The Deduced Sequence of the Novel Protransglutaminase E (TGase3) of Human and Mouse*

(Received for publication, November 18, 1992, and in revised form, February 11, 1993)

In-Gyu Kim‡, Jeffrey J. Gorman§, Sang-Chul Park¶, Soon-Il Chung∥, and Peter M. Steinert**

From the §Skin Biology Branch, National Institute of Arthritis and Musculoskeletal and Skin Diseases and \ Laboratory of Cellular Development and Oncology, National Institute of Dental Research, National Institutes of Health, Bethesda, Maryland 20892, the ¶Biomolecular Research Institute, Parkville, Victoria 3052, Australia, and the ⊥Department of Biochemistry, Seoul National University Medical School, Seoul 110-460, Republic of Korea

At least three transglutaminases are involved in terminal differentiation events in the epidermis and its derivatives, such as the hair follicle, presumably in cross-linking structural proteins and in the formation of the cornified cell envelope. Of these, only the transglutaminase 3 is a proenzyme, requiring activation by proteolytic cleavage, and is the least understood. Using oligonucleotides designed from the amino acid sequences of peptides of the guinea pig enzyme, we amplified mRNA and deduced the complete amino acid sequences of the mouse and human protransglutaminase 3 enzymes. Both proteins contain 692 amino acids of molecular mass about 77 kDa. Following expression in yeast, the proenzymes encoded by the full-length cDNA clones are active enzymes and can be further activated 15-fold on treatment with dispase. Although these proteins share 38–53% identity to other members of the transglutaminase family, surprisingly, the mouse, human, and guinea pig enzymes have not been highly conserved and show only 50–75% identity to each other. Much of the sequence variation occurs in the vicinity of the proteolytic activation site which lies at the most flexible and hydrophilic region of the molecule and is flanked by a sequence of 12 residues that are absent from other transglutaminases. We suggest that cleavage of this exposed flexible hinge region promotes a conformational change in the protein to a more compact form, resulting in activation of the enzyme. Expression of mouse and human protransglutaminase 3 mRNAs is regulated by calcium, as for other late differentiation products of the epidermis, suggesting that this enzyme is responsible for the later stages of cell envelope formation in the epidermis and hair follicle.

Transglutaminases (TGases) are calcium- and thiol-dependent enzymes that modify proteins by catalyzing the formation of an isodipeptide cross-link between an ε-NH2 of a lysine and the γ-amide of a glutamine residue (1–4). In mammals, five distinct TGases are known to exist: a membrane-associated activity first discovered in keratinocytes of about 92 kDa (5–7), TGase1, and is now known to be widely expressed (8–13); a ubiquitous “soluble” or “tissue” activity of about 80 kDa, termed TGase2 (14–18); a soluble pro-enzyme activity of about 77 kDa, known as the “epidermal” or “hair follicle” TGase3 (14, 19–21); an inactive TGase-like protein of about 75 kDa, band 4.2, which is a ubiquitous constituent of the subplasma membrane of most eukaryotic cells (22, 23); and the catalytic subunit of the blood clotting factor XIII of about 77 kDa (24–26). Curiously, all but the latter member of this family are expressed in terminally differentiating epidermis (27). Although the TGase2 enzyme is implicated in apoptosis and cell adhesion (4, 28, 29), the TGase1 and TGase3 enzymes are thought to be involved in the formation and assembly of the cornified cell envelope (CE) of the epidermis, hair follicle, and perhaps other stratified squamous epitheliums (1–4, 12, 30, 31), by cross-linking of the several known or putative CE protein constituents with isodipeptide bonds (30–33). However, while data on the protein (27, 34–36) and gene (15, 37–39) structures and expression characteristics (40, 41) of the TGase1 system are available, rather less is known about the TGase3 system. Moreover, the substrate specificities, if any, of these enzymes during CE assembly is not yet known.

The TGase3 activity was the first epidermal enzyme to be isolated and characterized (14, 15, 19, 20). Several early studies reported a soluble protein of about 50 kDa from both epidermal and hair follicle tissues (14, 15, 19, 20, 42, 43), but more rigorous biochemical and cell biological analyses revealed that it is in fact a proenzyme of molecular mass about 77 kDa and becomes highly active upon proteolytic cleavage into a 50 kDa (amino-terminal) and 27-kDa species (20, 21). Although newer work showed that these fragments are not normally separated upon activation (21), the fact that the isolated 50-kDa fragment can retain catalytic activity was the source of confusion in earlier studies. Furthermore, despite earlier work (42), it is now generally agreed that the epidermal and hair follicle pro-enzyme species are the same (12).

Biochemical analyses have revealed that purified guinea pig TGase3 is chemically and perhaps structurally different from the other members of the TGase family (21). Guinea pig TGase3 possesses an elongated shape that undergoes a conformational rearrangement to a more compact form upon proteolytic activation. Fractionation by ion-exchange chromatography showed that the mouse and guinea pig TGase3 enzymes are near neutral in net charge, whereas the TGase1...
and TGase2 enzymes are acidic (21). Activated TGase3 accounts for >75% of total TGase activity in mammalian epidermal and hair follicle tissues, although chromatographic experiments show that the amount of TGase3 protein is far less than for TGase1 or TGase2 (21).

We have undertaken a systematic analysis of the family of TGases involved in CE formation in the epidermis (13, 27). In this paper we now describe the length of cDNA and deduced amino acid sequences of both the mouse and human TGase3 enzymes and describe their unusual properties and structures.

### Table I

| Amino acid sequences of tryptic peptides of guinea pig TGase3 |
|-------------------------------------------------------------|
| **50-kDa fragment** |  |
| 1CLGVRSt* | 15 |
| 2VSQVeVGCGPSAVIaV | 15 |
| 3FQQCWFPAGTLNTVLe | 15 |
| 4EG/NDVLNFPVMPFiy | 15 |
| 5GDSVWNHFHWNVNAVFVR | 15 |
| 6LHTVFLLENPLQADDVYMS | 15 |
| **27-kDa fragment** |  |
| 4AOX5GREQAPSIRKFPVNGVLAVQFtE | 15 |
| 5TtAIc | 15 |
| 6ITYQYK | 15 |
| 7FEILPTr | 15 |
| 8D/HLVD/FGSC/LLR | 15 |
| 9DVLNDNPTLTEVLd | 15 |
| 10KPF/GN/VQCLFSNP/LDG | 15 |

*By comparison with other TGase sequences, these two peptides constitute the cleavage activation site of guinea pig TGase3. +This is the amino terminus of the 27-kDa fragment (21) and thus represents the cleavage activation site of guinea pig TGase3.

### Materials and Methods

**Determination of the Amino Acid Sequences of Selected Peptides of Guinea Pig TGase3—**The 50-kDa amino-terminal and 27-kDa carboxyl-terminal fragments of guinea pig TGase3, derived by dispase treatment, were fractionated and purified as described previously (21). Each portion was cleaved with trypsin (Boehringer sequencing grade) at 10 mg/ml in 0.1 M NH4HCO3 with a final enzyme to protein ratio of 1:50 and digested for a total of 4 h at 37 °C. Following drying, the peptides were redissolved in 0.1% aqueous trifluoroacetic acid, fractionated by HPLC, and well resolved peaks with absorbance at both 210 and 350 nm were selected for sequence analysis. Absorbance at 350 nm was taken as an indication of cysteine residues alkylated with 5-N-(iodoacetamidoethyl)amine in p-nitrophenyl sulfonic acid, possibly corresponding to active site peptides (21). Sequence analysis of selected peptides was then performed on an Applied Biosystems 470A protein sequenator using the automated Edman degradation method (21, 33, 44).

**Anchored PCR Cloning Strategies—**Initially, we constructed a series of degenerate oligonucleotide primers based on the available guinea pig TGase3 peptide sequences (see Table II for lists of primers) and used these to amplify DNA obtained from a random-primed cDNA library prepared from mouse epidermal mRNA (45). PCR was done with a commercial DNA amplification reagent kit (Perkin-Elmer Cetus) by following the manufacturer's specifications, using 25 pmol of primers and with conditions of 95 °C (5 min) and 35 cycles of denaturation at 94 °C (0.5 min), annealing at 42 °C (0.5 min), and elongation at 72 °C (1.5 min). The products were fractionated through low melting agarose, excised, and purified through Chroma spin 100 columns (Clontech). The ends of the amplified DNA were filled in with Klenow DNA polymerase (46), subcloned into the pGEM-3z vector (Promega), and then sequenced on both strands by the dideoxy chain termination method with Sequenase 2.0 (United States Biochemical Corp.). Although most sets of degenerate primers did not work, apparently because of the substantial nucleotide sequence differences between guinea pig and mouse TGase3 mRNAs.

### Table II

**Sequences of oligonucleotide primers used for anchored PCR experiments**

| Primer | How used |
|--------|----------|
| Mouse primers: | |
| P1+ | 5'-TGCTGTTTTT/GCCIGCGAC/CTCGAATAC/CAG |
| P3- | 5'-GCGGCGCG/ACAT/CTGAGAANAC/CTG |
| 1a+ | 5'-ACGTGCGTGCCCCACACATACA |
| 1b+ | 5'-AGAACAGCAGGCTATTCCAA |
| P10- | 5'-ACNCCA/GTAT/CTGTTAAGA/CTAC |
| P5+ | 5'-CAAGGGAGGATGCTGCTT/TACC |
| 2a- | 5'-TGTAGGAGGCGGATGATCGT |
| 2b- | 5'-CATGCAATGTCATGATACATGTCAC |
| 3a- | 5'-GTACGACGGGAAATCTGACCTC |
| 3b- | 5'-TTGTCTTCCGCGGTCTT |
| 4a+ | 5'-GGCTTGGGACAAACTCACAAC |
| 4b+ | 5'-GAAAGGAGGAGGCAAGCTT |
| 5a+ | 5'-GAGGAGATACCTGAAGACAGAGAG |
| 5b+ | 5'-ACATGATCGGAGATCCAG |
| Human primers: | |
| 6a+ | 5'-CCCTCCTGGAACATTGTGGT |
| 6b+ | 5'-CAAGGGGATGATGCTT |
| 7a- | 5'-GAAAATACCTCTCAGTAC |
| 7b- | 5'-GTCCAGGGGGTGGAGAAATA |
| 8a- | 5'-ACGGCGGAAAATCTGACCTC |
| 8b- | 5'-ATGCAATCCGAGTCTGTCG |
| 9a+ | 5'-GCAATCCTCATATAGACAGAC |
| 9b+ | 5'-GTACGCTTATGATAGAGAGA |
| TGase2+ | 5'-CCAGCCTGCTGAGGAGGC |
| TGase2- | 5'-CAGTGAGCTAGCAG |

*Corresponding mouse TGase sequences.*
FIG. 1. Strategy for generation of nucleotide information of mouse (top) human (bottom) TGase3 enzymes. In each case, the upper line represents the full-length sequence showing the initiation, termination, and polyadenylation signal sequences. Below are shown bars displaying the primers used (primer sequences are listed in Table II) and extent of sequence information obtained with each PCR step.

Expression of Full-length Human Coding Clone of TGase3 in Yeast—The human TGase3 sequence was initially obtained in three sections (see Fig. 1). The longer 1400-bp clone possessed unique NaeI and KpnI sites toward its 5'- and 3'-ends, respectively, that afforded simple joining of the 500-bp 5'-clone and the 900-bp 3'-clone into a full-length single clone in pGEM-3Z. This was then subcloned into the pYES 2.0 yeast expression vector system (Invitrogen, San Diego, CA), and yeast transformation was performed with the INVSc1 (MATa, his3-A1 leu2 trp-289 ura3-52) strain (47). The selected yeast colonies were grown in 1 liter of minimal medium (0.7% yeast nitrogen base and 0.1 mg/ml each of His, Leu, Trp) containing 2% glucose at 30 °C until an A600 = 0.2 had been achieved (2–3 days). The expression of TGase3 protein was induced by resuspension of yeast cells in fresh minimum medium containing 2% galactose and further incubation for 24 h. Following homogenization with glass beads, TGase activity was assayed in the yeast lysate using established methods (48). Dispase or ethanol activation was performed as described before (21, 49).

Northern Blotting Procedures—Total cellular RNA was prepared from human foreskin epidermis (90), newborn BALB/c mouse epi-
dermis (45), and human and mouse keratinocytes grown to confluence in the presence of low (0.1 mM) or high (0.6 mM) Ca\(^{2+}\) (51, 52). Dr. Ulrike Lichten kindly provided a preparation of hair follicles from 5-day-old mice, from which RNA was isolated. Northern gels using denaturing conditions were loaded with 25 \(\mu\)g of total cellular RNA, performed as described (39), and calibrated with standard RNA size markers (GIBCO/BRL). Northern slot blots were prepared as described (46). In this case, the blots were calibrated with 10-, 1.0-, 0.1-, and 0.01-fmol amounts of probes encoding the full-length TGase1 (27), a 1.0-kbp PCR fragment of 3'-noncoding region for TGase3 added here (see Fig. 1) and a 0.7-kbp 3'-noncoding region of the published sequences of TGase2 (18) (see Table II for the two primers used). Aliquots of 10 \(\mu\)g of the several RNA samples were tested separately with the three TGase-specific probes. All Northern filters were washed with a final stringency of 0.5 \(\times\) SSC at 65°C for 30 min. The resulting x-ray blots were exposed for varying amounts of time in order to facilitate quantification of the abundance of the specific mRNAs with a PhosphorImager (Molecular Dynamics Corp.).

**Computer Analyses of Sequences**—Nucleic acid and protein sequence homologies were performed using the University of Wisconsin software packages compiled by the Wisconsin Genetics Computer Group (53), the IBI Pestell sequence software (version 3.5, International Biotechnologies Inc.), and Geneworks sequence software (IntelliGenetics Inc.), based on published algorithms (54–56).

**RESULTS AND DISCUSSION**

Our initial attempts to locate clones for either mouse or human TGase3 in available \(\lambda\)gt11 libraries using low stringency hybridizations with TGase1 or TGase2 probes or active site probes (27) were unsuccessful. Accordingly, we made TGase3-specific degenerate oligonucleotide probes derived from the amino acid sequences of tryptic peptides of TGase3 isolated from guinea pig epidermis. The implicit assumption was that the guinea pig, mouse, and human TGase3 proteins would share high degrees of sequence homology, as found for the TGase1 (13, 34) and TGase2 (18) systems.

**Amino Acid Sequences of Guinea Pig TGase3 Tryptic Peptides**—Although the 27-kDa fragment resulting from dispase treatment yielded a clean amino acid sequence for 28 cycles, corresponding to its amino terminus and the activation site of proteolytic cleavage, no useful information on the larger catalytic 50-kDa portion was possible (21). Accordingly, using larger quantities, both portions were cleaved to completion with trypsin, and selected well resolved peptides, especially those containing cysteine residues, were chosen for sequencing. In this way, sequences from a total of 12 tryptic peptides (six from each of the 50- and 27-kDa portions) and the amino terminus of the 27-kDa portion were obtained (Table I). These represented 180 sequenced residues or about 25% of the total protein. Peptides 1 and 3 (order 3–1) are recognizable as constituting the active site region, based on comparisons with the known sequences of the TGase family members (4, 27). The minor amino acid substitutions in this active site region in relation to the other family members are diagnostic for the TGase3 system (see Table V).

**Cloning by Anchored PCR and Deduced Amino Acid Sequences of Mouse and Human TGase3 Proteins**—Degenerate oligonucleotide probes based on the above amino acid sequences of guinea pig TGase3 failed to identify positive clones

**Fig. 3. Expression of functional recombinant TGase3 in yeast and activation by dispase.** Aliquots (0.1 ml) of yeast cell extracts were assayed for TGase activity using \(^{14}\)C]putrescine and succinylated casein as substrate (48). Human trunk epidermal protein TGase3 was extracted and assayed as described (21). Open box = activity in cell lysates; shaded box = assayed in 20% ethanol; stippled box = pretreatment of lysates with dispase (1 mg/ml, 15 min). Yeast = pYES vector alone, no insert; rTGase3 = vector + full-length human cDNA; skin TGase3 = enzyme isolated from human epidermis.

**Fig. 4. Size, abundance, and expression characteristics of human and mouse TGase3 mRNAs.** A, Northern blots. Aliquots of 25 \(\mu\)g of total cellular RNA from human foreskin epidermis (lanes 1–4), newborn mouse epidermis (lane 5), or 5-day-old mouse hair follicles (lane 6) were probed with: a 58 nucleotides antisense degenerate oligonucleotide encoding active site sequences (lane 1) (27), a 175-bp 3'-noncoding probe of a TGase1 cDNA clone (lane 2) (27), a 0.7-kbp 3'-noncoding portion of TGase2 generated by PCR (lane 3) (see Table II), a 1-kbp cDNA probe encoding 3'-noncoding sequences of human TGase3 (lanes 4–6) (see Fig. 1). The individual strips were exposed for 14 days (lane 1), 2 days (lane 2), 4 days (lane 3), and 23 days (lanes 4–6). Positions of migration of RNA size markers are shown. B, Northern slot blots. Aliquots of 10 \(\mu\)g of RNA from the sources shown were probed with the above TGase-specific probes. X-ray films were exposed for several different times; this figure shows one exposure (for 6 days) only.
Sequence of Transglutaminase 3

| Table III |
|----------------|
| **Abundances of mouse and human TGase mRNAs (fumol/10 µg of total cellular RNA)** |
| The data are the average (±S.D.) of four different experiments, except for mouse hair follicles (one batch, a gift of Dr. Ulrike Lichti). These values were calculated from the Northern slot blots (Fig. 4B), based on calibrations of 10, 1, 0.1, and 0.01 fumol of known cloned probes. |
| TGase1 | TGase2 | TGase3 |
| Human foreskin epidermis | 0.45 ± 0.06 | 0.071 ± 0.023 | 0.078 ± 0.022 |
| Human, cultured, low Ca²⁺ | 0.91 ± 0.08 | 0.135 ± 0.015 | <0.002 |
| Human, cultured, high Ca²⁺ | 1.05 ± 0.11 | 0.112 ± 0.021 | 0.008 ± 0.002 |
| Mouse epidermis | 0.57 ± 0.01 | 0.053 ± 0.017 | 0.081 ± 0.019 |
| Mouse, cultured, low Ca²⁺ | 1.37 ± 0.16 | 0.121 ± 0.028 | <0.002 |
| Mouse, cultured, high Ca²⁺ | 2.53 ± 0.39 | 0.102 ± 0.026 | 0.010 ± 0.005 |
| Mouse hair follicles | 0.27 | 0.029 | 0.087 |

in available mouse or human αgt11 cDNA libraries. However, when such oligonucleotides were employed in primer extension experiments with poly(A)-enriched RNA from newborn mouse or human foreskin epidermis, weak signals corresponding to the presumed size of the TGase3 mRNAs (27) were found. Therefore, we used the oligonucleotide primers to amplify by PCR the DNA extended by primer P3-- (Table II). One pair of primers (P1+, P3--; Fig. 1, Table II) yielded a product of 292 bp and was subcloned into pGEM-3z. About 5% of such clones contained TGase-like sequences, including the active site region, which were identical to peptides 3'-1 of the available guinea pig tryptic peptides (Table I). This finding afforded confidence that we were indeed amplifying the mouse TGase3 mRNA system. However, when this 292-bp probe was used to screen our αgt11 epidermal foreskin cDNA library (27), no clones were found, possibly due to the very low abundance of its mRNA (and see below). Accordingly, we used this exact sequence data of the 292-bp probe to extend the mouse TGase3 sequence by use of RNA-mediated anchored PCR as described under "Materials and Methods." First, we used one set of specific nested primers and another degenerate primer from the guinea pig peptide information (1a+=P10→ and 1b+=P10--; P5+=/2a-- and P5+=/2b--) (Fig. 1, Table II). The remaining of the 5'-end up to the capsite was recovered by primer extension, tailing with dG, and PCR amplification in two steps with nested primers (oligo(dC)/3a--; then oligo(dC)/3b--) (Fig. 1). The 3'-end sequence information was recovered in two steps by use of primer extension with a random hexamer, followed by tailing with dG. The cDNA products were amplified by PCR in two steps with two sets of nested primers (4a+/oligo(dC); then 4b+/oligo(dC)) and (5a+/oligo(dC); then 5b+/oligo(dC)) (Fig. 1, Table II).

The human TGase3 sequence was generated in essentially the same manner in three steps (Fig. 1), except that an oligo(dT) primer was used to generate the full-length 3'-noncoding information.

A series of further RNA-mediated anchored PCR experiments was performed using primers that cross over those shown in Fig. 1 and Table II, in order to confirm and check the sequences for PCR-induced sequence mutations (lists of primers used are not shown). The nature of seven ambiguous nucleotides were resolved in additional PCR experiments.

The available nucleotide sequence information consists of 2297 nucleotides for mouse, including the entire 5'-noncoding information, but incomplete 3'-noncoding sequences (Fig. 2). The human data extend for 2645 nucleotides and are assumed to be near full-length because of the inclusion of the polyadenylation signal sequence (Fig. 2); thus its estimated mRNA size is about 2.8 kb. In both cases, there is open reading frame of 2079 bp, so that both proteins contain 692 amino acids of calculated molecular mass 77.1 kDa (mouse) and 76.6 kDa (human), which are very close to the values added for guinea pig TGase3 by analytical ultracentrifugation and SDS-polyacrylamide gel electrophoresis experiments (20, 21). Interestingly, mouse TGase3 is near neutral in charge (PI 6.5) compared with human TGase3 (PI 5.6), findings that are also consistent with earlier chromatographic observations (21). Recombinant Clones of TGase3 in Yeast Produce a Highly Active Enzyme—We performed experiments to confirm that the cDNA sequences adduced by RNA-mediated anchored PCR methods described above in fact encode a functional TGase3 enzyme system. An intact full-length cDNA for the human TGase3 was assembled into a pYES yeast expression vector and transfected into yeast. Following induction, cells were lysed and assayed for TGase activity (Fig. 3). Yeast clones with an empty expression vector did not express detectable levels of TGase activity (<50 cpm), consistent with the view that yeast do not contain TGases (1). However, clones containing the full-length human TGase3 cDNA had...
significant TGase activity, comparable with that which can be extracted from human trunk epidermis (21) (Fig. 3). Moreover, we found that the activity of the yeast expression product could be activated 7.5-fold by ethanol and 15-fold by dispase (Fig. 3), values which are also very comparable with those found in the activation of the native epidermal TGase3 pro-enzyme (Fig. 3; Ref. 21). Taken together, these data establish that the sequences determined in the present work encode a functional human pro-TGase3 enzyme system with expected properties similar to those of the native epidermal protein.

Amino Acid Sequences of the Human, Mouse, and Guinea Pig TGase3 Proteins Are Not Highly Conserved—In Table IV are listed the several tryptic peptides generated for guinea pig TGase3 that were found in the mouse and human TGase3 sequences. The comparisons further extend confidence for the correct identity of these sequences. In addition, the availability of the amino-terminal information of the 27-kDa fragment formed on proteolytic cleavage activation enabled identification of the activation region in the mouse and human TGase3 proteins as well (Table IV).

Previous studies have shown that the sequences of human TGase1 (13, 34) and TGase2 (18) enzymes have been very highly conserved; sequences show identities of about 95% and homologies of about 97%. In contrast, the data of Fig. 2 reveal that mouse and human TGase3 sequences have deviated more widely (Table V). Overall, the sequences show 75% identity and 84% homology, with the 27-kDa fragment generated following proteolytic activation somewhat less conserved: 71% identity and 81% homology. Interestingly, the amino acid sequences of the available tryptic peptides of the guinea pig TGase3 show far more variation from mouse and human (Table V), such that the 27-kDa fragment displays as little as 45% sequence identity in available comparable sequences. Most of the variations have occurred in the vicinity of the proteolytic activation site (Table IV), which may mean that the different species have evolved alternate mechanisms for proteolytic activation of the TGase3 pro-enzyme. These sequence variations can account for the difficulties we initially encountered in generating mouse and human sequence information using the guinea pig data.

Comparisons Show That Human TGase3 Is Distantly Re-
FIG. 5. Updated alignment of amino acid sequences of human TGase-like proteins. The sequences were aligned to maximize homologies according to the protocol of Pearson and Lipman (56). Alignments of the amino-terminal sequences are arbitrary. The arrowhead marks the presumed site of proteolytic cleavage required for activation of TGase3. Homology and identity scores (see Table 4.2) correspond to the positions of known human TGase-like proteins characterized to date. The approximately known intron boundaries conserved in TGase-like proteins.
FIG. 6. Predicted structural features of human TGase3. These were performed using the IBI Pustell and Intelligenetics Geneworks software packages and are based on the analytical methods of Refs. 53-56. The predicted secondary structure, flexibility, and hydrophilicity profiles are shown. The arrows centered at residue 475 demarcate the point of cleavage activation of the proenzyme and denote a prominent turn region of highest flexibility and hydrophilicity in the entire protein.

lytic activation. Overall homology and identity scores between the five TGases are shown in Table V. We have chosen to analyze only those sequences bounded by the conserved intron locations identified previously (13), which presumably delineate the conserved structural regions of the TGases. Each TGase chain deviates widely at its termini in both sequence and length, which thus does not admit meaningful comparisons (Fig. 5). The human TGase3 protein is most closely related to TGasel and TGase2, and more similar to band 4.2 than factor XIIIa, although the band 4.2 is least related to the other TGases. The evolutionary significance of these findings is not yet clear; however, analysis of gene structures has suggested to other investigators that the band 4.2 protein has either been least conserved during evolution, perhaps because it is not a functional enzyme (22, 23) or because it evolved before TGasel, TGase2, and factor XIIIa (38). Analyses of the gene structure and chromosomal location of human TGase3 are in progress and may clarify these evolutionary questions.

The TGase3 Proteins Consist of Two Globular Domains Separated by a Flexible Hinge at the Site of Activation—Secondary structural analyses (53-56) of the human and mouse TGase3 proteins reveal multiple interspersed regions of turns, sheet structures, and $\alpha$-helix, in both the 50-kDa amino-terminal and 27-kDa carboxyl-terminal portions (Fig. 6). In general, these features suggest a folded compact configuration (13, 27). However, the 12-residue insertion immediately following the cleavage site required for activation of the pro-enzyme describes a prominent protein turn that is surrounded by sequences that are the most hydrophilic and flexible in the entire protein (Fig. 6). Thus this sequence describes a flexible hinge region and is likely to be located near the surface of the molecule. From these observations, we can infer that the intact TGase3 molecules adopt an elongated shape consisting of two globular domains, a larger amino-terminal and a smaller carboxyl-terminal, that are separated by a flexible hinge corresponding to the activation site. This is flanked by highly polar residues, predicted to lie near the surface of the protein, that may be involved in recognition by and accessible to the activating protease(s). We predict that following cleavage, the hinge region collapses, promoting a more compact configuration that greatly enhances catalytic activity of the TGase3 molecule. These predictions are entirely consistent with the hydrodynamic observations that the guinea pig TGase3 pro-enzyme is an elongated molecule (21).

In contrast, no other members of the TGase family possess a flexible hinge region (Fig. 5), and all are predicted to adopt a compact globular form (13, 27). Further structural studies on expressed forms of human TGase3 are in progress to explore the implications of these predictions on its activation, enzymic function, and substrate specificity.

Acknowledgments—We are indebted to George Poy who synthesized all of the oligonucleotide primers used in this work and Gary Beddome for performing HPLC separation and step wise sequence analysis of tryptic peptides.

REFERENCES
1. Folk, J. E. (1983) Adv. Enzymol. Relat. Areas Mol. Biol. 54, 1-54
2. Lorand, L., and Conrad, S. M. (1984) Mol. Cell. Biochem. 58, 9-35
3. Polakowska, R. R., and Goldsmith, L. A. (1991) in Physiology, Biochemistry, and Molecular Biology of the Skin (Goldsmith, L. A., ed) pp. 168-201, Oxford University Press, New York
