

(19)



(11)

**EP 1 317 548 B1**

(12)

**EUROPEAN PATENT SPECIFICATION**

(45) Date of publication and mention of the grant of the patent:  
**18.11.2009 Bulletin 2009/47**

(51) Int Cl.:  
**C12N 15/54** <sup>(2006.01)</sup>      **C12N 9/10** <sup>(2006.01)</sup>  
**C12Q 1/48** <sup>(2006.01)</sup>      **C07K 16/18** <sup>(2006.01)</sup>  
**G01N 33/53** <sup>(2006.01)</sup>

(21) Application number: **01967485.2**

(86) International application number:  
**PCT/GB2001/004120**

(22) Date of filing: **14.09.2001**

(87) International publication number:  
**WO 2002/022830 (21.03.2002 Gazette 2002/12)**

**(54) TRANSGLUTAMINASE GENE PRODUCTS**

TRANSGLUTAMINASE GENPRODUKTE

PRODUITS GENIQUES DE TRANSGLUTAMINASE

(84) Designated Contracting States:  
**AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU  
MC NL PT SE TR**

(74) Representative: **Arth, Hans-Lothar et al**  
**ABK Patent Attorneys**  
**Jasminweg 9**  
**14052 Berlin (DE)**

(30) Priority: **15.09.2000 GB 0022768**  
**16.05.2001 GB 0111995**

(56) References cited:  
**WO-A-99/10507 US-A- 5 726 051**

(43) Date of publication of application:  
**11.06.2003 Bulletin 2003/24**

(73) Proprietor: **Zedira GmbH**  
**64293 Darmstadt (DE)**

- **AESCHLIEMANN ET AL: "ISOLATION OF A cDNA ENCODING A NOVEL MEMBER OF THE TRANSGLUTAMINASE GENE FAMILY FROM HUMAN KERATINOCYTES" JOURNAL OF BIOLOGICAL CHEMISTRY, AMERICAN SOCIETY OF BIOLOGICAL CHEMISTS, BALTIMORE, MD, US, vol. 273, no. 6, 6 February 1998 (1998-02-06), pages 3452-3460, XP002089593 ISSN: 0021-9258 cited in the application**

(72) Inventors:  
• **AESCHLIMANN, Daniel, Peter**  
**Cardiff CF10 3US (GB)**  
• **GRECARD, Pascale, Marie**  
**Cardiff CF10 3US (GB)**

**EP 1 317 548 B1**

Note: Within nine months of the publication of the mention of the grant of the European patent in the European Patent Bulletin, any person may give notice to the European Patent Office of opposition to that patent, in accordance with the Implementing Regulations. Notice of opposition shall not be deemed to have been filed until the opposition fee has been paid. (Art. 99(1) European Patent Convention).

## Description

[0001] The present invention relates to the identification of novel transglutaminase enzymes TG<sub>Z</sub> and TG<sub>Y</sub>.

[0002] Transglutaminases are a family of structurally and functionally related enzymes that catalyze the post-translational modification of proteins via a Ca<sup>2+</sup> dependent transferase reaction between the γ-carboxamide group of a peptide-bound glutamine residue and various primary amines. Most commonly, γ-glutamyl-ε-lysine cross links are formed in or between proteins by reaction with the ε-amino group of lysine residues. Analysis of the three-dimensional structure of the α-subunit of factor XIII showed that transglutaminases contains a central core domain containing enzymatic activity, and a N-terminal p-sandwich domain and two C-terminal β-barrel domains, which are thought to be involved in the regulation of enzyme activity and specificity.

[0003] Seven different transglutaminase genes have been characterised in higher vertebrates on the basis of their primary structure (Aeschlimann, D. and Paulsson, M (1994) *Thromb. Haemostasis* 71: 402-415 Aeschlimann, D. et al: (1998) *J. Biol Chem* 273, 3542-3460). Transglutaminases can be found throughout the body, but each transglutaminase is characterised by its own typical tissue distribution, although each may be present in a number of different tissue types often in combination with other transglutaminases. Transglutaminase gene products have specific functions in the cross linking of particular proteins or tissue structures. For review see Aeschlimann and Paulsson (1994) (*supra*) and Aeschlimann, D. and Thomazy V. (2000) *Connective Tissue Res.*41, 1-27. For example, factor XIIIa stabilises the fibrin clot in haemostasis, whereas prostate transglutaminase (TG<sub>P</sub>) is involved in semen coagulation. Other transglutaminases have adopted additional functions such as the tissue transglutaminase (TG<sub>C</sub>), which is involved in GTP-binding in receptor signalling, and band 4.2 protein which functions as a structural component of the cytoskeleton. Four transglutaminases have been shown to be expressed during the different stages of epidermal growth and differentiation. Three of these, keratinocyte transglutaminase (TG<sub>K</sub>), epidermal transglutaminase (TG<sub>E</sub>) and TG<sub>X</sub>, are associated with keratinocyte terminal differentiation and the cross-linking of structural proteins to form the cornified envelope. The fourth enzyme TG<sub>C</sub>, is expressed in skin primarily in the basal cell layer, and plays a role in the stabilisation of the dermo-epidermal junction. The importance of proper cross-linking of the cornified envelope is exemplified by the pathology seen in patients suffering from a severe form of the skin disease referred to as congenital ichthyosis, which has been linked to mutations in the gene encoding TG<sub>K</sub>.

[0004] All transglutaminase enzymes appear to be encoded by a family of closely related genes. Alignment of these genes demonstrates that all members of the transglutaminase family exhibit a similar gene organisation, with remarkable conservation of intron distribution. Furthermore, phylogenetic analysis indicates that an early gene duplication event subsequently gave rise to two different transglutaminase lineages; one comprising TG<sub>C</sub>, TG<sub>E</sub>, and band 4.2 protein; the other, factor XIIIa, TG<sub>K</sub> and possibly also TG<sub>P</sub> (Aeschlimann and Paulsson (1994) (*supra*)). The genes encoding TG<sub>K</sub> and factor XIIIa have been mapped to human chromosome 14q11.2 and chromosome 6p24-25 respectively, whereas TG<sub>C</sub> and TG<sub>E</sub> have been mapped to chromosome 20q11, and TG<sub>P</sub> has been mapped to chromosome 3p21-22.

[0005] Comparison of the structure of the individual transglutaminase genes shows that they may be divided into two subclasses, wherein the genes encoding TG<sub>C</sub>, TG<sub>E</sub>, TG<sub>P</sub> and band 4.2 protein comprise 13 exons and 12 introns, and the genes encoding factor XIIIa and TG<sub>K</sub> contain two extra exons. Exon IX of the former group is separated into two exons (X and XI) in TG<sub>K</sub> and factor XIIIa, and the amino-terminal extensions of TG<sub>K</sub> and factor XIIIa comprise an additional exon. However, except for the acquisition of an additional intro and the recruitment of an exon by the genes encoding factor XIIIa and TG<sub>K</sub>, the gene structure is remarkably conserved among all members of the transglutaminase gene family. Not only is the position of intron splice points highly conserved, but also the intron splice types. This similarity in gene structure and homology of the primary structure of the transglutaminases provides further support for the proposition that the different transglutaminase genes are derived from a common ancestral gene.

[0006] US 5,726,051 discloses a homologue of tissue transglutaminase (TG<sub>C</sub>) that is designated as TGase-H. However this homologue does not correspond to the transglutaminase sequences newly disclosed by the inventors in this application. It constitutes a splice variant derived from the TGM2 gene.

[0007] The inventors have previously isolated a cDNA encoding a novel member of the transglutaminase gene family TG<sub>X</sub>, from human foreskin keratinocytes (Aeschlimann et al (1998) *J. Biol. Chem.*, 273, 3452-3460). Two related transcripts with an apparent size of 2.2 and 2.8 kb were obtained. The deduced amino acid sequence for the full-length gene product encodes a protein with 720 amino acids and a molecular mass of 81kDa. A sequence comparison of TG<sub>X</sub> to the other members of the transglutaminase gene family revealed that the domain structure and the residues required for enzymatic activity and Ca<sup>2+</sup> binding are conserved and show an overall sequence identity of about 35%, with the highest similarity being found within the enzyme's catalytic domain.

[0008] The inventors subsequently determined that TG<sub>X</sub> is the product of a ~35kb gene located on chromosome 15, comprising 13 exons and 12 introns. The intron splice sites were found to conform to the consensus for splice junctions in eukaryotes. The transcription initiation site is localised to a point 159 nucleotides upstream of the initiator methionine and the likely polyadenylation site is localised ~600 nucleotides downstream of the stop codon. The two mRNA isoforms are the result of alternative splicing of exon III and give rise to 2 protein variants of TG<sub>X</sub> which comprise catalytic activity.

TG<sub>X</sub> is expressed predominately in epithelial cells, and most prominently during foetal development, in epidermis and in the female reproductive system.

**[0009]** The inventors have now localised the TGM5 gene to chromosome 15q15 by fluorescent *in situ* hybridization (Fig. 5A and B). Band 4.2 protein has previously been mapped to this chromosomal region (Sung L. A. et al (1992) Blood 79: 2763-2770; Najfeld V. et al (1992) Am. J. Hum. Genet 50: 71-75) and has subsequently been assigned to position 15q15.2 by expression mapping of the LGMD2A locus on chromosome 15 (Chiannikulchai N. et al (1995) Hum. Mol. Genet 4: 717-725). A short sequence encompassing the left arm of one of the YAC (yeast artificial chromosome) clones (926G10; CEPH-Généthon) used for expression mapping matched with the sequence of intron 12 of the TGM5 gene placing the genes encoding TG<sub>X</sub> and band 4.2 protein in close proximity on chromosome 15 (Fig. 5C). PCR with specific primers for 5' (exon I) or 3' (exon XIII) sequences of band 4.2 protein as well as southern blot analysis revealed that the BAC (bacterial artificial chromosome) clones containing the TGM5 gene also contained the EPB42 gene and that the 2 genes are arranged in tandem.

**[0010]** Further analysis by the inventors has recently led to the identification of two novel transglutaminase genes TGM7 and TGM6 which encode the proteins TG<sub>Z</sub> and TG<sub>Y</sub> respectively. Alternative mRNA sequences of the TGM7 gene are given in Fig. 6A and Fig. 6B. The TGM7 derived mRNA (Fig. 6A and Fig. 6B) comprises an open reading frame of 2130 nucleotides and a polyadenylation signal (AATAAA) 158 nucleotides downstream of the termination codon (TGA). The deduced protein for TG<sub>Z</sub> consists of 710 amino acids. The deduced protein for TG<sub>Z</sub> from Fig. 6A has a molecular mass of 79, 908 Da and an isoelectric point of 6.7. The deduced protein for TG<sub>Z</sub> from Fig. 6B has a molecular weight of 80,065 and an isoelectric point of 6.6.

**[0011]** The TGM6 full length transcript (Fig. 10A) comprises an open reading frame of 2109 nucleotides. The deduced protein for the long form of TG<sub>Y</sub> consists of 708 amino acids and has a calculated molecular mass of 79, 466 Da and an isoelectric point of 6.9. The transcript for the short form of TG<sub>Y</sub> (Fig. 10B) comprises an open reading frame of 1878 nucleotides and the deduced protein consists of 626 amino acids with a molecular mass of 70, 617 Da and an isoelectric point of 7.6.

**[0012]** To analyse the relationship between the different transglutaminase genes, the inventors calculated their amino acid similarity based upon sequence alignments, and calculated their evolutionary distances using different algorithms. All the algorithms used predicted a close relationship between TG<sub>X</sub>, TG<sub>Z</sub>, TG<sub>Y</sub>, TG<sub>E</sub>, band 4.2 protein and TG<sub>C</sub>, and factor XIIIa and TG<sub>K</sub>, respectively. The grouping of TG<sub>X</sub>, TG<sub>Z</sub>, TG<sub>Y</sub>, TG<sub>E</sub>, TG<sub>C</sub>, and band 4.2 protein in one subclass and factor XIIIa and TG<sub>K</sub> in another is supported by the results of this analysis and by the gene structure and genomic organisation of the different transglutaminase genes (Fig.9).

**[0013]** The inventors have therefore determined the structure of the human TGM5 gene, and its flanking sequences, and have mapped the gene to the 15q15 region of chromosome 15. Further, the inventors have determined that the human TGM5 gene comprises 13 exons separated by 12 introns spanning roughly 35kb, and that the structure of the TGM5 gene is identical to that of EPB42 (band 4.2 protein), TGM2 (TG<sub>C</sub>) and TGM3 (TG<sub>E</sub>) genes. Southern blot analysis has also shown that TGM5 is a single copy gene in the haploid genome. The inventors developed a method for detection and identification of transglutaminase gene products based on RT-PCR with degenerate primers and using this method have discovered the gene product of the TGM5 gene in keratinocytes (Aeschlimann et al (1998) J. Biol Chem. 273, 3452-3460). Using this method, the inventors have identified another new transglutaminase gene product in human foreskin keratinocytes and in prostate carcinoma tissue which has been designated TG<sub>Z</sub> or transglutaminase type VII. A full-length cDNA for this gene product was obtained by anchored PCR. Long range genomic PCR was used comprising different combinations of primers designed from the flanking sequences of the TGM5 - EPB42 gene sequence and the TG<sub>Z</sub> cDNA sequence to explore whether the gene encoding TG<sub>Z</sub> (TGM7) was present in close proximity to the other two transglutaminase genes. This placed the TGM7 gene approximately 9kb upstream of the TGM5 gene and demonstrated that the genes are arranged in tandem fashion (Fig. 5C). The inventors have therefore determined that the transglutaminase genes, TGM5 (TG<sub>X</sub>), TGM7 (TG<sub>Z</sub>) and EPB42 (band 4.2 protein) are positioned side by side within approximately 100 kb on chromosome 15. It has also been found that the mouse homologues of these genes are similarly arranged on mouse chromosome 2. Finally, the inventors have identified and determined the nucleotide and amino acid Sequences as well as tissue distribution for the novel transglutaminase gene products TG<sub>Z</sub> and TO<sub>Y</sub>.

**[0014]** According to a first aspect of the invention there is provided a nucleotide sequence comprising at least a portion of the nucleotide sequence of Fig. 10A or Fig. 10B; a nucleotide sequence which hybridise to the nucleotide sequence of Fig. 10A or Fig. 10B; a nucleotide sequence which is degenerate to the nucleotide sequence of Fig. 10A or Fig. 10B; all of which nucleotide sequences encode a polypeptide having transglutaminase activity.

**[0015]** Preferably the nucleotide sequence consists of the nucleotide sequence of Fig. 10A or Fig. 10B.

**[0016]** The first aspect of the present invention also provides a nucleotide sequence which hybridises under stringent conditions to the nucleotide sequences of Fig. 10A or Fig. 10B and which encodes a polypeptide having transglutaminase activity. Preferably the nucleotide sequence has at least 80%, more preferably 90% sequence homology to the nucleotide sequence shown in Fig. 10A or Fig. 10B. Homology is preferably measured using the BLAST program.

**[0017]** The invention further provides a method of expressing a polypeptide comprising inserting a nucleotide sequence

according to the first aspect of the present invention into a suitable host and expressing that nucleotide sequence in order to express a polypeptide having transglutaminase activity.

**[0018]** The invention also provides a vector comprising a nucleotide sequence according to the first aspect of the present invention.

**[0019]** According to another aspect of the invention there is provided a polypeptide having an amino acid sequence comprising at least a portion of the amino acid sequence of Fig. 10A or Fig. 10B, wherein the polypeptide has transglutaminase activity.

**[0020]** The invention also provides a polypeptide sequence which is at least 90% identical to the amino acid sequence of Fig. 10A or Fig. 10B and which has transglutaminase activity. The amino acid sequence of the polypeptide having transglutaminase activity may differ from the amino acid sequence given in Fig. 10A or Fig. 10B by having the addition, deletion or substitution of some of the amino acid residues. Preferably the polypeptide of the present invention only differs by about 1 to 20, more preferably 1 to 10 amino acid residues from the amino acid sequence given in Fig. 10A or Fig. 10B.

**[0021]** The invention also provides a composition comprising the polypeptide of the present invention for use in transamidation reactions on peptides and polypeptides.

**[0022]** According to a further aspect of the invention, there is provided a polypeptide comprising exons II through to exon IV or exons X through to exon XII of the sequence shown in Fig. 10A or Fig. 10B. The positions of the exons in the sequence shown in Fig. 10A or Fig. 10B can be determined from Fig. 8 where intron splice sites are marked with arrow heads.

**[0023]** According to another aspect of the invention there is provided a composition comprising the polypeptide according to the present invention for use in the cross-linking of proteins.

**[0024]** According to a further aspect of the invention there is provided a diagnostic method comprising detecting expression of the polypeptide according to the present invention in a subject or in cells derived from a subject.

**[0025]** The invention also provides an antibody directed against the polypeptide according to the present invention. The antibody may be any antibody molecule capable of specifically binding the polypeptide including polyclonal or monoclonal antibodies or antigen binding fragments such as Fv, Fab, F(ab')<sub>2</sub> fragments and single chain Fv fragments.

**[0026]** In accordance with a further aspect of the invention there is provided a method of diagnosis of autoimmune disease comprising taking a sample from a subject and testing that sample for the presence of a transglutaminase encoded by the nucleotide sequence of Fig. 10A or Fig 10B or portions thereof. Preferably the transglutaminase is detected by using an antibody having affinity for the transglutaminase.

**[0027]** The autoimmune disease to be diagnosed can be selected from Addison's disease, AI haemolytic anaemia, AI thrombocytopenic purpura, AI thyroid diseases, atrophic gastritis - pernicious anaemia, Crohn's disease, colitis ulcerosa, Goodpasture syndrome, IgA nephropathy or IGA glomerulonephritis, myasthenia gravis, partial lipodystrophy, polymyositis, primary biliary cirrhosis, primary sclerosing cholangitis, progressive systemic sclerosis, recurrent pericarditis, relapsing polychondritis, rheumatoid arthritis, rheumatism, sarcoidosis, Sjögren's syndrome, SLE, splenic atrophy, type 1 (insulin-dependent) diabetes mellitus, diabetes mellitus, Wegener granulomatosis, ulcerative colitis, vasculitis (both systemic and cutaneous), vitiligo.

**[0028]** The invention also provides a competitive protein binding assay for the differential diagnosis of autoimmune diseases comprising the detection of antibodies against the transglutaminase encoded by the nucleotide sequence of Fig. 10A or Fig 10B, or portions thereof.

**[0029]** Preferably the protein binding assay comprises using exogenous transglutaminase TGz or TGy, or both, as a competitive antigen.

**[0030]** The assay can be a competitive immunoassay selected from RIA, EIA/ELISA, LiA and FiA.

**[0031]** The invention will now be described with reference to the accompanying Figures 1 to 11, in which:

Fig. 1 is a representation of the genomic organisation of the human TGM5 gene. The human TGM5 gene is represented with the exons numbered I to XIII indicated by solid boxes separated by the introns 1 to 12. The sizes of the introns and exons are given in bp (base pairs). The 5'- and 3'- untranslated regions in exon I and XIII, respectively, are represented by hatched boxes with functional elements defining the transcript indicated. Additional sequence elements found in the TGM5 gene are indicated as follows: *Alu*, *Alu 7SL* derived retroposon; STS, sequence tagged site. Below the genomic map, a representation of the sequences present in the individual BAC clones is depicted.

Fig. 2 is a representation of the structure of the 5' untranslated region of the human TGM5 gene and mapping of the transcriptional start site. **A.** Primer extension analysis of poly (A<sup>+</sup>) RNA isolated from primary human keratinocyte prior to (lane 1) or after (lane 2) culture in suspension for 12h. Extension products were separated on denaturing polyacrylamide gel alongside a Sanger dideoxynucleotide sequencing reaction of the appropriate genomic DNA fragment primed with the same oligonucleotide. The transcriptional start site is indicated by the arrow. **B.** Nucleotide sequence of the proximal 5' region of the TGM5 gene, 5' ends of mRNA from primary keratinocytes mapped by

RACE are indicated by arrowhead. The major transcription start site identified by primer extension is highlighted with an *asterisk* (labeled +1). Consensus sequences for putative regulatory elements are underlined.

Fig. 3 is a representation of the structure of the 3' untranslated region of the human TGM5 gene. 3'-flanking sequence is shown with sequences homologous to known consensus sequences for 3' processing of transcripts (AATAAA, CAYTG and YGTGTTY) underlined. The termination points of cDNA's isolated from human keratinocytes (Aeschlimann et al (1998) J. Biol Chem 273, 3453-3460) by 3' RACE are indicated by arrowhead. A pair of inverted long repeat sequences is highlighted in italics.

Fig. 4 is a southern blot analysis of human genomic DNA hybridised to genomic TG<sub>X</sub> probes. Human genomic DNA was digested with *Bam*HI, *Eco*RI, and *Hind*III restriction enzymes and hybridised with short <sup>32</sup>P-labelled DNA fragments corresponding to intron 2 and flanking sequences of exon II and III (left panel) and exon X (right panel), respectively. The migration positions of the *Hind*III DNA size markers is indicated on the left.

Fig. 5 shows the chromosomal localisation of the human TGM5 gene by fluorescence in situ hybridisation. **A.** representative picture of fluorescein-labelled genomic DNA of BAC-228(P20) (fluorescence, arrows) hybridised to metaphase spreads of human chromosome stained with propidium iodide. **B.** An ideogram of banded chromosome 15 showing the localisation of the fluorescent signal on 13 chromosomes. **C.** Is a schematic map of the respective locus showing the organisation of the genes encoding TG<sub>X</sub> (TGM5), band 4.2 protein (EPB4.2) and TG<sub>Z</sub> (TGM7) as well as other genes [mitochondrial ATPase subunit D pseudogene; D-type cyclin-interacting protein 1 (DIP1); EST Genbank AA457639, AA457640], L1 repetitive element and genetic markers.

Fig. 6 shows A. the nucleotide sequence and deduced amino acid sequence for human TG<sub>Z</sub>. B. an alternative nucleotide sequence and deduced amino acid sequence for human TG<sub>Z</sub>. The initiation and termination codons as well as polyadenylation signal (AATAAA) are underlined.

Fig. 7 is a representation of the different tissue expression patterns for TG<sub>X</sub>, band 4.2 protein TG<sub>Y</sub> and TG<sub>Z</sub> in different fetal and mature human tissues. Human tissue Northern dot blot normalised for average expression of 9 different housekeeping genes probed with a fragment corresponding to the C-terminal β-barrel domains of TG<sub>X</sub> (**A**), TG<sub>Z</sub> (**B**), (**C**) TG<sub>Y</sub> and band 4.2 (**D**). A diagram showing the type of poly (A)<sup>+</sup> RNA dotted onto the membrane is shown in panel E.

Fig. 8 is a comparison of the structure of the different human transglutaminase genes. A. is an alignment of the nine characterised human gene products (TG<sub>X</sub>, TG<sub>Y</sub>, TG<sub>Z</sub> (shown in Fig. 6A), TG<sub>C</sub>, TG<sub>E</sub>, band 4.2, factor XIII a-subunit, TG<sub>K</sub>, TG<sub>P</sub>) is shown, with dashes indicating gaps inserted for optimal sequence alignment and underlined residues representing amino acids conserved in at least five gene products. The sequences are arranged to reflect the transglutaminase domain structure, based on the crystal structure of factor XIII a-subunit. N-terminal propeptide domain (**d1**), β-sandwich domain (**d2**), catalytic core domain (**d3**) and β-barrel domains 1 (**d4**) and 2 (**d5**) (from top to bottom). Known intron splice sites are marked by †. B. is an alignment of the nine characterised human gene products (TG<sub>X</sub>, TG<sub>Y</sub>, TG<sub>Z</sub> (shown in Fig. 6B), TG<sub>C</sub>, TG<sub>E</sub>, band 4.2, factor XIII a-subunit, TG<sub>K</sub>, TG<sub>P</sub>) is shown, with dashes indicating gaps inserted for optimal sequence alignment and underlined residues representing amino acids conserved in at least five gene products. The sequences are arranged to reflect the transglutaminase domain structure, based on the crystal structure of factor XIII a-subunit. N-terminal propeptide domain (**d1**), β-sandwich domain (**d2**), catalytic core domain (**d3**) and β-barrel domains 1 (**d4**) and 2 (**d5**) (from top to bottom). Known intron splice sites are marked by arrowheads.

Fig. 9 is a phylogenetic tree of the transglutaminase gene family and genomic organisation of the genes in man and in mouse. Sequences were aligned to maximise homology as shown in Fig. 8 except including sequences from different species as available: h, human; m, mouse; r, rat. Note, the mouse sequence for TG<sub>X</sub>3 is at present incomplete and no information is available for the N-terminal domain. In panel A5, a hypothetical pedigree for the gene family is given that is consistent with the data on the sequence relationship of the individual gene products (**A**) as well as with the data on the gene structure and genomic organisation (**B**). Phylogenetic trees based on the amino acid sequence homology of the gene products have been constructed using the NJ method (Saitou and Nei, 1987 Mol. Biol. Evol. 4, 406-425) of the PHYLIP software package for (**A1**) the N-terminal β-sandwich domain (**A2**) the catalytic core domain, (**A3**) the C-terminal β-barrel domains, and (**A4**) the entire gene products (**C**). The table given in C. Shows the similarity of TG<sub>X</sub> to the other transglutaminase gene products. The domain structure is based on the x-ray crystallographic structure of the factor XIII a-subunit dimer and inferred on the other gene products based upon the sequence alignment shown in Fig. 8. The numbers reflect % sequence identity.

Fig. 10. shows the nucleotide sequence and deduced amino acid sequence of TG<sub>γ</sub>. **A.** Shows the nucleotide and deduced amino acid sequence for the long form of TG<sub>γ</sub>. **B.** Shows the nucleotide sequence and deduced amino acid sequence for the short form of TG<sub>γ</sub>.

5 Fig. 11 is a schematic representation of the organisation of the identified transglutaminase gene clusters in the human genome.

#### Isolation and Determination of the Structure of the Human TGM5 Gene.

10 **[0032]** A unique insertion sequence of about 30 amino acids between the catalytic core domain and β-barrel domain 1 found in TG<sub>z</sub> was used as a template to design specific primers for the screening of a human genomic library. The characterisation of several genes of the transglutaminase gene family showed that the positions of the introns has been highly conserved and a comparison of the TG<sub>x</sub> sequence to the sequences of the other transglutaminases indicated that this unique sequence is present within an exon, exon X (see Fig. 8, aa 460-503) in TG<sub>x</sub>. A PCR reaction from human  
15 genomic DNA using oligonucleotides P1 and P2 that match sequences at either end of this unique segment yielded a DNA fragment of expected size which was confirmed to be the correct product by sequencing (results not shown). Screening of a human genomic DNA BAC library by PER using these oligonucleotides revealed two positive clones, BAC-33(p5) and BAC-228(P20) that were subsequently shown by Southern blotting with different cDNA probes to contain sequences spanning at least exon II to exon X of the TGM5 gene (results not shown).

20 **[0033]** Restriction analysis further indicated that each of the BAC clones contained substantially more than 50kb of human genomic DNA.

**[0034]** The similarity in the gene structure of the different transglutaminase genes prompted us to approach the characterisation of introns by PCR amplification using oligonucleotide primers corresponding to the flanking exon sequences at the presumptive exon/intron boundaries. All intron/exon boundaries were sequenced from the PCR products  
25 obtained in at least two independent PCR reactions, where applicable from both BAC clones, to exclude mutations introduced by *Taq* DNA polymerase, and the results compared. When sequences of PCR products comprising adjacent introns had no overlap, the intervening sequence (exon sequence) was determined by direct sequencing from isolated BAC plasmid DNA to confirm the absence of additional introns. Similarly, the 3'-untranslated region was obtained by step-wise extension of the known sequence using direct sequencing of BAC plasmid DNA. Both BAC clones terminated  
30 short of exon I and all attempts at isolating clones spanning exon I by screening of BAC and P 1 libraries failed. Exon I and intron 1 sequences were finally derived by nested PCR from human genomic DNA using conditions optimised for long range genomic PCR.

**[0035]** We established that the TGM5 gene comprises approximately 35kb of genomic DNA and contains 13 exons and 12 introns (Fig. 1). All intron/exon splice sites conform to the known GT/AG donor/acceptor site rule and essentially  
35 to the consensus sequence proposed by Mount S.M (1982) *Nucleic Acids Res.* 10, 459-472. (Table I). A sequence homologous to the branch point consensus CTGAC (Keller E. B and Noon. W. A (1984) *Proc. Natl. Acad. Sci. USA* 81: 7417-7420) was found 24 to 44 nucleotides upstream of the 3' splice site in introns 1, 3-6, and 9-12. The size of the introns varied considerably, ranging from 106bp to more than 6kb (Fig. 1, Table Is. The sequence obtained from the two different BAC clones matched with the exception of a deletion spanning the sequence from intron 6 to intron 8 in  
40 BAC-33(P5) (Fig. 1).

**[0036]** Further, we also resequenced the entire coding sequence of TG<sub>x</sub> and found 3 point mutations as compared to the previously reported cDNA sequence (Aeschlimann. et al (1998) *J. Biol Chem* 273 3452-3460) One of the nucleotide exchanges is silent, the other two result in an amino acid exchange (Table III). The first two mutations were found in  
45 both BAC clones, the third was only present in BAC-228(P20) due to the deletion in the other BAC clone. These differences may result from sequence polymorphism in the human gene pool as there was no ambiguity of the cDNA-derived sequence in this position determined from multiple independently amplified PCR products. However, the fact that a serine and alanine residue are changed into a proline and glycine residue that constitute the conserved amino acid in these positions in the transglutaminase protein family (see Fig. 8, aa 67 and aa 352 in TG<sub>x</sub>) suggested that these may have been PCR related mutations in the DNA sequence. To clarify this issue, we have prepared cDNA from human  
50 foreskin keratinocytes from different individuals, amplified full-length cDNA with high fidelity DNA polymerase, and sequenced the respective portions of the cloned cDNAs. The data confirmed that allelic variants exist with differences in these positions (Table III).

**[0037]** The isolation and sequencing of cDNAs encoding TG<sub>x</sub> and Northern blotting with TG<sub>x</sub> DNA probes revealed expression of at least two differentially spliced mRNA transcripts for TG<sub>x</sub> in human keratinocytes. Solving the gene  
55 structure confirmed the short form of TG<sub>x</sub> to be the result of alternative splicing of exon III as predicted. A third isolated cDNA that differed also at the exon III/exon IV splice junction turned out to be the result of incomplete or absent splicing out of intron 3 as the sequence upstream of exon IV in the cDNA matched with the 3' sequence of intron 3. Exon 3 encodes part of the N-terminal β-barrel domain of TG<sub>x</sub> and the absence of the sequence encoded by exon 3 is expected

to result in major structural changes in at least this domain of the protein. Nevertheless, expression of TG<sub>x</sub> in 293 cells using the full-length cDNA resulted in synthesis of two polypeptides with a molecular weight consistent with the predicted products from the alternatively spliced transcripts (results not shown).

**[0038]** Initially, 5' RACE was used to determine the 5' end of TG<sub>x</sub> cDNAs. Transcripts starting 77,96 and 157 nucleotides upstream of the initiator ATG were isolated in addition to the previously described shorter transcript (Fig. 2B arrowheads). All of these transcripts were recovered repeatedly in independent experiments. Finally, primer extension experiments located the major transcription initiation site used in keratinocytes 157 nucleotides upstream of the translation start codon (Fig. 2). The proximal promoter region was analysed for potential binding sites of transcription factors using MatInspector (Genomatix, Munich Germany) and GCG (Genetics Computer Group, Inc., Madison, Wisconsin) software packages. No classical TATA-box sequence was found but a number of other potential transcription factor binding sites could be identified (Fig. 2B), suggesting that TG<sub>x</sub> promoter is a TATA-less promoter. Interaction of CBP (may bind to CAAT-box), nuclear factor I (NF1) and upstream stimulatory factor (USF) to form a core proximal promoter has been demonstrated in a number of TATA-less genes. c-Myb is found in TATA-less proximal promoters of genes involved in hematopoiesis and often interacts with Ets-factors, and these sites maybe operative in the expression of TG<sub>x</sub> in hematopoietic cells, e.g HEL cells AP1, Ets and SP1 elements are typically found in keratinocyte-specific genes and may be involved in transcriptional regulation in keratinocytes. Several AP1 sites are present within 2.5kb of the upstream sequence and could interact with the proximal AP1 factor for activation. SP1 sites are properly positioned upstream of the start points of the shorter transcripts raising the possibility that these could also be functional, though to a lesser degree.

**[0039]** The last exon, exon XIII, contained a consensus polyadenylation signal AATAAA-600bp downstream of the termination codon (Fig. 3). This is in good agreement with the size of the mRNA (2.8kb) encoding full-length TG<sub>x</sub> expressed in human keratinocytes as detected by Northern blotting considering the length of the coding sequence (2160bp.). A CAYTG signal that binds to U4 snRNA which is identical for 4 out of 5 nucleotides is present in tandem in 3 copies 7 nucleotides downstream of the polyadenylation signal. A close match (YCTGTTY) of another consensus sequence YGTGTTY that is found in many eukaryotic transcripts and provides a signal for efficient 3' processing is present 46 nucleotides downstream of the polyadenylation signal. However, we have previously reported that all cDNAs isolated by RT-PCR with an oligo(dT) oligonucleotide from human keratinocytes ended within 9 to 34 nucleotides downstream of the pentanucleotide AATAA at position 2169. It has been shown that this pentanucleotide functions as a polyadenylation signal and these shorter transcripts are selectively enhanced by PCR amplification because of the smaller size of the PCR product.

#### Chromosomal Localisation of the TGM5 Gene.

**[0040]** To address the genomic organisation and identify the chromosomal localisation of the TGM5 gene or genes in the human genome, we performed Southern blot analysis of human genomic DNA cut with *Bam*HI, *Eco*RI and *Hind*III restriction enzymes using probes derived from intron 2 as well as from the sequence encoded by exon X that is unique to TG<sub>x</sub> (Fig.4). Bands of 4.5, 6.0, 10.5, 4.3, 9.3 and 2.6kb were revealed with the respective probes. The simple pattern of restriction fragments hybridising with the probes indicated that the haploid human genome contains only one TGM5 gene.

**[0041]** The TGM5 gene was subsequently localised to chromosome 15 by fluorescent *in situ* hybridisation of human metaphase chromosome spreads using genomic DNA derived from either BAC clone as a probe (Fig. 5A). A comparison of the probe signal to the DAPI banding pattern localised the TGM5 gene to the 15q15 region. The localisation was subsequently refined by determining the distance of the fluorescent signal to the centromere as well as to either end of the chromosome on 13 copies of chromosome 15 and expressing it as a fractional distance of the total length of the chromosome. These measurements placed the TGM5 gene close to the centre of the 15q15 region, *i.e.* to the 15q15.2 locus (Fig. 5B).

#### TGM5 is part of a cluster of transglutaminase genes.

**[0042]** The EPB42 gene has previously been assigned to locus 15q15.2 on chromosome 15. This raised the possibility that the EPB42 gene may be arranged in tandem with the TGM5 gene. Indeed, PCR with specific primers for sequences derived from the 5' and 3' of the EPB42 gene yielded products of appropriate size from both, BAC-33(P5) and BAC-228 (P20) (results not shown) and sequencing confirmed the identity of the PCR products. Southern blotting of BAC plasmid DNA with cDNA probes comprising the 5' or 3' end of the EPB42 gene and subsequent comparison of the pattern of labelled restriction fragments with that of the TGM5 gene allowed us to map this locus in more detail. The EPB42 gene and TGM5 are arranged in the same orientation being spaced apart by ~11kb (Fig. 5C) approximately 30% of which was sequenced to characterise the 3' and 5' flanking UTR of the TGM5 and EPB42 gene, respectively.

**[0043]** To analyse the relationship between the most closely homologous genes of the transglutaminase gene family (TGM7, TGM5 and EPB42 on human chromosome 15q15.2 and TGM2 and TGM3 on chromosome 20q11/12), we

mapped the respective mouse genes using radiation hybrid mapping. All genes mapped to the distal part of mouse chromosome 2. The genes for *tgm7*, *tgm5*, and *epb42* showed a best fit location for the segment defined by D2Mit104 proximal and D2Mit 396 distal, with a highest LOD of 14.5, 16.6 and 15.5 to the anchor marker D2Mit395 (66.9cM) and an LOD of >20 to D2Ert616e (69.0cM). This is in good agreement with the assigned locus 67.5cM distal from the centromere (White, R.A et al., 1992 Nat. Genet. 2, 80-83). This *tgm3* gene showed a best fit location for the segment defined by D2Mit447 proximal and D2Mit258 distal, with a highest LOD of 14.8 and 122 to D2Mit258 (78.0cM) and D2Mit338 (73.9cM), respectively The *tgm2* gene showed a best fit location for the segment defined by D2Mit139 proximal (86.0cM) and D2Mit225 distal (91.0cM), with a highest LOD of 17.0 to the anchor marker. D2Mit287, consistent with its assigned locus 89.0cM from the centromere (Nanda, N., et al., 1999. Arch. Biochem. Biophys. 366, 151-156).

**[0044]** We developed a method for detection and identification of transglutaminase gene products based on RT-PCR with degenerate primers and using this method discovered the gene product of the TGM5 gene in keratinocytes. Using this same method, we have identified another new transglutaminase gene product in human foreskin keratinocytes and in prostate carcinoma tissue which we designated TG<sub>z</sub> or transglutaminase type VII A full-length cDNA for this gene product was obtained by anchored PCR (see below). We used long range genomic PCR with different combinations of primers designed flanking sequences of the TGM5 - EPB42 gene segment and the TG<sub>z</sub> cDNA sequence to explore whether the gene encoding TG<sub>z</sub>, TGM7, was present in close proximity to the other two transglutaminase genes. This placed the TGM7 gene approximately 9kb upstream of the TGM5 gene and demonstrated that the genes are arranged in tandem fashion (Fig. 5C). The 5' UTR of the TGM5 gene was sequenced (Fig. 2). The genes encoding TG<sub>C</sub>, TG<sub>E</sub>, which are more closely related to the TGM5, TGM7, and EPB42 genes than the other transglutaminase genes based on amino acid sequence comparison and similarity in gene structure (Fig. 9) have been mapped to human chromosome 20q11 (Gentile V. et al (1994) Genomics 20, 295-297; Kim I. G. et al (1994) J. Invest. Dermatol 103, 137-142). The syntenic regions of the 15q15 and 20q11 locus in the mouse are present in a short segment, 2F1-G, of chromosome 2, which puts all five transglutaminase genes in proximity (Fig. 9B). The mouse homologue of band 4.2 protein has been mapped to this region of mouse chromosome 2 (White R. A. et al (1992) Nat. Genet 2, 80-83). We have isolated a BAC clone containing the gene encoding the mouse homologue of TG<sub>x</sub> and shown that the *tgm5* gene is located next to the *epb42* gene in tandem fashion, similar to the organisation in the human genome. Furthermore, this clone was shown to contain the gene encoding the mouse homologue of TG<sub>z</sub>. Genomic sequences derived from this BAC clone and also cDNA sequences derived from cDNA prepared from mouse uterus showed that the mouse and human gene products are 85% identical on the nucleotide level. To analyse the relationship between the transglutaminase genes in more detail, we calculated the amino acid similarity (Fig. 9C) based on the sequence alignment shown in Fig. 8 and calculated evolutionary distances using different algorithms (Fig. 9A). All algorithms predicted a close relationship between TG<sub>x</sub> and TG<sub>E</sub>, and band 4.2 protein and TG<sub>C</sub>, raising the possibility that a single transglutaminase gene initially locally duplicated to generate a cluster of 3 genes, followed by a duplication of a larger segment of the chromosomal region, gave rise to the organisation of the genes in mouse. In humans these chromosomal regions were apparently redistributed to two different chromosomes. This hypothesis led us to speculate on the existence of an additional gene on human chromosome 20q11. Careful analysis of the chromosomal sequences of this locus derived by the human genome project revealed the presence of a candidate gene, TGM6, located approximately 45kb downstream of the TGM3 gene consistent with our hypothesis (Fig. 11). To confirm that this is in fact a functional gene and not a pseudogene, we screened a large number of cell lines for expression of a respective gene product by PCR. A corresponding gene product, TG<sub>y</sub>, or transglutaminase type VI could be identified in a small cell lung carcinoma cell line and a full-length cDNA was subsequently derived by anchored PCR.

#### Determination of cDNA and amino acid sequences of TGM6 (TG<sub>y</sub>) and TGM7 (TG<sub>z</sub>) gene products

**[0045]** A full-length cDNA sequence for TG<sub>z</sub> was obtained by anchored PCR using oligo(dT)-Not I primed cDNA prepared from human foreskin keratinocytes, prostate carcinoma tissue and human carcinoma cell line PC3, essentially following the strategy previously described (Aeschlimann et al (1998) J. Biol Chem 273: 3245-3460). The oligo(dT)-Not I primer was used as the anchoring primer to obtain the 3' end of the cDNA. 5' RACE was used to determine the 5' end of the cDNA. The obtained sequence information (Fig. 6) contained an open reading frame 2130 nucleotides and a polyadenylation signal (AATAAA) 158 nucleotides downstream of the termination codon (TGA). The deduced protein consists of 710 amino acids. The cDNA and amino acid sequence in Fig. 6A was first determined and the deduced protein has a calculated molecular mass of 79,908 Da and an isoelectric point of 6.7. The cDNA and amino acid sequence in Fig. 6B was then determined. This sequence differs by a few nucleotides and amino acids from the sequence given in Fig. 6A. The protein deduced from the sequence given in Fig. 6B has a calculated molecular mass of 80,065 and an isoelectric point of 6.6. A number of aberrantly spliced gene products were isolated which lacked part of exon IX (5' end) or-retained the whole or part of intron 11. These products are unlikely to be of physiological significance but may point out that splicing of certain introns in this gene is a difficult and inefficient process.

**[0046]** A full-length cDNA sequence for TG<sub>y</sub> was obtained by PCR using oligo(dT) primed cDNA prepared from the

lung small cell carcinoma cell line H69, and using sequence specific primers based on the presumptive transcribed genomic sequence. 5' RACE was used to determine the 5' end of the cDNA. The obtained sequence information for the long form of TG<sub>Y</sub> (Fig. 10A) contained an open reading frame of 2109 nucleotides. The deduced protein for the long form of TG<sub>Y</sub> consists of 708 amino acids and has a calculated molecular mass of 79,466 Da and an isoelectric point of 6.9. A shorter transcript was also isolated which apparently resulted from alternative splicing of the sequence encoded by exon XII. The absence of exon XII results in a frame shift and thereby in premature termination within exon XIII. The obtained sequence information for the short form of TG<sub>Y</sub> (Fig. 10B) contained an open reading frame of 1878 nucleotides. The deduced protein for the short form of TG<sub>Y</sub> consists of 626 amino acids and has a calculated molecular mass of 70,671 Da and an isoelectric point of 7.6. The sequence alterations due to the splicing result in a short protein which terminates just after the first C-terminal β-barrel domain. The β-barrel domains have been implicated in the regulation of enzyme-substrate interaction, and the lack of the second C-terminal β-barrel domain (see Fig. 8, d5) is likely to be of biological significance.

**[0047]** The catalytic mechanism of transglutaminases has been solved based on biochemical data available for several transglutaminases and the X-ray crystallographic structure of the factor XIII α-subunit dimer. The reaction center is formed by the core domain and involves hydrogen bonding of the active site Cys to a His and Asp residue to form a catalytic triad reminiscent of the Cys-His-Asa triad found in the papain family of cysteine proteases. The residues comprising the catalytic triad are conserved in TG<sub>Y</sub> (Cys276, His335, Asp358) and TG<sub>Z</sub> (Cys227, His336, Asp359) (Fig. 8) and the core domain shows a high level of conservation as indicated by a sequence identity of about 50% between these gene products and the other transglutaminases (Fig. 9). A Tyr residue in barrel 1 domain of the α subunit of factor XIII is hydrogen bonded to the active site Cys residue and it has been suggested that the glutamine substrate attacks from the direction of this bond to initiate the reaction based on analogy to the cysteine proteases. In TG<sub>Y</sub>, the Tyr residue is conserved (Tyr 540) while in TG<sub>Z</sub> the Tyr residue has been replaced by His538 similar to TG<sub>X</sub> (Fig. 8). This is expected to be a conservative change which is supported by our data demonstrating that recombinant TG<sub>X</sub> from 293 cells has transglutaminase activity. Crystallization experiments with factor XIIIa further indicated that 4 residues are involved in binding of Ca<sup>2+</sup>-ion, including the main chain carbonyl of Ala457 and the side chain carboxyl groups of Asp438, Glu485, and Glu490. All three acidic residues are conserved in TG<sub>Y</sub> and in TG<sub>Z</sub> (Fig. 8). None of the residues critical to enzyme function are affected by the alternative splicing of TG<sub>Y</sub>. Based on the preservation of critical residues for enzyme function and domain folding and the extensive overall similarity of the TG<sub>Y</sub> and TG<sub>Z</sub> isoforms to the other members of the transglutaminase protein family, it can be predicted that the characterized cDNAs are encoding active transglutaminases.

#### Tissue Expression Patterns for TG<sub>X</sub>, TG<sub>Y</sub> and TG<sub>Z</sub>.

**[0048]** We have previously shown that TG<sub>X</sub> is expressed in a number of different cell types (Aeschlimann et al (1998) J. Biol. Chem. 273, 3452-3460). To obtain a more complete picture on the expression of TG<sub>X</sub> and the novel gene products, we performed a dot blot Northern blot analysis of more than 50 adult and fetal human tissues. Band 4.2 protein was expressed at high level in bone marrow and fetal spleen and liver, consistent with its role in hematopoietic cells, and virtually undetectable in all other tissues. In contrast, TG<sub>X</sub>, TG<sub>Y</sub> and TG<sub>Z</sub> showed widespread expression at low level, with highest levels of TG<sub>X</sub>, TG<sub>Y</sub> and TG<sub>Z</sub> mRNA present in the female reproductive system, in the central nervous system, and in testis, respectively (Fig. 7).

**[0049]** RT-PCR analysis on human cell lines and tissues shows that TG<sub>Z</sub> is expressed in osteosarcoma cells (MG-63), dermal fibroblasts (TJ6F, HCA2), erythroleukemia cells (HEL), in primary keratinocytes, mammary epithelium carcinoma cells (MCF7), HELA cells, skin, brain, heart, kidney, lung, pancreas, placenta, skeletal muscle, fetal liver, prostate and in prostate carcinoma tissue. A similar analysis for TG<sub>Y</sub> revealed expression only in a lung small cell carcinoma cell line (H69) and extremely low levels of expression in tissues.

**[0050]** In conclusion, TG<sub>Z</sub> is expressed widely in cells and tissues and expression levels are not apparently affected by cellular differentiation, (i.e keratinocyte differentiation or fibroblast senescence). TG<sub>Y</sub> expression, on the other hand, was very restricted and expression was only found in H69 cell line. This cell line has characteristics of neuronal cells such as the expression of neuron-specific enolase and brain isozyme of creatine kinase which together with widespread expression in tissues of the nervous system suggests that TG<sub>Y</sub> expression may be specific to neuronal cells. Transglutaminase action has been implicated in the formation of aberrant protein complexes in the central nervous system leading to nerve cell degeneration, e.g in Alzheimers and Huntington's disease. Based on its expression pattern, Toy is a logical candidate to bring about the underlying transglutaminase-related pathological changes.

#### Reagents

**[0051]** Oligonucleotides were from Oligos. Etc. Inc. (Wilsonville, OR) or life technologies and restriction enzymes from Promega Corp. (Madison, WI).

**Genomic Library Screening**

[0052] A human BAC library established in a F-factor-based vector, pBeloBAC 11, and maintained in *E. coli* DH10B was screened by PCR (Genome Systems, Inc., St. Louis, MO). A 147bp DNA fragment unique to TG<sub>X</sub> was amplified from 100ng of genomic DNA in 100μl of 10mM Tris/HCl, pH 8.3, 50mM KCl containing 2mM MuCl<sub>2</sub>, 0.2mM dNTPs using 2.5 units of *Taq* DNA polymerase (Fisher Scientific Corp. Pittsburgh, PA) and 50pmol of upstream primer P1, 5'-CCA-CATGTTGCAGAAGCTGAAGGCTAGAAGC and downstream primer P2, 5'-CCACATGTCCACATCACTGGGTC-GAAGGGAAGG. PCR cycles were 45 sec at 94°C (denaturation), 2 min at 60°C (annealing), and 3 min at 72°C (elongation) for a total of 37 cycles, with the first cycle containing an extended denaturation period (6 min) during which the polymerase was added (hot start), and the last cycle contained an extended elongation period (10 min). Two positive clones were identified, BAC-33(P5) and BAC-228(P20) (Genome Systems), and their identity verified by Southern blotting. Plasmid DNA was prepared using a standard alkaline lysis protocol. 2μg plasmid DNA was restricted with *Bam*HI, *Eco*RI, and *Spe*I and probed with a <sup>32</sup>P-labelled-500bp *Nco*I/*Bsp*HI and ~600bp *Bsp*HI/*Nde*I cDNA fragment of TG<sub>X</sub>, respectively, as described below.

**Amplification of TGM5 Intron Sequences**

[0053] PCRs were carried out with 2.5 units of *Taq* DNA polymerase (Fisher Scientific) and 100-200ng of plasmid DNA from BAC clones in 100μl of 10mM Tris/HCl, pH 8.3 50mM KC1 containing 2mM MgCl<sub>2</sub>, 0.2mM dNTPs and 50pmol of the desired oligonucleotide primers. The PCR cycles were 45 sec at 94°C (denaturation), 1 min at 60°C (annealing), and 5 min at 72°C (elongation). A total of 32 cycles were carried out, with the first cycle comprising an extended denaturation period (6 min) during which the polymerase was added (hot start) and the last cycle comprised an extended elongation period (10 min). The following oligonucleotides were used as upstream and downstream primers, respectively, in the individual reactions:

intron 2, 5'-GGACCACCTGCTTGTTCGCCGGGG,5'-AGGGGCTGGGGCTGTGATGGCGTG;  
 intron 3,5'-ACCTCTTGAAAATCCACATCGACTCCT, 5'-CAGTTCTTGCTGCCTTGGTAGATGAAGCC;  
 intron 4,5'-GACAGTGAACCCAGAGGCAGGAG, 5'-TCTGTGGCTGGGTCAGTCTGGAAGTGCA (P3);  
 intron 5,5'-GCCTGCACTTCCAGACTGACCCAGCCACA, 5'-TCCAGTTTCCATTGAGCACCCCA;  
 intron 6,5'-TGCTGGGTCTTTGCTGCCGTCATGTGC, 5'-TCCTTCTTCTTATCCCCAAAATCCTGCC;  
 intron 7, 5'-TAGATGAGTATTATGACAACACAGGCAGG, 5'-GCGTCCAGCACCTGCCAGCCTCC;  
 intron 8,5'-TGAGTGCTGGATGGCCCGGAAGG, 5'-CCCGCTCGTCACTCTGGATGCTC;  
 intron 9, 5'-TTCACCAGGACACGAGTTCTGTTGGCA, P2 (see above);  
 intron 10, P1 (see above), 5'-TCAGGACTGCTTTCTCTTACCC;  
 intron 11, 5'-ACCCCTGCAAAATCTCCTATTCCC, 5'-AATATCACCTGTATGGAGAGTGGCTGG;  
 intron 12, 5'-TTGAGGACTGTGTGCTGACTGTGGM 5'AATGATGCTTGTGGGG.

[0054] PCR's were carried out with 1.25 units of *Pfu* Turbo DNA polymerase (Stratagene) and 260ng genomic DNA in a total of 100μl of supplied reaction buffer supplemented with 0.2mM dNTPs, 2μl DMSO and 50 pmol primers. The PCR cycles were 45 sec at 94°C (denaturstion), 1 min at 68°C (annealing), and 2 min at 72°C (elongation). A total of 37 cycles were made, with the first cycle containing an extended denaturation period (6 min) during which the polymerase was added (hot start), and the last cycle containing an extended elongation period (10 min).

**Rapid Amplification of 5'-mRNA End**

[0055] A modified RACE protocol was used to determine the transcription start site and obtain additional sequence information of exon L Double stranded cDNA was prepared from poly(A<sup>+</sup>) RNA of cultured normal human keratinocytes (Aeschlimann et al (1998) J. Biol Chem. 273, 3452-3460) with the Copy Kit (Invitrogen, San Diego, CA). The cDNA was purified from nucleotides using the GlassMax DNA Isolation Kit (Life Technologies, Inc.) and tailed in the presence of 200μM dCTP with 10 units of terminal deoxynucleotidyl transferase (Promega) for 30 min at 37°C to anchor the PCR at the 5'-end. The PCR reaction was anchored by performing a total of 5 cycles of one-sided PCR at a lower annealing temperature (37°C) with the abridged anchor primer (Life Technologies, Inc.) only Following transfer of 25% of this reaction at 94°C to a new tube containing abridged anchor primer and TG<sub>X</sub>-specific primer P3 (see above), the first round of amplification was carried out for a total of 37 cycles under the conditions described above except for annealing which was carried out at 55°C. Nested PCR was done with the universal amplification primer (Life Technologies, Inc.) and TG<sub>X</sub>-specific primer P4, 5'-TGAAGTACAGGGTGAGGTTGAAGG, as described above (annealing at 60°C) using 1.0 μl from the first round PCR

### Primer Extension Analysis

[0056] Oligonucleotide P5 5'-CATGGTAGCTGCCTCCGGTTCCTG containing a 5'-infrared label (IRD 800) was purchased from MWG Biotech (Ebersberg, Germany). Primer P5 (5.3pmol) was hybridised to 1 ug of poly (A<sup>+</sup>)RNA from primary keratinocytes (Aeschlimann et al (1998) J. Biol. Chem. 273, 3452-3460) and reverse transcription performed with 200 units of Superscript II RNase H reverse transcriptase (Life Technologies) in a total of 20µl for 90 min at 42°C according to the manufacturer's instructions. Enzyme was heat inactivated and primer extension products extracted with phenol chloroform, precipitated with ethanol, and then analysed on a 4.5% denaturing polyacrylamide gel adjacent to dideoxynucleotide chain termination sequencing reactions (Thermo Sequenase Cycle Sequencing Kit; Amersham) derived from a double-stranded genomic DNA fragment using the same primer.

### DNA Preparation and Sequencing

[0057] Plasmid DNA from BAC clones was further purified for direct sequencing by digestion with 200µg/ml of RNase A (Sigma, St. Louis, MO) for 1h at 37°C and by subsequent micro-dialysis using Spectra/Por 2 membranes (Spectrum Medical Industries, Inc. Laguna Hills, CA). PCR products were gel purified using the QIA quick Gel Extraction Kit (Qiagen, Inc. Chatsworth, CA) for sequencing. Cycle sequencing was performed by the dideoxy chain termination method using the Cyclist Exo-Pfu DNA Sequencing Kit (Stratagene, LaJolla, CA) and pre-cast 6% polyacrylamide gels with the Cast-Away Sequencing System (Stratagene) or using the dRhodamine Terminator Cycle Sequencing Ready Reaction Kit (PE Biosystems) and an ABI 310 automated sequencer.

### Southern Blotting

[0058] 18µg human genomic DNA was digested with *Bam*HI, *Eco*RI, and *Hind*III restriction enzymes, separated in a 0.8% agarose gel and transferred to a Zeta-probe membrane (Bio-Rad, Labs. Hercules, CA). The gel was calibrated using the Lambda DNA/*Hind*III markers (Promega). <sup>32</sup>P-labelled probes were prepared by random prime labelling using the Multiprime DNA Labelling System (Amersham, Int. Amersham, UK) and PCR products corresponding to intron 2, intron 12, and exon X (see above) as DNA templates. Probes were hybridised to the blot overnight at 65°C in 500 mM NaH<sub>2</sub>PO<sub>4</sub>, pH 7.5, containing 1mM EDTA and 7% SDS. The membrane was washed at 65°C to a final stringency of 40mM NaH<sub>2</sub>PO<sub>4</sub>, pH 7.5, 1mM EDTA, and 1% SDS, and the result developed by exposure of the membrane to BioMax MR film (Eastman Kodak, Rochester, NY).

### Chromosomal Localisation

[0059] Human peripheral blood lymphocytes were used to prepare metaphase chromosome spreads (Bebbington C. R. and Hentschel, C. C. G. (1987) in DNA cloning (Volume III) 184-188, IRL Press, Oxford UK). Cells were cultured in PB-Max Karyotyping medium (Gibco, BRL, Gaithersburg, MD) for 72h, and synchronised by culture in the presence of 10<sup>-7</sup>M amethophterin (Fluka) for another 24h. Cells were released from the mitotic block by extensive washing and subsequent culture in the above medium containing 10<sup>-5</sup>M thymidine for 5h. Cells were subsequently arrested in metaphase by addition of colcemid to a final concentration of 0.1 µg/ml (Gibco BRL). Harvester cells were incubated in 0.075M KCl for 25 min at 37°C, fixed in methanol/acetic acid (3:1) solution, and chromosome spreads prepared by dropping the cells onto the glass slides. After air drying, chromosomes were treated with 100µg/ml of RNase A in 2x SSC for 1h at 37°C, denatured in 70% (v/v) formamide in 2x SSC for 3 min at 75°C, and dehydrated in a graded ethanol series. DNA probes were prepared by random prime labelling of plasmid DNA of BAC-33(P5) and BAC-228(P20) with fluoresceine-conjugated dUTP using the Prime-It Fluor Fluorescence Labelling Kit (Stratagene). Probes were denatured at 75°C for 10 min in hybridisation buffer consisting of 50% formamide (v/v) and 10% dextran sulphate (w/v) in 4x SSC and prehybridised at 42°C for 20 min to 0.2 µg/ml human competitor DNA (Stratagene) to block repetitive DNA sequences. Probes were subsequently hybridised to the chromosome spreads at 37°C overnight, followed by washing to a final stringency of 0.1x SSC at 60°C. Spreads were mounted in phosphate-buffered glycerol containing 200 ng/ml propidium iodide to counterstain chromosomes. Slides were examined by epifluorescence microscopy using a 100x objective and images captured with a DC-330 CCD camera (DAGE-MTI, Inc. Michigan City, IN) using a LG-3 frame grabber board (Scion Corp. Frederick, MD) in a Macintosh 8500 workstation and a modified version of the NIH image 1.6 software (Scion Corp.). Images representing fluoresceine-labelling and propidium iodide staining of the same field were superimposed using Adobe Photoshop 3.0 (Adobe Systems, Inc. Mountain View, CA) to map the gene to a chromosomal region.

### Cloning of Novel Transglutaminase Gene Products by Anchored PCR

[0060] For cloning of TG<sub>γ</sub>, poly(A)<sup>+</sup>RNA was prepared from about 10<sup>6</sup> H69 cells (American Type Culture Collection,

Rockville, MD) by oligo(dT)-cellulose column chromatography using the Micro-Fast Track Kit (Invitrogen, San Diego, CA) and recovered in 20 $\mu$ l 10mM Tris/HCl, pH 7.5. The poly(A)+RNA (5.0 $\mu$ l) was reverse transcribed into DNA in a total volume of 20 $\mu$ l using the cDNA Cycle Kit (Invitrogen) with 1.0 $\mu$ l oligo(dT) primer (0.2 $\mu$ g/ $\mu$ l). Overlapping fragments of TG $_{\gamma}$  were amplified by PCR using oligonucleotides 5'-ATCAGAGTCACCAAGGTGGAC, 5'-AGAAACACATCGTC-  
 5 CTCTGCACACC (P6), 5'-CAGGCTTTCCTCTCACCGCAAACAC, 5'-CGTACTTGACTGGCTTGTACCTGCC, 5'-TC-  
 TAGGTCACCAGGGTCATCAGTGC, 5'-GCCTGTTCACCGCCTTGCTGT, 5'-CATCACTGACCTCTACAAGTATCC, 5'-  
 ACGGCGTGGGATTTCATGCAGG, 5'-CATCCTCTATACCGCAAAGCC, and 5'-AGGTTGAGGCAGGATTAAGTACG-  
 GCCTC. PCRs were carried out with 1.25 units of AmpliTaq Gold DNA polymerase (PE Biosystems) and 2.0 $\mu$ l cDNA  
 10 in a total of 50 $\mu$ l of supplied reaction buffer supplemented with 2mM MgCl $_2$ , 0.2mM dNTPs and 25 pmol of the appropriate  
 gene-specific primers. A total of 40 PCR cycles were made, with an elevated annealing temperature of 65°C for the  
 initial 5 cycles and an annealing temperature of 60°C for the remaining cycles. The 5'-end of the cDNA was isolated by  
 5'-RACE as described above with the exception of using the gene-specific oligonucleotides P6, 5'-GATGTCT-  
 GGAACACAGCTTTGG, and 5'-TCACAGTCCAGGGCTCTGCTCAG. The PCR-products were either directly sequenced  
 15 or when desired, cloned by taking advantage of the 3' A-overhangs generated by *Taq* DNA polymerase using the Original  
 TA-Cloning Kit (Invitrogen).

**[0061]** For cloning of TG $_z$ , we used a series of degenerate and gene-specific oligonucleotides to isolate overlapping  
 DNA fragments, essentially following our previously described strategy. TG $_z$ -specific oligonucleotide primers were 5'-  
 CAACCTTGCGGCTTGAGTCTGTCTG, 5'-CAGCAGCTCTGACGGCTTGGGTC (P7), 5'-ATCACCTTTGTGGCTGA-  
 GACCG, 5'-CAAGGGTTAAAAAGTAGGATGAAAGTTC, 5'-CACAGTGTGACTTACCGCTG, 5'-CATAACAC-  
 20 CACGTGTTCCGCTG, 5'-CTTAAAGAACCCGGCCAAAGACTG, 5'-CGATGGTCAAGTTCCTATCCAXGTTG, 5'-  
 TGTTGTTTCCAATTTCCGTTCCGC, 5'-TCTGGCACCTCTGGATACGCAG, 5'-CTTAGGGATCAGCCAGCGCAGC,  
 5'-GCGGATGAACCTGGACTTTGG, 5'-GGGTGACATGGACTCTCAGCG, 5'-TGGGCAAGGCGCTGAGAGTCCATG,  
 5'-GCTGGAGGGCGGGTCTCAGGGAGC, and 5'-AGGACAGAGGTGGAGCCAAGACGACATAGCC. Preparation of  
 cDNA from human foreskin keratinocytes and prostate carcinoma tissue has been described previously. The PCRs were  
 25 performed under the conditions described above or for PCR with degenerate primers as described previously. Nested  
 PCRs were done by replacing the cDNA with 1.0 $\mu$ l from the first PCR reaction. The 5'-end of the cDNA was isolated by  
 5'-RACE as described above with the exception of using the gene-specific oligonucleotides P7, 5'-TGAAGCTCAGCCG-  
 GAGGTAGAAG, and 5'-GACAGACTCAAGCCGCAAGGTTG.

### 30 Northern Hybridization

**[0062]** A human RNA Master Blot containing poly(A)<sup>+</sup> mRNA of 50 different tissues was obtained from Clontech  
 Laboratories, Inc (Palo Alto, CA). <sup>32</sup>P-labeled probes were prepared by random prime labelling of DNA fragments of the  
 different transglutaminase gene products using the Multiprime DNA Labelling System (Amersham, Int, Amersham, UK).  
 35 DNA fragments of 500-700bp comprising the 3'-end of TG $_{\chi}$ , TG $_z$ , and band 4.2 protein, were generated by restriction  
 with *Pst* I and *Acc* I, *Nco* I and *Not* I (exon XII and XIII), and *Xho* I, respectively. The cDNA encoding human band 4.2  
 protein (Korsgren et al. 1990 Biochemistry 87,613-617) was kindly provided by Dr Carl M. Cohen, Boston, MA. A ~  
 220bp <sup>32</sup>P-labeled fragment of TG $_{\gamma}$  was generated by PCR using oligonucleotides 5'-CAGCCTCAGTCACCGCCATC-  
 CGC and 5'-GATACTTGTAGAGGTCAGTGATG. Hybridization was performed under the conditions recommended by  
 40 the manufacturer. The labeled membrane was exposed to BioMax MR film (Eastman Kodak) and films developed after  
 15 to 24h for first exposure and 3 to 5 days for second exposure.

### Amplification of TG $_{\gamma}$ and TG $_z$ from different tissues

**[0063]** cDNA from various cell lines and human tissue was prepared as previously described. A panel of cDNAs from  
 45 human tissue (Multiple Tissue cDNA Panel I) were also obtained from Clontech Laboratories. A 365 or 287bp fragment  
 of TG $_z$  was amplified by PCR using oligonucleotides 5'-TGGGCAAGGCGCTGAGAGTCCATG and 5'-CTCTGGAG-  
 GGCGGTCTCAGGGAGC or 5'-AGGACAGAGGTGGAGCCAAGACGACATAGCC, respectively, with an annealing  
 temperature of 60°C. A 218 or 170bp fragment of TG $_{\gamma}$  was amplified by PCR using oligonucleotides 5'-CAGCCTCAGT-  
 50 CACCGCCATCCGC and 5'-GATACTTGTAGAGGTCAGTGATG or 5'-GTGAAGGACTGTGCGCTGATG and 5'-CG-  
 GGAAGTGAGGGCTTACAAG, respectively, and identical conditions as above.

### Mapping of Transglutaminase Genes in Mouse Genome

**[0064]** The 100 radiation hybrid (RH) clones of the T31 mouse/hamster RH panel (McCarthy et al., (1997), Genome  
 Res., 7, 1153-1161) (Research Genetics, Huntsville, AL) were screened by PCR. A 139bp fragment of the *tgm5* gene  
 was amplified with primers 5'-TGAGGACTGTGTGCTGACCTTG (f) and 5'-TCCTGTGTCTGGCCTAGGG (r), a 149bp  
 55 fragment of the *epb42* gene with primers 5'-CAGGAGGAGTAAGGGGAATTGG (f) and 5'-TGCAGGCTACTGGAATC-

## EP 1 317 548 B1

CACG (r), a 400bp fragment of *tgm7* with primers 5'-GGGAGTGGCCTCATCAATGG (f) and 5'-CCTTGACCTCACT-GCTGCTGA (r), a -600bp fragment of *tgm3* with primers 5'-TCGGTGGCAGCCTCAAGATTG (f) and 5'-AGACATCAAT-GGGCAGGCAGTGG (r), and 655bp and 232bp fragments of *tgm2* with primers 5'-TTGGGGAGCTGGAGAGCAAC (f) and 5'-ATCCAGGACTCCACCCAGCA (r) and primers 5'-(GCGGCCGCTAGT)CCACATTGCAGGGCTCCTGACT (f) and 5'-GCTAGCCTGTGCTCACCATGAGG (r), respectively. PCRs were carried out in a GeneAmp 9600 thermalcycler with 0.035 units/ $\mu$ l AmpliTaq Gold polymerase in standard reaction buffer containing 2mM MgCl<sub>2</sub>, 0.2mM dNTPs, 0.4 $\mu$ M of each primer and 2.5 ng/ $\mu$ l genomic DNA in a total reaction volume of 25 $\mu$ l. PCR conditions were: polymerase activation for 10min at 95°C, annealing at 60°C for 45sec, extension at 72°C for 1min and denaturation at 94°C for 30sec for 35 cycles with a final extension of 3.5min at 72°C. PCR reactions were analyzed by agarose gel electrophoresis using 1% or 1.5% gels. The hybrid cell panel was analyzed at least twice in each case to exclude PCR related errors. The data was submitted to the Jackson Laboratory Radiation Hybrid Database for analysis and mapped relative to known genomic markers (<http://www.jax.org/resources/documents/cmdata.rhmap>).

5

10

15

20

25

30

35

40

45

50

55

**Table 1. Splice donor and acceptor sequences in the human TGM5 gene. Residues consistent with the splice site consensus sequence (MAG/GTRAG and YAG/G) are underlined.**

Intron number	Donor sequence	Acceptor sequence
1	M A Q GCTACCATGGCCCAAGGtagggaaagcccccctgtggccactggagtt	G L E V A ttttgtctaaccctggctgccccattgagGGGTAGAAAGTGGCC
2	F V V E T TTCGTGGTTGAAACTGgtaagaaccccagctggctcacaggggctg	G P L P D tgaggggcctcagctctacttccctcctctagGACCCGCTGCCAGAC
3	N P W C P AATCCCTGGTCCCAAGgtaaggctgggtgcccaggggtgctcct	E D A V Y tgcttcgtgccctcccactctggttcctctagAGGATGCTGTCTAC
4	W N Y G Q TGGAACTATGGACAGgtagtctcagccctgcttatggcccatcc	F E D K I tgccttccctctctgcctctccccccgaagTTTGAAGACAAAATC
5	V V C A M GTGGTGTGCCATGgtgaggtccctggctgcccggggagggagg	I N S N D ctcacactctctatataggcttctctctcagATCAACAGCAATGAT
6	A V M C T GCCGTCATGTGCACAGgtaggtagaagaaaggacctcacaaaaagg	V M R C L acaggtgattttttgtgcccctttttgtagTGATGAGGTGCTGTG
7	K D T I W AAGGATACTATCTGgtgagaacaacctctcaacctatttctag	N F H V W caacgtctcccctggctctgtttgatacagGAACCTCCATGTCTGG
8	Q E M S N CAGGAGATGAGCAACGgtgaggctctccagaagaaggcaggccccc	G V Y C C gcccaccgaggctcccctgttctccttcagGGTCTACTGCTGT
9	Y K Y E E TACAAGTATGAAGAAGgtagtaagcaagccagccctactcagagc	G S L Q E cagctggtgctgctctcccccaacttcagGATCCCTCCAGGAG
10	L S P K E CTCTCTCCTAAAGAAGgtacgcatgtgcacagtttgtgtacgcaga	A K T Y P tctcaccccafccttgtgttcttcttcttagCAAAAGACCTACCCC
11	S I T I N AGCATCAGGATTAAtgtaggaggtcctgcaaatggcttgtgg	V L G A A taattctccttccccctcctggtctgtttagGTTCTAGGAGCAGCC
12	Q Q K V F CAGCAGAAAGTCTTgtaagtgtgcaagtgtcagccctctcct	L G V L K ttttctgacatgctccattctctgttggcagCCTTGGAGTCTCAAA

**Table II. Intron sizes and splice types in the human TGM5 gene.** Sizes of introns are estimated to be within about a 100bp unless indicated to be sequenced entirely.

Intron number	Splice type	Size	Method
1	1	6,300bp	PCR
2	1	102bp	Sequencing
3	1	3,300bp	PCR
4	0	2,900bp	PCR
5	0	600bp	PCR
6	1	~11,800bp	PCR and Restriction Analysis
7	2	1,600bp	PCR
8	1	106bp	Sequencing
9	1	2,900bp	PCR
10	1	545bp	Sequencing
11	0	1100bp	PCR
12	2	209bp	Sequencing

**Table III. Apparent polymorphisms in the cDNA and genomic DNA sequences for TG<sub>x</sub>.** The positions with nucleotide and amino acid variations are underlined.

Residue	cDNA (a)	Gene	cDNA (b)
67	<u>S</u>	<u>TCA</u>	<u>P</u>
220	<u>Y</u>	<u>TAC</u>	<u>Y</u>
352	<u>A</u>	<u>GCA</u>	<u>A</u>

a. Aeschlimann et al., 1998  
b. additional sequence variant isolated in this work

**Claims**

1. A nucleotide sequence which has at least 80% homology to the nucleotide sequence of Fig. 10A or Fig. 10B and which nucleotide sequences encode a polypeptide having transglutaminase activity.
2. A nucleotide sequence according to claim 1 consisting of the nucleotide sequence of Fig. 10A or Fig. 10B.
3. A nucleotide sequence which hybridises under stringent conditions to the nucleotide sequence of Fig. 10A or Fig. 10B and which encodes a polypeptide having transglutaminase activity.
4. A method of expressing a polypeptide comprising inserting a nucleotide sequence according to any one of claims 1 to 3 into a suitable host cell and expressing that nucleotide sequence in order to express a polypeptide having transglutaminase activity.
5. A vector comprising a nucleotide sequence according to any one of claims 1 to 3.
6. A polypeptide having an amino acid sequence which is at least 90% identical to the amino acid sequences of Fig. 10A or Fig. 10B and which has transglutaminase activity.
7. A polypeptide according to claim 6 wherein the amino acid sequence differs from that given in Fig. 10A or Fig. 10B by about 1 to 20 amino acid additions, deletions or substitutions.
8. A polypeptide according to any one of claims 6 to 7 comprising exons II through to exon IV of the sequence shown in Fig. 10A or Fig. 10B.

9. A polypeptide according to any one of claims 6 to 7 comprising exon X through to exon XII of the sequence shown in Fig. 10A or Fig. 10B.
10. A composition comprising a polypeptide according to any one of claims 6 to 9 suitable for use in cross-linking proteins.
11. A composition comprising a polypeptide according to any one of claims 6 to 9 suitable for use in a transamidation reaction on peptides and polypeptides.
12. A diagnostic method comprising detecting expression of a polypeptide according to any one of claims 6 to 9 in cells derived from a subject.
13. An antibody directed against a polypeptide according to any one of claims 6 to 9.
14. A method of diagnosis of autoimmune disease comprising testing a sample isolated from a subject for the presence of a transglutaminase encoded by the nucleotide sequences of Fig. 10A or Fig. 10B, or portions thereof.
15. A method according to claim 14 wherein the autoimmune disease to be diagnosed is selected from Addison's disease, AI haemolytic anaemia, AI thrombocytopenic purpura, AI thyroid diseases, atrophic gastritis - pernicious anaemia, Crohn's disease, colitis ulcerosa, Goodpasture syndrome, IgA nephropathy or IGA glomerulonephritis, myasthenia gravis, partial lipodystrophy, polymyositis, primary biliary cirrhosis, primary sclerosing cholangitis, progressive systemic sclerosis, recurrent pericarditis, relapsing polychondritis, rheumatoid arthritis, rheumatism, sarcoidosis, Sjögren's syndrome, SLE, splenic atrophy, type I (insulin-dependent) diabetes mellitus, diabetes mellitus, Wegener granulomatosis, ulcerative colitis, vasculitis (both systemic and cutaneous), vitiligo.
16. A method of diagnosis of autoimmune diseases comprising the detection in a sample isolated from a subject of antibodies against a transglutaminase encoded by the nucleotide sequences of Fig. 10A or Fig. 10B or portions thereof.
17. A method according to claim 16 comprising non-endogenous transglutaminase TG<sub>Y</sub> (Fig 10A or 10B) as a competitive antigen.
18. A method according to claim 17 also comprising non-endogenous transglutaminase TG<sub>Z</sub> (Fig. 6A or 6B) as a competitive antigen.
19. A method according to claim 16 to 18, wherein the method is a competitive immunoassay selected from RIA, EIA/ELISA, LiA and FiA.

#### Patentansprüche

1. Eine Nukleotidsequenz, die mindestens 80% Homologie zu der Nukleotidsequenz in Fig. 10A oder Fig. 10B aufweist und deren Nukleotidsequenzen für ein Polypeptid codieren, das Transglutaminase-Aktivität aufweist.
2. Eine Nukleotidsequenz gemäß Anspruch 1, die aus der Nukleotidsequenz aus Fig. 10A oder Fig. 10B besteht.
3. Eine Nukleotidsequenz, die unter stringenten Bedingungen an die Nukleotidsequenz aus Fig. 10A oder Fig. 10B hybridisiert und die für ein Polypeptid, das Transglutaminase-Aktivität aufweist, codiert.
4. Ein Verfahren zur Expression eines Polypeptids umfassend eine Nukleotidsequenz gemäß einem der Ansprüche 1 bis 3 in eine geeignete Wirtszelle einzubringen und diese Nukleotidsequenz zu exprimieren, um ein Polypeptid zu exprimieren, das Transglutaminase-Aktivität aufweist.
5. Ein Vektor, der eine Nukleotidsequenz gemäß einem der Ansprüche 1 bis 3 aufweist.
6. Ein Polypeptid, das eine Aminosäuresequenz aufweist, die zu mindestens 90% identisch mit der Aminosäuresequenz aus Fig. 10A oder Fig. 10B ist und das Transglutaminase-Aktivität aufweist.
7. Ein Polypeptid gemäß Anspruch 6, wobei sich die Aminosäuresequenz von derjenigen, die in Fig. 10A oder Fig.

10B angegeben ist, in ungefähr 1 bis 20 zusätzlichen Aminosäuren, Aminosäuredeletionen oder -substitutionen unterscheidet.

- 5 8. Ein Polypeptid gemäß einem der Ansprüche 6 oder 7, das die Exons II durchgehend bis Exon IV der Sequenz umfasst, die in Fig. 10A oder Fig. 10B angegeben ist.
9. Ein Polypeptid gemäß einem der Ansprüche 6 oder 7, das Exon X durchgehend bis Exon XII der Sequenz umfasst, die in Fig. 10A oder Fig. 10B angegeben ist.
- 10 10. Eine Zusammensetzung, die ein Polypeptid gemäß einem der Ansprüche 6 bis 9 umfasst, die zur Verwendung bei der Kreuzvernetzung von Proteinen geeignet ist.
11. Eine Zusammensetzung, die ein Polypeptid gemäß einem der Ansprüche 6 bis 9 umfasst, die zur Verwendung bei einer Transamidierungsreaktion an Peptiden und Polypeptiden geeignet ist.
- 15 12. Ein diagnostisches Verfahren, daß die Detektion der Expression eines Polypeptids gemäß einem der Ansprüche 6 bis 9 in von einer Person gewonnenen Zellen umfasst.
13. Ein Antikörper, der gegen ein Polypeptid gemäß einem der Ansprüche 6 bis 9 gerichtet ist.
- 20 14. Ein Verfahren zur Diagnose einer Autoimmunerkrankung, das das Testen einer von einer Person gewonnenen Probe auf die Gegenwart einer Transglutaminase umfasst, die von den Nukleotidsequenzen aus Fig. 10A oder Fig. 10B oder Abschnitten daraus kodiert wird.
- 25 15. Ein Verfahren gemäß Anspruch 14, wobei die Autoimmunerkrankung, die diagnostiziert werden soll, aus Morbus Addison, autoimmunvermittelter hämolytischer Anämie, autoimmunvermittelter thrombozytopenischer Purpura, autoimmunvermittelter Schilddrüsenerkrankungen, atrophischer Gastritis - perniziöser Anämie, Morbus Crohn, Colitis ulcerosa, Goodpasture-Syndrom, IgA-Nephropathie oder IgA-Glomerulonephritis, Myasthenia gravis, partieller Lipodystrophie, Polymyositis, primärer biliärer Zirrhose, primär sklerosierender Cholangitis, progressiver systemischer Sklerose, rezidivierender Perikarditis, rezidivierender Polychondritis, rheumatoider Arthritis, Rheumatismus, Sarkoidose, Sjögren-Syndrom, SLE, Milzatrophy, Diabetes mellitus Typ 1 (Insulin-abhängig), Diabetes mellitus, Wegener-Granulomatose, Colitis ulcerosa, Vaskulitis (sowohl systemisch als kutan), Weißfleckenkrankheit ausgewählt wird.
- 30 16. Ein Verfahren zur Diagnose von Autoimmunerkrankungen, das die Detektion von Antikörpern gegen eine Transglutaminase, die von den Nukleotidsequenzen aus Fig. 10A oder Fig. 10B oder Abschnitten davon kodiert wird, in einer von einer Person gewonnenen Probe umfasst.
- 35 17. Ein Verfahren gemäß Anspruch 16, das die nicht-endogene Transglutaminase TG<sub>γ</sub> (Fig. 10A oder Fig. 10B) als kompetitives Antigen umfasst.
- 40 18. Ein Verfahren gemäß Anspruch 17, das zudem die nicht-endogene Transglutaminase TG<sub>z</sub> (Fig. 6A oder Fig. 6B) als kompetitives Antigen umfasst.
- 45 19. Ein Verfahren gemäß Anspruch 16 bis 18, wobei das Verfahren ein kompetitiver Immunoassay ist, der aus RIA, EIA/ELISA, LiA und FiA ausgewählt wird.

### Revendications

- 50 1. Une séquence nucléotidique ayant au moins 80% d'homologie avec la séquence nucléotidique de fig. 10A ou fig. 10B et où lesdites séquences nucléotidiques encodent pour un polypeptide ayant activité transglutaminase.
2. Une séquence nucléotidique selon la revendication 1 consistant en la séquence nucléotidique de fig. 10A ou fig. 10B.
- 55 3. Une séquence nucléotidique qui s'hybride avec la séquence nucléotidique de fig. 10A ou fig. 10B dans des conditions stringentes et qui encode pour un polypeptide ayant activité transglutaminase.

## EP 1 317 548 B1

4. Un procédé pour exprimer un polypeptide comprenant l'insertion d'une séquence nucléotidique selon l'une des revendications 1 à 3 dans une cellule hôte appropriée et l'expression de cette séquence nucléotidique pour exprimer un polypeptide ayant activité transglutaminase.
5. Un vecteur comprenant une séquence nucléotidique selon l'une des revendications 1 à 3.
6. Un polypeptide ayant une séquence d'acides aminés qui est identique au moins à 90% avec la séquence d'acides aminés de la fig. 10A ou de la fig. 10B et qui a activité transglutaminase.
7. Un polypeptide selon la revendication 6 dans lequel la séquence d'acides aminés diffère de celle donnée en fig. 10A ou fig. 10B par à peu près 1 à 20 additions, délétions ou substitutions d'acides aminés.
8. Un polypeptide selon l'une des revendications 6 à 7 comprenant les exons II jusqu'à l'exon IV de la séquence donnée en fig. 10A ou fig. 10B.
9. Un polypeptide selon l'une des revendications 6 à 7 comprenant l'exon X jusqu'à l'exon XII de la séquence donnée en fig. 10A ou fig. 10B.
10. Une composition comprenant un polypeptide selon l'une des revendications 6 à 9 appropriée pour l'usage dans la liaison transversale de protéines.
11. Une composition comprenant un polypeptide selon l'une des revendications 6 à 9 appropriée pour une réaction de transamidation sur des peptides et polypeptides.
12. Un procédé diagnostique comprenant la détection de l'expression d'un polypeptide selon l'une des revendications 6 à 9 dans des cellules dérivées d'une personne.
13. Un anticorps dirigé contre un polypeptide selon l'une des revendications 6 à 9.
14. Un procédé de diagnostic d'une maladie auto-immune comprenant l'analyse d'un prélèvement isolé d'une personne sur la présence d'une transglutaminase encodée par la séquence nucléotidique de fig. 10A ou fig. 10B ou sections de laquelle.
15. Un procédé selon la revendication 14 dans laquelle la maladie auto-immune à diagnostiquer est choisie de la maladie d'Addison, anémie hémolytique auto-immune, purpura thrombocytopénique auto-immune, maladies de thyroïde auto-immunes, atrophie gastrique - anémie pernicieuse, maladie de Crohn, colite ulcéreuse, syndrome de Good-pasture, néphropathie à IgA ou glomérulonéphrite à IgA, myasthénie grave, lipodystrophie partielle, polymyosite, cirrhose biliaire primitive, cholangite sclérosante primitive, sclérose systémique progressive, péricardite récurrente, polychondrite récidivante, polyarthrite rhumatoïde, rhumatisme, sarcoïdose, syndrome de Goujerot-Sjögren, LED, atrophie splénique, diabète de type 1 (insulino-dépendante), diabète sucré, granulomatose de Wegener, rectocolite hémorragique, vascularite (systémique et cutanée), vitiligo.
16. Un procédé de diagnostic de maladies auto-immunes comprenant la détection d'anticorps contre une transglutaminase encodée par la séquence nucléotidique de fig. 10A ou fig. 10B ou sections de laquelle dans un prélèvement isolé d'une personne.
17. Un procédé selon la revendication 16 comprenant la transglutaminase non-endogène TG<sub>γ</sub> (fig. 10A ou fig. 10B) comme antigène compétitif.
18. Un procédé selon la revendication 17 comprenant en outre la transglutaminase non-endogène TG<sub>z</sub> (fig. 6A ou fig. 6B) comme antigène compétitif.
19. Un procédé selon les revendications 16 à 18 dans lequel le procédé est une analyse immunologique choisie de RIA, EIA/ELISA, LiA et FIA.



**Fig. 2A**

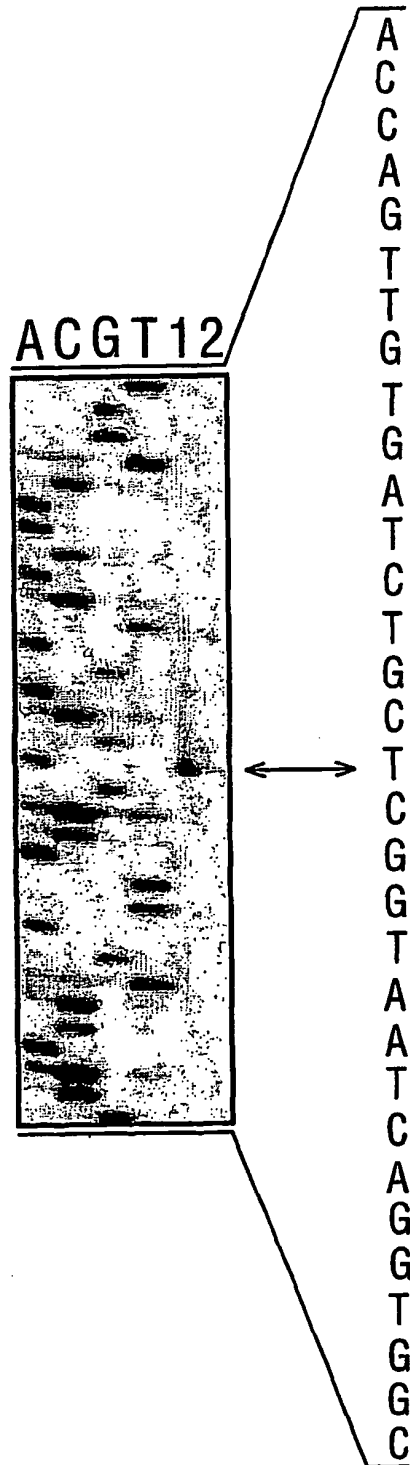


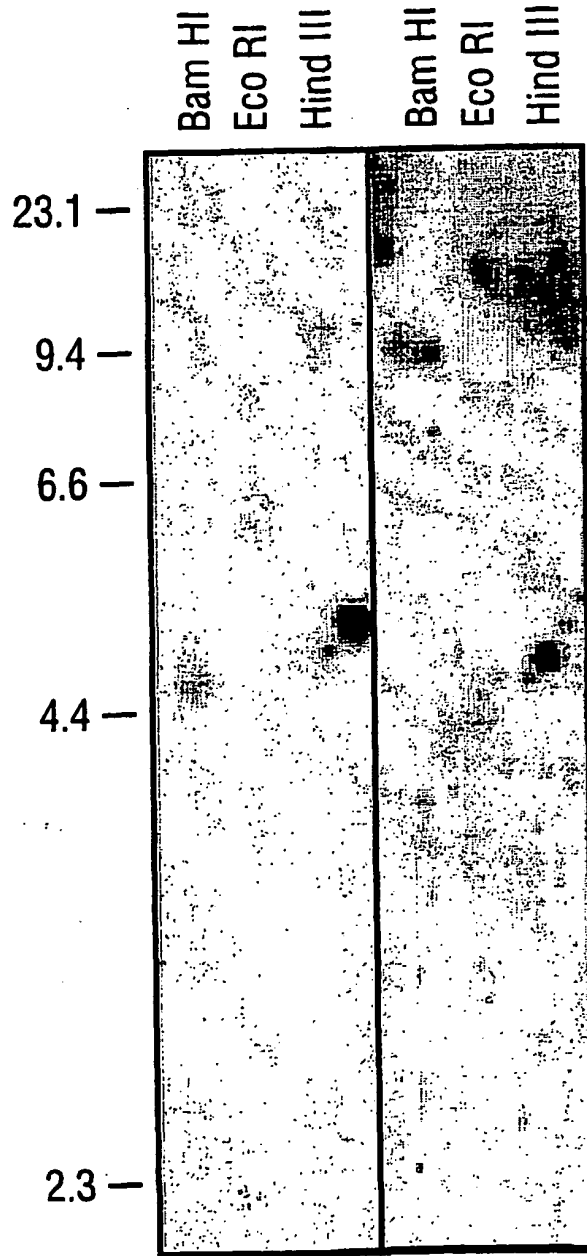


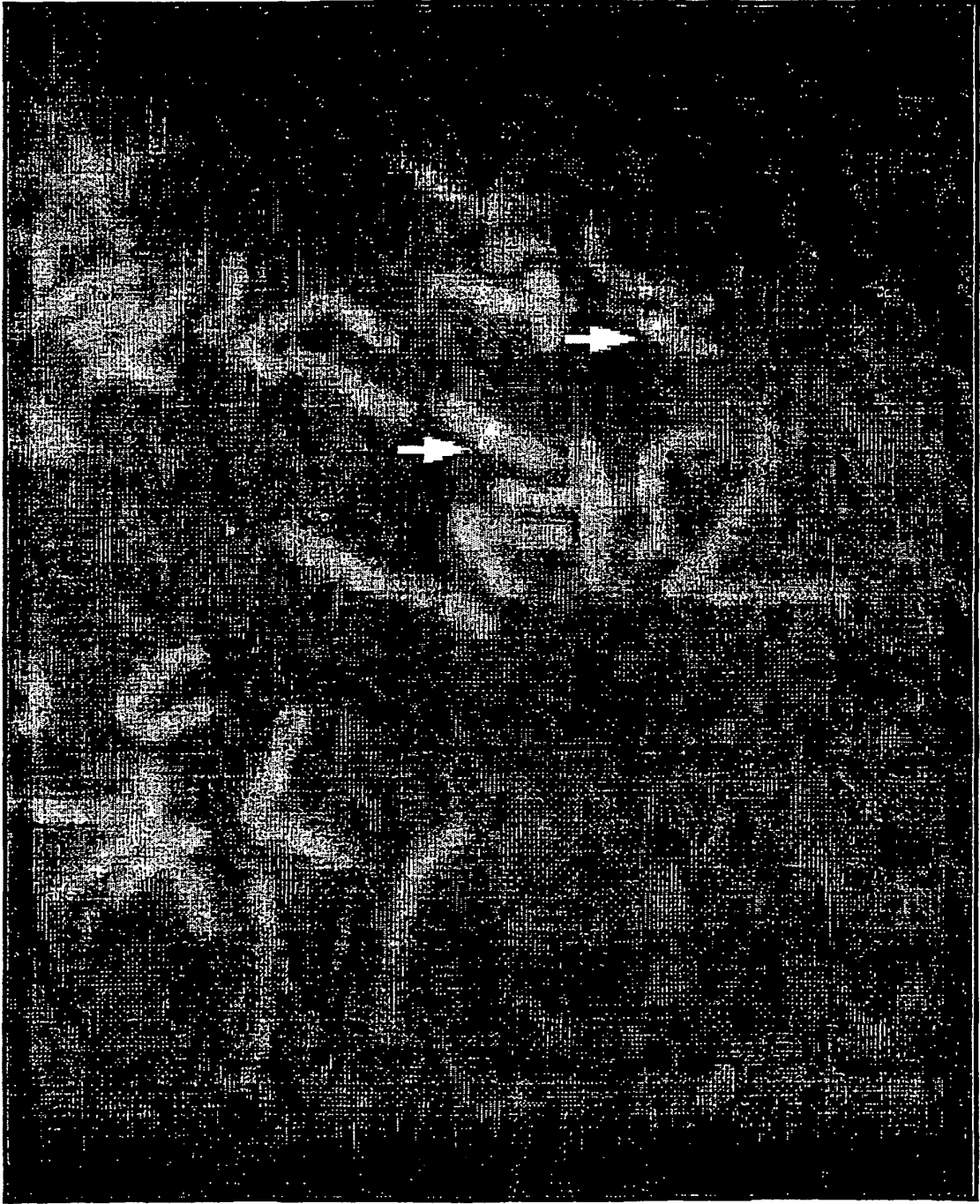
Fig. 3

V D F A L och

2303 GTAGACTTTGCATTAaaattctggaac<sup>▼</sup>aacgcgc<sup>▼</sup>agacgtgtgaattc<sup>▼</sup>aagcttcagg  
2364 aaaaggagcaagttcaaatgcaagctgcgccattccccaccacaacagaggcttcacagggctc  
2428 cagcaagagccacagaggggatgacgtgttcatttctgtctctcctgactccactagaaattt  
2492 aagctccatgaggcaagaactttgctttgtttactaccctatactcagaaccatttctggca  
2556 tatgctaggcactcaacaaatattttgaaatgaatgaggactccagcatccagagaaac  
2620 aggtaggaaatgtctatggatggatatttccctggaccatttgcacagctcccctggactcttt  
2684 tcagggcccaggattccactgtgtcccatccagagattccaggattccagtcacctatccaga  
2748 agcgtgatttggcacagaggtcagaggatactggtaggacttggccatgacttaactgccccct  
2812 gcccagatatccaggaagaaagacagggctgaacagctcactgttgtttgttgttgcgaa  
2876 agctaattccctagtagtaaaacttcagacccttgctcctctcctcatgtagtcatcact  
2940 ttcctatctgttcccttggatagctgcagctcccatccaaaaaagtcatttattgagtgcc  
3004 tagcatagccagaagtgttctgagtgtaggggtacaaagtaaacaaagcaagtcctgccc  
3068 tcatggagcgcacatttctcagtgagg

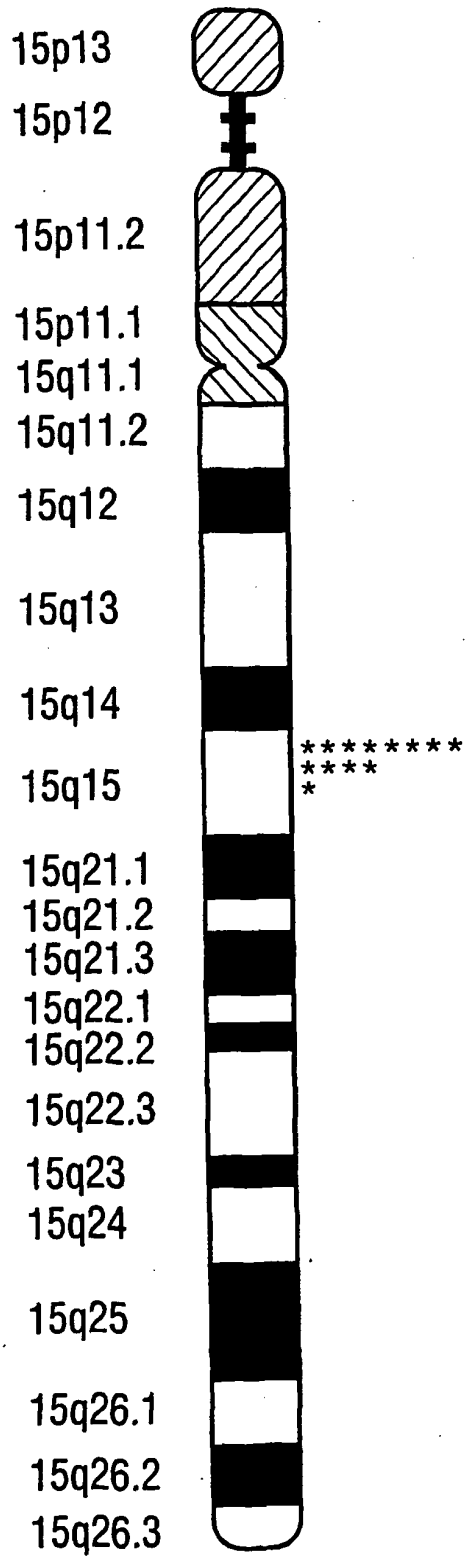
**Fig. 4**



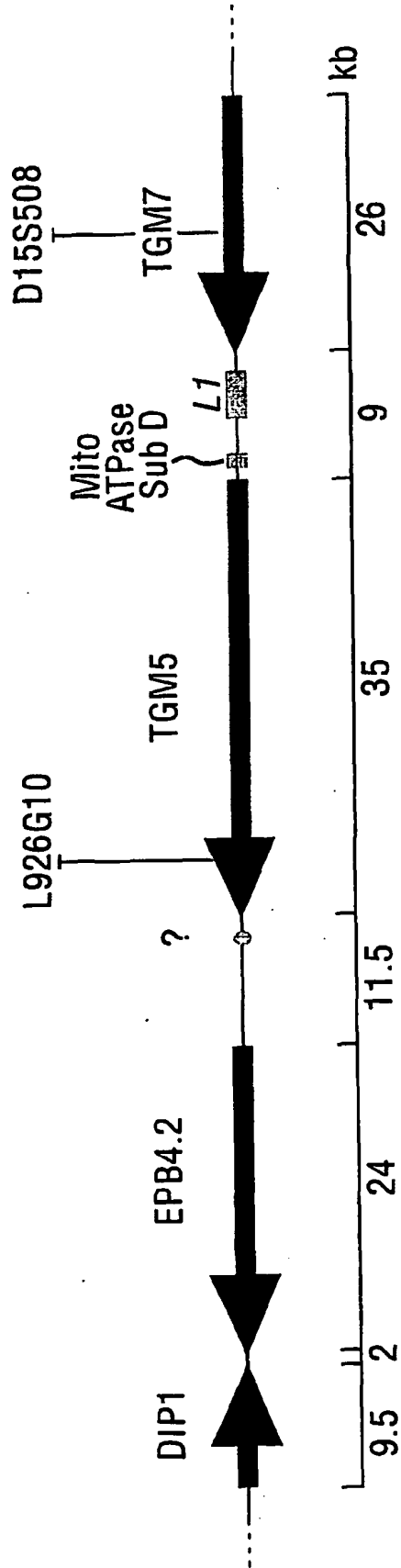


*Fig. 5A*

**Fig. 5B**



**Fig 5C**





**Fig. 6A(contd.)**

G	Q	P	V	K	Y	G	Q	C	W	V	F	A	S	V	M	C	T
		878			887			896			905			914			923
GTA	ATG	AGA	TGC	TTA	GGT	GTT	CCA	ACC	CGT	GTT	GTT	TCC	AAT	TTC	CGT	TCC	GCG
V	M	R	C	L	G	V	P	T	R	V	V	S	N	F	R	S	A
		932			941			950			959			968			977
CAC	AAC	GTG	GAT	AGG	AAC	TTG	ACC	ATC	GAT	ACG	TAC	TAT	GAC	CGA	AAT	GCC	GAG
H	N	V	D	R	N	L	T	I	D	T	Y	Y	D	R	N	A	E
		986			995			1004			1013			1022			1031
ATG	CTG	TCA	ACT	CAG	AAA	CGA	GAC	AAA	ATA	TGG	AAC	TTC	CAC	GTC	TGG	AAT	GAG
M	L	S	T	Q	K	R	D	K	I	W	N	F	H	V	W	N	E
		1040			1049			1058			1067			1076			1085
TGC	TGG	ATG	ATC	CGG	AAA	GAT	CTC	CCA	CCA	GGA	TAC	AAC	GGG	TGG	CAG	GTT	CTG
C	W	M	I	R	K	D	L	P	P	G	Y	N	G	W	Q	V	L
		1094			1103			1112			1121			1130			1139
GAC	CCC	ACT	CCC	CAG	CAG	ACC	AGC	AGT	GGG	CTG	TTC	TGC	TGT	GGC	CCT	GCC	TCT
D	P	T	P	Q	Q	T	S	S	G	L	F	C	C	G	P	A	S
		1148			1157			1166			1175			1184			1193
GTG	AAG	GCC	ATC	AGG	GAA	GGG	GAT	GTC	CAC	CTG	GCC	TAT	GAC	ACC	CCT	TTT	GTG
V	K	A	I	R	E	G	D	V	H	L	A	Y	D	T	P	F	V
		1202			1211			1220			1229			1238			1247
TAT	GCC	GAG	GTG	AAC	GCC	GAT	GAA	GTC	ATT	TGG	CTC	CTT	GGG	GAT	GGC	CAG	GCC
Y	A	E	V	N	A	D	E	V	I	W	L	L	G	D	G	Q	A
		1256			1265			1274			1283			1292			1301
CAG	GAA	ATC	CTG	GCC	CAC	AAC	ACC	AGT	TCC	ATC	GGG	AAG	GAG	ATC	AGC	ACT	AAG
Q	E	I	L	A	H	N	T	S	S	I	G	K	E	I	S	T	K
		1310			1319			1328			1337			1346			1355
ATG	GTG	GGG	TCA	GAC	CAG	CGC	CAG	AGC	ATC	ACC	AGC	TCC	TAC	AAG	TAC	CCA	GAA
M	V	G	S	D	Q	R	Q	S	I	T	S	S	Y	K	Y	P	E
		1364			1373			1382			1391			1400			1409
GGA	TCC	CCT	GAG	GAG	AGA	GCT	GTC	TTC	ATG	AAG	GCT	TCT	CGG	AAA	ATG	CTG	GCC
G	S	P	E	E	R	A	V	F	M	K	A	S	R	K	M	L	G
		1418			1427			1436			1445			1454			1463
CCC	CAA	AGA	GCT	TCT	TTG	CCC	TTC	CTG	GAT	CTC	CTG	GAG	TCT	GGG	GGT	CTT	AGG
P	Q	R	A	S	L	P	F	L	D	L	L	R	S	G	G	L	R
		1472			1481			1490			1499			1508			1517
GAT	CAG	CCA	GCG	CAG	CTG	CAG	CTT	CAC	CTG	GCC	AGG	ATA	CCC	GAG	TGG	GGC	CAG
D	Q	P	A	Q	L	Q	L	H	L	A	R	I	P	E	W	G	Q
		1526			1535			1544			1553			1562			1571
GAC	CTG	CAG	CTG	CTG	CTG	CGT	ATC	CAG	AGG	GTG	CCA	GAC	AGC	ACC	CAC	CCT	CGG
D	L	Q	L	L	L	R	I	Q	R	V	P	D	S	T	H	P	R
		1580			1589			1598			1607			1616			1625
GGG	CCC	ATC	GGA	CTG	GTG	GTG	CGC	TTC	TGT	GCA	CAG	GCC	CTG	CTG	CAT	GGG	GGT
G	P	I	G	L	V	V	R	F	C	A	Q	A	L	L	H	G	G
		1634			1643			1652			1661			1670			1679
GGT	ACC	CAG	AAG	CCC	TTC	TGG	AGG	CAC	ACA	GTG	CGG	ATG	AAC	CTG	GAC	TTT	GGG
G	T	Q	K	P	F	W	R	H	T	V	R	M	N	L	D	F	G
		1688			1697			1706			1715			1724			1733
AAG	GAG	ACA	CAG	TGG	CCG	CTC	CTC	CTG	CCC	TAC	AGC	AAT	TAC	AGA	AAC	AAG	CTA
K	E	T	Q	W	P	L	L	L	P	Y	S	N	Y	R	N	K	L

1742            1751            1760            1769            1778            1787  
 ACG GAC GAA AAG CTC ATC CGC GTG TCT GGC ATC GCG GAG GTT GAA GAG ACA GGG  
 T D E K L I R V S G I A E V E E T G

1796            1805            1814            1823            1832            1841  
 AGG TCC ATG CTG GTC CTA AAA GAT ATC TGT CTG GAG CCT CCC CAC TTG TCT ATT  
 R S M L V L K D I C L E P P H L S I

1850            1859            1868            1877            1886            1895  
 GAG GTG TCT GAG AGG GCT GAG GTG GGC AAG GCG CTG AGA GTC CAT GTC ACC CTC  
 E V S E R A E V G K A L R V H V T L

1904            1913            1922            1931            1940            1949  
 ACC AAC ACC TTA ATG GTG GCT CTG AGC AGC TGC ACG ATG GTG CTG GAA GGA AGC  
 T N T L M V A L S S C T M V L E G S

1958            1967            1976            1985            1994            2003  
 GGC CTC ATC AAT GGG CAG ATA GCA AAG GAC CTT GGG ACT CTG GTG GCC GGA CAC  
 G L I N G Q I A K D L G T L V A G H

2012            2021            2030            2039            2048            2057  
 ACC CTC CAA ATT CAA CTG GAC CTC TAC CCG ACC AAA GCT GGA CCC CGC CAG CTC  
 T L Q I Q L D L Y P T K A G P R Q L

2066            2075            2084            2093            2102            2111  
 CAG GTT CTC ATC AGC AGC AAC GAG GTC AAG GAG ATC AAA GGC TAC AAG GAC ATA  
 Q V L I S S N E V K E I K G Y K D I

2120            2129            2138            2147            2156            2165  
 TTC GTC ACT GTG GCT GGG GCT CCC TGA GAC CCG CCC TCC AGC TGC CCT CCC TGG  
 F V T V A G A P \*

2174            2183            2192            2201            2210            2219  
 CAC CCC TGC CCC ACC TGG CTC CTT TCT ACT CCF GGC TAT GTC GTC TTG GCT CCA

2228            2237            2246            2255            2264            2273  
 CCT CTG TCC TCT CTC TAG CCT GCC TGG GAA TGA ATG AAG CTC TGT TAG AAA CAC

2282            2291            2300            2309  
 CGT GTG CTT TGG GAA GAG ACA ATA AAG ATG TCT TTA TTT ATC AC

**Fig. 6A(contd.)**

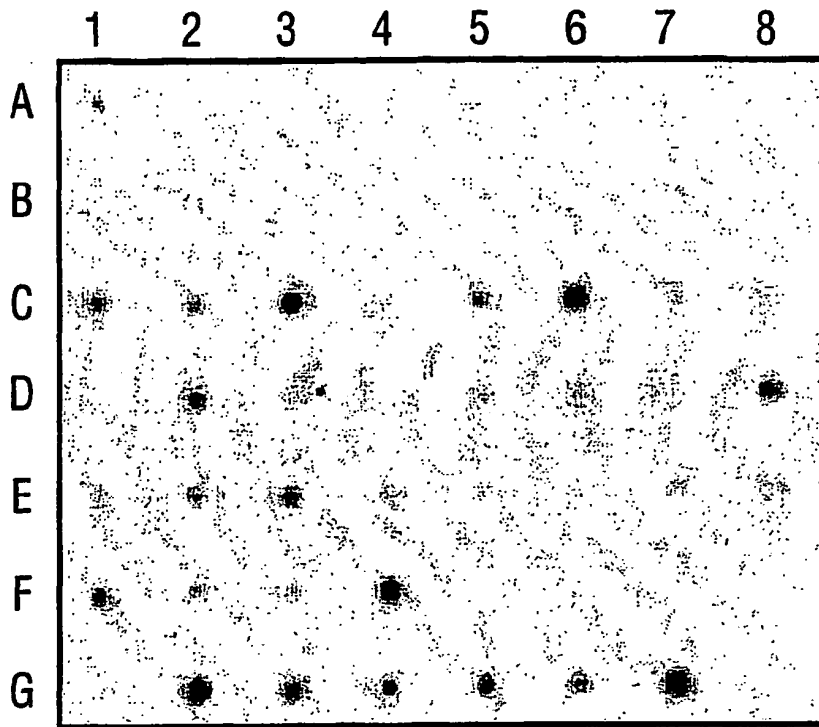
Fig. 6B

ATG	GAT	CAG	GTG	GCA	ACC	TTG	CGG	CTT	GAG	TCT	GTC	GAC	CTG	CAG	AGC	TCC	AGG	AAC	AAC	AAG	GAG	CAC	GG	GAG	5
M	D	Q	V	A	T	L	R	L	E	S	V	D	L	Q	S	S	R	N	N	K	E	H			74
CAC	ACG	CAG	GAG	ATG	GGC	ATC	AAG	CGG	CTC	ACT	GTG	CGC	CGC	GGC	CAG	CCC	TTC	TAC	CTC	CGG	CTG	AGC			23
H	T	Q	Q	E	M	G	V	K	R	L	T	V	R	R	G	Q	P	F	Y	L	R	L	S		46
TTT	AGC	CGA	CCC	TTT	CAG	TCC	CAG	AAC	GAC	CAC	ATC	ACC	TTT	GTG	GCT	GAG	ACC	GGA	CCC	AAG	CCG	TCA			212
F	S	R	P	F	Q	S	Q	N	D	H	I	T	F	V	A	E	T	G	P	K	P	S			69
GAG	CTG	CTG	GGG	ACC	CGA	GCC	ACA	TTT	CTC	ACC	CGG	CTC	ACA	CCA	CCC	GGG	AAT	GTC	TGG	AGC	GCT	TCT			281
E	L	L	G	T	R	A	T	F	L	T	R	V	Q	ACA	CCA	GCC	AAT	GCA	GTT	ATT	GGC	CAT			92
GAT	TTT	ACC	ATT	GAC	TCC	AAC	TCT	CTC	CAA	GTT	TCC	CTT	TTT	CTC	ACA	GCC	AAT	GCA	GTT	ATT	GGC	CAT			350
D	F	T	I	D	S	N	S	L	Q	V	S	L	F	T	P	A	N	A	V	I	G	H			115
TAC	ACT	CTG	AAA	ATA	GAG	ATC	TCT	TCT	CAG	GGC	CAA	GGT	CAC	AGT	CTG	ACT	TAC	CCG	CTG	GGA	ACT	TTC	ATC		419
Y	T	L	L	K	I	E	I	S	Q	G	Q	G	H	S	T	Y	P	L	G	T	F	I			138
CTA	CTT	TTT	AAC	CCT	TGG	AGT	CCA	GAG	GAC	GAC	GTC	TAC	CTG	CCA	AGT	GAA	ATA	CTG	CTG	CAG	GAG	TAT			488
L	L	F	N	P	W	S	P	E	D	D	V	Y	L	P	S	E	I	L	L	Q	E	Y			161
ATC	ATG	CGA	GAT	TAT	GGC	TTT	GTT	TAC	AAG	GGT	CAT	GAA	AGA	TTT	ATC	ACC	TCC	TGG	CCC	TGG	AAC	TAC			557
I	M	R	D	Y	G	F	V	Y	K	G	H	E	R	F	I	T	S	W	P	W	N	Y			184
GGG	CAG	TTT	GAA	GAG	GAC	ATC	ATA	GAC	ATC	TGC	TTT	GAG	ATC	CTG	AAC	AAG	AGC	CTG	TAT	CAC	TTA	AAG			626
G	Q	F	E	E	D	I	I	D	I	C	F	E	I	L	N	K	S	L	Y	H	L	K			207
AAC	CCG	GCC	AAA	GAC	TGT	TCC	CAG	CGG	AAC	GAC	GTG	GTG	TAT	GTG	TGC	AGG	GTG	GTG	AGT	GCC	ATG	ATC			695
N	P	A	K	D	C	S	Q	R	N	D	V	V	Y	V	C	R	V	V	S	A	M	I			230
AAC	AGC	AAC	GAT	GAC	AAT	GGC	GTG	CTG	CAG	GGG	AAC	TGG	GGC	GAG	GAC	TAC	TCC	AAA	GGG	GTC	AGT	CCT			764
N	S	N	D	D	N	G	V	L	Q	G	N	W	G	E	D	Y	S	K	G	V	S	P			253
CTG	GAG	TGG	AAG	GGC	AGT	GTG	GCC	ATC	CTA	CAG	CAG	TGG	TCA	GCC	AGG	GGC	GAG	CCT	GTG	AAG	TAC			833	
L	E	W	K	G	S	V	A	I	L	Q	Q	W	S	A	R	G	G	Q	P	V	K	Y			276
GGA	CAG	TGC	TGG	GTG	CTC	TTC	GCC	TCT	ATT	ATG	TGC	ACC	GTA	ATG	AGA	TGC	TTA	GGT	GTT	CCA	ACC	CGT	GTT		902
G	Q	C	W	V	F	A	S	V	M	C	T	V	M	R	C	L	G	V	P	T	R	V			299
GTT	TCC	AAT	TTT	CTC	TCC	GCG	CAC	AAC	GTG	GAT	AGG	AAC	TTG	ACC	ATC	GAT	ACG	TAC	TAT	GAC	CGA	AAT			971
V	S	N	F	R	S	A	H	N	V	D	R	N	L	T	I	D	T	Y	Y	D	R	N			322
GCC	GAG	ATG	CTG	TCA	ACT	CAG	AAA	CGA	GAC	AAA	ATA	TGG	AAC	TTT	CAC	GTC	TGG	AAT	GAG	TGC	TGG	ATG			1040
A	E	M	L	S	T	Q	K	R	D	K	I	W	N	F	H	V	W	N	E	C	W	M			345
ATC	CGG	AAA	GAT	CTC	CCA	GGA	TAC	AAC	GGG	TGG	CAG	GTT	CTG	GAC	CCC	ACT	CCC	CAG	CAG	ACC	AGC				1109
I	R	K	D	L	P	P	G	Y	N	G	W	Q	V	L	D	P	T	P	Q	Q	T	S			368

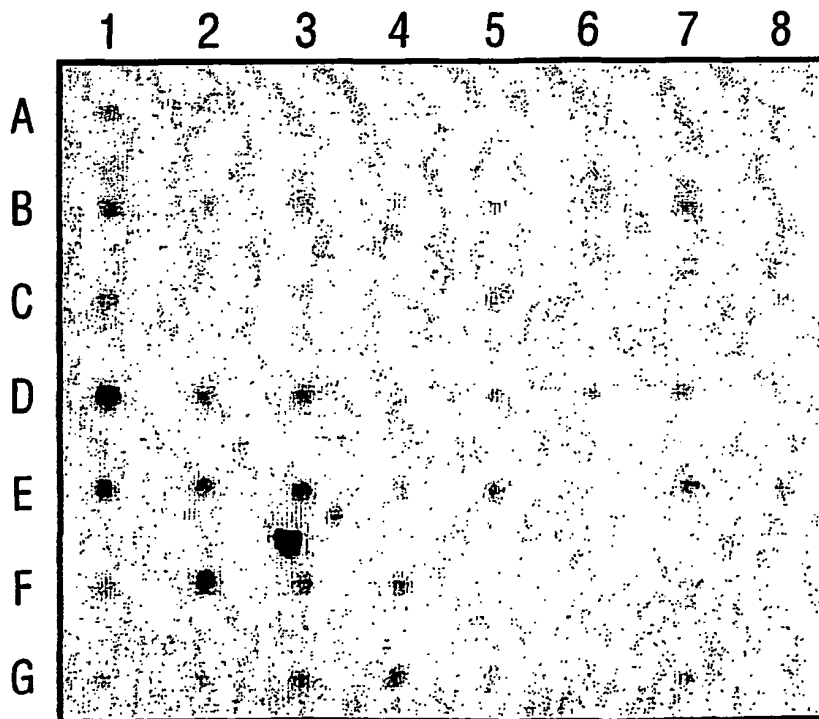
AGT GGG CTG TTC TGC TGT TGC CCT GGC CCT GGC TCT GTG AAG GCC ATC AGG GAA GGG GAT GTC CAC CTG GCC TAT 1178  
 S G L F C C C G P A S V K A I R E G D V H L A Y 391  
 GAC ACC CCT TTT GTG TAT GCC GAG GTG AAC GCC GAT GAA GTC ATT TGG CTC CTT GGG GAT GGC CAG GCC 1247  
 D T P F V Y A E V N A D E V I W L L G D G Q A 414  
 CAG GAA ATC CTG GCC CAC AAC ACC AGT TCC ATC GGG AAG GAG ATC AGC ACT AAG ATG GTG GGG TCA GAC 1316  
 Q E I L A H N T S S I G K E I S T K M V G S D 437  
 CAG CGC CAG AGC ATC ACC AGC TCC TAC AAG TAC CCA GAA GGA TCC CCT GAG GAG AGA GCT GTC TTC ATG 1385  
 Q R Q S I T S S Y K Y P E G S P E E R A V F M 460  
 AAG GCT TCT CGG AAA ATG CTG GGC CCC CAA AGA GCT TCT TTG CCC TTC CTG GAT CTC CTG GAG TCT GGG 1454  
 K A S R K K M L G P Q R A S L P F L D L L E S G 483  
 GGT CTT AGG GAT CAG CCA GCG CAG CTG CAG CTT CAC CTG GCC AGG ATA CCC GAG TGG GGC CAG GAC CTG 1523  
 G L R D Q P A Q L Q L H L A R I P E W G Q D L 506  
 CAG CTG CTG CGT ATC CAG AGG GTG CCA GAC AGC ACC CAC CCT CGG GGG CCC ATC GGA CTG GTG GTG 1592  
 Q L L L R I Q R V P D S T H P R G P I G L V V 529  
 CGC TTC TGT GCA CAG GCC CTG CTG CAT GGG GGT ACC CAG AAG CCC TTC TGG AGG CAC ACA GTG CGG 1661  
 R F C A Q A L L H G G T Q W P L L L P Y S N Y R N K 575  
 ATG AAC CTG GAC TTT GGG AAG GAG ACA CAG TGG CCG CTC CTC CTG CCC TAC AGC AAT TAC AGA AAC AAG 1730  
 M N L D F G K E T Q W P L L L P Y S N Y R N K 575  
 CTA ACG GAC GAA AAG CTC ATC CGC GTG TCT GGC ATC GCG GAG GTT GAA GAG ACA GGG AGG TCC ATG CTG 1799  
 L T D E K L I R V S G I A E V E T G R S M L 598  
 GTC CTA AAA GAT ATC TGT CTG GAG CCT CCC CAC TTG TCT ATT GAG GTG TCT GAG AGG GCT GAG GTG GGC 1868  
 V L K D I C L E P P H L S I E V S E R A E V G 621  
 AAG GCG CTG AGA GTC CAT GTC ACC CTC ACC AAC ACC TTA ATG GTG GCT CTG AGC AGC TGC ACG ATG GTG 1937  
 K A L R V H V T L T N T L M V A L S S C T M V 644  
 CTG GAA GGA AGC GGC CTC ATC AAT GGG CAG ATA GCA AAG GAC CTT GGG ACT CTG GTG GCC GGA CAC ACC 2006  
 L E G S G L I N G Q I A K D L G T L V A G H T 667  
 CTC CAA ATT CAA CTG GAC CTC TAC CCG ACC AAA GCT GGA CCC CGC CAG CTC CAG GTT CTC ATC AGC AGC 2075  
 L Q I Q L D L Y P T K A G P R Q L Q V L I S S 690  
 AAC GAG GTC AAG GAG ATC AAA GGC TAC AAG GAC ATA TTC GTC ACT GTG GCT GGG GCT CCC TGA GAC CCG 2144  
 N E V K E I K G Y K D I F V T V A G A P \* 710  
 CCC TCC AGC TGC CCT CCC TGG CAC CCC TGC CTC CTT TCT ACT CCT GGC TAT GTC TTC 2213  
 GCT CCA CCT CTG TCC TCT CTC TAG CCT GGC TGG GAA TGA ATG AAG CTC TGT TAG AAA CAC CGT GTG CTT 2282  
 TGG GAA GAG ACA ATA AAG ATG TCT TTA TTT 2312

Fig. 6B (contd.)

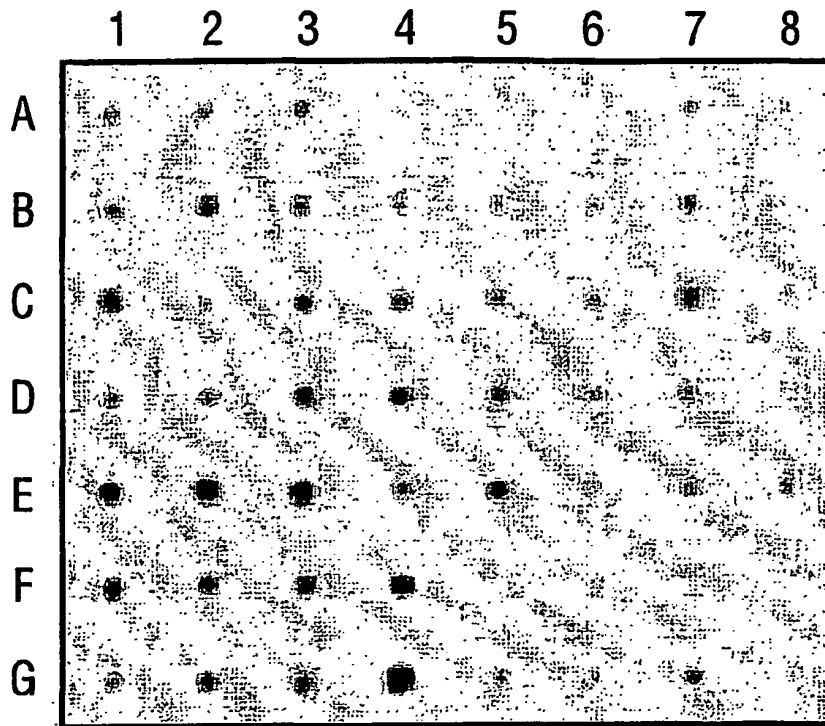
**Fig. 7A**



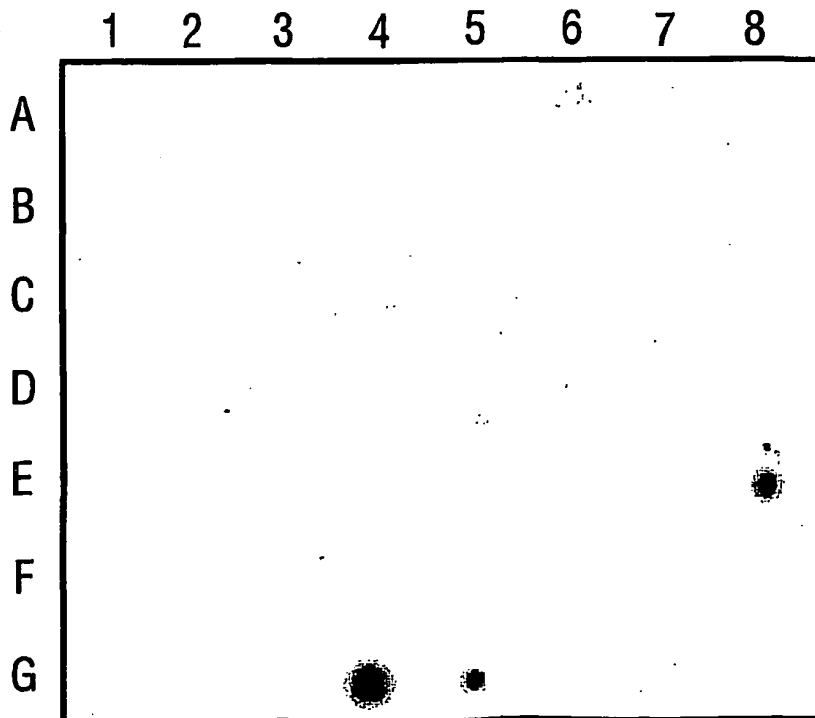
**Fig. 7B**



**Fig. 7C**



**Fig. 7D**



**Fig. 7E**

brain	amygdala	caudate nucleus	cerebellum	cerebral cortex	frontal lobe	hippocampus	medulla oblongata
occipital lobe	putamen	substantia nigra	temporal lobe	thalamus	Sub-thalamic nucleus	spinal cord	
heart	aorta	skeletal muscle	colon	bladder	uterus	prostate	stomach
testis	ovary	pancreas	pituitary gland	adrenal gland	thyroid gland	salivary gland	mammary gland
kidney	liver	small intestine	spleen	thymus	peripheral leukocyte	lymph node	bone marrow
appendix	lung	trachea	placenta				
fetal brain	fetal heart	fetal kidney	fetal liver	fetal spleen	fetal thymus	fetal lung	



503 d3  
 487  
 471  
 494  
 479  
 469  
 515  
 573  
 469  
 617 d4  
 606  
 585  
 608  
 592  
 584  
 628  
 686  
 584  
 719 d5  
 710  
 690  
 709  
 692  
 686  
 731  
 814  
 683

TGX SDRERDITENTKYBE GSIQERQVFLKALQKAKSPHSGORGAELQPSRPTLSQDSPRSLLTFLSARPSDV  
 TGZ SDORQSITSSVYKYPE GSPERRAVFMAKSRKMLGQRASLFFLDLLESGLRD-----  
 B4.2 SDRCEDIQNYKYPE GSIQBEVLRVVEKEMHEEDNGORRPPSLBTA-----  
 TGY SDSRVDITDLYKYPE GSRKEQVYSKAVNRLFOYASGARRIWRAGGRCLHDDDLLEPAFK-----  
 TGB SHAMQVYDKYKYPE GSDQERQVFOKALGKLPKPHPPAASFSGKLSIEBQF-----  
 TGC RDRERDITZHTYKYPE GSSEREAPFRANHLNKLAEKE-----  
 FXIII GDCMNDITDVKYQF GDBERLALFALAMYGAKKPLNTEGVKSR-----  
 TGA SHAREDITVLYKYPE GSDAREKAVEFAAAGSKPZYVYANRGA-----  
 TGP QDRERDITVEYKYPE GSSEREQVMDHAFLLLSBEBHRRPVKE-----  
 TGX VOVLKFKALDPPNMGDICTVYLALNHSSQFD-----LKVMSAQOLLEDGSLSPY-WODTAPITLSENE AKTYECK--ISYSQISOZLSTOKLIRISALGEBKSSPKLIVNKKIITLS  
 TGZ QPAQLQLHARIPENGDQLLARIQRPDSTHPRPIGAVVAPFCQALHGGGTQPF-WRHTVNNLDFGK EQORPLI--LPZSNVYRNKLTDBKLIWVSGIAFVSETOGRSHLVAKDICTLE  
 B4.2 SPYVLLKAPSSLELGDQISVTVVHNSQEK-----IQALICVQAVHNGVLAACL-WRKKLHLLTSANU EXLITIG--LFFSNFERNPENTFLQJANATHSSSNLSCFAQEDIALC  
 TGY PSTAGKAVLEPPMLQBDLALALCLANLTSBAQF-----VWVMSGATLITRRPVABEILHESHAVNLCQOP ERRTPIF--ISYKYEKEDIEDKILLANCLVYK-GRKLLVEKDYTLE  
 TGB PSISGKLVAGGLAVGRVNLVLLKMLSDIKY-----VYVWNTAWTIIHNGLVHBY-WKDSATNSLDQBE EASHPIK--ISYAQIERYLIGSDNHWIRIYAVCKVPD-BSEVVVERDIILD  
 TGC TGNANRIRVQSMNHGSDPZVFAHTHTNABEYV-----CRLALICARTVSVHNGILGEPGOTKYLNLNLTREFS RESVPLC--LIVYKRYDCLTESMLINVRALLVPERVIVHSYLLAERDLYLE  
 FXIII SHVDNDQFV-ENAVLGRDFKLSITFRNNSHRYT-----ITAVLSANITPTGTVPKAEF-KRETEDVYLEPLS PUKBAVL--IQAGEYMGQLLEQASLHPFVYARINETHEDVLARQKSTVLF  
 TGA EDVAMQVEA-QDAVHGDMLAVSVMLINHSSRRF-----VKLHLVLSVFTGVSQIF-KETKKEVELAPGA SDRVTWF--VAKYERPELVDOGAMLLNWSGHWKESGVVLARQHTPRER  
 TGP NPLHMSVQS-DDVLLGHSVNFVYILKRTKALQV-----VNILGSEFELQVYCKKMAKL-CDLNKTSGIQSQ VSBVTILDSEYTHNSLAILDDPEPVIEGPIIAEIVESKEIMASSEVYTFSEQ  
 TGX YPBITIN VLGAAVVRQPLSIOVIFSHPLSQVBCVLFVEGGGLPKQKVP LGV-LKFORQASIILETVPFRSGORQIOANMNSHFKDIKGYRNVYVDFAL  
 TGZ PPHLSIE YSERAEVGHARVHTNPLNVALSCTNVLGEGSLNGQIAND LGS-JVAGHTIQLOLQYTKAGPRQLQVLISSNEVKEIKGYKDIPTVAGAP  
 B4.2 RPHLAJK MPEKAGQYPLTASVSIQNSLDAPMBCVLSLGRQLHRRRSYR PRS-VWPEMFKAKPOPTPHVGLQRLTVEVDCNHPQNLTYHRSVTVVAPBLSA  
 TGY -DPTIK VLGPAVGVAVTVVAVNPLIBRVKQCALNVEGSLQBEQLSID VPT-LRPOERASVQFDITPSKSGPQLQVDIVSPHPPDIKGFVIHVYATAK  
 TGB NPTLSTB VLNBARVYKPVNVQMLZSNLDEPVEDCVLWVEGSLILGNLKD VPT-LGPEERGRVRFDIIPSRSCYKQLLADPFSCHKFPPIAKAMLSIDVAB  
 TGC NPEIKIR IIGEPKOKRXLVAEVSLOQLPVALGCTFTVEGAGLTFBEQKVEI PDVFBAGEVYKVRKDI VPLHMGLHKLWVNPESDKLKA VEGFRNVIIGPA  
 FXIII IPEITIK VSGTVVQSDNVTVQETHLNETLNNVNHLDQFVYR-PPKMKFR E-IRENFTVQWESVCHPFWSGHRELIASMSDSEHHTVGELOVQIQRRPSM  
 TGA TPDLSLT VLGAAVVRQRCVQVFNPLPVTLNNVFRLEGSLQ-RPKILAVG D-IGGNVTVLRSVYVAPFQQLIASLDSQLSQVGVQVQDVAAPAGDGGFFSDAGSDSHLGTTPMASRGGGA  
 TGP YPEPSIE LNTORTIGOLLVNCNLCFHTLAIPLTDVXPSLESIGIS-SLAFSDHG T-VQCEIIGSQIKCTPIKTKHKEIVKLSQKVEIQNOKIYITK

Fig. 8A (contd.)

# Fig. 8B

FXIII 38 d1  
T6x 100  
SEFSPTAFCGRRAVPPHNSNAARDLPTVZLQGVPRG  
MDGPRSDVGRWGNPLQPPPTFSPPEFGRSRRGGSSFWARCOCGCCRNAAADDWDPPEBDBRGRGSSGTEPRGSRSDSRFVPSRSGSVNARAGDO

T6x 139 d2  
AQ GERVALTDLQSSR---NIVSHEHEEITVDHLIYERGOAENLTYEYH---BSFQPOLDNIIPVVEF QPLDLDLQCF---RAVPSLARHHSPPHIAWLETNGASTKVSICAPPTAAVGRYLKIHIDSPQGSVTA--YQJGEFILLF  
T6x 141  
DQ VAYTRLESVDLQSSR---NWKHEHTQEMGVKRLIYERGOPIYHJBSF---EPFQSOQNDHITVYAEF QENKHELLQF---RATFPIRFPQPNVHSADFTIDSHSIOVSLPTANAVIGHYLXLEISDQGCHSVT---YPLGCFILLF  
B4.2 137  
GQ ALGITSQDFQAAH---NNEBEHTKALSSRILYERGOPTIYERAPVAFAPALKKVALTAQF QEQPSEKINF---QAPFPISSIGDRNHSAVVEBERDAQSHVTSVTPADAVIGHSYLLQVGBRKO-----LILQDZTILF  
T6x 138  
TQ GTRVAVDQGRSR---RGAHHTQEQYPCPEAVYERGOFSSTLELS---RAL-DCREILIFVEF QERASELHF---RAVQYSELEREGGHTARAZAQMEKFTVLSLASPSSAVIGRYLISIRLSSRKHSHN---RRLGSEVYILE  
T6x 134  
A ALGYSOSINQKAL---NRQAHHTDQFSSQELIYERGOVFILIMN---IGL-GSNERLBFIDYI QEYFSEBAMT---KAVPELQNG-SSGGHSAVYQJASHNGHTITIS1662A6APIGRYTMALQIFSDGGGISS---VKLGFILLF  
T6x 138  
AE ELVLERCDLELEF---NCRDHTADICREKLYERGOFPWLTJHEFG---RHYEASVDSIJTESVVI QEAFSQAQF---KARPELDAVBERDVTATVVDQDQCTL61QJLTPANAPICGLYLSLEASTGYQSSS---FVLGHFILLF  
FXIII 184  
VHLQ EFNVTSVHLEKREHDFKVDHHDYENKLIYERGOYVQIDFS---EYDPRRDLFRVYVVI QEYQENKCF---YIPVPIVSELQSKHGAIKYHREBRSVRLSIOSSPFCIVGKPRMYAVVMTFVGLRFSRNPEPTDYILF  
T6x 244  
TIR EGMVYVGVLDLSSSDQRRBEBHHDYEVDBLIYERGOPEHMLJLS---EY-SSDRITLLELI GNNPEVCKCF---HVIYIPVKG-GSGGKAOVVKASQOHMLARVHTSPNAIIGKQFPTVTRTQSDAGBEFQLPPOPRMEIYILE  
T6x 137  
MDASK EIQVZHTIDFANQD---NAVSEHTWBFQTSPPFRGQVPEHRLVJN---QPL-QSXYHQJLRFEST QENESIAKHILVLDZRPFS-DHVNQOALQNESKRYAVVTSSEPHILGKYQIVVYTONHILK-----SEENILYILE

T6x 281 d3  
NPKCF EDAVYLDSEPOREYVMHDYGFYQGSKNWIRPCPNYGO FEDKIIDICLKLKDLKLFQDPATDCALZGSEPVYVAVCAM INSNDDNGVZLNGHNSNTDGMNPAHTGSAVLKQHNATGCPVRYGQCHVFAAVNCT V  
T6x 289  
NPESP EDDVYLDSEYLLQELINRDYGFYKGERFITSNPNYGO FEBDIDICPEILNKLYLHNPAKDCSQENDVVIYCVVSAH INSNDDNGVZLNGHNGEDYSKGVSLERKGSVALLQOHSARGQOPVYGQCHVFAAVNCT V  
B4.2 277  
NPNR EDAVYLDSEPOREYVMHDYGFYKGERFITSNPNYGO FEBDIDICPEILNKLYLHNPAKDCSQENDVVIYCVVSAH INSNDDNGVZLNGHNGEDYSKGVSLERKGSVALLQOHSARGQOPVYGQCHVFAAVNCT V  
T6x 286  
NPKCA EDDVYLDSEYLLQELINRDYGFYKGERFITSNPNYGO FEBDIDICPEILNKLYLHNPAKDCSQENDVVIYCVVSAH INSNDDNGVZLNGHNGEDYSKGVSLERKGSVALLQOHSARGQOPVYGQCHVFAAVNCT V  
T6x 282  
NPLN VDSVFNQHAREBEIYVQEDAGIYVQSTNRIGHIWHNCFQ FEBDIDICPEILNKLYLHNPAKDCSQENDVVIYCVVSAH INSNDDNGVZLNGHNGEDYSKGVSLERKGSVALLQOHSARGQOPVYGQCHVFAAVNCT A  
T6x 286  
NAMCP ADAVYLDSEYLLQELINRDYGFYKGERFITSNPNYGO FEBDIDICPEILNKLYLHNPAKDCSQENDVVIYCVVSAH INSNDDNGVZLNGHNGEDYSKGVSLERKGSVALLQOHSARGQOPVYGQCHVFAAVNCT V  
FXIII 323  
NPKC DDAVYLDSEYLLQELINRDYGFYKGERFITSNPNYGO FEBDIDICPEILNKLYLHNPAKDCSQENDVVIYCVVSAH INSNDDNGVZLNGHNGEDYSKGVSLERKGSVALLQOHSARGQOPVYGQCHVFAAVNCT F  
T6x 383  
NPKCP EDIYVVDHEDNQEYVYVHESGRIYVQSTNRIGHIWHNCFQ FEBDIDICPEILNKLYLHNPAKDCSQENDVVIYCVVSAH INSNDDNGVZLNGHNGEDYSKGVSLERKGSVALLQOHSARGQOPVYGQCHVFAAVNCT V  
T6x 276  
NPKC EDKVPMPDEBERKEYLINDTGCCHVGAARSIKCPNFCQ FEBNVDCCISLITBS-----SIAKPTDRDPLVAVCRANCAM MSPRGGQVLIQHNWTDGDEGTPAPYKWTGSAPILQOYUHTK-OAVCFQCHVFAAVNCT V

T6x 432 d3  
MRCUGIPRVTNFPDGHDTDGNLIDYVDNCGRILGHKKKNDIY NPHVNECHMAHNDLPPGCGHQVLDATFQENEN GYVCCGPFASVPAIHEGVDLNDYDPEVFSK-VNADCHSHIYO---GGKQKLIH---QDTSVGNFISTKSIQ  
T6x 435  
MRCUGIPRVTNFPDGHDTDGNLIDYVDNCGRILGHKKKNDIY NPHVNECHMAHNDLPPGCGHQVLDATFQENEN GYVCCGPFASVPAIHEGVDLNDYDPEVFSK-VNADCHSHIYO---GGKQKLIH---QDTSVGNFISTKSIQ

B4.2 433  
LRCUGIPRVTNFPDGHDTDGNLIDYVDNCGRILGHKKKNDIY NPHVNECHMAHNDLPPGCGHQVLDATFQENEN GYVCCGPFASVPAIHEGVDLNDYDPEVFSK-VNADCHSHIYO---GGKQKLIH---QDTSVGNFISTKSIQ

T6x 432  
LRCUGIATRVVSNENSAHDTQDNLSDVRYVDSFGRTLEDLTDENW NFEVNECHMAHNDLPPGCGHQVLDATFQENEN GYVCCGPFASVPAIHEGVDLNDYDPEVFSK-VNADCHSHIYO---GGKQKLIH---QDTSVGNFISTKSIQ

T6x 428  
LRSIGLPSVITNENSAHDTQDNLSDVRYVDSFGRTLEDLTDENW NFEVNECHMAHNDLPPGCGHQVLDATFQENEN GYVCCGPFASVPAIHEGVDLNDYDPEVFSK-VNADCHSHIYO---GGKQKLIH---QDTSVGNFISTKSIQ

T6x 431  
LRCUGIPRVTNFPDGHDTDGNLIDYVDNCGRILGHKKKNDIY NPHVNECHMAHNDLPPGCGHQVLDATFQENEN GYVCCGPFASVPAIHEGVDLNDYDPEVFSK-VNADCHSHIYO---GGKQKLIH---QDTSVGNFISTKSIQ

FXIII 470  
LRCUGIPARIWNYFSAHDNDALQNDIYFASBEGGNVNSKLIYDSVH NFEVNECHMAHNDLPPGCGHQVLDATFQENEN GYVCCGPFASVPAIHEGVDLNDYDPEVFSK-VNADCHSHIYO---GGKQKLIH---QDTSVGNFISTKSIQ

T6x 530  
LRCUGIATRVVSNENSAHDTQDNLSDVRYVDSFGRTLEDLTDENW NFEVNECHMAHNDLPPGCGHQVLDATFQENEN GYVCCGPFASVPAIHEGVDLNDYDPEVFSK-VNADCHSHIYO---GGKQKLIH---QDTSVGNFISTKSIQ

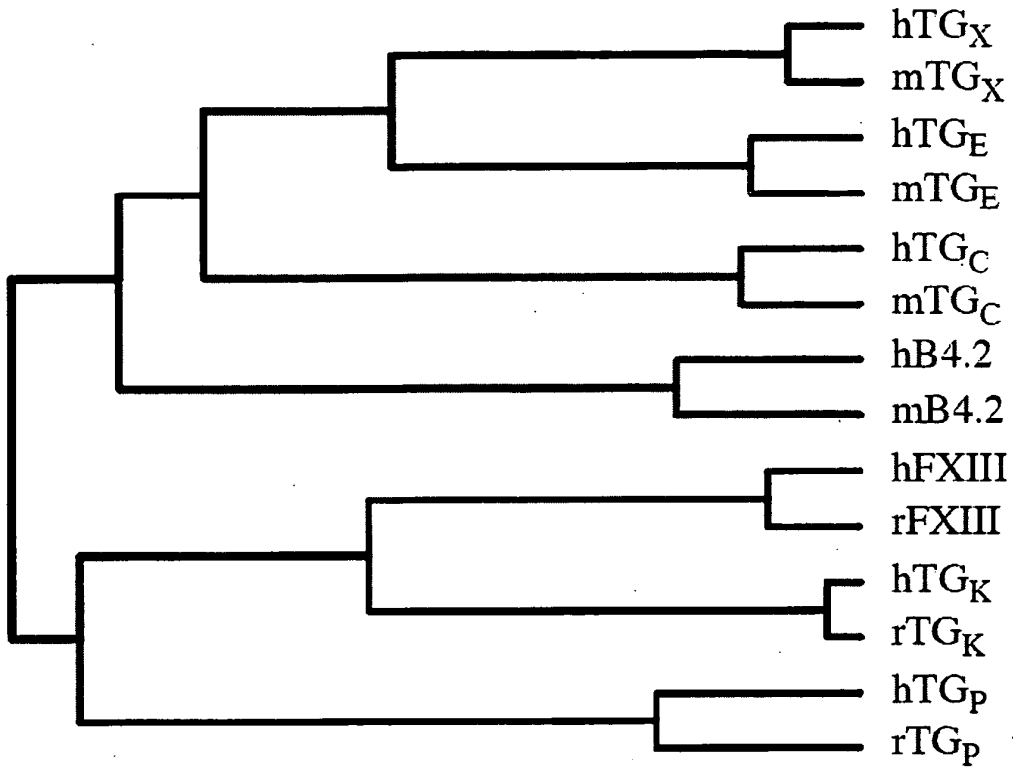
T6x 426  
LRLALGIPARESITGDSAHDTQDNLSDVRYVDSFGRTLEDLTDENW NFEVNECHMAHNDLPPGCGHQVLDATFQENEN GYVCCGPFASVPAIHEGVDLNDYDPEVFSK-VNADCHSHIYO---GGKQKLIH---QDTSVGNFISTKSIQ

TCY	SDRRDDITVYKYBE	GSIQERQVEFKALQXKAKRBFHGSQGAELQPSRPTSLISQDSPRSLETSPSRPDSV	503	d3
TCZ	SDQRQSIITSSYKYBE	OSPBERAVFVKASRKMLQPRASLFPDLLEGGLRD	487	
B4.2	SDRCBDITQNKYBE	GSIQERQVEFKALQXKAKRBFHGSQGAELQPSRPTSLISQDSPRSLETSPSRPDSV	471	
TCY	SDSRVDITDLYKYBE	GSRBERQVYKAVNKLFCVBEASGRVTHIRACGRCLHDDLLLEPAFK	494	
TCZ	SNARMVDYKYBE	GSDQERQVFKALGHLKKNVFPFAATSSMGLTBEQE	479	
TCG	RDERBDITHYKYBE	GSSEREAFTRAHNKLAKERBE	469	
FXIII	GDMKMDITVYKFOE	GQBERLALFALMYGAKKPLATFGVKNKR	515	
TCK	SNMREDITVYKHPK	GSDAEKAVETAHAHGSXPNVIANRGS	573	
TCP	QDRRDIPTVYKYBE	GSSEREQVMDHAPLLSSERERREPYKE	469	
TCY	VQNSLXFKLLDPPNMGQDICFVLLALNMSQPKD	-----LKNVLSAQSLHDGSPLEPF-WODTAFITLSPKE AKTYPK--ISYSQYSLSTDKLIRISALGEKSSPERKILVYKXITLIS	617	d4
TCZ	QPAQQLHARIPENQDLOLLARIQVDPDSHPRGPIGVVRFCAQALLHGCTQXP	WRHTVNMILDPK ETOWPUL--LPSYVNRKLTDEKLTIRVSGIARVBEETGRSMVLYKXICLE	606	
B4.2	SPLYLLKAPSEPLRGDAQISVTVVHSEQEKA	-----VOLAIQVAVHNGVLAAL-WRKKLHLTISANL EXIITIG--LFFSNFERMPENTFLRLTAMATHSESLSQCPAQEDIAIC	585	
TCY	PSIAGRFVLEPPHGHDLRALCLANLTSRAQR	-----YVNVLSGATILITRRPVARI-JHESHAVRIGQOE EKRIPII--ISYSKYKEDLTHEDKLILAAANCLVTK-GEKLLVEKDIPLB	608	
TCZ	PSISGKLVAKGLAVGKREVNVLVLLANLSRDYKY	-----YVNVHAWYIINGLVHEV-WDSATHSLDEBE EABHPK--ISYAOYBYIYLSQDNHILIRIYVCKVDP-ESVYVVERDIID	592	
TCG	YGMAMRIRVQSMNNGSDFDVFABITNNTAEYV	-----CKLILCARVYSINGILGPECFKYLMLLLEPPS EKSVPIC--ILYBYKROCLTESMLIKVVALVBEVVIHNSYLLAERDIYLE	584	
FXIII	SNVDNDPEV-ENAVLGKDFKLSITPRNNSHRYT	-----ITAYLSANITFTGVPKAEF-KKETFQVTLIPLS FKXNAVL--IQAGEYMQQLBQASLHFFVVTARINETRDVLAKQKSTVLT	628	
TCK	EDVAKQVEA-QDAVHGQDLAVSVMLINHSSSRK	-----VKLHLVLSVFTYGVSGRIP-KETKKEVEIAEQA SDRVTNP--VAYKRYRPHLYDQGAMLLNVSCHVKSQVLAQKQBTFRER	686	
TCP	NPLEMSVOS-DDVILGNSVHFTVILKRYTAALQN	-----VNILGSPFLOLITCKAKMARL-CDLNKTSQIQG VSEVTLDSKYINSLAILDDBEVINGPILARIYVESKEINASEVFTSFQ	584	
TCY	YPSITIN VLGAAVYHQPISQVIFSNPLSEQVDCVITVEGSGLEPKQAVF	LGV-LFEPQOASILLESVTFPSQEQIQANRBSNKFKDILGYRNYVYDPAL	719	d5
TCZ	YPHLSIE VSERAEYKALRVHTLNTLMAVLSCTWVLEGGSLINGQIAND	LGT-LVAGHTLQIQDLAVYKAGPQLOVYLSSENEVKEIKGYRDIPTVYVAGAP	710	
B4.2	RPHLAIN NPERAEQYQPLTASVSLONSIDAPMEDCVISILGRGLXRRRSYK	FRG-VWFEYTKCAKEQFTFHVGLQRLAVYVDCNMFQNLTHYKSVTVYVAPLSA	690	
TCY	-DPTIK VLGPAVYAVTVVAVVWVPLIBRAXDCALWVGGGLLQEQLEID	VPT-LBEPQERASVQDITPSSKSPQLOVLDVSPHPPDIKGFVIVHATAK	709	
TCZ	HETLLE VLINEARVREPVVQMLPSNPLDEPVRDCLAVWVGGGLLGNLXID	VPT-LGPKERSRVRFDILLESSTYKOLLADYFCHKRPFAIKAMLSIDYAE	692	
TCG	NPEKIR ILGEPKQREIVAEVSLQNEPLVAVLEGCTFTVEGLTPEBQKTVBI	PDPVBAQBEVKVVRMDVPLHMGHLKLVVWFESDKLAKAVGFRVVIIGPA	686	
FXIII	YPEIILK YNGQVYSDMTYVQVFNPLKEVIRNVVHLDGQYV-RPKKMPK	B-IRPNSTVQWESVCHPWSGHRLLASMSDSLRRVHVGELDVOIQARPSN	731	
TCK	YDLSLT LIGAAVQGBQCYVIFANPLPVPVFNVFRLESSLQ-RPKLNVG	D-IGNHVTVLRQSFVVRPQRQLIASLDSQPSQVHVIQVADVAPAGDGGFPDAGGSHLGETIPNARSQGA	814	
TCP	YPEPSIE LPNTRIGOLVNCVCFKNTLALPLADVWKSLSLQIS-SLQTSDBQ	Y-VQCEHIQSQIKCIEKTEGPKAVIYKLSKQVKNNAOKVLLTK	683	

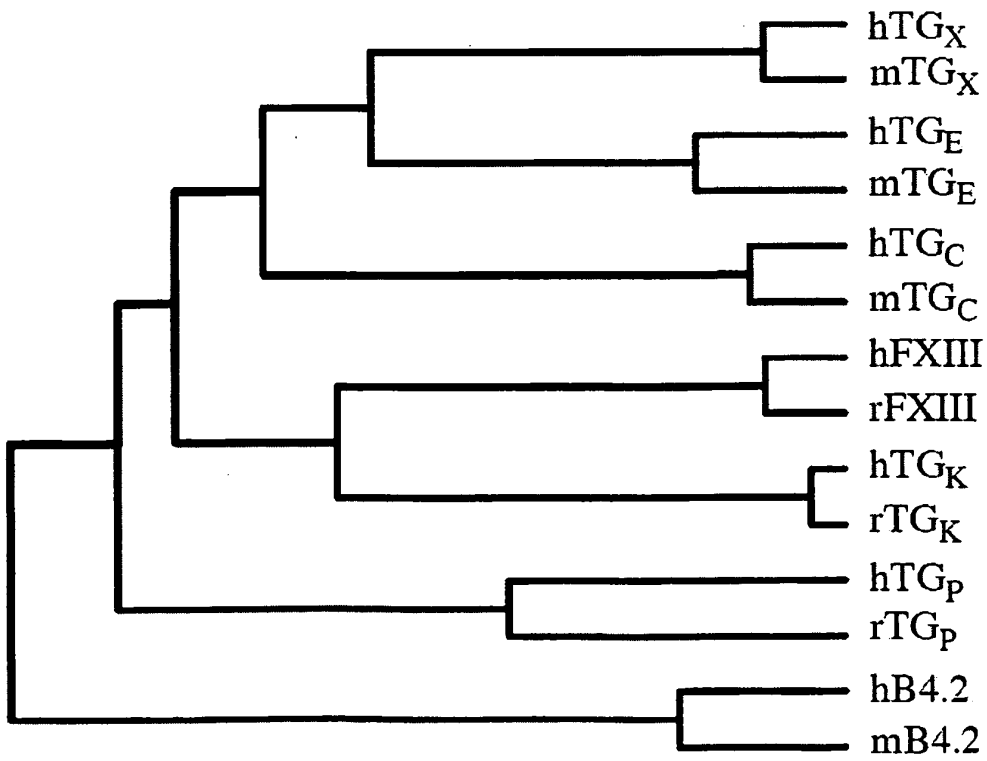
Fig. 8B (contd.)



**Fig. 9A(3)**



**Fig. 9A(4)**



**Fig. 9A(5)**

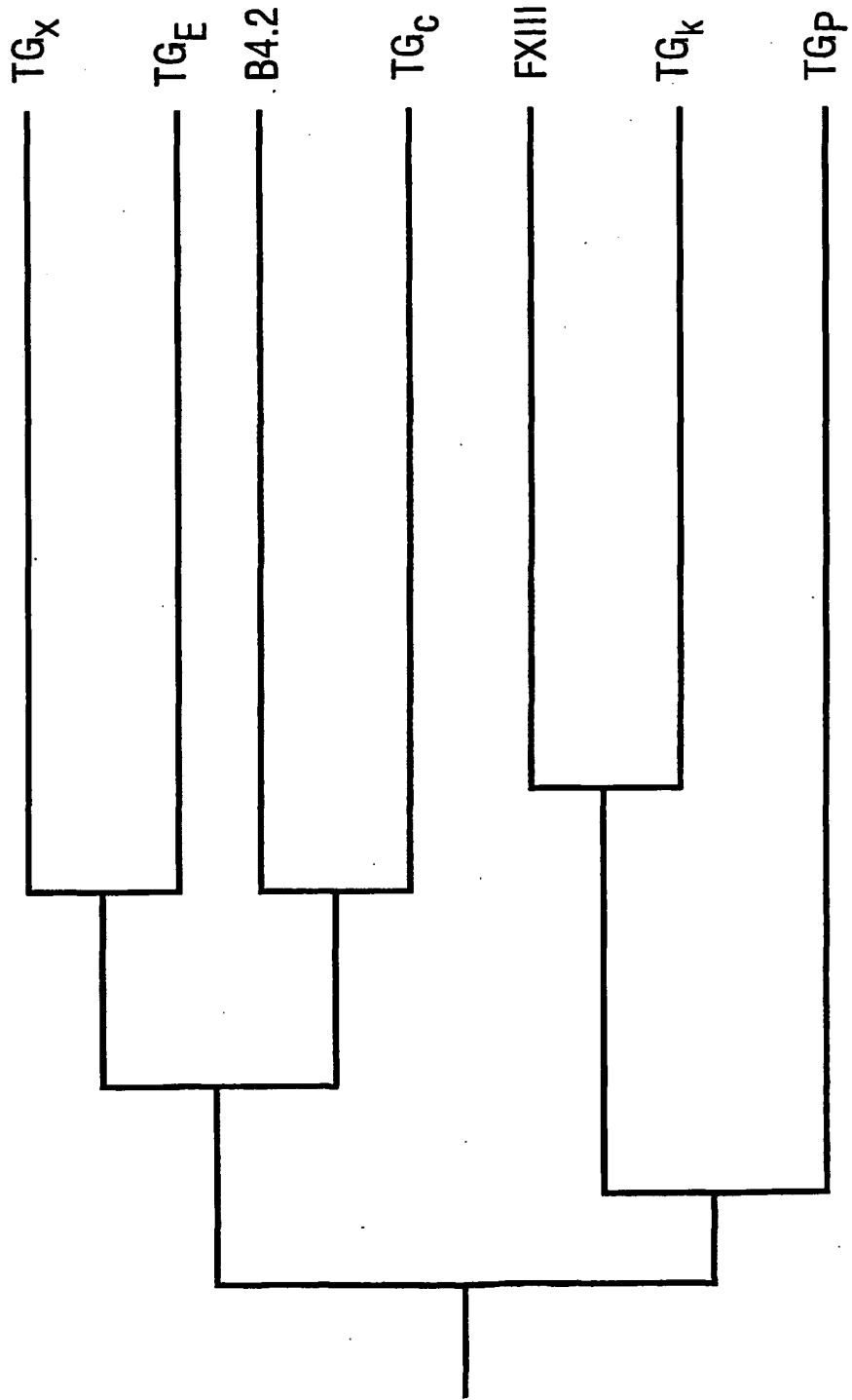


Fig. 9B

	TGM5 (TGx)	TGM7 (TGz)	EPB42 (band 4.2 protein)	TGM2 (TGc)	TGM6 (TGy)	TGM3 (TGe)	TGM4 (TGp)	F13A1 (factor XIII a-subunit)	TGMI (TGk)
<u>Chromosomal localization</u>	15q15	15q15	15q15 (a) 2, 67.5cM	20q11-12(b) 2, 89cM	20q11(c)	20q11 (c)	3p21-22 (d)	6p24-25 (e)	14q11.2 (f)
<u>Gene size</u>	~35kb	~22kb	~20kb (h) ~22kb (m)	~37kb (b) ~34kb (o)		~43kb (i)	~35kb (n)	~160kb (k)	~14kb (l)
<u>Number of exons</u>	13	13	13 13	13 13	13	13	13	15	15
human									
mouse									

- (a) Sung et al., 1992. Blood 79, 2763-2770; Najfeld et al., 1992. Am J Hum Genet 50, 71-75  
 (b) Gentile et al., 1994. Genomics 20, 295-297; Kim et al., 1994. J Invest Dermatol 103, 137-142; GenBank™/EMBL Data Bank with accession number AL031651  
 (c) Kim et al., 1994. J Invest Dermatol 103, 137-142; GenBank™/EMBL Data Bank with accession number AL031678  
 (d) Dubbink et al., 1996. Biochem J 315, 901-908; Gentile et al., 1995. Genomics 27, 219-220  
 (e) Board et al., 1988. Cytogenet Cell Genet 48, 25-27

- (f) Kim et al., 1992. *J Biol Chem* 267, 7710-7717
- (g) White et al., 1992. *Nat Genet* 2, 80-83
- (h) Korsgren and Cohen, 1991. *Proc Natl Acad Sci USA* 88, 4840-4844
- (i) Kim et al., 1994. *J Invest Dermatol* 103, 137-142
- (k) Ichinose and Davie, 1988. *Proc Natl Acad Sci USA* 85, 5829-5833
- (l) Phillips et al., 1992. *J Biol Chem* 267, 2282-2286; Kim et al., 1992. *J Biol Chem* 267, 7710-7717; Polakowska et al., 1992. *Proc Natl Acad Sci USA* 89, 4476-4480
- (m) Korsgren and Cohen, 1994. *Genomics* 21, 406-425
- (n) Dubbink et al., 1998. *Genomics* 51, 434-444
- (o) Nanda et al., 1999. *Arch Biochem Biophys* 366, 151-156

**Fig. 9C**

Gene Product	Overall	Protein Domains			
		$\beta$ -sandwich	catalytic core	$\beta$ -barrel 1	$\beta$ -barrel 2
TG <sub>x</sub>	100.0				
TG <sub>z</sub>	48.1	49.6	63.1	31.6	36.3
B4.2	31.6	33.1	41.1	16.7	23.5
TG <sub>y</sub>	45.2	41.0	56.6	34.2	44.1
TG <sub>B</sub>	42.3	29.5	56.5	31.6	38.2
TG <sub>C</sub>	40.1	35.3	55.6	20.2	31.4
FXIIIa	32.7	23.7	46.8	18.4	25.5
TG <sub>k</sub>	34.9	25.9	49.6	18.4	29.4
TG <sub>p</sub>	31.0	23.7	47.1	11.4	20.6



**Fig. 10A(contd)**

824	833	842	851	860	869
AAG TAC GGC CAG TGC TGG GTC TTC GCC GGA GTC CTG TGC ACA GTC CTC AGG TGC					
K Y G Q C W V F A G V L C T V L R C					
878	887	896	905	914	923
TTG GGG ATA GCC ACA CGG GTC GTG TCC AAC TTC AAC TCA GCC CAC GAC ACA GAC					
L G I A T R V V S N F N S A H D T D					
932	941	950	959	968	977
CAG AAC CTG AGT GTG GAC AAA TAC GTG GAC TCC TTC GGG CGG ACC CTG GAG GAC					
Q N L S V D K Y V D S F G R T L E D					
986	995	1004	1013	1022	1031
CTG ACA GAA GAC AGC ATG TGG AAT TTC CAT GTC TGG AAT GAG AGC TGG TTT GCC					
L T E D S M W N F H V W N E S W F A					
1040	1049	1058	1067	1076	1085
CGG CAG GAC CTA GGC CCC TCT TAC AAT GGC TGG CAG GTT CTG GAT GCC ACC CCC					
R Q D L G P S Y N G W Q V L D A T P					
1094	1103	1112	1121	1130	1139
CAG GAG GAG AGT GAA GGT GTG TTC CGG TGC GGC CCA GCC TCA GTC ACC GCC ATC					
Q E E S E G V F R C G P A S V T A I					
1148	1157	1166	1175	1184	1193
CGC GAG GGT GAT GTG CAC CTG GCT CAC GAT GGC CCC TTC GTG TTT GCG GAG GTC					
R E G D V H L A H D G P F V F A E V					
1202	1211	1220	1229	1238	1247
AAC GCC GAC TAC ATC ACC TGG CTG TGG CAC GAG GAT GAG AGC CGG GAG CGT GTA					
N A D Y I T W L W H E D E S R E R V					
1256	1265	1274	1283	1292	1301
TAC TCA AAC ACG AAG AAG ATT GGG AGA TGC ATC AGC ACC AAG GCG GTG GGC AGT					
Y S N T K K I G R C I S T K A V G S					
1310	1319	1328	1337	1346	1355
GAC TCC CGC GTG GAC ATC ACT GAC CTC TAC AAG TAT CCG GAA GGG TCC CGG AAA					
D S R V D I T D L Y K Y P E G S R K					
1364	1373	1382	1391	1400	1409
GAG AGG CAG GTG TAC AGC AAG GCG GTG AAC AGG CTG TTC GGC GTG GAA GCC TCT					
E R Q V Y S K A V N R L F G V E A S					
1418	1427	1436	1445	1454	1463
GGA AGG AGA ATC TGG ATC CGC AGG GCT GGG GGT CGC TGT CTC TGG CGT GAC GAC					
G R R I W I R R A G G R C L W R D D					
1472	1481	1490	1499	1508	1517
CTC CTG GAG CCT GCC ACC AAG CCC AGC ATC GCT GGC AAG TTC AAG GTG CTA GAG					
L L E P A T K P S I A G K F K V L E					
1526	1535	1544	1553	1562	1571
CCT CCC ATG CTG GGC CAC GAC CTG AGA CTG GCC CTG TGC TTG GCC AAC CTC ACC					
P P M L G H D L R L A L C L A N L T					

1580            1589            1598            1607            1616            1625  
 TCC CGG GCC CAG CGG GTG AGG GTC AAC CTG AGC GGT GCC ACC ATC CTC TAT ACC  
 S R A Q R V R V N L S G A T I L Y T

1634            1643            1652            1661            1670            1679  
 CGC AAG CCA GTG GCA GAG ATC CTG CAT GAA TCC CAC GCC GTG AGG CTG GGG CCG  
 R K P V A E I L H E S H A V R L G P

1688            1697            1706            1715            1724            1733  
 CAA GAA GAG AAG AGA ATC CCA ATT ACA ATA TCT TAC TCT AAG TAT AAA GAA GAC  
 Q E E K R I P I T I S Y S K Y K E D

1742            1751            1760            1769            1778            1787  
 CTG ACA GAG GAC AAG AAG ATC CTG TTG GCT GCC ATG TGC CTT GTC ACC AAA GGA  
 L T E D K K I L L A A M C L V T K G

1796            1805            1814            1823            1832            1841  
 GAG AAG CTT CTG GTG GAG AAG GAC ATT ACT CTA GAG GAC TTC ATC ACC ATC AAG  
 E K L L V E K D I T L E D F I T I K

1850            1859            1868            1877            1886            1895  
 GTT CTG GGC CCA GCC ATG GTG GGA GTG GCA GTT ACA GTG GAA GTG ACA GTA GTC  
 V L G P A M V G V A V T V E V T V V

1904            1913            1922            1931            1940            1949  
 AAC CCC CTC ATA GAG AGA GTG AAG GAC TGT GCG CTG ATG GTG GAG GGC AGC GGC  
 N P L I E R V K D C A L M V E G S G

1958            1967            1976            1985            1994            2003  
 CTT CTC CAG GAA CAG CTC AGC ATC GAC GTG CCT ACC CTG GAG CCT CAG GAG AGG  
 L L Q E Q L S I D V P T L E P Q E R

2012            2021            2030            2039            2048            2057  
 GCC TCA GTC CAG TTT GAC ATC ACC CCC TCC AAA AGT GGC CCA AGG CAG CTG CAG  
 A S V Q F D I T P S K S G P R Q L Q

2066            2075            2084            2093            2102            2111  
 GTG GAC CTT GTA AGC CCT CAC TTC CCG GAC ATC AAG GGC TTT GTG ATC GTC CAT  
 V D L V S P H F P D I K G F V I V H

2120            2129            2138            2147            2156            2165  
 GTG GCC ACT GCC AAG TGA TGG ATC ATG AGG GAC TGA GAG GGG TGG ATT TGG CCC  
 V A T A K \*

2174            2183            2192            2201            2210            2219  
 CTG TCC TCC TCC TGC CCA TTC TTT GTC TCT TCC ACA TGG GAG CCA GGA GGC CTC

2228            2237  
 AGT TAA TCC TGC CTC AAC CT

*Fig. 10A(contd)*



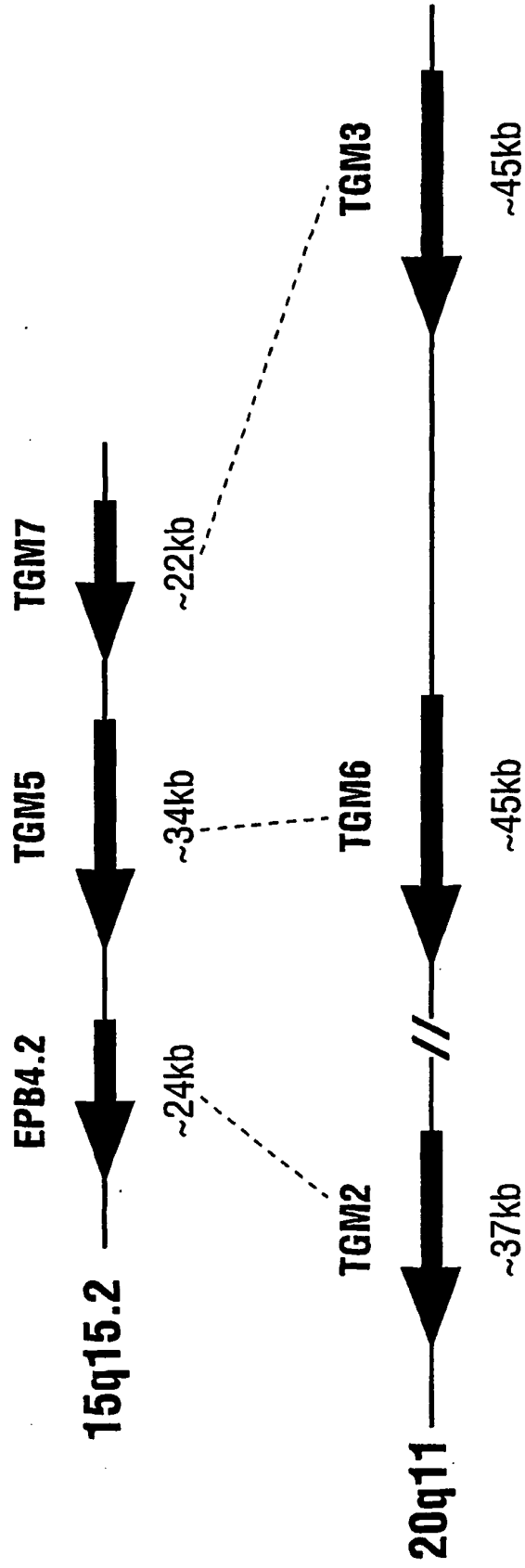
**Fig. 10B(contd)**

AAG	TAC	824	GGC	CAG	TGC	833	TGG	GTC	TTC	842	GCC	GGA	GTC	851	CTG	TGC	ACA	860	GTC	CTC	AGG	869	TGC	
K	Y	G	Q	C	W	V	F	A	G	V	L	C	T	V	L	R	C							C
TTG	GGG	878	ATA	GCC	ACA	887	CGG	GTC	GTG	896	TCC	AAC	TTC	905	AAC	TCA	GCC	914	CAC	GAC	ACA	923	GAC	
L	G	I	A	T	R	V	V	S	N	F	N	S	A	H	D	T	D							D
CAG	AAC	932	CTG	AGT	GTG	941	GAC	AAA	TAC	950	GTG	GAC	TCC	959	TTC	GGG	CGG	968	ACC	CTG	GAG	977	GAC	
Q	N	L	S	V	D	K	Y	V	D	S	F	G	R	T	L	E	D							D
CTG	ACA	986	GAA	GAC	AGC	995	ATG	TGG	AAT	1004	TTC	CAT	GTC	1013	TGG	AAT	GAG	1022	AGC	TGG	TTT	1031	GCC	
L	T	E	D	S	M	W	N	F	H	V	W	N	E	S	W	F	A							A
CGG	CAG	1040	GAC	CTA	GGC	1049	CCC	TCT	TAC	1058	AAT	GGC	TGG	1067	CAG	GTT	CTG	1076	GAT	GCC	ACC	1085	CCC	
R	Q	D	L	G	P	S	Y	N	G	W	Q	V	L	D	A	T	P							P
CAG	GAG	1094	GAG	AGT	GAA	1103	GGT	GTG	TTC	1112	CGG	TGC	GGC	1121	CCA	GCC	TCA	1130	GTC	ACC	GCC	1139	ATC	
Q	E	E	S	E	G	V	F	R	C	G	P	A	S	V	T	A	I							I
CGC	GAG	1148	GGT	GAT	GTG	1157	CAC	CTG	GCT	1166	CAC	GAT	GGC	1175	CCC	TTC	GTG	1184	TTT	GCG	GAG	1193	GTC	
R	E	G	D	V	H	L	A	H	D	G	P	F	V	F	A	E	V							V
AAC	GCC	1202	GAC	TAC	ATC	1211	ACC	TGG	CTG	1220	TGG	CAC	GAG	1229	GAT	GAG	AGC	1238	CGG	GAG	CGT	1247	GTA	
N	A	D	Y	I	T	W	L	W	H	E	D	E	S	R	E	R	V							V
TAC	TCA	1256	AAC	ACG	AAG	1265	AAG	ATT	GGG	1274	AGA	TGC	ATC	1283	AGC	ACC	AAG	1292	GCG	GTG	GGC	1301	AGT	
Y	S	N	T	K	K	I	G	R	C	I	S	T	K	A	V	G	S							S
GAC	TCC	1310	CGC	GTG	GAC	1319	ATC	ACT	GAC	1328	CTC	TAC	AAG	1337	TAT	CCG	GAA	1346	GGG	TCC	CGG	1355	AAA	
D	S	R	V	D	I	T	D	L	Y	K	Y	P	E	G	S	R	K							K
GAG	AGG	1364	CAG	GTG	TAC	1373	AGC	AAG	GCG	1382	GTG	AAC	AGG	1391	CTG	TTC	GGC	1400	GTG	GAA	GCC	1409	TCT	
E	R	Q	V	Y	S	K	A	V	N	R	L	F	G	V	E	A	S							S
GGA	AGG	1418	AGA	ATC	TGG	1427	ATC	CGC	AGG	1436	GCT	GGG	GGT	1445	CGC	TGT	CTC	1454	TGG	CGT	GAC	1463	GAC	
G	R	R	I	W	I	R	R	A	G	G	R	C	L	W	R	D	D							D
CTC	CTG	1472	GAG	CCT	GCC	1481	ACC	AAG	CCC	1490	AGC	ATC	GCT	1499	GGC	AAG	TTC	1508	AAG	GTG	CTA	1517	GAG	
L	L	E	P	A	T	K	P	S	I	A	G	K	F	K	V	L	E							E
CCT	CCC	1526	ATG	CTG	GGC	1535	CAC	GAC	CTG	1544	AGA	CTG	GCC	1553	CTG	TGC	TTG	1562	GCC	AAC	CTC	1571	ACC	
P	P	M	L	G	H	D	L	R	L	A	L	C	L	A	N	L	T							T

	1580		1589		1598		1607		1616		1625						
TCC	CGG	GCC	CAG	CGG	GTG	AGG	GTC	AAC	CTG	AGC	GGT	GCC	ACC	ATC	CTC	TAT	ACC
S	R	A	Q	R	V	R	V	N	L	S	G	A	T	I	L	Y	T
	1634		1643		1652		1661		1670		1679						
CGC	AAG	CCA	GTG	GCA	GAG	ATC	CTG	CAT	GAA	TCC	CAC	GCC	GTG	AGG	CTG	GGG	CCG
R	K	P	V	A	E	I	L	H	E	S	H	A	V	R	L	G	P
	1688		1697		1706		1715		1724		1733						
CAA	GAA	GAG	AAG	AGA	ATC	CCA	ATT	ACA	ATA	TCT	TAC	TCT	AAG	TAT	AAA	GAA	GAC
Q	E	E	K	R	I	P	I	T	I	S	Y	S	K	Y	K	E	D
	1742		1751		1760		1769		1778		1787						
CTG	ACA	GAG	GAC	AAG	AAG	ATC	CTG	TTG	GCT	GCC	ATG	TGC	CTT	GTC	ACC	AAA	GGA
L	T	E	D	K	K	I	L	L	A	A	M	C	L	V	T	K	G
	1796		1805		1814		1823		1832		1841						
GAG	AAG	CTT	CTG	GTG	GAG	AAG	GAC	ATT	ACT	CTA	GAG	GAC	TTC	ATC	ACC	ATC	AAG
E	K	L	L	V	E	K	D	I	T	L	E	D	F	I	T	I	K
	1850		1859		1868		1877		1886		1895						
CGT	GCC	TAC	CCT	GGA	GCC	TCA	GGA	GAG	GGC	CTC	AGT	CCA	GTT	<u>TGA</u>	CAT	CAC	CCC
R	A	Y	P	G	A	S	G	E	G	L	S	P	V	*			
	1904		1913		1922		1931		1940		1949						
CTC	CAA	AAG	TGG	CCC	AAG	GCA	GCT	GCA	GGT	GGA	CCT	TGT	AAG	CCC	TCA	CTT	CCC
	1958		1967		1976		1985		1994		2003						
GGA	CAT	CAA	GGG	CTT	TGT	GAT	CGT	CCA	TGT	GGC	CAC	TGC	CAA	GTG	ATG	GAT	CAT
	2012		2021		2030		2039		2048		2057						
GAG	GGA	CTG	AGA	GGG	GTG	GAT	TTG	GCC	CCT	GTC	CTC	CTC	CTG	CCC	ATT	CTT	TGT
	2066		2075		2084		2093		2102		2105						
CTC	TTC	CAC	ATG	GGA	GCC	AGG	AGG	CCT	CAG	TTA	ATC	CTG	CCT	CAA	CCT		

**Fig. 10B(contd)**

**Fig. 11**



## REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

## Patent documents cited in the description

- US 5726051 A [0006]

## Non-patent literature cited in the description

- **Aeschlimann, D ; Paulsson, M.** *Thromb. Haemostas*, 1994, vol. 71, 402-415 [0003]
- **Aeschlimann, D. et al.** *J. Biol Chem*, 1998, vol. 273, 3542-3460 [0003]
- **Aeschlimann, D. ; Thomazy V.** *Connective Tissue Res.*, 2000, vol. 41, 1-27 [0003]
- **Aeschlimann et al.** *J. Biol. Chem.*, 1998, vol. 273, 3452-3460 [0007] [0048] [0056]
- **Sung L. A. et al.** *Blood*, 1992, vol. 79, 2763-2770 [0009]
- **Najfeld V. et al.** *Am. J. Hum. Genet*, 1992, vol. 50, 71-75 [0009]
- **Chiannikulchai N. et al.** *Hum. Mol. Genet*, 1995, vol. 4, 717-725 [0009]
- **Aeschlimann et al.** *J. Biol Chem.*, 1998, vol. 273, 3452-3460 [0013] [0055]
- **Aeschlimann et al.** *J. Biol Chem*, 1998, vol. 273, 3453-3460 [0031]
- **Saitou ; Nei.** *Mol. Biol. Evol.*, 1987, vol. 4, 406-425 [0031]
- **Mount S.M.** *Nucleic Acids Res.*, 1982, vol. 10, 459-472 [0035]
- **Keller E. B ; Noon. W. A.** *Proc. Natl. Acad. Sci. USA*, 1984, vol. 81, 7417-7420 [0035]
- **Aeschlimann et al.** *J. Biol Chem*, 1998, vol. 273, 3452-3460 [0036]
- **White, R.A et al.** *Nat. Genet.*, 1992, vol. 2, 80-83 [0043]
- **Nanda, N. et al.** *Arch. Biochem. Biophys*, 1999, vol. 366, 151-156 [0043]
- **Gentile V. et al.** *Genomics*, 1994, vol. 20, 295-297 [0044]
- **Kim I. G. et al.** *J. Invest. Dermatol*, 1994, vol. 103, 137-142 [0044]
- **White R. A. et al.** *Nat. Genet*, 1992, vol. 2, 80-83 [0044]
- **Aeschlimann et al.** *J. Biol Chem*, 1998, vol. 273, 3245-3460 [0045]
- **Bebbington C. R. ; Hentschel, C. C. G.** DNA cloning. IRL Press, 1987, vol. III, 184-188 [0059]
- **Korsgren et al.** *Biochemistry*, 1990, vol. 87, 613-617 [0062]
- **McCarthy et al.** *Genome Res.*, 1997, vol. 7, 1153-1161 [0064]

