 7 codons upstream, respectively. The human sequence exhibits an ATG codon at the same location as in the rat sequence and a second one immediately upstream (the likely translation start site). Termination signals are not observed in the 53-nucleotide region upstream, which presumably is not translated.

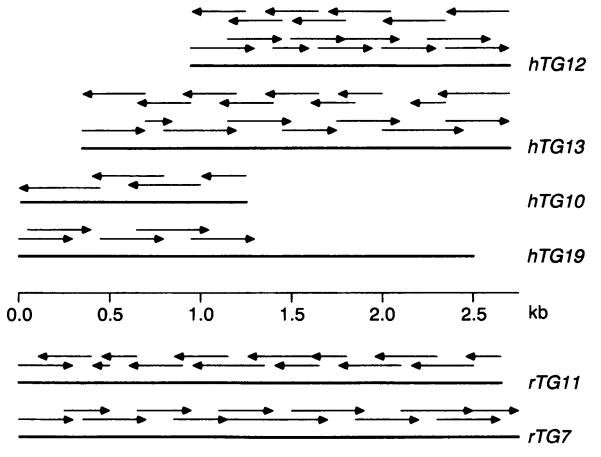

FIG. 1. cDNA clones of keratinocyte transglutaminase. The various clones (numbered solid lines) and subclones (arrows indicating sequencing direction) of the human (hTG) and rat (rTG) enzymes are given (in kb) as shown. Several redundant subclones are not shown.

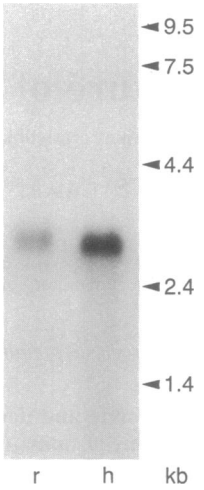

FIG. 2. Northern blot of keratinocyte RNA. After electrophoresis and transfer to nylon, samples of poly(A)$^+$ RNA from cultured human epidermal (h) and rat esophageal (r) epithelial cells were hybridized with a mixture of ^{32}P-labeled human and rat cDNA probes. A commercial ladder of synthetic poly(A)-tailed RNA (BRL) was run and detected by methylene blue stain; location of bands and sizes appear at right. Total RNA gave the same result as shown (bands of 2.9–3.0 kb) when compared with commercial markers, but a value of 3.7 kb was calculated when the internal 28S and 18S ribosomal RNA markers were used instead for standardization. Probing total human RNA with the ^{32}P-labeled oligonucleotide used in primer extension gave the same results as with cDNA.

To estimate the total length of 5′-untranslated sequence in the human transglutaminase message, cDNA synthesis was primed from total RNA of cultured keratinocytes with a 32-base oligonucleotide complementary to positions 88–119 in the known mRNA sequence. The products of reverse transcription, sized on a sequencing gel, were found to be two discrete oligonucleotides of 154 and 158 nucleotides (nt), or ≈39 nt beyond that represented in the cloned sequence. The 5′-untranslated region, thus, amounts to an estimated 92 bases, similar to the 110 bases measured for tissue transglutaminase by the same method (12), whereas the corresponding rat sequence may be slightly longer.

The deduced amino acid sequence of the human keratinocyte transglutaminase is shown in Fig. 4 in alignment with the human factor XIII catalytic subunit, guinea pig liver tissue transglutaminase, and human erythrocyte band-4.2 protein. The keratinocyte enzyme is the longest of these proteins (788 residues), extending 105 residues beyond the N terminus and 28 residues beyond the C terminus of the tissue transglutaminase. The transglutaminases are clearly homol-

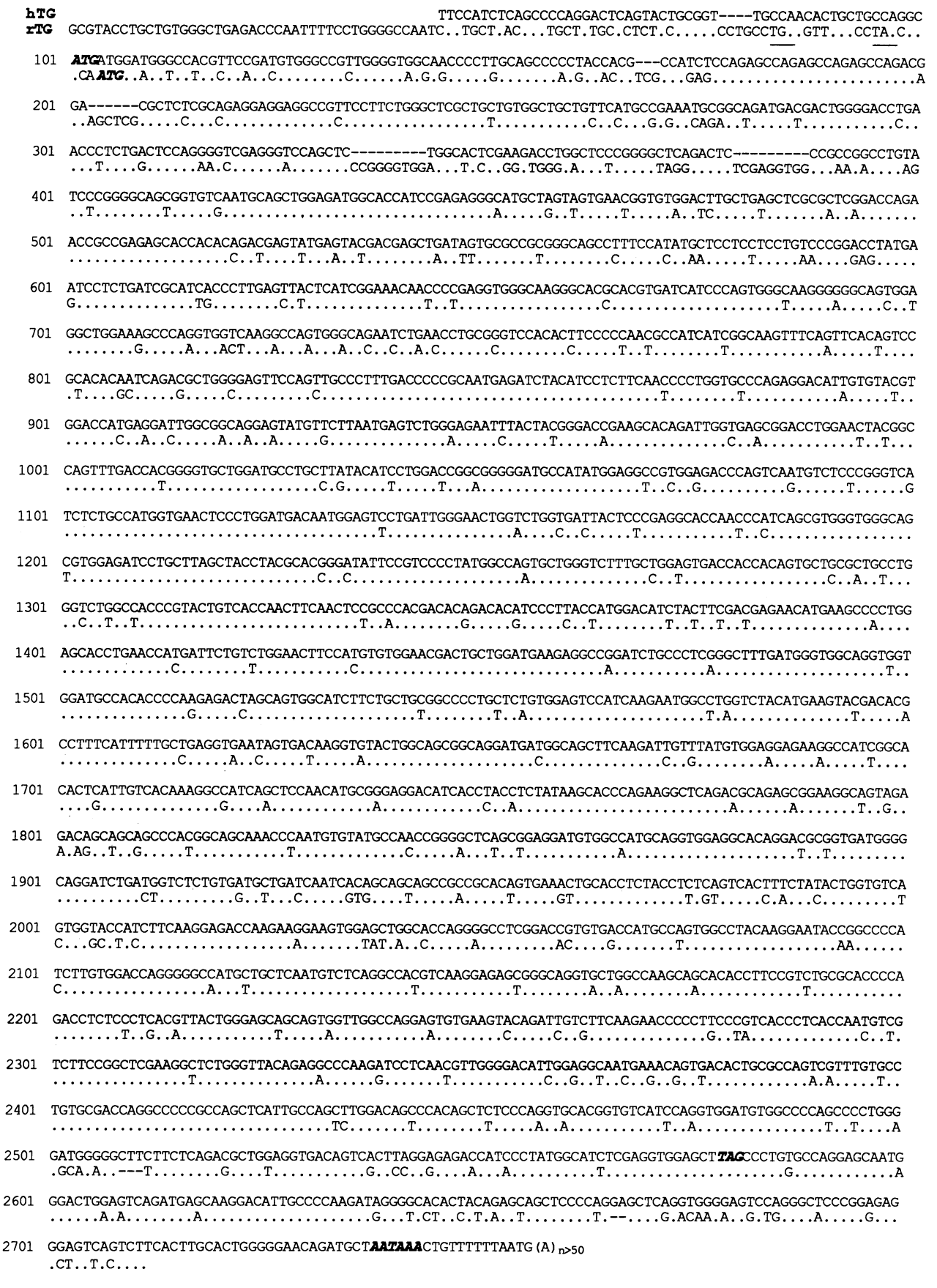

FIG. 3. Composite sequences of human (hTG) and rat (rTG) keratinocyte transglutaminase cDNAs. Each nucleotide is given for the human sequence; nucleotide substitutions in the rat sequence are shown, while identical nucleotides are indicated by periods. Gaps to maximize identity are indicated by hyphens. Codons corresponding to initiation (ATG) and termination (TAG) of translation and the polyadenylylation signal (AATAAA) are shown in boldface. Upstream in-frame termination codons in the rat sequence are underlined. Nucleotide positions in the alignment are numbered as indicated at left.

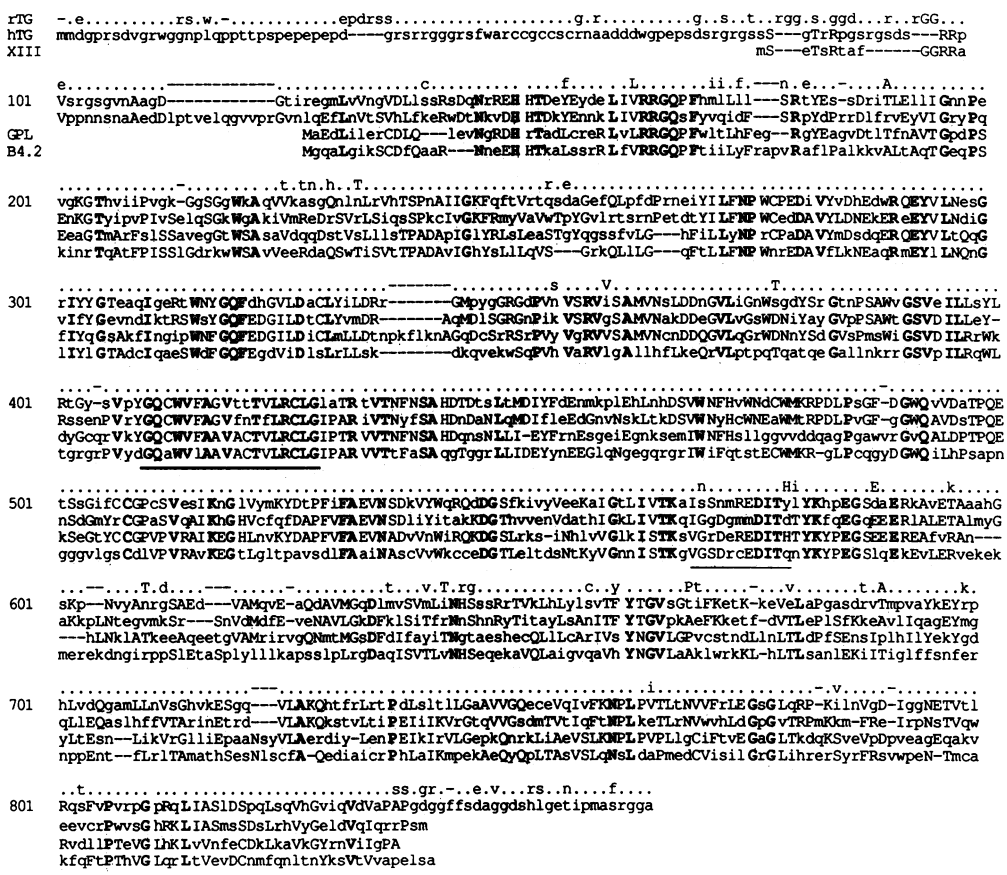

FIG. 4. Amino acid sequences of human (hTG) and rat (rTG) keratinocyte transglutaminases. The human sequence is given in its entirety, whereas only amino acid substitutions in the rat sequence are shown explicitly; identical amino acids in the latter are indicated by periods. By using the GAP program of the University of Wisconsin Genetics Computer Group computer system (27), the human factor XIII catalytic subunit (XIII), guinea pig liver tissue transglutaminase (GPL), and human erythrocyte band-4.2 (B4.2) were aligned for comparison; for these, each amino acid is shown. Gaps to maximize homology in alignments are indicated by hyphens. The transglutaminase active-site region showing high homology (positions 410–427) and a putative Ca^{2+}-binding region (positions 568–578; numbering at left) are underlined. Identical amino acids are capitalized and, where all are identical, printed in boldface type.

ogous, with conspicuous identity around the enzyme active site, which was the basis for the original selection of a keratinocyte enzyme cDNA clone with a synthetic oligonucleotide corresponding to this region (15). By contrast, homology with the Ca^{2+}-binding region proposed for the factor XIII catalytic subunit (9) is less obvious in the keratinocyte enzyme.

In the common alignment (Fig. 4), identical amino acids in each of the four sequences occur at 109 of ≈710 positions (15%). When identical amino acids are seen at a given position in only three of the sequences, the fourth sequence in 58% (74/127) of such cases is the band 4.2 protein, indicating the latter is more dissimilar to the transglutaminases than any of the others is to the group. Corresponding numbers for the keratinocyte, factor XIII, and tissue transglutaminases are much lower (12, 19, and 22 positions, respectively). When identical amino acids are present at a given position in only two of the four sequences, it is evident that the keratinocyte transglutaminase and the factor XIII catalytic subunit are most similar to each other (Table 1). The same conclusion follows from percentages of amino acid identity after aligning the sequences two at a time (Table 1). By contrast, the tissue transglutaminase appears about as closely related to the band 4.2 protein as to the other transglutaminases. This analysis is compatible with divergence of the band 4.2 protein from the group first, with subsequent separation of the tissue transglutaminase from the lineage leading to the factor XIII catalytic subunit and the keratinocyte enzyme.

The N terminal extension of the keratinocyte transglutaminase has a high content of proline, serine, and charged residues, including abundant arginine but no lysine (Fig. 4). Features of functional importance in this region presumably would be conserved between rat and human sequences. In this regard, a repeating 5-residue motif is evident, most conspicuously in the rat sequence (Table 2). This serine and arginine-containing motif is reminiscent of known phosphorylation sites for protein kinase C, often flanked by multiple basic (especially arginine) residues (28). The membrane-anchorage region of the keratinocyte transglutaminase, near one end of the protein, is subject to phorbol ester-stimulated

Table 1. Amino acid identities for protein pairs in the transglutaminase family within the region spanned by tissue transglutaminase

	Identities, total (%) or unique (number)			
	hTG	XIII	GPL	B4.2
hTG		96	40	31
XIII	45%		42	28
GPL	37%	36%		74
B4.2	29%	25%	33%	

Absolute numbers give positions identical in only two of the four sequences in Fig. 4; percentages are for total identities in pairwise comparisons. hTG, human keratinocyte transglutaminase; XIII, human factor XIII catalytic subunit; GPL, guinea pig liver tissue transglutaminase; B4.2, human erythrocyte band-4.2 protein.

Table 2. Tandem repeats of a 5-residue motif in the N-terminal extension

Species	Amino acids	Nucleotides				
Rat (residues 71–104)	RSRGT	AGA	AGC	CGA	GGG	ACC
	SSRGG	AGC	TCC	CGG	GGT	GGA
	GSRGG	GGC	TCC	CGG	GGT	GGG
	DSRGR	GAC	TCT	CGG	GGT	AGG
	DSRGG	GAC	TCT	CGA	GGT	GGC
	RRP ESRGS	GAG	TCT	CGG	GGC	AGT
Consensus	**DSRGG**	**GRC**	**TCY**	**CGG**	**GGT**	**RGS**
Human (residues 68–101)	RGRGS	AGG	GGT	CGA	GGG	TCC
	SS...	AGC	TCT	—	—	—
	GTRRP	GGC	ACT	CGA	AGA	CCT
	GSRGS	GGC	TCC	CGG	GGC	TCA
	DS...	GAC	TCC	—	—	—
	RRP VSRGS	GTA	TCC	CGG	GGC	AGC
Consensus	**GSRGS**	**GGC**	**TCY**	**CGR**	**GGC**	**TCC**

The extra RRP in the human sequence (italics) matches the equivalent in the rat protein and is not included in place of a gap in the previous repeat because the nucleotide sequence (CGC CGG CCT) is not a good match for the consensus. Nucleotides S = C or G, R = G or A, and Y = T or C.

phosphorylation of serine(s) separable from the actual site of anchorage by mild trypsin cleavage at arginine(s) (8).

Hydropathy plots (29) of the keratinocyte transglutaminase yield patterns of hydrophobic and hydrophilic regions (data not shown), quite similar to those seen for the three other known proteins in the transglutaminase family (14). In each enzyme, the most hydrophobic region is near the active site. Thus, as with the soluble transglutaminases, no markedly hydrophobic region that could serve alone as membrane anchor is evident in the keratinocyte enzyme. Previous results have shown that the latter is anchored within 10 kDa of the N or C terminus by fatty acid in a hydroxylamine-sensitive linkage (5). In other intracellular proteins so far analyzed, this type of linkage is a thioester (30). Indeed, a pair of cysteine residues very near the N terminus of the neuronal protein GAP-43 appears to serve in membrane anchorage upon palmitoylation (31). Inspection of the deduced amino acid sequence in the present instance reveals a cluster of 5 cysteines in a stretch of 7 amino acids ≈50 residues from the N terminus (the only exception to the marked hydrophilicity of the 105-residue extension) that could serve this function admirably.

The deduced evolutionary relations among the transglutaminases suggest that extension of the N terminus for specialized functions occurred to yield the factor XIII catalytic subunit and the keratinocyte enzyme after divergence of the tissue transglutaminase from a common lineage. Sequencing of factor XIII catalytic subunit genomic DNA has shown the activation peptide to be encoded in a separate exon (32), compatible with this scenario. Due to degree of gene divergence, whether the activation peptide and the N-terminal extension in the keratinocyte enzyme themselves have a common origin was not evident from current sequence information. Alignments of (*i*) the activation peptide amino acid (or nucleotide) sequence with the N-terminal extension in the keratinocyte transglutaminase or (*ii*) the nucleotide sequence of the keratinocyte extension and 5′-untranslated sequences in factor XIII mRNA demonstrate no detectable similarity beyond that which could have arisen by chance.

The basis for association of the erythrocyte band-4.2 protein with plasma membrane has not been determined but has been speculated to involve N-terminal myristoylation of the penultimate glycine after removal of the initial methionine (13). If so, this erythrocyte protein may be an example of convergence by a different means toward hydrophobic character. Alternately, the band-4.2 protein may associate with an intrinsic membrane protein by a catalytically inactive transglutaminase substrate-like binding site (14). Similarly, the keratinocyte enzyme could initially become membrane-associated through its affinity for certain substrates, thereafter becoming more firmly anchored by the action of membrane-bound fatty acid transacylases.

We thank Dr. Norbert Riedel for expert advice and Dr. Jerome Zeldis for providing the priming oligonucleotide. This research was supported by Public Health Service Grant AR 27130.

1. Green, H. (1979) *Harvey Lect.* **74**, 101–139.
2. Simon, M. & Green, H. (1985) *Cell* **40**, 677–683.
3. Thacher, S. M. & Rice, R. H. (1985) *Cell* **40**, 685–695.
4. Simon, M. & Green, H. (1984) *Cell* **36**, 327–334.
5. Chakravarty, R. & Rice, R. H. (1989) *J. Biol. Chem.* **264**, 625–629.
6. Rice, R. H. & Green, H. (1979) *Cell* **18**, 681–694.
7. Rice, R. H., Rong, X. & Chakravarty, R. (1990) *Biochem. J.* **265**, 351–357.
8. Chakravarty, R., Rong, X. & Rice, R. H. (1990) *Biochem. J.* **271**, 25–30.
9. Takahashi, N., Takahashi, Y. & Putnam, F. W. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 8019–8023.
10. Grundmann, U., Amann, E., Zettlemeissl, G. & Kupper, H. A. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 8024–8028.
11. Ichinose, A., Hendrickson, L. E., Fujikawa, K. & Davie, E. W. (1986) *Biochemistry* **25**, 6900–6906.
12. Ikura, K., Nasu, T., Yokota, H., Tsuchiya, Y., Sasaki., R. & Chiba, H. (1988) *Biochemistry* **27**, 2898–2905.
13. Korsgren, C., Lawler, J., Lambert, S., Speicher, D. & Cohen, C. M. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 613–617.
14. Sung, L. A., Chien, S., Chang, L.-S., Lambert, K., Bliss, S. A., Bouhassira, E. E., Nagel, R. L., Schwartz, R. S. & Rybicki, A. C. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 955–959.
15. Floyd, E. E. & Jetten, A. M. (1989) *Mol. Cell. Biol.* **9**, 4846–4851.
16. Folk, J. E. (1980) *Annu. Rev. Biochem.* **49**, 517–531.
17. Allen-Hoffman, B. L. & Rheinwald, J. G. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 7802–7806.
18. Heimann, R. & Rice, R. H. (1983) *J. Cell. Physiol.* **117**, 362–367.
19. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* **18**, 5294–5299.
20. Huynh, T. V., Young, R. A. & Davis, R. W. (1985) *DNA Cloning* (IRL, Oxford, England), pp. 49–78.
21. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
22. Poncz, M., Soloweijczyk, D., Ballantine, M., Schwartz, E. & Surrey, S. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 4298–4302.
23. Fourney, R. M., Miyakoshi, J., Day, R. S., III, & Patterson, M. C. (1988) *Focus* **10**, 5–7.
24. Rice, R. H., Chakravarty, R., Chen, J., O'Callahan, W. & Rubin, A. L. (1988) *Adv. Exp. Med. Biol.* **231**, 51–61.
25. Ando, Y., Imamura, S. & Kannagi, R. (1990) *Arch. Dermatol. Res.* **282**, 65–67.
26. Park, S. C., Kim, S. Y., Kim, H. C., Thacher, S. M. & Chung, S. I. (1988) *J. Cell Biol.* **107**, 139 (abstr.).
27. Devereux, J., Haeberli, P. & Smithies, O. (1984) *Nucleic Acids Res.* **12**, 387–395.
28. House, C., Wettenhall, R. E. H. & Kemp, B. E. (1987) *J. Biol. Chem.* **262**, 772–777.
29. Kyte, J. & Doolittle, R. F. (1982) *J. Mol. Biol.* **157**, 105–132.
30. Towler, D. A. & Gordon, J. I. (1988) *Annu. Rev. Biochem.* **57**, 69–99.
31. Skene, J. H. P. & Virag, I. (1989) *J. Cell Biol.* **108**, 613–624.
32. Ichinose, A. & Davie, E. W. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 5829–5833.