Structure of the Gene for Human Transglutaminase 1*

(Received for publication, February 27, 1992)

Kiyofumi Yamashishi, Juhji Inazawa, Foo-Min Liew, Kyoko Nonomura, Takeshi Ariyama, Hirokazu Yasunot, Tatsuo Abe, Hiroshi Doj, Jiro Hirano, and Shoji Fukushima

From the Departments of Dermatology and Hygiene, Kyoto Prefectural University of Medicine, Kamigyo-ku, Kyoto 602 and the Advanced Skin Research Institute, Kohoku-ku, Yokohama 223, Japan

Transglutaminase 1 (TG1) is an enzyme that is expressed and activated during terminal differentiation of keratinocytes and synthesizes cornified envelope by a cross-linking reaction. The gene encoding human TG1 was isolated from human genomic DNA and characterized. It spans 14.3 kilobase pairs and is composed of 15 exons. All exon-intron junctional sequences conformed to the canonical GT-AG rule. The translation start was located in the second exon. The active site Cys residue of the enzyme was in exon 7. The coding sequence for human TG1 was comprised of 2454 nucleotides identical with the published human TG1 cDNA sequence (Yamanishi, K., Liew, F.-M., Konishi, K., Yasuno, H., Doi, H., Hirano, J., and Fukushima, S. (1991) Biochem. Biophys. Res. Commun. 175, 906–913). The sizes of exons from 3 to 14 were markedly conserved between the genes for the human TG1 and factor XIIIa, another member of the transglutaminase family. The one major and two minor transcription initiation sites of the TG1 gene were determined by primer extension. The 5'-flanking region of the human TG1 gene showed features of a housekeeping gene and contained potential regulatory motifs, including elements found in keratinocyte-related genes. The chromosome sublocalization of the TG1 gene was assigned to 14q11.2.

Transglutaminases (EC 2.3.2.13) catalyze acyl transfer reactions between γ-carboxamide groups of peptide-bound glutamine residues and the primary amino groups of various amines, including the ε-amino group of lysine in peptides to form ε-(γ-glutamyl) lysine cross-linking (1, 2). This family of enzymes share the common structure of an active site, Tyr-Gly-Glu-Cys-Trp, operated by a cysteine-thiol active center. The nucleotide sequence of the enzyme has been deduced from the nucleotide sequences of the clones. The enzyme is conserved between the genes for the human TG1 and factor XIIIa, but exons 2 and 15 are less homologous. We also analyzed the transcriptional regulatory motifs of the TG1 gene determined by primer extension. The 5'-flanking region of the human TG1 gene showed features of a housekeeping gene and contained potential regulatory motifs, including elements found in keratinocyte-related genes. The chromosome sublocalization of the TG1 gene was assigned to 14q11.2.

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

** To whom correspondence should be addressed: Dept. of Dermatology, Kyoto Prefectural University of Medicine, Kamigyo-ku, Kyoto 602, Japan. Tel./Fax: 81-75-251-5872.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBankTM/EMBL Data Bank with accession number(s) D10339-D10353.

1 The abbreviations used are: TG1, transglutaminase 1; kb, kilobase pairs; NHEK, normal human epidermal keratinocytes; SDS, sodium dodecyl sulfate; TPA, 12-0-tetradecanoylphorbol-13-acetate.

2 The abbreviation used is: TG1, transglutaminase 1; kb, kilobase pairs; NHEK, normal human epidermal keratinocytes; SDS, sodium dodecyl sulfate; TPA, 12-0-tetradecanoylphorbol-13-acetate.

3 The abbreviation used is: TG1, transglutaminase 1; kb, kilobase pairs; NHEK, normal human epidermal keratinocytes; SDS, sodium dodecyl sulfate; TPA, 12-0-tetradecanoylphorbol-13-acetate.

4 The abbreviation used is: TG1, transglutaminase 1; kb, kilobase pairs; NHEK, normal human epidermal keratinocytes; SDS, sodium dodecyl sulfate; TPA, 12-0-tetradecanoylphorbol-13-acetate.

5 The abbreviation used is: TG1, transglutaminase 1; kb, kilobase pairs; NHEK, normal human epidermal keratinocytes; SDS, sodium dodecyl sulfate; TPA, 12-0-tetradecanoylphorbol-13-acetate.

6 The abbreviation used is: TG1, transglutaminase 1; kb, kilobase pairs; NHEK, normal human epidermal keratinocytes; SDS, sodium dodecyl sulfate; TPA, 12-0-tetradecanoylphorbol-13-acetate.

7 The abbreviation used is: TG1, transglutaminase 1; kb, kilobase pairs; NHEK, normal human epidermal keratinocytes; SDS, sodium dodecyl sulfate; TPA, 12-0-tetradecanoylphorbol-13-acetate.

MATERIALS AND METHODS

Cell Culture—Secondary cultures of normal human epidermal keratinocytes (NHEK) derived from foreskin were obtained from Kur-
are numbered

Screening of Human Genomic DNA Libraries—A human cosmid library (catalog No. HL1095m) was obtained from Clontech. The source of the genomic DNA was human placenta. DNAs ranging from 35 to 50 kb were inserted into the BamHI site of cosmid pWE15. The host bacterial strain was Escherichia coli NM545. A 2.7-kb human TG1 cDNA (0.1 μg) was prepared by BamHI-EcoRI digestion of pHETG-M (14) and labeled with 32P by nick translation. A library of 1 × 10^6 clones was screened by filter hybridization with the probe at 65 °C for 16 h in 1 M NaCl, 50 mM Tris-HCl (pH 8.0), 10 mM EDTA, 0.1% SDS, and 5 × Denhardt's solution (1 × Denhardt's solution: 0.02% polyvinylpyrrolidone, 0.02% Ficoll, 0.02% bovine serum albumin). Subsequently, the filters were washed three times for 10 min at room temperature in 2 × SSC (1 × SSC: 0.15 M NaCl, 0.015 M sodium citrate), 0.05% SDS, and twice for 1 h at 68 °C in 1 × SSC, 0.1% SDS. The filters were exposed to Fuji AIF RX film.

Subcloning of Cosmid DNA and Restriction Analysis—The isolated clone, pHETG, was characterized by restriction digestion with BamHI, EcoRI, and PstI, followed by Southern hybridization with the probe as described above, or with oligonucleotides used for sequencing TG1 cDNA (14). A 5.4-kb EcoRI fragment was subcloned into plasmid pGEM3Z (Promega Biotec) to generate pE6, as well as an EcoRI-BamHI fragment of 5.7 kb and a PstI fragment of 9.7 kb to yield plasmid pEB5 and pP9, respectively. To establish a restriction map, the inserts were digested by the same restriction enzymes.

Exon Mapping by Polymerase Chain Reaction—The exons were mapped within the TG1 cDNA. Oligonucleotide primers corresponding to sequences previously published for sequencing the TG1 cDNA (14). The 17 sets of primers were selected for amplification of each subclone plasmid DNA as the template. The polymerase chain reaction products were mapped by digestion with PstI. The size of the amplified DNA was determined by 1% agarose gel electrophoresis and computer-aided gel image analysis.

DNA Sequencing—In addition to the oligonucleotide primers used for sequencing TG1 cDNA, further primers were designed and synthesized for walking primer sequencing of the genomic DNA subclones. All DNA fragments were sequenced on both strands except the probe region. The isolated subclones, pE6, pEB5, and pP9 were further analyzed by restriction digestion and polymerase chain reaction with the primers including these oligonucleotides. A map of the human TG1 gene shown in Fig. 1, comprised 15 exons and 14 introns and spans approximately 14.3 kb. The DNA sequences of the exons, exon-intron boundaries, and 5′-upstream region were determined. As shown in Fig. 2, the first exon contained the 5′-untranslated region. In the second exon, there were nucleotide sequences encoding an initial 106-amino acid stretch that was unique to TG1 (14). The active site region of the enzyme, Tyr-Gly-Gln-Cys-Trp, was in exon 7. The last exon contained 229 nucleotides of a coding sequence followed by the translation stop codon (TAG) and a 175-nucleotide 3′-untranslated region. In this region, a possible polyadenylation signal (AATAAA) was positioned at nucleotides 2221 and 2226 (28). The exon-intron border DNA sequences of the TG1 gene are summarized in Fig. 3. The 5′ intron border invariably starts with GT, and the 3′ border terminates with AG. The rule is consistent with reported donor and acceptor consensus sequences (29).

The genomic structure of human factor XIIIa (plasma transglutaminase) has been characterized by Ichinose and Davie (15). The amino acid and nucleic acid homology between the human TG1 and factor XIIIa is 42 and 54%, respectively. The span of the factor XIIIa gene is different from that of the human TG1. The factor XIIIa gene extends from 160 kb, in contrast to 14.3 kb of the human TG1. However, the number of exons are also 15, and the sizes of

RESULTS AND DISCUSSION

Identification and Structure of the Human TG1 Gene—A human genomic cosmid library was screened with nick-translated fragments of the human TG1 cDNA representing all the coding regions from pHETG-M (14). Of 1 × 10^6 screened clones, one clone was identified, pHETG, containing the complete TG1 gene. Cosmid pHETG was subcloned by restriction enzyme digestion and Southern hybridization with the cDNA fragment and defined oligonucleotides corresponding to regions throughout the TG1-cDNA as the probes (14). The isolated subclones, pE6, pEB5, and pP9 were further analyzed by restriction digestion and polymerase chain reaction with the primers including these oligonucleotides. A map of the human TG1 gene shown in Fig. 1, comprised 15 exons and 14 introns and spans approximately 14.3 kb. The DNA sequences of the exons, exon-intron boundaries, and 5′-upstream region were determined. As shown in Fig. 2, the first exon contained the 5′-untranslated region. In the second exon, there were nucleotide sequences encoding an initial 106-amino acid stretch that was unique to TG1 (14). The active site region of the enzyme, Tyr-Gly-Gln-Cys-Trp, was in exon 7. The last exon contained 229 nucleotides of a coding sequence followed by the translation stop codon (TAG) and a 175-nucleotide 3′-untranslated region. In this region, a possible polyadenylation signal (AATAAA) was positioned at nucleotides 2221 and 2226 (28). The exon-intron border DNA sequences of the TG1 gene are summarized in Fig. 3. The 5′ intron border invariably starts with GT, and the 3′ border terminates with AG. The rule is consistent with reported donor and acceptor consensus sequences (29).

The genomic structure of human factor XIIIa (plasma transglutaminase) has been characterized by Ichinose and Davie (15). The amino acid and nucleic acid homology between the human TG1 and factor XIIIa is 42 and 54%, respectively. The span of the factor XIIIa gene is different from that of the human TG1. The factor XIIIa gene extends from over 160 kb, in contrast to 14.3 kb of the human TG1. However, the number of exons are also 15, and the sizes of

FIG. 1. Structure of the human TG1 gene. The structure of the gene (TGM1) is schematically represented by the bar at the top of the diagram. Exons are indicated by boxes numbered 1 to 15; untranslated regions are shown in plain coding regions by filled boxes. Intron regions are numbered I to XIV. Restriction map and subcloning strategy are also shown. E, EcoRI; B, BamHI; P, PstI.
exons from 3 to 14 were completely conserved between the factor XIIIa and TG1 genes except for minor differences in exons 4 and 12 (Table I). Furthermore, the nucleotide sequences encoding the active sites of these enzymes are both in exon 7, and the positions are perfectly conserved in the 26–40 nucleotides upstream from the end of the exon. The amino acid sequences encoded by the neighboring exons are also relatively conserved. On the other hand, exons 2 and 15 differ

**Human Transglutaminase 1 Gene**

**Fig. 2. Nucleotide sequence of the human TG1 gene.** The number of the first nucleotide in each line is shown to the left. The major transcription start site obtained in primer extension (see Fig. 5) is assigned base +1, representing the first base of exon 1. The deduced amino acid sequence is shown below the nucleotide sequence. The asterisk indicates a termination codon. The sites for introns 1–14 are indicated by slashes. The length of each intron and the nucleotide sequences found at the 5' and 3' borders are shown. The putative polyadenylation signal is underlined. The polyadenylation site is shown by a dot above the letter.
The consensus sequences are given in the sequences of the intron-exon boundaries from Fig. 2 are summarized.

With TPA, 2.8-kb TG1 transcripts were markedly induced. These findings suggest that these two transglutaminases originate from a common lineage and then developed separately.

The consensus sequences are given in the bottom line.

### TABLE I

Comparison of exon structures between TG1 and factor XIII genes

<table>
<thead>
<tr>
<th>Exon</th>
<th>Exon size</th>
<th>Homology</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TG1*</td>
<td>FXIII*</td>
</tr>
<tr>
<td></td>
<td>base pairs</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>94</td>
<td>ND(^c)</td>
</tr>
<tr>
<td>2</td>
<td>321</td>
<td>148</td>
</tr>
<tr>
<td>3</td>
<td>189</td>
<td>189</td>
</tr>
<tr>
<td>4</td>
<td>249</td>
<td>252</td>
</tr>
<tr>
<td>5</td>
<td>119</td>
<td>119</td>
</tr>
<tr>
<td>6</td>
<td>108</td>
<td>108</td>
</tr>
<tr>
<td>7</td>
<td>175</td>
<td>175</td>
</tr>
<tr>
<td>8</td>
<td>139</td>
<td>139</td>
</tr>
<tr>
<td>9</td>
<td>104</td>
<td>104</td>
</tr>
<tr>
<td>10</td>
<td>89</td>
<td>89</td>
</tr>
<tr>
<td>11</td>
<td>154</td>
<td>154</td>
</tr>
<tr>
<td>12</td>
<td>282</td>
<td>288</td>
</tr>
<tr>
<td>13</td>
<td>161</td>
<td>161</td>
</tr>
<tr>
<td>14</td>
<td>137</td>
<td>137</td>
</tr>
<tr>
<td>15</td>
<td>404</td>
<td>408</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\) From Fig. 2.  
\(^b\) Ichinose and Davie (15).  
\(^c\) ND, not determined.  
\(^d\) —, less homology.

Fig. 3. Intron-exon boundaries of the human TG1 gene. The sequences of these boundaries from Fig. 2 are summarized. These consensus sequences are given in the bottom line.

Fig. 4. Induction of the TG1 mRNA by TPA in cultured NHEK. NHEK were cultured as described under "Materials and Methods" and treated with (right) or without (left) 2 nM TPA. After 20 h of incubation, total RNAs were prepared and Northern hybridized using \(^32\)P-labeled oligonucleotide R5'-2 as the primer and 30 \(\mu\)g of yeast tRNA in lane 1 and total RNA of TPA-treated NHEK in lane 2. One major (large arrow) and two minor products (small arrow) were detected (for location, see Figs. 2 and 6). As a sizing reference, pE6 was sequenced using the same primer.

Fig. 5. Determination of the transcription initiation sites by primer extension. Primer extension was performed using \(^32\)P-labeled oligonucleotide R5'-2 as the primer and 30 \(\mu\)g of yeast tRNA in lane 1 and total RNA of TPA-treated NHEK in lane 2. One major (large arrow) and two minor products (small arrow) were detected (for location, see Figs. 2 and 6). As a sizing reference, pE6 was sequenced using the same primer.

Fig. 6. Nucleotide sequence of the 5'-end region of the human TG1 gene. TATA-like sequences are shown by boldface letters. Putative control element consensus sequences of FPI and CCAAT box and binding motifs of S1 (reverse complements), CTF/NF1, KER1, AP1, and KTF-1 are underlined. The sequence of the FPI core consensus is shown by dots above the letters. The major transcription start site is indicated also by boldface letters (nucleotide +1). Two minor start sites are shown by asterisks above the letters.

-820 GAACTCGAAGTTGGAAAAAGAGATCCTGCTCAGACTGAGAGAGCTCTTCGTCAAGT
-760 GAGAGAAGGTTGCTCTCTTGAGAGGGCTACCTGCCACCCCGCTGTGAGGGTGG
-700 AAAAGCCCTAACCCTTACCTGCTCTGACCTGACCGGCTGTAGCTCAATG
-640 ATTCTCGCCCTCTGTGAGTGTCGACTGCGAGGCTGTATGCTCCCTGCGG
-590 CACCTGAGACGAGGAGTGACTGACTGAGGCTGTAGCTCAATGCTCCCTGCGG
-580 CTCTGAGACGAGGAGTGACTGACTGAGGCTGTAGCTCAATGCTCCCTGCGG

\(^a\) From Fig. 2.  
\(^b\) Ichinose and Davie (15).  
\(^c\) ND, not determined.  
\(^d\) —, less homology.

in size and nucleotide sequence between these two genes. Exon 2 of the factor XIIIa gene encodes the activation peptide region that is released by thrombin, whereas that of the TG1 gene encodes a unique hydrophilic region that is rich in Arg and Ser. Though the function of the region is unknown, this comparison suggests an important role of the region in cell envelope formation in keratinocytes, a specific function of the epidermal enzyme. Exon 15 of the factor XIIIa gene has the 1.5-kb 3'-untranslated region. Even the coding region of the exon is less homologous to that of the TG1 gene. As shown in Fig. 1, exon 15 of TG1 is located 4.8-kb distant from exon 14. These structural features imply that exon 15 is also involved in a keratinocyte-specific function of the enzyme. These findings suggest that these two transglutaminases originated from a common lineage and then developed separately.
Chromosome 14 at the q11.2 region. with the addition of exons 2 and 15 for specialized functions. 

Transcribed RNA Message—The coding sequence of the human TG1 comprised 2454 nucleotides, including the stop codon. There was perfect sequence matching between the genomic data shown here and our cDNA data (14). Apart from minor differences (12), the same was true of the cDNA sequences reported previously from other laboratories (11, 12). Constitutive expression levels of the TG1 mRNA were very low in cultured normal human epidermal keratinocytes (NHEK) (Fig. 4). In mouse epidermal cells, TPA reportedly induces TG1 activity (30). However, the effect of TPA on the mRNA expression of TG1 had not been examined. As shown in Fig. 4, we found that TPA markedly induced the mRNA level of TG1 in NHEK when the cells were treated with TPA for 20 h. Therefore, the site of transcription initiation of the TG1 gene was determined by primer extension using RNA prepared from TPA-treated NHEK. The extension products were analyzed using 32P-labeled primer R5'-2 (20-mer). The product was sized by comparison with bands generated by dideoxy sequencing of pE6 with the R5'-2 primer. As a result, one major and two minor extension products were obtained (Fig. 5). The major product was assigned the number 1 (G) in the genomic DNA sequence of the TG1 in Figs. 2 and 6. The positions of the other sites were at positions −10 and +16, as indicated in Fig. 6. No products were detected in control experiments in which yeast tRNA was substituted for the human RNA.

Putative Regulatory Sequence Elements—To find sequence elements that might be involved in the control of human TG1 gene expression, the 5′-flanking region was analyzed using a computer program that identified known regulatory motifs. A sequence closely resembling the AP1 binding consensus (TGAGTCAG) was located at position −536 (Fig. 6), where 7 of 8 nucleotides were identical to the AP1 consensus motif. AP1 sites have been identified in a number of TPA-responsive genes, such as those of collagenase, stromelysin, and interleukin-2. Indeed, the expression of the TG1 mRNA is markedly induced by TPA in cultivated human keratinocytes (Fig. 4). The AP1-like sequence identified in the 5′-upstream region of TG1 may be responsive to the signals generated by TPA. TPA directly activates protein kinase C and functions as an analog of the physiological protein kinase C activator, diacylglycerol (39). Recently, Osada et al. have identified nPKCγ, a new member of the protein kinase C family, that is predominantly expressed in the lung and skin (40). Therefore, nPKCγ may mediate the induction of the TG1 mRNA by TPA.

The sequence between −14 to −27 of the TG1 gene containing CATAAA was 71% homologous to the transcriptional element FPI that has been identified in human papilloma virus 18 (HPV18) (41). FPI has a core sequence, TTAGTCA, that binds a family of AP1. The sequence is necessary to enhance HPV18 expression in cooperation with KRF-1, a keratinocyte-specific transcription factor. The sequence TGAGTCA, in the FPI-like sequence of TG1 gene, was closely homologous to the core element of FPI. These findings suggest that a common transcription factor of the AP1 family is involved in the expression of the TG1 gene, as well as that of HPV18 that replicates in differentiating keratinocytes.

TG1 is an enzyme expressed specifically in keratinocytes. Therefore, other transcriptional elements related to keratinocyte-specific expression were examined. KER1 is a keratinocyte-specific transcription factor in human K14 keratin gene (42). The factor binds a 10-base pair palindrome, GCCTGAGGCT. In the 5′-upstream region of the TG1 gene, there were two sites at positions −266 and −516 that were 80% homologous to the KER1 consensus motif. K14 keratin is not differentiation-specific, but is rather expressed in basal keratinocytes (43). Therefore, these sequences homologous to the KER1-binding motif might be involved not in the differentiation-specific but in the tissue-specific expression of TG1 gene. At position −678, a sequence resembling a motif of KTF-1 binding was also found. The KTF-1 binding sequence is ACCCTGAGGCT. This factor is a Xenopus embryo nuclear factor that binds a keratin gene expressed at midblastula transition in the outer ectoderm (44). Sequences related to these factors might be required in tissue-specific expression of the TG1 gene, during the process of development and maturation of the epidermis.

Chromosomal Mapping of the TG1 Gene—To determine the chromosomal location of the TG1 gene, fluorescence in situ hybridization was performed using pHETG-M bearing human TG1-cDNA (14). A total of 46 metaphase cells was examined, and 26 of them showed either single or double spots of fluorescent signals on band q11.2 of chromosome 14 (Fig. 7). The distribution pattern of the signals on 14q11.2 was as follows: double spots on one homolog with or without a single spot on the other (23.1%) or a single spot on both homologs and on only one homolog (76.9%). These results localized the human TG1 gene to 14q11.2. Recently, Poakowska et al. (17) assigned the human TG1 gene (TGMI) to chromosome 14 using a mapping panel of somatic cell hybrids. Our data

Fig. 7. Fluorescence in situ hybridization of the human TG1 gene. Double fluorescent spots were detected on R-banded chromosome 14 at the q11.2 region.
obtained by fluorescence in situ hybridization clearly indicate the sublocalization of TGMI to the region 14q11.2.

In this study, we characterized the structure of human TG1 gene including the 5′-upstream region. The TG1 gene is a key to help understand the molecular mechanisms of terminal keratinocyte differentiation. We intend to clarify the regulation of the gene expression of the TG1 by identifying unique elements determining tissue- and differentiation-specific transcription of the gene.

Acknowledgment—We thank Prof. Akio Iwashima, Department of Biochemistry, Kyoto Prefectural University of Medicine, for encouragement.

REFERENCES