Amino Acids (2013) 44:215–225
DOI 10.1007/s00726-011-1194-6

ORIGINAL ARTICLE

Polymorphism of transglutaminase 2: unusually low frequency of genomic variants with deficient functions

Róbert Király · Endre Barta · László Fésüs

Received: 19 September 2011 / Accepted: 30 November 2011 / Published online: 13 December 2011

Abstract Transglutaminase 2 (TG2) is a multifunctional member of an enzyme family: it modifies glutamine residues by cross-linking proteins and incorporating primary amines into them, has protein disulphide isomerase and protein kinase activities, mediates trans-membrane signal transduction and interactions between cell surface proteins and the extracellular matrix. These unusual multiple roles encoded into one polypeptide chain suggest that genomic variations in the TGM2 gene should be limited. Indeed, the available information in databases shows that unlike in the case of most other transglutaminases there are no common single nucleotide polymorphisms in exons of human TGM2. We collected data on and produced some of the rare genetic variants of TGM2 by site directed mutagenesis and found that some were less stable than the most abundant (wild type) enzyme variant and the majority had deficient transamidating activity. Further studies are required to clarify the pathologic significance of these rare TGM2 alleles in the human population.

Keywords Human transglutaminase 2 · Wild type sequence · SNPs · No common variants · Rare variants · Deficient activities

Introduction

Transglutaminase research started more than five decades ago when Waelsch and co-workers observed a Ca^{2+}-dependent activity in liver extracts incorporating primary amines into proteins (Sarkar et al. 1957; Mycek et al. 1959). In the sixties, there were two important developments. First, the chemical nature of the cross-link formed in fibrin (Laki and Lóránd 1948) by blood coagulation factor XIIIa, a plasma transglutaminase (Loewy 1968), was revealed (Pisano et al. 1968; Matacić and Loewy 1968; Lorand et al. 1968). Second, the purification of tissue transglutaminase (later named transglutaminase 2; TG2) from guinea pig liver could be achieved (Folk and Cole 1966) in the NIH laboratory of Jack Folk, opening the way to the characterization of its enzymatic properties in details; the findings were summarized in the early seventies by Folk and Chung providing a basis of further developments in the field (Folk and Chung 1973).

In the pioneering TG2 experiments, and even later works with human transglutaminase 2 obtained from endothelial, red blood or other types of cells, it was assumed that the studied enzyme preparation represented a reproducibly obtainable homogenous protein population with the same amino acid sequence. After the completion of the human genome project and the availability of the HapMap (Thorisson et al. 2005) and other databases, it became evident that the vast majority of our proteins are polymorphic due mainly—but not exclusively—to single nucleotide polymorphism (SNP). The data presented by the 1000 Genomes Project Consortium (2010) have established the normal de novo mutation rate to be $\sim 1.1 \times 10^{-8}$ bp for a human haploid genome. Therefore, differences amounting to several million SNPs exist between the genome of any two individuals, of which tens of thousands are in protein coding exons. As a consequence, proteins encoded by the two parental chromosomes may differ significantly due to their individual polymorphism. The inherited rare genetic variations (most often SNPs) in the Mendelian disorder genes

R. Király · E. Barta · L. Fésüs (✉)
Department of Biochemistry and Molecular Biology, Apoptosis and Genomics Research Group of Hungarian Academy of Sciences, Medical and Health Science Center, University of Debrecen, Egyetem tér 1, Debrecen 4012, Hungary
e-mail: fesus@dote.hu

result in defective proteins leading to severe pathologic conditions in either their homo- or heterozygous forms. Examples in the transglutaminase protein family include the life threatening bleeding syndrome as a consequence of any of the more than 70 reported disease causing mutations of FXIIIa (Duckert et al. 1960; Karimi et al. 2009) and lamellar ichtiosis in cases of TGM1 mutations (Huber et al. 1995). No such genetic mutation of TG2 has been reported so far. Most of our genes also have common variants and the significance of these SNPs have been the subject of a large number of genome wide association studies to reveal their contribution to or role in multi-factorial complex diseases. For example, the Val34Leu SNP of FXIIIa (25% prevalence in whites) provides a moderate protection against coronary artery disease (Muszbek et al. 2010). Regarding TG2 systemic analysis of its polymorphisms has not been carried out yet.

In this paper, based on population genetic data (NCBI, Uniprot, HapMap, Ensemble, 1000 Genom datadases, Complete Genomics), we present the sequence of the most abundant (wild type) TG2 protein sequence and provide an overview of its so far found SNPs. Several of the TGM2 mutant forms have been produced and tested in biochemical assays. We compare the prevalence of TGM2 SNPs to those in other members of the transglutaminase family and discuss the possible reasons of the surprisingly low frequency of genetic variations in the protein coding sequence of TGM2.

Materials and methods

Databases and bioinformatic background

We used the standard websites of the NCBI, UniProt, HapMap, Ensembl (Galperin and Cochrane 2011), 1000 Genome and Complete Genomics databases for browsing the data and retrieving the relevant sequences. The 1000 Genomes data frequencies were extracted from the VCF file using the vcftools program (Danecek et al. 2011). For analyzing the Complete Genomics data (Drmanac et al. 2010), the REFSEQ gene variation tables of the high coverage sequencing of 46 individuals (diversity panel) were downloaded. The relevant diversity data were extracted from these tables using BASH shell scripts.

Transglutaminase enzyme preparations

The cloned (Gentile et al. 1991) and variant recombinant human TG2s were expressed in N-terminally $(His)_6$-tagged form and were purified by Ni-NTA affinity chromatography as described previously (Király et al. 2009). TG2 variants were constructed based on the QuikChange Site-Directed Mutagenesis Kit (Stratagene, La Jolla, CA, USA) and were checked by restriction analysis and DNA sequencing (ABI PRISM). The protein concentration was determined using the Bradford method (Bio-Rad, München, Germany). The purity of proteins was checked by Coomassie BB staining of SDS-polyacrylamide gels and by western blots.

Antibody-binding assays

Antigenicity and functional purity of TG2 proteins were determined by direct ELISA using G92 (G92.1.2; Trejo-Skalli et al. 1995; 1:3,000), CUB7402 (NeoMarkers, Fremont, CA, USA; 1:2,000), TG100 (NeoMarkers; 1:2,000) mouse monoclonal anti-TG2 and human coeliac antibodies as described earlier (Király et al. 2009). The wells were coated with 0.5 μg protein for 1 h at 4°C. After washing, the antibody solutions were incubated in the wells for 1 h at room temperature. The bound antibodies were detected using anti-mouse IgG/HRP or anti-human IgA IgG/HRP antibodies and TMB substrate.

Fibronectin binding

The fibronectin-binding property of the variants was tested using a direct ELISA assay. The wells were coated with 0.3 μg fibronectin in carbonate buffer pH 9.6 at 4°C for 1 h. After washing, 0.5 μg TG2 variant was incubated in the wells for 1 h at room temperature. The bound TG2 variants were detected using G92 (1:3,000) or CUB7402 (1:2,000) monoclonal antibodies and then anti-mouse IgG/HRP (1:5,000, Sigma) and TMB substrate.

Measurement of transglutaminase activity

The microtiter plate assay based on the incorporation of 5-(biotinamido)pentylamine (Molecular Probes, Invitrogene) into immobilized *N*,*N*-dimethylated casein (DMC) was used as described before (Király et al. 2009). The fibronectin bound TG2 activity assay was performed as published earlier (Király et al. 2006). Shortly, the wells were coated with 100 μl of 1 mg/ml human fibronectin in carbonate buffer pH 9.6 for 1 h at room temperature. After washing, the wells were coated with 3 μg recombinant enzyme in buffer A, then incubated with 200 ml of 1 mM 5-(biotinamido)-pentylamine substrate in 0.1 M Tris–HCl buffer, pH 8.5 containing 5 mM $CaCl_2$, 10 mM DTT. Amine incorporation was detected as in the conventional method.

Results and discussion

The reference sequence of wild type transglutaminase 2

Transglutaminase 2 was first cloned in the laboratory of Peter Davies (Gentile et al. 1991). The reference TGM2

sequence in the Ensemble database (obtained after the completion of the genome project) is shown in Fig. 1a. Comparing the two, five non-synonymous differences are seen in the former and we realized that these variants are also listed in the UniProt database as sequence conflicts (Fig. 1b). The TGM2 cDNA construct used in our and many other laboratories is the firstly cloned cDNA originating from the Davies' lab. After sequencing this cDNA (ABI PRISM) several times, we found that it has the reference nucleotide sequence and not the conflicting one at four of the five disputed sites. The lower reliability of the classical sequencing technique used 20 years ago probably explains this discrepancy. Until this time there has been no reported genome sequence which contains any of the four non-synonymous alleles. There is one exception, however, it is Gly224 in the Gentile sequence as opposed to Val224 in the reference sequence. We have validated this variant by re-sequencing the Gentile cDNA several times and suspect that it is either a mutation in the genome of the person from where TG2 was originally cloned or a mutation generated during the cloning procedures. It should be clarified whether this difference has any functional significance since to our knowledge many of the published transfection, recombinant and structural studies in the last two decades have been carried out using this cDNA, which is not the most frequent variant of the enzyme. Silent changes may in turn alter the expression level of TG2. Therefore, we have also analyzed the validity of such variants of the Gentile cDNA and found only one silent nucleotide change (in the triplet coding for Leu291), which could be confirmed as different from the reference sequence (Fig. 1b).

Due to the above mentioned differences, we also have analyzed the available data (allele frequency of the single nucleotide changes) of the 1000 Genome database and Complete Genomics high coverage sequencing diversity panel results. These databases provide the newest, high fidelity sequence data. The comparison of these sequences confirms that the TGM2 reference sequence, which can be found in the Ensemble database, is indeed by far the most common haplotype of the TGM2 gene (see also data and discussion in the next and the fourth section).

SNPs in the TGM2 gene

At the time of writing the present article, 365 SNPs have been reported in the databases for the 38 kb TGM2 gene. Of these 34 SNPs are in the coding region, 18 are non-synonymous mutations (Table 1), 14 are synonymous and 2 would lead to frameshift and shorter protein. Only eight missense SNPs have frequency data (2 others were found

A

```
1MAEELVLERC DLELETNGRD HHTADLCREK LVVRRGQPFW LTLHFEGRNY50
51EASVDSLTFS VVTGPAPSQE AGTKARFPLR DAVEEGDWTA TVVDQQDCTL100
101SLQLTTPANA PIGLYRLSLE ASTGYQGSSF VLGHFILLFN AWCPADAVYL150
151DSEEERQEYV LTQQGFIYQG SAKFIKNIPW NFGQFEDGIL DICLILLDVN200
201PKFLKNAGRD CSRRSSPVYV GRVVSGMVNC NDDQGVLLGR WDNNYGDGVS250
251PMSWIGSVDI LRRWKNHGCQ RVKYGQCWVF AAVACTVLRC LGIPTRVVTN300
301YNSAHDQNSN LLIEYFRNEF GEIQGDKSEM IWNFHCWVES WMTRPDLQPG350
351YEGWQALDPT PQEKSEGTYC CGPVPVRAIK EGDLSTKYDA PFVFAEVNAD400
401VVDWIQQDDG SVHKSINRSL IVGLKISTKS VGRDEREDIT HTYKYPEGSS450
451EEREAFTRAN HLNKLAEKEE TGMAMRIRVG QSMNMGSDFD VFAHITNNTA500
501EEYVCRLLLC ARTVSYNGIL GPECGTKYLL NLNLEPFSEK SVPLCILYEK550
551YRDCLTESNL IKVRALLVEP VINSYLLAER DLYLENPEIK IRILGEPKQK600
601RKLVAEVSLQ NPLPVALEGC TFTVEGAGLT EEQKTVEIPD PVEAGEEVKV650
651RMDLLPLHMG LHKLVVNFES DKLKAVKGFR NVIIGPA687
```

B

	AA position	Reference gene (Ensemble)	PDB database	First published (Gentile et al. 1991)	Gentile clone sequenced in Debrecen, Hungary
non-synonymous	51	Glu (gag)	Gln	Gln (cag)	Glu (gag)
	186	Glu (gaa)	Gln	Gln (caa)	Glu (gaa)
	224	**Val (gtg)**	**Gly**	**Gly (ggt)**	**Gly (ggt)**
	533	Asn (aac)	Thr	Thr (acc)	Asn (aac)
	655	Leu (ctg)	Val	Val (gtg)	Leu (ctg)
synonymous	213	Arg (cgc)	Arg	Arg (cgg)	Arg (cgc)
	291	Leu (ctg)	Leu	Leu (cta)	Leu (cta)
	367	Gly (ggg)	Gly	Gly (gga)	Gly (ggg)
	532	Leu (ctc)	Leu	Leu (cta)	Leu (ctc)
	654	Leu (ctg)	Leu	Leu (ctc)	Leu (ctg)

Fig. 1 Protein sequence of TG2 based on the reference sequence (**a**) and its comparison with the published sequences (**b**). The TGM2 cDNA that we use in our laboratory has only one amino acid difference compared to the reference gene. However, the origin of our TG2 cDNA was the vector, which was cloned and constructed by Gentile et al. We sequenced the Kojima laboratory's (RIKEN Advanced Science Institute, Japan) TG2 construct and its sequence was the same as our clone (the origin of this clone is also Gentile). The *shaded rows* in the table show the non-synonymous nucleotide changes which result amino acids changes

Table 1 Single nucleotide changes in the TGM2 coding sequence, which lead to amino acid changes

Amino acid position	Alleles (codon) and protein residues	Variation ID	Average heterosigosity in NCBI	Genotyped individuals in 1KG	Heterozygote in 1KG	Homozygote in 1KG	NCBI	Ensemble	1KG	HapMap	Uniprot	SIFT/ PolyPhen scores
51	(*g*ag) Glu–(*c*ag) Gln	rs1062735	N.D.	151	0	0	+	+	+	−	Sequence conflict	Tolerated/ benign
56	(*a*gt) Ser–(*g*gt) Gly	rs112985165	0.5	1	1	0	+	+	+	−	−	Tolerated/ benign
76	(c*g*t) Arg–(c*a*t) His	rs41274720	0.021	1,258	7	0	+	+	+	−	+	Tolerated/ benign
78	(*c*ca) Pro–(*t*ca) Leu	rs61733919	0.025	Failed variations			+	+	Failed	−	−	Tolerated/ benign
153	insertion:-–(*a*)Glu	rs34175510	N.D.	No data			+	+	+	−	−	−
170	deletion: (*c*)–-Gly	rs35700485	N.D.	No data			+	+	+	−	−	−
214	(c*g*c) Arg–(c*a*c) His	rs45530133	0.011	722	8	0	+	+	+	−	+	Deleterious/ probably damaging
314	(*g*ag) Glu–(*a*ag) Lys	rs138412064	N.D.	629	2	0	+	+	+	−	−	Deleterious/ probably damaging
324	(c*a*g) Gln–(c*g*g) Arg	rs45567334	0.010	95	1	0	+	+	+	−	+	Tolerated/ benign
330	(a*t*g) Met–Arg (a*g*g)	Porzio et al. (2007)	−	−	−	−	−	−	−	−	Mutations in early-onset type 2 or MODY patients	−
331	(a*t*c) Ile–Asn (a*a*c)											
333	(a*a*c) Asn–Ser (a*g*c)	Bernassola et al. (2002)										
377	(c*g*t) Arg–(c*a*t) His	rs78619991	0.113	629	5	0	+	+	With-drawn	−	−	Deleterious/ possibly damaging
383	(*g*ac) Asp–(*a*ac) Asn	rs137876403	N.D.	629	2	0	+	+	+	−	−	Tolerated/ benign
389	(*g*at) Asp–(*a*at) Asn	rs41274716	N.D.	No data			+	+	+	−	−	Deleterious/ probably damaging
436	(*c*gg) Arg–(*t*gg) Trp	rs45629036	0.011	718	5	0	+	+	+	−	+	Deleterious/ probably damaging
469	(g*a*g) Glu–(g*g*g) Gly	rs139974885	N.D.	629	3	0	+	+	+	−	−	Tolerated/ benign
523	(*g*ag) Glu–(*a*ag) Lys	rs112578056	0.5	1	1	0	+	+	+	−	−	Tolerated/ benign

Table 1 continued

Amino acid position	Alleles (codon) and protein residues	Variation ID	Average heterosigosity in NCBI	Genotyped individuals in 1KG	Heterozygote in 1KG	Homozygote in 1KG	NCBI	Ensemble	1KG	HapMap	Uniprot	SIFT/ PolyPhen scores
536	(*c*ct) Pro–(*t*ct) Ser	rs45556333	0.021	722	2	0	+	+	+	−	+	Deleterious/ possibly damaging
542	(*g*tt) Val–(*t*tt) Phe	rs115436227	0.033	736	7	0	+	+	+	−	−	Deleterious/ probably damaging
549	(g*a*g) Glu–(ggg) Gly	rs148045565	N.D.	629	1	0	+	+	+	−	−	Deleterious/ possibly damaging
556	(*a*cg) Thr–(*g*cg) Ala	rs143549454	N.D.	629	1	0	+	+	+	−	−	Deleterious/ possibly damaging
660	(ggc) Gly–(*g*tc) Val	COSM32992	−	−	−	−	−	−	−	−	Somatic mutation in human cancer	Deleterious/ probably damaging

Based on the available data of the 1000 Genomes as of September 1, 2011 and completed using NCBI, Ensemble, HapMap and Uniprot databases

+ Presence in the given database

The *underlined amino acid numbers* label the variants cloned and tested (see Figs. 2, 3, 4, and 5)

The *italicized bases* mark the single nucleotide change in the codon

377Aa: This variation has been flagged as failed. Variation submission has been withdrawn by the 1000 genomes project due to high false positive rate

78 Aa: None of the variant alleles match the reference allele

1KG: One Thousand Genome database

SIFT (Sorting Intolerant From Tolerant) is a program that predicts whether an amino acid substitution affects protein function (Ng and Henikoff 2003)

PolyPhen: Variant Effect Predictor (Adzhubei et al. 2010)

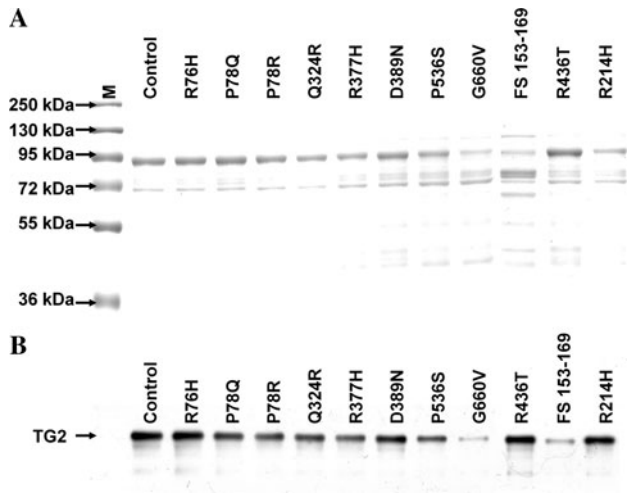

Fig. 2 Purity and immunoreactivity of TG2 variants. **a** 4 μg protein sample/lane was visualized with Coomassie BB staining after 10% SDS-PAGE and **b** immunoblot of 0.5 μg protein sample/lane. Data are representative of two independent experiments. Control: un-modified recombinant TG2

only in single genome): they occur in low frequency in the population and all are heterozygous. The analysis of the 1000 Genomes database also confirms that the frequency of the non-synonymous changes in TGM2 is very low while it has synonymous SNPs with average frequency.

It should be noted that the databases contain some controversial data. Some of the SNPs are listed differently in the databases and sometimes the frequency value of the same SNP varies from one database to the other. Some SNPs have identification number but their heterozygosity value is 0.000 and in the population genetic data there is no chromosome which contains the particular variant (e.g. in the NCBI database rs146621816, rs139974885).

In the literature, three missense mutations (Asn333Ser, Met330Arg and Ile331Asn) were described in patients of Maturity Onset Diabetes of the Young (MODY) or early onset type 2 diabetes. These variants have impaired transglutaminase activity (TGase) suggesting that enzymatic activity of TG2 can play a role in disorders of glucose metabolism (Porzio et al. 2007; Bernassola et al. 2002). However, there is no direct evidence that these mutations contribute to the development of diabetes in these patients. In a British population, eight non-coding SNPs across the TGM2 gene were genotyped and it was suggested that the detected variants were associated with schizophrenia and TGM2 may have a role in the development of this mental disease (Bradford et al. 2009). In another study, a major fraction of human protein coding genes was examined in breast and colorectal cancer (Sjöblom et al. 2006) to determine the spectrum and extent of somatic mutations; TGM2 was one of the genes with a somatic mutation (Gly660Val).

Functional consequence of TG2 mutations

Based on the data obtained from the databases between 2009 and 2010, 11 variants or mutants were produced in our laboratory by site directed mutagenesis. The studied variants were selected on the basis of some predictions for the potential effect of the amino acid change by comparing TGM2 to other transglutaminase family members and expecting that some SNPs in regions which are well conserved in the family may lead to serious functional defects while others, for example in the N terminal part, could cause just small deviances from the wild type protein. We predicted that SNPs close to the Ca^{2+}-binding sites can also cause serious functional defects.

After protein expression and one step Ni-NTA affinity purification, the yield and purity of the variants were determined (Fig. 2). Although each variant could be expressed in a bacterial system, the yield and purity of the protein preparations were different. The Gly660Val mutant and FS153–169 double frame shift variants show lower purity and yield. In the case of the FS153–169 variant, this is not surprising because in our previous work, when this region was modified by mutations, the stability and purity of the expressed proteins were also less compared to the wild type enzyme and other mutants (Király et al. 2009). This region of TG2 could be responsible for the stabilization of the protein or its proper folding.

Before studying the functional properties of the TG2 variants, three monoclonal antibodies were used to compare their folding and binding properties on a microtiter plate (Fig. 3a). The G92 antibody recognizes the N-terminal part of TG2 (Trejo-Skalli et al. 1995), the CUB7402 (Epitope: aa447-478) and TG100 (Epitope: aa447-538) reacts with the catalytic core domain of TG2. In the case of G92 and CUB7402 antibodies, we could not see significant decrease in their binding to the mutant TG2. For the TG100 antibody, the Arg436Thr and Pro536Ser variants showed slight, while the Gly660Val mutant and FS153–169 variant slight but significant decrease in the antibody binding. The structural change in the molecules may alter the antibody-binding region or can decrease the stability of the variants.

It is known that TG2 plays a role in coeliac disease where it is the main antigen. The coeliac antibody binding of the mutants was also tested by ELISA (Fig. 3b). The Arg214His and Pro536Ser variants show slight while the Gly660Val and FS153–169 ones significant decrease in antibody binding. The lowest antibody binding was found in the case of the FS153–169 variant. There is a sequence overlap between the FS153–169 variant and the S4 Ca^{2+}-binding site mutants produced in our previous work (Király et al. 2009) where we also observed significant decrease in coeliac antibody binding. The present observation confirms

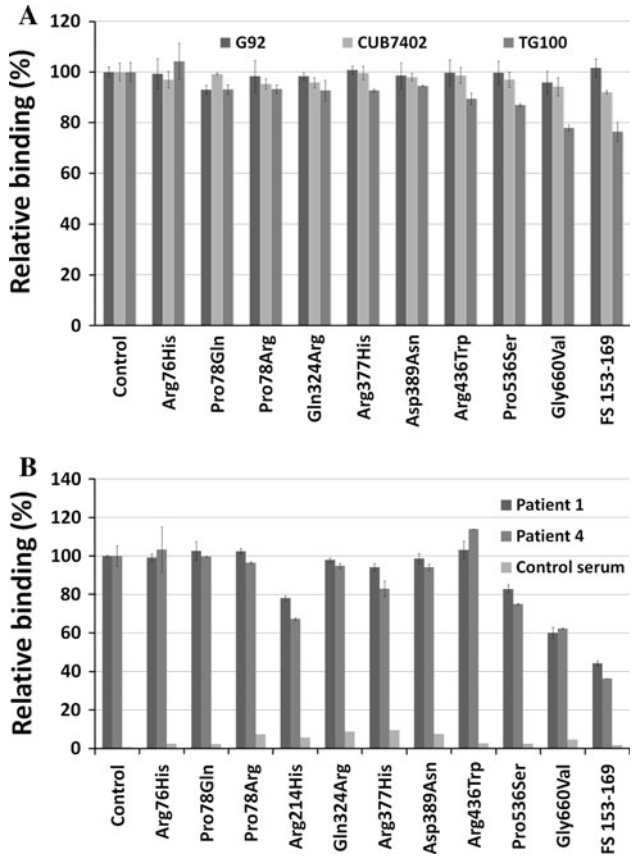

Fig. 3 Monoclonal and coeliac antibody binding of TG2 variants. **a** The applied monoclonal antibodies recognize the TG2 variants. **b** Patient 1 and 4 are untreated coeliac patients, control is a healthy donor. Dilution of the sera is 1:150. Binding to the un-modified TG2 (control) is 100%. Data are presented as means with ±standard deviations from two separate experiments performed in duplicate

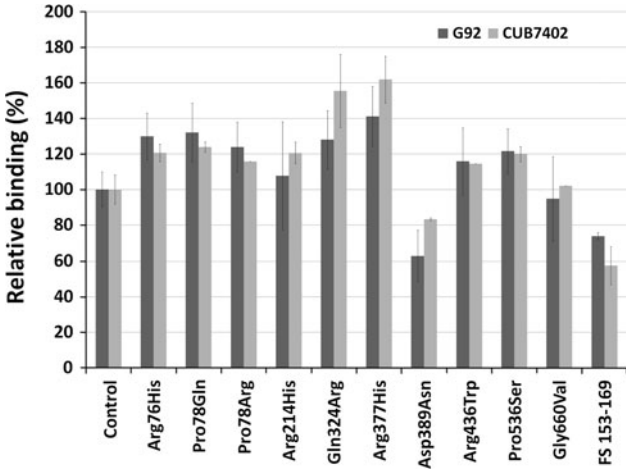

Fig. 4 Fibronectin-binding property of the TG2 variants. The fibronectin binding was detected using two different monoclonal antibodies. Binding of the un-modified TG2 (control) is 100%. Data are presented as means with ±standard deviations from two separate experiments performed in duplicate

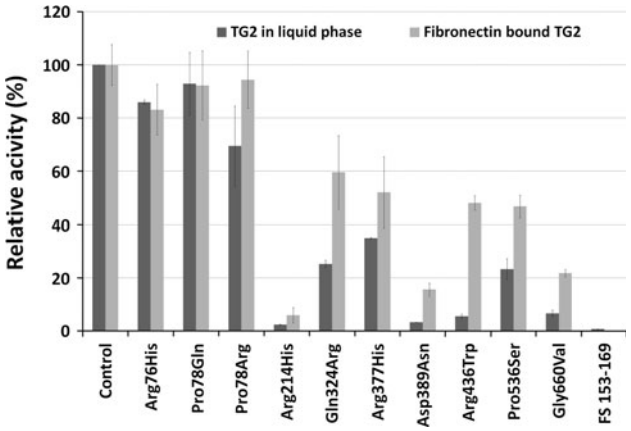

Fig. 5 Transglutaminase activity of liquid phase and fibronectin bound TG2 variants. Transglutaminase activity is shown as a percentage of activity of the un-modified (control) TG2 (specific activity of liquid phase un-modified TG2: 6.5 ΔAbs405/min/mg protein). Data are presented as means with ±standard deviations from two separate experiments performed in triplicate

that this region contains an important anchor point for the autoantibodies to bind TG2.

Since TG2 has a significant role in scaffolding processes at the extracellular side of the cells and its interaction with fibronectin has a crucial role in this function, the fibronectin-binding property of the variants was also tested by an ELISA method (Fig. 4). The variants bound fibronectin similarly to the un-modified enzyme, except the Asp389Asn and FS153–169 variants which exhibited decreased but not totally lost binding.

Finally, we measured transglutaminase activity of the variants in either their free or fibronectin bound form (Fig. 5). The Arg76His, Pro78Gln, Pro78Arg variants showed activity comparable to the un-modified enzyme. The Arg214His and FS153–169 variants lost their transglutaminase activity in both assays. Asp389Asn, Arg436Thr and Gly660Val variants had very low transglutaminase activity (under 10% compared to wild type) using the conventional microtiter plate method but they had significantly higher transglutaminase activity when the enzyme was bound to fibronectin. Actually, in case of the other variants also higher transamidase activity was observed when they were bound to fibronectin. This suggests that fibronectin binding may improve stability of the active conformation of TG2 variants.

The results confirmed our functional prediction since the functionally defective variants have amino acid change in well conserved regions or close to the Ca^{2+}-binding sites of TG2. Since most TG2 variants with non-synonymous changes in exon sequences result in defective proteins it is likely that their appearance in affected individuals may have pathologic consequences. They may disturb cellular and tissue functions when are manifested in both TGM2 alleles.

Table 2 Frequent variants of the transglutaminase family in the NCBI and Ensemble database

Gene	Examples for frequent variants (polymorphism/% allele frequency or heterozygosity)
F13A1	V35L/20.9%; P565L/19.5% or 0.328*; V651I/0.123*; E652Q/20.2% or 0.318*
TGM1	There is no non-synonimous SNP over 5% frequency
TGM2	There is no non-synonimous SNP over 2% frequency
TGM3	T13K/0.393*; I163L/0.333*; S249 N/0.244*; G654R/0.431*
TGM4	E313K/35.9%; R372C/27.6%; V409I/27.4%
TGM5	S15C/20.9%; A352G/31%; V504 M/25.8%
TGM6	M58V/24.7% or 0.278*; A141E/10.5%; A372T/0.375*
TGM7	M229T/25.6%; V251A/25.9%
EPB42	T595P/0.241*

The source of the heterozygosity (*) is the NCBI database and the allele frequencies (%) were calculated based on the Ensemble database

No cases like that have been described so far. The pathologies observed in TG2 knockout mice, e.g. diabetes (Bernassola et al. 2002) and autoimmunity (Szondy et al. 2003), may give a clue to what to expect in such genetic cases if ever occurred. Somatic mutations may also lead to TG2 deficiency in cellular lineages or tumors contributing to tissue and disease specific malfunctions. Heterozygosity may also contribute to the development of complex diseases, for example, the missense mutations to the onset of MODY. Even at the present stage of genomic technologies, it is a challenging task to prove association of rare frequency mutations with diseases; large cohorts, in the range of tens of thousands participants are needed to assess their contribution to pathologic phenotypes. However, recent developments in fast and precise DNA sequencing methods and the personal medicine approaches will open new possibilities to clarify the role and significance of rare SNPs, including the TG2 variants reported here, in various human pathologies.

Comparison of the frequency of genetic variations in genes of the transglutaminase family

In the Ensemble and 1000 Genomes database more than 100 single nucleotide changes are reported for FXIIIA and TGM1 and most of these are also listed as rare mutations in The Human Gene Mutation Database (http://www.hgmd.org) having causative roles in severe diseases such as FXIIIa deficiency with bleeding disorders (F13A1 mutants) and lamellar ichtyosis (TGM1) (Table 2). In the transglutaminase family F13A1, TGM3 and TGM6 have some SNPs (3–4) with high frequency occurrence (over 10%) in the population while TGM2 and TGM1 have very low polymorphism. Regarding common variants the significance of Val34Leu polymorphism of Factor XIIIa have been assessed and was found to provide protection against myocardial infarction (Muszbek et al. 2010). The enzyme activity of the Val34Leu variant does not change but its thrombin activation is quicker than that of the most frequent sequence variant (Kohler et al. 1998; Var et al. 2009). Another frequent variant of Factor XIIIa is Pro564Leu (Anwar et al. 1999; Saha et al. 2000) with higher enzymatic activity but no physiologic or pathologic association. The numbers of variants of other transglutaminases in the databases are fewer than 40. Among them there are two interesting mutations in TGM5 which show association with the Acral Peeling Skin Syndrome (Gly113Cys, Cassidy et al. 2005; Trp255Arg, Lys445Asn, Kharfi et al. 2009).

In order to compare the diversity in the coding sequences of the transglutaminase family genes, we used the Complete Genomics diversity panel high coverage sequence data (Table 3). We counted the number of the observed diverse loci in each reference gene transcript and normalized the results based on the length of the proteins. We have also counted the number of individuals out of 46 examined genomes who are homo- or heterozygous on the given locus. The data show clearly that among the examined gene transcripts the TGM2 NM_198951 transcript has the lowest missense number per 100 amino acids, while the compatible (silent or synonymous) mutation rate of this transcript is at average. This proves that the TGM2 protein is under a great purifying selective evolutionary pressure, which eliminates most of the missense mutations from the population. The reason for this strong evolutionary constraint is most likely the multiple, important functions of the protein in various cell compartments and outside of cells (Fesus and Piacentini 2002). It is also worth to note, however, that both the TGM2 and the EPB42 genes, which have the lowest mutation rate (Table 3), have two annotated transcript and protein form. This would cause increased selectional pressure on different parts of the proteins and thus can explain partly the lower than average mutational rate.

Conclusions

Analysis of the available genomic databases clearly show that the protein coding sequence of human transglutaminase 2 has very low polymorphism in the population and it has the lowest variability when compared to other members

Table 3 Mutation analysis of the Complete Genomics diversity panel (46 individuals) data on the transglutaminase family reference transcripts

	Length of the protein	Number of loci	Mutation frequency /1 kbp	Missense						Compatible (silent)					
				No. of mutations	No. of loci	Mutations/100 AA	No. of individuals having mutations	No. of homozygous mutations	No. of heterozygous mutations	No. of mutations	No. of loci	Mutations/100 AA	No. of individuals having mutations	No. of homozygous mutations	No. of heterozygous mutations
TGM2-NM_198951	547	31	18.9	1	1	0.18	1	0	1	36	30	6.58	23	6	24
TGM2-NM_004613	688	38	18.4	3	3	0.44	3	0	3	41	35	5.96	26	6	29
TGM1-NM_000359	818	26	10.6	5	5	0.61	5	0	5	24	21	2.93	18	3	18
TGM3-NM_003245	694	114	54.8	170	109	24.50	46	61	48	6	5	0.86	4	1	4
F13A1-NM_000129	733	67	30.5	59	53	8.05	33	6	47	14	14	1.91	14	0	14
TGM4-NM_003241	685	228	111	252	178	36.79	43	74	104	69	50	10.07	41	19	31
TGM5-NM_004245	639	78	40.7	44	36	6.89	26	8	28	60	42	9.39	36	18	24
TGM5-NM_201631	721	80	37	47	38	6.52	27	9	29	60	42	8.32	36	18	24
TGM6-NM_198994	707	98	46.2	95	58	13.44	45	37	21	49	40	6.93	29	9	31
TGM7-NM_052955	711	86	40.3	9	9	1.27	7	0	9	118	77	16.60	37	41	36
EPB42-NM_001114134	691	7	3.38	3	3	0.43	3	0	3	5	4	0.72	4	1	3
EPB42-NM_000119	721	9	4.16	6	5	0.83	5	1	4	5	4	0.69	4	1	3

of the human transglutaminase family. According to functional studies of the rare protein variants, which so far have been found encoded only as heterozygous forms in genomes, the amino acid changes lead to defective proteins. This raises the possibility that they contribute to the development of pathologic conditions that will be very difficult to find considering their very low frequency of occurrence. Based on these conclusions, there is another important issue to address: how it is possible that while there is such a strong evolutionary pressure on human TGM2, the homozygous deletion of the mouse TGM2 does not result in life threatening phenotypes and these mice are fertile. Obviously, this and similar questions will generate exciting studies in the future resulting in important novel findings in human physiology and pathology.

Acknowledgments This work has been supported by grants from the Hungarian Scientific Research Fund (OTKA NI 67877), TAMOP-4.2.2-08/1/2008-0015 and TÁMOP 4.2.1./B-09/1/KONV-2010-0007 projects implemented through the New Hungary Development Plan, co-financed by the European Social Fund, an ETT grant from the Hungarian Ministry of Health, European Union MRTNCT-2006-036032 and MRTN-CT 2006-035624.

Conflict of interest The authors declare that they have no conflict of interest.

References

Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR (2010) A method and server for predicting damaging missense mutations. Nat Methods **7**:248–249

Anwar R, Gallivan L, Edmonds SD, Markham AF (1999) Genotype/phenotype correlations for coagulation factor XIII: specific normal polymorphisms are associated with high or low factor XIII specific activity. Blood 93:897–905

Bernassola F, Federici M, Corazzari M, Terrinoni A, Hribal ML, De Laurenzi V, Ranalli M, Massa O, Sesti G, McLean WH, Citro G, Barbetti F, Melino G (2002) Role of transglutaminase 2 in glucose tolerance: knockout mice studies and a putative mutation in a MODY patient. FASEB J 16:1371–1378

Bradford M, Law MH, Stewart AD, Shaw DJ, Megson IL, Wei J (2009) The TGM2 gene is associated with schizophrenia in a British population. Am J Med Genet B Neuropsychiatr Genet 150B:335–340

Cassidy AJ, van Steensel MA, Steijlen PM, van Geel M, van der Velden J, Morley SM, Terrinoni A, Melino G, Candi E, McLean WH (2005) A homozygous missense mutation in TGM5 abolishes epidermal transglutaminase 5 activity and causes acral peeling skin syndrome. Am J Hum Genet 77:909–917

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R, 1000 Genomes Project Analysis Group (2011) The variant call format and VCFtools. Bioinformatics 27:2156–2158

Duckert F, Jung E, Shmerling DH (1960) A hitherto undescribed congenital haemorrhagic diathesis probably due to fibrin stabilizing factor deficiency. Thromb Diath Haemorrh 5:179–186

Drmanac R, Sparks AB, Callow MJ, Halpern AL, Burns NL, Kermani BG, Carnevali P, Nazarenko I, Nilsen GB, Yeung G, Dahl F, Fernandez A, Staker B, Pant KP, Baccash J, Borcherding AP, Brownley A, Cedeno R, Chen L, Chernikoff D, Cheung A, Chirita R, Curson B, Ebert JC, Hacker CR, Hartlage R, Hauser B, Huang S, Jiang Y, Karpinchyk V, Koenig M, Kong C, Landers T, Le C, Liu J, McBride CE, Morenzoni M, Morey RE, Mutch K, Perazich H, Perry K, Peters BA, Peterson J, Pethiyagoda CL, Pothuraju K, Richter C, Rosenbaum AM, Roy S, Shafto J, Sharanhovich U, Shannon KW, Sheppy CG, Sun M, Thakuria JV, Tran A, Vu D, Zaranek AW, Wu X, Drmanac S, Oliphant AR, Banyai WC, Martin B, Ballinger DG, Church GM, Reid CA (2010) Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. Science 327:78–81

Fesus L, Piacentini M (2002) Transglutaminase 2: an enigmatic enzyme with diverse functions. Trends Biochem Sci 27:534–539

Folk JE, Cole PW (1966) Mechanism of action of guinea pig liver transglutaminase. I. Purification and properties of the enzyme: identification of a functional cysteine essential for activity. J Biol Chem 241:5518–5525

Folk JE, Chung SI (1973) Molecular and catalytic properties of transglutaminases. Adv Enzymol Relat Areas Mol Biol 38:109–191

Galperin MY, Cochrane GR (2011) The 2011 nucleic acids research database issue and the online molecular biology database collection. Nucl Acids Res 39(Database issue):D1–D6

Gentile V, Saydak M, Chiocca EA, Akande O, Birckbichler PJ, Lee KN, Stein JP, Davies PJ (1991) Isolation and characterization of cDNA clones to mouse macrophage and human endothelial cell tissue transglutaminases. J Biol Chem 266:478–483

Huber M, Rettler I, Bernasconi K, Frenk E, Lavrijsen SP, Ponec M, Bon A, Lautenschlager S, Schorderet DF, Hohl D (1995) Mutations of keratinocyte transglutaminase in lamellar ichthyosis. Science 267:525–528

Karimi M, Bereczky Z, Cohan N, Muszbek L (2009) Factor XIII deficiency. Semin Thromb Hemost 35:426–438

Kharfi M, El Fekih N, Ammar D, Jaafoura H, Schwonbeck S, van Steensel MA, Fazaa B, Kamoun MR, Fischer J (2009) A missense mutation in TGM5 causes acral peeling skin syndrome in a Tunisian family. J Invest Dermatol 129:2512–2515

Király R, Vecsei Z, Deményi T, Korponay-Szabó IR, Fésüs L (2006) Coeliac autoantibodies can enhance transamidating and inhibit GTPase activity of tissue transglutaminase: dependence on reaction environment and enzyme fitness. J Autoimmun 26:278–287

Király R, Csosz E, Kurtán T, Antus S, Szigeti K, Simon-Vecsei Z, Korponay-Szabó IR, Keresztessy Z, Fésüs L (2009) Functional significance of five noncanonical Ca^{2+}-binding sites of human transglutaminase 2 characterized by site-directed mutagenesis. FEBS J 276:7083–7096

Kohler HP, Stickland MH, Ossei-Gerning N, Carter A, Mikkola H, Grant PJ (1998) Association of a common polymorphism in the factor XIII gene with myocardial infarction. Thromb Haemost 79:8–13

Laki K, Lóránd L (1948) On the solubility of fibrin clots. Science 108:280

Loewy AG (1968) Enzymatic control of insoluble fibrin solution. In: Laki K (ed) Fibrinogen. Marcel Dekker, New York, pp 185–224

Lorand L, Downey J, Gotoh T, Jacobsen A, Tokura S (1968) The transpeptidase system which crosslinks fibrin by gamma-glutamyle-episilon-lysine bonds. Biochem Biophys Res Commun 31:222–230

Matacić S, Loewy AG (1968) The identification of isopeptide crosslinks in insoluble fibrin. Biochem Biophys Res Commun 30:356–362

Muszbek L, Bereczky Z, Bagoly Z, Shemirani AH, Katona E (2010) Factor XIII and atherothrombotic diseases. Semin Thromb Hemost 36:18–33

Mycek MJ, Clarke DD, Neidle A, Waelsch H (1959) Amine incorporation into insulin as catalysed by transglutaminase. Arch Biochem Biophys 84:528–540

Ng PC, Henikoff S (2003) SIFT: predicting amino acid changes that affect protein function. Nucl Acids Res 31:3812–3814

Pisano JJ, Finlayson JS, Peyton MP (1968) Cross-link in fibrin polymerized by factor 13: epsilon-(gamma-glutamyl)lysine. Science 160:892–893

Porzio O, Massa O, Cunsolo V, Colombo C, Malaponti M, Bertuzzi F, Hansen T, Johansen A, Pedersen O, Meschi F, Terrinoni A, Melino G, Federici M, Decarlo N, Menicagli M, Campani D, Marchetti P, Ferdaoussi M, Froguel P, Federici G, Vaxillaire M, Barbetti F (2007) Missense mutations in the TGM2 gene encoding transglutaminase 2 are found in patients with early-onset type 2 diabetes. Hum Mutat 28:1150

Saha N, Aston CE, Low PS, Kamboh MI (2000) Racial and genetic determinants of plasma factor XIII activity. Genet Epidemiol 19:440–455

Sarkar NK, Clarke DD, Waelsch H (1957) An enzymically catalyzed incorporation of amines into proteins. Biochim Biophys Acta 25:451–452

Sjöblom T, Jones S, Wood LD, Parsons DW, Lin J, Barber TD, Mandelker D, Leary RJ, Ptak J, Silliman N, Szabo S, Buckhaults P, Farrell C, Meeh P, Markowitz SD, Willis J, Dawson D, Willson JK, Gazdar AF, Hartigan J, Wu L, Liu C, Parmigiani G, Park BH, Bachman KE, Papadopoulos N, Vogelstein B, Kinzler KW, Velculescu VE (2006) The consensus coding sequences of human breast and colorectal cancers. Science 314:268–274

Szondy Z, Sarang Z, Molnar P, Nemeth T, Piacentini M, Mastroberardino PG, Falasca L, Aeschlimann D, Kovacs J, Kiss I, Szegezdi E, Lakos G, Rajnavolgyi E, Birckbichler PJ, Melino G, Fesus L (2003) Transglutaminase 2−/− mice reveal a phagocytosis-associated crosstalk between macrophages and apoptotic cells. Proc Natl Acad Sci USA 100:7812–7817

The 1000 Genomes Project Consortium (2010) A map of human genome variation from population-scale sequencing. Nature 467:1061–1073

Thorisson GA, Smith AV, Krishnan L, Stein LD (2005) The International HapMap Project Web site. Genome Res 15:1591–1593

Trejo-Skalli AV, Velasco PT, Murthy SN, Lorand L, Goldman RD (1995) Association of a transglutaminase-related antigen with intermediate filaments. Proc Natl Acad Sci USA 92:8940–8944

Var A, Utük O, Akçali S, Sanlidağ T, Uyanik BS, Dinç G (2009) Impact of hemostatic gene single point mutations in patients with non-diabetic coronary artery disease. Mol Biol Rep 36:2235–2243