THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY (PHD)

Amino acid variants in human transglutaminase 2 and their biological relevance

by Kiruphagaran Thangaraju

Supervisor: Prof. Dr. László Fésüs

UNIVERSITY OF DEBRECEN

DOCTORAL SCHOOL OF MOLECULAR CELL AND IMMUNE BIOLOGY

Debrecen, 2016

CONTENTS

ABBREVIATIONS

Atg	Autophagy-related protein
ATP	Adenosine triphosphate
BSA	Bovine serum albumin
Ca^{2+}	Calcium ions
CID	Collision-induced dissociation
DMEM	Dulbecco's Modified Eagle Medium
DTT	Dithiothreitol
ECM	Extracellular matrix
EDTA	Ethylenediaminetetraacetic acid
ELISA	Enzyme-linked immunosorbent assay
ELM	Eukaryotic linear motif
ESI	Electrospray ionization
ESP	Exome Sequencing Project
ExAC	Exome Aggregation Consortium
F13a	Factor XIII A
FN	Fibronectin
GDI	Gene damage index scores
GDP	Guanosine diphosphate
GFP	Green fluorescent protein
GST	Glutathione S-transferase
HD	Huntington's, disease
HPLC	High-performance liquid chromatography
HTT	Huntingtin
IDRs	Intrinsically disordered regions
kDa	Kilo Dalton
LC3	Microtubule-associated protein light chain 3
LIRs	LC3 interacting region
LOF	Loss-of-function
MAF	Minor allele frequency
mHTT	Mutant huntingtin
mM	Milli molar
MOPS	3-(N-morpholino)propanesulfonic acid
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
nsSNVs	Non-synonymous single nucleotide variants
NES	Nuclear export signal sequence
NII	Neuronal intranuclear inclusions
NLS	Nuclear localization signal sequence
p62	Nucleoporin 62
PANTHER	Protein ANalysis THrough Evolutionary Relationships
PBS	Phosphate-buffered saline
PC12	Pheochromocytoma
PDB	Protein Data Bank
PMSF	Phenylmethylsulfonyl fluoride
PolyPhen	Polymorphism Phenotyping
RFU	Relative fluorescence units
RGD	Arginine–glycine–aspartic motif

RIPA	Radioimmunoprecipitation assay buffer
SARs	Selective autophagy receptors
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SH	Src homology
SIFT	Sorting Intolerant From Tolerant
SLiMs	Short linear motifs
SUMO	Small Ubiquitin-like Modifier
TG	Transglutaminase
TMB	Tetramethylbenzidine
TRAF-2	TNF receptor-associated factor 2
TTBS	Tris-buffered saline
USP7	Ubiquitin-specific-processing protease 7
UV	Ultra violet

1. INTRODUCTION

1.1. Comparative genomics as a tool to study evolution and disease

Comparative genomics is a large-scale powerful approach to identify similarities and differences between genomes and to understand biology of the individual genomes by comparing two or more genomes (Wei et al., 2002). It offers insights in to genome structure including genome rearrangement at both DNA and gene level. The gene regulation, expression, complex interaction, gene content, orthologs, and paralogs can also be studied. It is also applied in analyzing the coding regions including the protein content and noncoding regions including the prediction of regulatory elements (Wei et al., 2002). In this way it helps to identify genes pertaining to inborn errors and defense mechanisms that have protected us and our ancestors from extinction (Wei et al., 2002). Comparative studies also give insight in to the evolutionary forces that have moulded our species, including underlying mutational processes and selective constraints.

In the past decade, development in whole-genome comparative analysis and associated tools, and advances in sequencing has revolutionized the field of comparative genomics. Large-scale genomic sequences, comparison programmes, algorithms and visualization tools like BLASTN, MEGABLAST, GLASS, MUMer, PatternHunter, VISTA, WABA, PipMarker, DIALIGN are widely used (Wei et al., 2002; Pennacchio et al., 2003). Databases and browsers for example, UCSC Browser, Ensembl, MapView, UniProt, KEGG, PEDANT, and STRING are being used (Wei et al., 2002; Pennacchio et al., 2003). Genome-to-genome comparisons and large scale analysis of individual genomes have become feasible due to the public availability of whole genome sequences of numerous organisms. The initial sequencing and comparative analysis of the mouse genome in 2002 (Waterston et al., 2002) and the nucleotide sequence of the euchromatic part of the human genome was published in 2004 under human genome project (International human genome sequencing consortium, 2004). This was followed by the publication of the initial nucleotide sequence of the chimpanzee (Pan troglodytes) (The chimpanzee sequencing and analysis consortium, 2005) and sequencing of the genome of the rhesus macaque (Macaca mulatta) (Gibbs et al., 2007).

Human, mouse and other mammals diverged about 75 million years ago. Mouse is a premier genetic model organism for probing human biology and disease because of extensive genetic studies, availability of scores of inbred strains, and techniques to gentically modify them (Chinwalla et al., 2002). In both organisms, comparative genomics helps in mapping the genes involved in a common pathway or diseases. Some mapping studies show that

evolutionary conserved regions between rodents and humans are mapped to common phenotypes like obesity, autoimmunity, hypertension, and coronary disease (Korstanje and DiPetrillo, 2004; Stoll et al., 2000; Sugiyama et al., 2001; Wang and Paigen, 2005; Jacob and Kwitek, 2002; Rollins et al., 2006). Comparative mapping will facilitate to identify better animal models for human diseases at both genotypic and phenotypic levels. The African apes (chimpanzees, bonobos, and gorillas) are the closest living evolutionary relatives of humans. Of all, the chimpanzees, differ from humans by an average of 1.1-1.4 % in overall genomic DNA sequences. Comparative primate genomics is gaining ground and by studying non-human primate genomes, origin, history and mechanism of human genome can be reconstructed and eventually human diseases can be analyzed.

Evolution of species is shaped by two types of selective forces namely, purifying selection (conservation of existing phenotype are favoured) and other is positive or Darwinian selection (emergence of new phenotypes are promoted) (Vallender and Lahn, 2004). The positively selected genes involved in brain development and cognitive abilities may offer valuable molecular clues about the evolution of our species and the selective pressure that drives it. Positively selected genes can also play a potential role in human medicine as many diseases are specific for humans and the therapeutic strategies based on animal models fail in humans. This is due to the differences between the humans and other species and genes under positive selection may be fundamental for some of these dissimilarities. These gene variants in humans may participate directly in the disease pathogenesis. Rapid population growth in the last few thousand years coupled with inefficient and weak natural selection lead to excessive damaging rare variants in our genome. In the recent decade considerable efforts are underway in human genetics research to study the allelic variations in human genes and relate it to evolutionary forces which shape these variations.

1.2. Genomic polymorphism and ongoing evolution in the human genome

Evolutionary history of a population can be followed closely by studying the polymorphism within a species (Lawrie and Petrov, 2014). After the completion of human genome project it has become evident that vast majority of our proteins are polymorphic to single nucleotide variations (SNVs). Exome sequencing has become a powerful approach for discovering rare variations and has facilitated the genetic dissection of unsolved Mendelian disorders and the study of human evolutionary history. Multi-center efforts like NHLBI-GO Exome Sequencing

Project (ESP), and Exome aggregation consortium (ExAC) have made data from exome projects easily accessible for analysis.

Single nucleotide variation research aids in understanding the genetics of human phenotype variation and complex diseases (Ramensky et al., 2002). Genomic polymorphism is broadly classified into two types i. those affecting coding regions and ii. those affecting non-coding regions. Coding variants are classified in to synonymous and non-synonymous and non-synonymous variants are again classified in to missense and nonsense variants. The non-synonymous single nucleotide variants (nsSNVs) lead to change in the amino acid sequence which can affect protein function. Out of 13,000 exonic variants present per person, 58 % are non-synonymous (Tennessen et al., 2012). The nsSNVs along with variants in the regulatory region have highest impact on the phenotype (Ramensky et al., 2002) and former is implicated in both Mendelian and complex diseases. Core of the protein and protein-protein interface are enriched with disease causing nsSNVs (Tennessen et al., 2012) and compared to neutral nsSNVs, they are more likely to affect the phosphorylation and glycosylation sites (Tennessen et al., 2012). Moreover, protein misfolding or instability caused by nsSNVs also influence protein-protein interactions.

The human genome creates mutations within its sequence and preserves the genetic variation in order for human adaptability. Continuous genetic changes caused by mutations will lead to increase in allele frequencies and thus evolution of the human genome. Allele frequencies are an important indicator of evolutionary changes as increase in allele frequency of a gene is favored and decrease in allele frequency is selected aganist. Increase in mutated allele frequency rate signifies ongoing selection, given the selection is recurrent and positive selection acts on protein-coding genes (Nielsen et al., 2007). The demographic history of human population during the past 10,000 years can be learnt from the frequencies of genetic variants but the population should contain variants from that same time span. Frequency of many of such variants are < 0.1 % and mostly are rare in the whole population and requires sequencing of larger sample of individuals (Keinan and Clark, 2012). Loads of rare variations are generated by recent rapid population growth and there is a growing body of evidence that rare SNVs with < 5 % minor allele frequency (MAF) might play an important role in common human disease etiology (Gorlov et al., 2011).

Rare human knockouts with homozygous loss-of-function (LOF) variants have been on focus for past few years in the field of exome sequencing research (MacArthur et al., 2010). For decades, researchers have used mouse models to study the function of a gene as it provides an opportunity to test possible therapeutic agents and evaluate their precise effects. As mouse

models do not provide sufficient answers related to human average mutation rate, current biomedical research focuses towards homozygous LOF variants or human knock-outs (Sulem et al., 2015; Lek et al., 2015; MacArthur et al., 2012). Thousands or millions of people genomes are scanned for naturally occurring mutations that knockout a particular gene pointing to disease phenotypes and novel approaches towards their prevention or treatment. Studies have speculated that LOF variants have been advantageous in human evolution but their mode of segregation and fixation through the human lineage should be evaluated in detail (Alkuraya, 2015).

The evolutionary history and importance of functional elements can be determined with more potency using an unified framework of multi-species polymorphism and divergence data. By combining polymorphism and comparative genomics the fraction of deleterious mutations can be estimated with greater efficacy. Large number of sequencing projects and polymorphism data of our closest relatives are further needed to gain an accurate picture of the functional elements in the human genome and its importance across it.

1.3. The transglutaminase family

First transglutaminase (TG; EC 2.3.2.13) was identified and reported in 1957 by Clarke and co-workers from guinea pig liver extracts (Sarkar et al., 1957). Since then, proteins exhibiting transglutaminase activity have been described in unicellular organisms, invertebrates, amphibians, fish, plants and mammals (reviewed in Eckert et al., 2014; Nurminskaya and Belkin, 2012; Lorand and Graham, 2003). Family of closely related genes encode the transglutaminase enzymes with high degree of sequence similarity and all of them exhibiting a similar gene organization (reviewed in Eckert et al., 2014; Nurminskaya and Belkin, 2012; Lorand and Graham, 2003). Amino acid similarity and evolutionary distances calculated by sequence alignments and algorithms placed vertebrate TGs into two different lineages: F13a (Factor XIII) and TG1 belongs to lineage 1 and others comprising lineage 2 (Figure 1). Mammalian transglutaminases seem to be derived from papain-like superfamily of cysteine proteases with a good structural homology (similar gene, primary protein, and three-dimensional structure and catalytic mechanism) (Grenard et al., 2001). Nine TGs are present in humans of which 8 are catalytically active (FXIII, TG1-TG7) and one is inactive (erythrocyte membrane protein band 4.2) (Table 1). The primary sequences of the human TGs differ but the active site is conserved, with the exception of band 4.2. Conserved catalytic

triad, consisting of a cysteine, histidine, and aspartate residues and conserved tryptophan residue is critical to TGs enzymatic activity.

These structurally and functionally related enzymes catalyze a variety of Ca^{2+}- and thiol-dependent post-translational modifications like crosslinking, amine incorporation, acylation (all transamidation reactions), esterification and hydrolysis either by deamidation or isopeptide cleavage (isopeptidase activity) (Figure 2) (reviewed in Eckert et al., 2014; Nurminskaya and Belkin, 2012; Lorand and Graham, 2003). Apart from this, TG2, TG4, and TG5 bind and hydrolyze GTP. The most studied and important activity is the crosslinking reaction leading to formation of isopeptide bonds, highly resistant to mechanical stress and proteolytic degradation. But the isopeptidase activity of TGs is still needed to be explored. Functions of TGs include blood coagulation, fibrin clot stabilization, wound healing, bone synthesis (Factor XIII), skin barrier formation, extracellular matrix assembly (TG1 and TG3), formation of cornified envelopes (TG1, TG3, and TG5) prostate function (TG4), apoptosis and extracellular-matrix assembly (TG2) (reviewed in Eckert et al., 2014; Nurminskaya and Belkin, 2012; Lorand and Graham, 2003). The functions of TG6 and TG7 are still unclear. In addition to their structural role, enzymatic and non-enzymatic functions of TGs regulate a plethora of biological processes. TG3 plays a pivotal role in hair fiber morphogenesis by crosslinking trichohyalin and intermediate keratin filaments (reviewed in Eckert et al., 2014; Nurminskaya and Belkin, 2012; Lorand and Graham, 2003). In hypertensive patients, F13a catalyzed crosslinking of AT1 receptor accelerates atherogenesis by promoting monocyte adhesion (reviewed in Eckert et al., 2014). Crosslinking of β_3 integrin to VEGFR-2 by F13a results in VEGFR-2 activation and angiogenesis in the extracellular environment (reviewed in Eckert et al., 2014). Detailed view on TG2 functions are given in the next section.

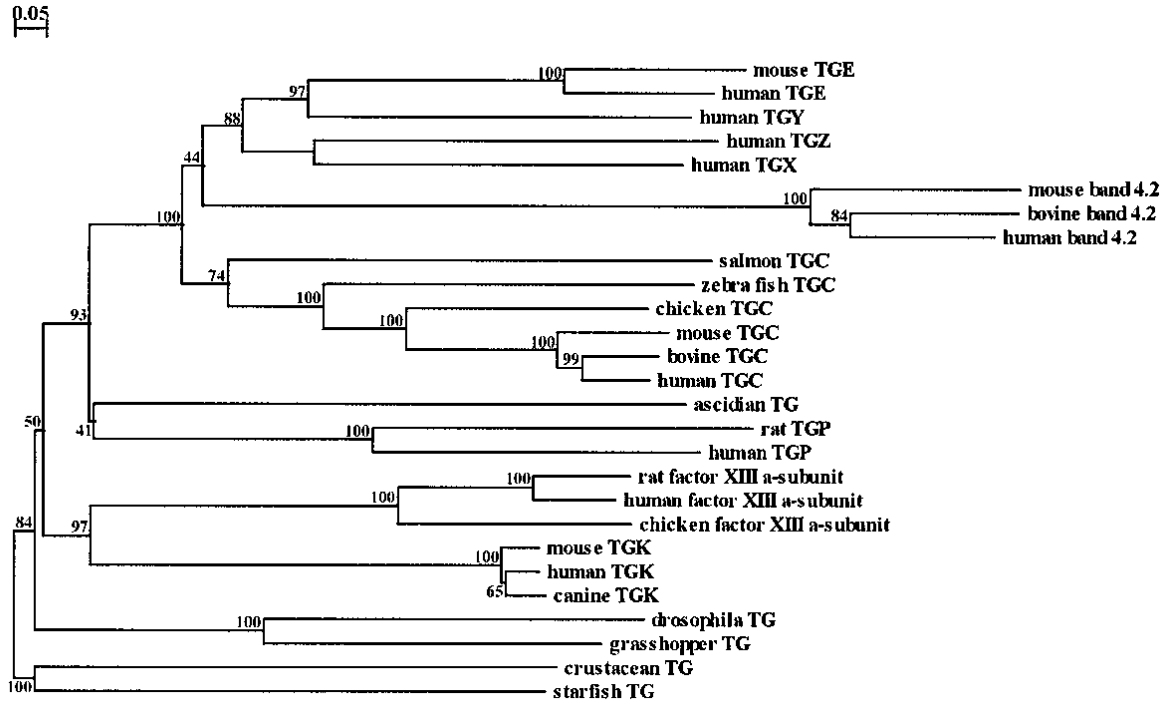
Due to the involvement in many essential processes, human transglutaminase mutations are linked to many deficiencies and disease pathologies (reviewed in Iismaa et al., 2009). TG1 mutations are associated with lamellar ichthyosis, characterized by abnormal cornification of the epidermis (reviewed in Eckert et al., 2014; Iismaa et al., 2009). Missense mutations in TG2 are implicated in early onset Type 2 diabetes (Porzio et al., 2007). TG5 and TG6 inactivating mutations result in skin peeling syndrome and spinocerebellar ataxia 35 respectively (reviewed in Iismaa et al., 2009). Mutations in F13a lead to bleeding syndrome and impaired wound healing (reviewed in Iismaa et al., 2009). When compared to other members of the transglutaminase family, human transglutaminase 2 protein coding sequence has low polymorphism in the population with lowest variability (Király et al., 2013). Knock-out mice studies have indicated a wide role for TG2 and F13a in pathophysiology (reviewed

in Iismaa et al., 2009). TG2 is connected to gluten sensitivity diseases, neurodegeneration, cataract and bone development, and tissue remodeling/repair associated with heart, liver, and kidney disease (reviewed in Eckert et al., 2014; Iismaa et al., 2009). F13a is involved in thrombosis, bone development, tissue remodeling in cancer and wound repair (reviewed in Iismaa et al., 2009).

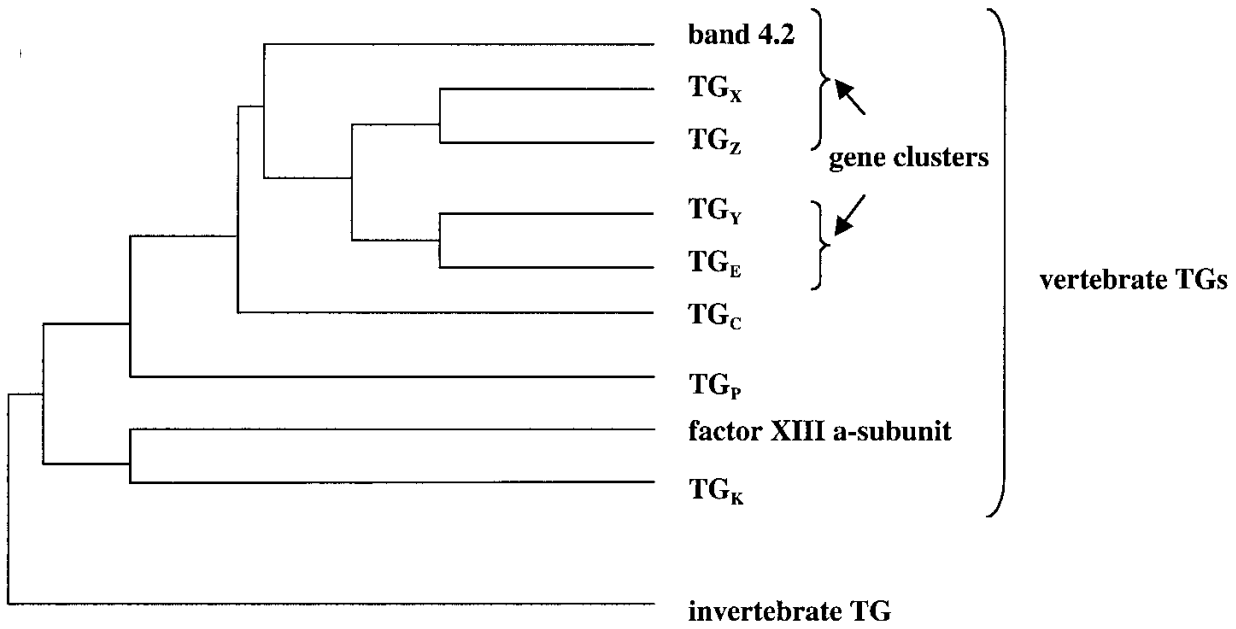
Table 1. Properties of Transglutaminase proteins (taken from Eckert et al., 2014)

Gene	Protein	Chromosomal Location	Molecular Mass, kDa	Main Function	Tissue Distribution	Alternate Names
TGM1	TG1	14q11.2	90	Cell envelope formation during keratinocyte differentiation	Membrane-bound keratinocytes	TG_k, keratinocyte TG, particulate TG
TGM2	TG2	20q11-12	80	Apoptosis, cell adhesion, matrix stabilization, signal transduction	Many tissues: cytosolic, nuclear, membrane, and extracellular	Tissue TG, TG_c, liver TG, endothelial TG, erythrocyte TG, $Gh\alpha$
TGM3	TG3	20q11-12	77	Cell envelope formation during keratinocyte differentiation	Hair follicle, epidermis, brain	TG_E, callus TG, hair follicle TG, bovine snout TG
TGM4	TG4	3q21-22	77	Reproduction, especially in rodents as a result of semen coagulation	Prostate	TG_p, androgen-regulated major secretory protein, vesiculase, dorsal prostate protein 1
TGM5	TG5	15q15.2	81	Cell envelope formation in keratinocytes	Foreskin keratinocytes, epithelial barrier lining, skeletal muscular striatum	TG_x
TGM6	TG6	20q11	78	Not known	Testis and lung	TG_y
TGM7	TG7	15q15.2	81	Not known	Ubiquitous but predominately in testis and lung	TG_z
F13A1	FXIIIa	6q24-25	83	Blood clotting, wound healing, bone synthesis	Platelets, placenta, synovial fluid, chondrocytes, astrocytes, macrophages, osteoclasts and osteoblasts	Fibrin-stabilizing factor, fibrinoligase, plasma TG, Laki-Lorand factor
EPB42	Band4.2	15q15.2	72	Membrane integrity, cell attachment, signal transduction	Erythrocyte membranes, cone marrow, spleen	B4.2, ATP-binding erythrocyte membrane protein band 4.2

A


0.05
mouse TGE
human TGE
human TGY
human TGZ
human TGX
mouse band 4.2
bovine band 4.2
human band 4.2
salmon TGC
zebra fish TGC
chicken TGC
mouse TGC
bovine TGC
human TGC
ascidian TG
rat TGP
human TGP
rat factor XIII a-subunit
human factor XIII a-subunit
chicken factor XIII a-subunit
mouse TGK
human TGK
canine TGK
drosophila TG
grasshopper TG
crustacean TG
starfish TG

B


band 4.2
TGX
TGZ
gene clusters
TGY
TGE
TGC
TGP
factor XIII a-subunit
TGK
vertebrate TGs
invertebrate TG

C

	TGM5 (TG_X)	TGM7 (TG_Z)	EPB42 (band 4.2 protein)	TGM2 (TG_C)	TGM6 (TG_Y)	TGM3 (TG_E)	TGM4 (TG_P)	F13A1 (factor XIII a-subunit)	TGM1 (TG_K)
Chromosomal localization									
human	15q15.2	15q15.2	15q15.2 (27, 28, a)	20q11-12 (14,26, b)	20q11 (14, c)	20q11 (14, c)	3p21-22 (29, 30)	6p24-25 (25)	14q11.2 (18)
mouse	2, 67-69 cM	2, 67-69 cM	2, 67-69 cM (46, a)	2, 89-91 cM (47, a)		2, 74-78 cM			
Gene size									
human	~35kb	~26kb	~20kb (15)	~37kb (14, 26, b)	~45kb	~43kb (14)	~35kb (16)	~160kb (17)	~14kb (12, 18, 19)
mouse			~22kb (62)	~34kb (47)					
Number of exons									
human	13	13	13	13	13	13	13	15	15
mouse			13	13					

(a) this study
(b) GenBank™/EMBL Data Bank with accession number AL031651
(c) GenBank™/EMBL Data Bank with accession number AL031678

Figure 1. Phylogenetic tree of the transglutaminase gene family and genomic organization of the genes in man and mouse (taken from Grenard et al., 2001). Phylogenetic trees for the individual domains and the whole gene was constructed based on amino acid sequence homology and neighbour joining method (Saitou and Nei 1987) (PHYLogeny Inference Package or Treecon 1.3b software packages) (Van de Peer et al, 1994). Alignment of sequences are done with ClustalX and to maximize the homology, sequences were adjusted with hand (nonhomologous N- and C-terminal extensions were excluded). **(A)** Based on the full length sequences a representative tree is shown (Treecon, distance estimation according to Tanja and Nej). Based on the sequence relationship data of the individual gene products **(B)** and on the gene structure and genomic organization **(C)** a hypothetical pedigree for the gene family is shown.

1.4. Transglutaminase 2

Transglutaminase 2 (TG2) is a unique member of the transglutaminase family with several enzymatic, non-enzymatic activities and interacting partners. Human TG2 gene is localized in chromosome 20q11-12 with a 37 kb exon (Gentile et al., 1994). The molecular mass of TG2 protein is 77.3-kD protein, consisting of 686 amino acids and abundant in cytosol but also present in nucleus, mitochondria and recycling endosomes (reviewed in Fesus and Piacentini 2002; Király et al 2011; Gundemir et al., 2012). It is also present on the cell surface and by non-classical mechanisms gets secreted to the extracellular matrix (ECM). Four TG2 spliced isoforms: TG2-S, TG2-H2, TG2V1, and TG2V2 with alternative C- terminus compared to the wild type enzyme have been reported (reviewed in Lai and Greenberg, 2013).

TG2 structure could explain the uniqueness and multifunctionality exhibited by this enzyme. GDP- and GTP-bound crystal forms of human TG2 (PDB IDs: 1KV3 and 4PYG), ATP (PDB ID: 3LY6), and four different irreversible inhibitors bound human TG2 (PDB IDs: 2Q3Z; 3S3J; 3S3P; 3S3S) have been characterized (reviewed in Demény et al., 2015). The crystal structure of TG2 reveals four sequential and structurally distinct domains: N-terminal β-sandwich, catalytic core, and C-terminal β-barrel 1, and β- barrel 2. The β-sandwich houses three fibronectin (FN), and one integrin binding site, and catalytic core has a catalytic triad (Cys277, His335, and Asp358), five non-canonical calcium binding sites, and a conserved tryptophan residue essential for catalytic activity (Figure 4A) (reviewed in Fesus and Piacentini, 2002; Király et al., 2011; Gundemir et al., 2012). β-barrel 1 domain binds and hydrolyzes GTP/ATP and contains a 14-amino acid nuclear localization signal sequence (NLS) (^{466}AEKEETGAMRIRV^{479}) (Shrestha et al., 2015). β-barrel 2 has a leucine-rich nuclear export signal sequence (NES) (^{657}LHMGLHKL^{664}) (Shrestha et al., 2015), and a phospholipase C binding sequence. Cell adhesion and migration function is regulated by the N-terminal domain, and the core, β1, and β2 domains together act as a G-protein to regulate the signal transduction (reviewed in Fesus and Piacentini 2002; Gundemir et al., 2012). The presence of a unique guanine-binding site differentiates TG2 from other transglutaminases. The binding of calcium ion is necessary for the catalytic crosslinking activity of TG2. This provides a conformational diversity with GTP/GDP bound form renders TG2 in a closed conformation (catalytically inactive) and upon Ca^{2+} binding TG2 acquires an open conformation (catalytically active). Five potential Ca^{2+} binding sites were identified by mutagenesis studies (Király et al., 2009) but the crystal structure of Ca^{2+} bound form of TG2 is still unresolved. So under physiological conditions the crosslinking activity of TG2 is tightly modulated by Ca^{2+}, guanine nucleotides, and redox potential. Because of the low calcium concentration inside the living cells, TG2 is predominantly present in GTP/GDP-bound form (crosslinking-inactive) (reviewed in Király et al., 2011). Extracellular TG2 is present in inactive form because of oxidative conditions prevailing in extracellular space.

Because of its ubiquitous presence in various cellular compartments and due to its diverse enzymatic and non-enzymatic functions TG2 has been implicated in versatile biological roles such as cell growth, differentiation, adhesion, migration, apoptosis, angiogenesis, ECM stabilization, wound healing, receptor signaling and signal transduction (reviewed in Fesus and Piacentini 2002; Eckert et al., 2014). In ECM, TG2 interacts with fibronectin, syndecan and at least 3 different β-subunits of the integrin family. Fibronectin is a substrate of TG2 and it mediates the formation of FN matrix fibrils independent of its crosslinking activity (Lorand

et al., 1988; Turner and Lorand 1989). The noncovalent interaction between N-terminal domain of TG2 and fibronectin, stabilizes ECM, crosslinks extracellular matrix proteins, and promotes stable cell adhesion and migration (reviewed in Eckert et al., 2014; Kanchan et al., 2015). Cell adhesion receptor integrin and fibronectin interact by integrin binding motif arginine–glycine–aspartic (RGD) (Takada et al., 2007; reviewed in Kanchan et al., 2015; Corti and Curnis, 2011). Studies have shown that additional binding sites for the FN–integrin interaction is enhanced by TG2 and thus facilitating the cell attachment to the matrix and integrin signaling (reviewed in Eckert et al., 2014; Nurminskaya and Belkin, 2012; Lorand et al., 2003). Another important TG2 binding partner syndecan-4, collaborates with integrins to enhance cell adhesion to fibronectin (Signorini et al., 1988, reviewed in Eckert et al., 2014). Thus syndecan interaction with fibronectin-integrin-TG2 complex promotes wound healing and ECM degradation.

In cytoplasm by crosslinking IκBα, TG2 regulates TGF-β signaling pathway via nuclear transcription factor *NF-$_k$B* signaling mechanism independent of IκB kinase (reviewed in Eckert et al., 2014; Nurminskaya and Belkin, 2012; Gundemir et al., 2012). In cystic fibrosis, cytoplasmic TG2 acts as a central link between inflammation and oxidative stress by crosslinking anti-inflammatory peroxisome proliferator-activated receptor (reviewed in Eckert et al., 2014; Nurminskaya and Belkin, 2012; Gundemir et al., 2012). Cytoplasmic TG2 inhibits autophagy by crosslinking beclin-1 involved in EGF/EGFR signaling in epithelial cancer cells and acts as atypical GTPase and mediator of GPCR-induced signaling (reviewed in Eckert et al., 2014; Nurminskaya and Belkin, 2012; Gundemir et al., 2012). Nuclear TG2 transamidation and protein kinase activities have impact on post-translational modification of histones and gene expression of proteins like SP1 (reviewed in Eckert et al., 2014; Nurminskaya and Belkin, 2012; Gundemir et al., 2012). Depending on the cell type, mitochondrial TG2 either has pro-apoptotic or anti-apoptotic role (reviewed in Eckert et al., 2014; Nurminskaya and Belkin, 2012; Gundemir et al., 2012). By regulating differentiation in phagocytes, monocytes, neutrophils, and T cells, TG2 contributes to immune response and believed to play an important role in cell-mediated immunity (reviewed in Eckert et al., 2014; Nurminskaya and Belkin, 2012; Gundemir et al., 2012). Evidences indicate that by modulating rate of neurite outgrowth TG2 might be involved in neuronal differentiation (Tee et al., 2010).

1.4.1. Importance of the catalytic activities of transglutaminase 2 – the neglected isopeptidase activity

At various cellular locations, distinct biochemical functions are mediated by TG2. Because of its physiological importance diverse biochemical activities are correlated with its cellular functions. The Ca^{2+} dependent transamidation reaction is an acyl-transfer reaction involving the crosslinking of γ-carboxamide group of glutamine residue and either the ε-amino group of a lysine residue or primary amines like histamine or polyamines (Fesus and Piacentini, 2002; Király et al., 2011). TG2 undergoes a conformational change upon Ca^{2+} binding and the cysteine residue at the active site is exposed. The carboxamide of the glutamine substrate is attacked by the thiol group of the cysteine residue releasing ammonia and producing a thioester intermediate or acyl-enzyme complex. Subsequently this complex reacts with primary amines like lysine to form γ-glutamyl-amino-cross-links or stable isopeptide bond between the two substrates (reviewed in Eckert et al., 2014; Nurminskaya and Belkin, 2012; Gundemir et al., 2012). On the other hand, acyl-enzyme complex can be replaced by water (hydrolysis) resulting in site specific deamidation of glutamine to glutamic acid (reviewed in Eckert et al., 2014; Nurminskaya and Belkin, 2012; Gundemir et al., 2012). In the availability of biogenic amines/polyamines as substrates, TG2 also catalyse other type of transamidation reactions like aminylation by incorporating primary amino group and forming γ-glutamyl-amine bonds (reviewed in Lai et al., 2016).

The previously formed Nε-(γ-glutamyl) lysine as well as γ-glutamylamine derivatives are hydrolyzed by the isopeptidase activity of TG2. Endogenous GTPase activity of TG2 is involved in the regulation of transmembrane signaling (Murthy et al., 1999) and it is shown to function as a G-protein. Studies by Mishra and coworkers provided evidence about the intrinsic kinase activity exhibited by TG2 (Mishra and Murphy 2004). Furthermore TG2 can act as a protein disulphide isomerase (Hasegawa et al., 2003), DNA hydrolase (Takeuchi et al., 1998) and a regulator of mitochondrial function (Mastroberardino et al., 2006).

The level of intracellular Ca^{2+} is elevated in certain cell types favoring TG2 mediated transamidation (reviewed in Király et al 2011). In NB4 APL acute promyelocytic leukemia cell line, TG2 is induced when treated with all-trans retinoic acid and its transamidase activity partially contributes to adherent, migratory phagocytic capacity of the neutrophil phenotype (Balajthy et al., 2006; Csomos et al., 2010). Apoptosis is induced by Ca^{2+} in multiple ways (Peitsch et al., 1993). During apoptosis TG2 is activated in liver and other tissues and by forming crosslinked protein polymers it might help in preventing inflammatory and autoimmune reactions by keeping intracellular components inside dying cells (Fesus et al.,

1987; Fesus et al., 1989; Szondy et al., 2003). Moreover TG2 suppresses apoptosis by crosslinking caspase-3 under hypoxic conditions (Jang et al., 2010). Cell type, intracellular localization and death stimuli decide the pro-apoptotic and anti-apototic nature of TG2 and transamidation activity is necessary for its pro-death and nuclear localization functions (Fesus & Szondy 2005; Gundemir and Johnson 2009). Apart from this TG2 transamidation activity is reported to play a role in many other physiological and pathological processes.

The deamidation reaction of TG2 might influence the structure, conformation, and activity of the target protein. In celiac disease TG2 catalyze the deamidation of specific glutamines and generate T cell stimulatory gluten peptides and play a role in pathogenesis of the disease (Shan et al., 2002; Solid and Jabri 2006). The deamidation of glutamine or lysine residues in Hsp20 protein, β-amyloid peptide, βB2 and βB3 crystallins are also catalyzed by TG2 (Boros et al., 2006; Boros et al., 2008; Schmid et al., 2011). In vivo role of protein disulfide isomerase activity of TG2 came to light when TG2 knock out mice displayed defect in the mitochondrial respiratory chain and ATP production (Bernassola et al., 2002). TG2 is able to phosphorylate Insulin-like growth factor-binding protein 3 in breast cancer cells by its protein kinase activity and later it was confirmed using purified TG2 (Mishra and Murphy, 2004).

Till date only few studies focused on isopeptidase activity of TG2 and other transglutaminases. So far no enzyme has been reported to split the proteinase resistant isopeptide bonds in proteins formed by transamidation reaction as discussed above. The proteases degrade the crosslinked proteins by proteasomal system or in lysosomes but the isopeptide bond is not affected (Guilluy et al., 2007). An enzyme γ- glutamylamine cyclotransferase was able to cleave the N^{ε}-(γ-glutamyl) lysine bond in dipeptide state but its activity is significantly confined to kidney and intestine (Raczyński et al., 1975). Factor XIIIa, a member of the transglutaminase family, was reported to reverse crosslinking of α2-plasmin inhibitor to fibrinogen and fibrin potentially regulating the fibrinolytic processes (Ichinose et al., 1982; Mimuro et al., 1986). Parameswaran and coworkers proposed that due to its similarity to papain superfamily, transglutaminases could play a vital role in N^{ε}-(γ-glutamyl) lysine isopeptide bond cleavage (Parameswaran et al., 1997) and confirmed their claim using peptides cross-linked with small primary amines. The human red blood cell transglutaminases and guinea pig TG2 was able to release a cadaverine linked quencher from oligopeptides (Parameswaran et al., 1997; Folk et al., 1967). In other studies TG2 was able to hydrolyze the isopeptide bonds between gliadin-histamine conjugates (Qiao et al., 2005) and cleave the isopeptide bond between small acyl-donor and acyl-acceptor peptides (Stamnaes et al., 2008).

Mucin 2 is a main structural component of the mucus in the gastrointestinal tract. This protein is encoded by MUC2 gene and its secretion in the gut lumen protects the environment by forming insoluble mucous barrier. About 5000 amino acids make the mucin 2 and in later stages of its secretory pathway it forms reduction-insensitive bonds. Recently it was shown that TG2 has affinity to mucin 2 CysD2 domain and can catalyze the crosslinking of this network. The isopeptidase activity of TG2 was utilized to show the cleavage of the reduction-insensitive bonds (Recktenwald and Hansson, 2016). This raised the possibility that isopeptidase activity of TG2 could also play an important role in regulation of biological processes. Deubiquitinases reverse the isopeptide bond formed between ubiquitin and proteins and regulate fates of many proteins. Like deubiquitinases, the isopeptidase activity of TG2 might also impact critical cellular functions, but due to the lack of a protein based methods for its detection and extensive research, potential applications of TG2 isopeptidase activity remains unexplored.

Recently to measure the transamidase activity of TG2 a real-time fluorescence polarization assay was published (Kenniston et al., 2013). It is based on increase in anisotropy of the enzymatically crosslinked product formed after TG2 mediated crosslinking of fluorescently labelled TG2 specific dodecapeptide (FLpepT26) into bovine serum albumin (BSA). We hypothesize that with an appropriate lysine donor substrate, after the crosslinking reaction, isopeptidase activity of TG2 (cleavage of the isopeptide bond) can be monitered using the same feature. We suppose it can be attained by following the decrease in anisotropy, similar to assays measuring protease activity, for example cleavage of isopeptide bond between ubiquitin and its substrate protein by deubiquitinating enzymes (Bolger et al., 1994), (Reyes-Turcu et al., 2009). Besides, based on the previous data we hypothesize that TG2 might have a still unknown regulatory mechanism which might activate the transamidase activity independent of isopeptidase activity and vice versa. Disconnecting transamidase and isopeptidase activity may have particular use in tissue fibrosis affecting different organs (Johnson et al., 2007) or in neurodegenerative diseases were accumulation of crosslinked proteins in intraneuronal inclusions with increased intracellular TG2 activity was observed (Martin et al., 2013).

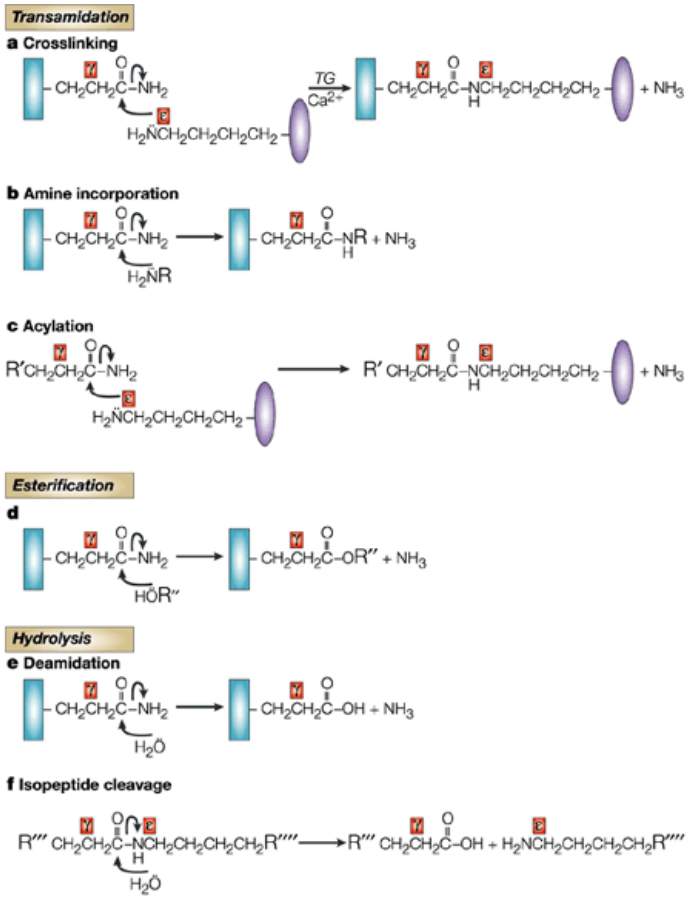


Figure 2. Biochemical activities of TG2 (taken from Lorand and Graham, 2003). Crosslinking reaction catalyzed by TG2. (a) Inter- or intra-molecular crosslinking (b) amine incorporation (c) acylation. (d) esterfication reaction. Hydrolysis reaction catalyzed by TG2 (e) deamidation (f) isopeptidase activity

1.4.2. Transglutaminase 2 in disease; its role in neurodegenerative disorders

Progressive neuronal loss, formation of insoluble aggregates and the aggregation of disease specific pathogenic proteins are characteristic features of neurodegenerative disorders. TG2 is highly expressed in human central nervous system and TG2 transamidating activity plays a role in covalently crosslinking pathogenic proteins leading to formation of protein polymers (Mastroberardino & Piacentini, 2010). TG2 is implicated in neurodegenerative disorders like Alzheimer's, Huntington's, Parkinson's and amyotrophic lateral sclerosis as well as in nervous system injuries (reviewed in Grosso et al., 2012). Several factors like energy depletion, loss of calcium homeostasis, low GTP levels, role of autophagy and increase in reactive oxygen species connect TG2 to neurodegenerative disorders (reviewed in Eckert et al., 2014). Cross-linking of Alzheimer's disease-related proteins, tau and Aβ amyloid by TG2 is demonstarted by in vivo and cell culture studies (reviewed in Eckert et al., 2014). TG2

plays a pathogenic role in both Parkinson's disease and Dementia with Lewy Bodies based on several lines of evidence (Nemes et al., 2009). Some studies indicate that TG2 is also involved in the pathophysiology of Progressive Supranuclear Palsy (Zemaitaitis et al., 2003) and in spinocerebellar ataxia-1 (D'Souza et al., 2006).

TG2 involvement in Huntington's disease (HD) is extensively investigated (reviewed in Eckert et al., 2014; Mastroberardino & Piacentini, 2010). HD is an age related neurodegenerative disorder caused by a poly-glutamine repeat expansion in huntingtin protein (HTT) resulting in formation of protein aggregates (neuronal intranuclear inclusions) (NII) (reviewed in Ross and Tabrizi 2011). HD is one of the trinucleotide repeat disorders caused by expansion of DNA base CAG coding for amino acid glutamine and disease is manifested when the repeat number is > 35. HD is characterized by loss of neurons in striatum, cortex and aggregation of expanded polyglutamine tract in the nuclei of vulnerable neurons (reviewed in Ross and Tabrizi, 2011).

The activity, expression and amounts of TG2 is increased in HD and the conditions favouring the activation of this enzyme like high Ca^{2+} concentration and low GTP levels have also been observed. Several studies have shown that polyglutamine repeats in HTT are excellent substrates for TG2 crosslinking (reviewed in Eckert et al., 2014; Mastroberardino & Piacentini, 2010). This resulted in a hypothesis that TG2 has a pathogenic role in HD (reviewed in Eckert et al., 2014; Mastroberardino & Piacentini, 2010). Purified TG2 incubated with polypeptides containing expanded polyglutamine repeats formed insoluble aggregates containing epsilon-gamma glutamyl lysine (GGEL) bonds characteristic of TG2 (Kahlem et al., 1996). HTT extracted from HD patients brain incubated with TG2 lead to decrease in load of monomeric protein but high molecular weight aggregates were detectable and this is probably due to the conversion of monomers into cross-linked polyproteins by TG2 (Kahlem et al., 1998). Similarly, in postmortem HD brain tissue, TG2 was shown to crosslink intranuclear inclusions (Zainelli et al., 2003 and 2005). In contrast, one study argued that TG2 only selectively modifies mutant huntingtin (mHTT) associated proteins and it does not contribute to the mHTT aggregates formation (Chun et al., 2001). Invivo mouse studies showed neuronal intranuclear inclusions enhanced in TG2 absence. TG2 knock-out crossed with HD R6/1 transgenic mice showed reduced cell death, improved motor performance and survival compared to R6/1 transgenic mice (Mastroberardino et al., 2002).

Misfolded proteins are either cleared by the ubiquitin-proteasome or by autophagy lysosome system and impairment of both the pathways are reported in many neurodegenerative disorders (reviewed in Cortes et al., 2014; Ortega and Lucas, 2014). TG2 is involved in

autophagosome maturation and in SH-SY5Y cells it is shown to crosslink beclin-1, a mammalian ortholog of the yeast autophagy-related gene 6 (D'Eletto et al., 2009; Verhaar et al., 2013). TG2 is also involved in the clearance of ubiquitinated proteins by interacting with autophagy cargo protein nucleoporin 62 (p62) (D'Eletto et al, 2012). Taken together, the role of TG2 in HD is still not clear with contradictory results from numerous studies. Insitu studies have concentrated only on TG2 transamidation activity using wild type and active site mutant TG2. The role of isopeptidase activity in the clearance of protein aggregates and the reversal of crosslinked proteins in intraneuronal inclusions in pathological conditions remains to be elucidated.

2. AIM OF THE STUDY

1) By using comparative genomics and protein structural analysis to study the impact of evolutionary changes in TG2 on its structure and function.

2) To screen the large-scale public databases for transglutaminase genomic variants and to compare TG2 variants with those appearing in other members of the family

3) To study the impact of TG2 non-synonymous SNVs on its structure, function and biochemical activities

4) To separate the transamidase and isopeptidase activities of TG2 by producing mutants and to study the possible role of both these activities on cell aggresome system

5) To develop a real-time kinetic, protein based method for monitoring the isopeptidase activity of TG2.

3. MATERIALS and METHODS

3.1. Materials

3.1.1. Cell culture media

Dulbecco's Modified Eagle Medium - high glucose (D5671) – (Sigma, St Louis, MO, United States)

Fetal bovine serum (FBS) – (Gibco, Waltham, MA, United States)

Heat inactivated horse serum – (Gibco, Waltham, MA, United States)

3.1.2. Antibodies/Inhibitors

GFP antibody (Biolegend, San Diego, California, United States)

MAB 2166 huntingtin (Millipore, Billerica, Massachusetts, United States)

MAB 1574 1C2 (Millipore, Billerica, Massachusetts, United States)

TG100 (Santa Cruz Biotechnology, California, United States)

TG2 CUB7402 (Neo markers, Fremont, California)

p62/SQSTM1 (Novus biologicals, Littleton, Colorado, United States)

LC3 (Novus biologicals, Littleton, Colorado, United States)

Atg 5 (Cell signaling, Danvers, Massachusetts)

Actin (Sigma, St Louis, MO, United States)

Tubulin (Sigma, St Louis, MO, United States)

Anti-mouse (Advansta, Menlo Park, California, United States)

Anti-rabbit (Advansta, Menlo Park, California, United States)

Anti-rat (Advansta, Menlo Park, California, United States)

MG132 (Sigma, St Louis, MO, United States)

Z-DON (Darmstadt, Germany)

TheFLpepT26 peptide was obtained as published in (Biri et al., 2016). All other materials were purchased from Sigma (St Louis, MO, USA) unless otherwise indicated.

3.2. Methods

3.2.1. Databases

UniProt database sequences (UniProt Consortium, 2015) served as the source for the comparative genomics study (http://www.uniprot.org/). The UniProt ID for the sequences used in the study: TGM2: *human* (P21980), *mouse* (P21981), *rat* (Q6P6R6), *guinea pig* (P08587), *chimpanzee* (*Pan troglodytes*) (K7D7G1), *rhesus macaque* (*Macaca mulatta*)

(I2CVT6), *gorilla* (*Western lowland gorilla*) (G3RCE0), *sumatran orangutan* (*Pongo abelii*) (H2P1W1), *northern white-cheeked gibbon* (*Nomascus leucogenys*) (G1R3A6), *white-tufted-ear marmoset* (*Callithrix jacchus*) (U3BQR1), *green monkey* (*Chlorocebus sabaeus*) (A0A0D9RMQ2), *olive baboon* (*Papio anubis*) (A0A096NXG3), *horse* (F7DA17), *dog* (F1Q435), *domestic pig* (A0A0B8RZZ6), and *zebrafish* (Q66L63). UniProt IDs for other human/mouse transglutaminases: F13a (P00488/Q8BH61), TGM1 (P22735/Q9JLF6), TGM3 (Q08188/Q08189), TGM4 (P49221/Q8BZH1), TGM5 (O43548/Q9D7I9), TGM6 (O95932/Q8BM11), and TGM7 (Q96PF1/A2ART8).

Eukaryotic Linear Motif (ELM) resource was used to identify short linear motifs (SLiMs) (Dinkel et al., 2014), (http://elm.eu.org/search/). SLiMs were predicted for wild-type human TG2 and then for modified human TG2, after replacing each human amino acid cluster individually with the corresponding mouse sequence. For posttranslational modification predictions same procedure was followed. NetPhos2.0 (Blom et al., 1999) (http://www.cbs.dtu.dk/services/NetPhos/) was applied for phosphorylation sites and NetNGlyc 1.0 (Gupta et al., 2002) was used for N-glycosylation sites prediction (http://www.cbs.dtu.dk/services/NetNGlyc/).

Exome Aggregation Consortium browser (beta version) was used for single nucleotide variation data (Lek et al., 2016; Exome aggregation consortium (ExAC), Cambridge, MA (URL: http://exac.broadinstitute.org/) accessed August 2016). Exome sequencing data from a wide variety of large-scale sequencing projects are collated by group of investigators to make data available for wider scientific community. About 17 projects contribute to the data and the data set provided in the ExAC spans 60,706 unrelated individuals sequenced as part of various disease-specific and population genetic studies.

3.2.2. Bioinformatics tools

Secondary structures of wild type and mutant human and mouse TG2 proteins were predicted by GORIV method (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_gor4.html; Accessed 15 March 2016). FoldX program was applied for stability analyses using default parameters of the program (Guerois et al., 2002). Using crystal structures of opened (PDB ID: 2Q3Z) and closed (PDB ID: 1KV3) conformational human TG2, and homology models of the mouse TG2, changes in the overall stability of the proteins [ΔΔG [kcal/mol] upon mutations were calculated. Modeller 9v13 software was used to perform homology modelling (Sali and Blundell, 1993) using crystal structures of human TG2 (PDB ID: 1KV3) and human cellular coagulation Factor XIII (PDB

ID: 1F13) proteins as templates in the case of the closed conformational mouse TG2. While the template for the opened conformation was based on crystal structure of human TG2 (PDB ID: 2Q3Z). To prepare the cluster mutant proteins, the human and mouse TG2 clusters were replaced with the corresponding mouse and human sequences, respectively. To prevent crystallization causing unfavorable geometries the PDB structure was subjected to energy minimization and mutations were built into the optimized structures to carry out stability calculations. Stability analysis was done by Dr. Mónika Fuxreiter and Mr. Viktor Attila Ambrus, MTA‑DE Momentum Laboratory of Protein Dynamics, Department of Biochemistry and Molecular Biology, Faculty of Medicine, University of Debrecen, Debrecen, Hungary. Bioinformatics predictors: PONDR VSL1 (Romero et al., 2001) and IUPred (Dosztányi et al., 2005) (Kanchan et al., 2015) were used to predict intrinsically disordered regions (IDRs).

Polymorphism Phenotyping (PolyPhen) and Sorting Intolerant From Tolerant (SIFT) analyses were applied in ExAc database to predict the possible impact of single amino acid substitutions on the structure and function of a human protein. A PANTHER (Protein ANalysis THrough Evolutionary Relationships) prediction was carried out to determine the damaging nsSNVs (Thomas et al., 2003) if more than one change is present at an amino acid residue and both having same scores. The deleterious, possibly or probably damaging nsSNVs are mentioned as damaging in the text. PANTHER analysis was carried out by Dr. János András Mótyán, Laboratory of Retroviral Biochemistry, Department of Biochemistry and Molecular Biology, Faculty of Medicine, University of Debrecen, Debrecen, Hungary.

3.2.3. Homology modelling of TG2

Homology models were prepared by SWISS-MODEL (Arnold et al., 2006) using crystal structures of either FXIIIa (PDB ID: 4KTY) (Stieler et al., 2013) or TG3 (PDB ID: 1L9M) (Ahvazi et al., 2002) as templates to represent an one and a three Ca^{2+}-bound states of TG2. The models were then repaired for energy minimization using the FoldX forcefield. The R222 residue was mutated to glutamine using FoldX and the structures were visualized with the help of the FoldX plugin in YASARA (Krieger et al., 2002; Schymkowitz et al., 2005). Figures were prepared using PyMol Molecular Graphics System (version 1.8 Schrödinger LLC). Homology modelling was done by Dr. Máté A. Demény, MTA-DE Stem cell, Apoptosis and Genomics Research Group of Hungarian Academy of Sciences, Faculty of Medicine, University of Debrecen and Dr. János András Mótyán, Laboratory of Retroviral

Biochemistry, Department of Biochemistry and Molecular Biology, Faculty of Medicine, University of Debrecen, Debrecen, Hungary.

3.2.4. Expression and purification of proteins

Most of the studies in the transglutaminase field for more than two decades used the cDNA of TG2 containing Gly at amino acid position 224 (Gentile et al., 1991). But so far sequenced human TG2 alleles have Val at amino acid position 224 based on the large genomic databases (Király et al., 2013). Both the Gly and Val containing wild-type TG2 displayed significant difference in their calcium binding affinity and sensitivity in *in vitro* and *in situ* transglutaminase activity assays (Kanchan et al., 2013). Based on this data Gly was changed to Val at 224 position and we used Val 224 containing wildtype in our TG2 constructs.

QuikChange Site-Directed Mutagenesis Kit II Manual (Stratagene, La Jolla, CA, USA) was used to construct TG2 mutants and variants and were checked by restriction analysis and DNA sequencing (Capillary sequencing runs were performed by Genomic Medicine and Bioinformatics Core Facility at University of Debrecen). The Val224 containing recombinant human TG2 (Uniprot code: P21980) mutants and the homozygous variants were expressed in N-terminally (His)6-tagged form (pET-30 Ek/LIC-TG2; Mw: 82,745 Da) and purified by Ni-NTA affinity chromatography as described previously (Király et al., 2013). The expression level, yield and stability of the mutant proteins and wild type TG2 were comparable. The expression and purification protocol was refined and the following improved protocol was applied after introducing the Val224 TG2: Rosetta 2 strains were grown in LB broth at 25°C to an OD600 of 0.6 to 0.8 after transformation. The cultures were grown overnight at 16°C in the presence of 0.05 mmol/L isopropyl β-D-thiogalactoside to induce the expression of His-tagged proteins. The cells were harvested by centrifugation at 4°C and then resuspended in 25 ml ice cold binding buffer (50 mM Tris-HCl, pH 7.2, 1 mM EDTA, 5 mM imidazole, 1 mM PMSF). Cells were lysed by sonication (6 times 30 seconds with 1 minute breaks, on ice), and centrifuged (20000g, 30 minutes, 4°C) supernatant containing His-tagged protein was loaded onto an equilibrated (20 ml binding buffer without dithiothreitol (DTT) and then 10 ml binding buffer) His GraviTrap column (GE Healthcare, Little Chalfont, UK). After loading, the column was washed with 10 ml binding buffer, then again with 10 ml of 20 mM imidazole containing binding buffer. Finally, 10 ml of 500 mM imidazole containing binding buffer was used to eluate the protein and the eluate was concentrated using Amicon Centricon-YM 50 MW (Millipore, Billerica, MA). Then the buffer was exchanged to 20 mM Tris-HCl, pH 7.2, with 150 mM NaCl, 1 mM DTT, 1 mM EDTA, and 10% by volume glycerol. Bradford

method was used to determine the protein concentration (Bio-Rad, München, Germany). Finally, Coomassie brilliant blue staining of SDS-PAGE and Western blots were used to check the purity of the protein.

Rosetta 2 competent cells (Novagen, Darmstadt, Germany) were used to express N-terminal GST-tagged S100A4 (pETARA-S100A4; Uniprot code: P26447; Mw: 39,559 Da). The LB medium containing 50 μg/mL ampicillin and 34 μg/mL chloramphenicol was inoculated with 20:1 ratio of overnight culture and was grown at 25°C until the OD reached 0.6–0.8 at 600 nm. Isopropyl β-D-thiogalactoside (0.1 mM) was added to induce the expression for overnight at 18°C. Cells were harvested and Buffer A (20 mM Tris-HCl pH 7.2, 150 mM NaCl, 1 mM EDTA, 1 mM DTT) containing 1% Triton X-100, 10% glycerol, 1 mM PMSF and protease inhibitor cocktail was added to dissolve the pellets. Cells were lysed by sonication and after centrifugation at 20000g, 4°C, 25 min the separated supernatant was loaded onto Glutathione Sepharose 4B resin (GE Healthcare, UK). Buffer A was used to wash the S100A4 (GST) protein and then elution was done using 10 mM reduced glutathione in Tris-HCl pH 8.0 and dialyzed overnight in buffer A. Protein concentration was determined by Bradford method (Bio- Rad Protein Assay, Bio-Rad, München, Germany).

3.2.5. Large scale production of the crosslinked FLpepT26-S100A4 (GST)

In the presence of 5 mM Ca^{2+}, mixture of 5 μM FLpepT26, 12.8 μM S100A4 (GST) and 5 nM TG2 was incubated for 1 hour in the reaction buffer 20 mM Tris-HCl pH 7.5 containing 150 mM NaCl, 5 mM DTT and 0.01% Tween 20. EDTA (10 mM final concentration) was added to stop the reaction and to prevent unwanted modification of the crosslinked molecules during their separation. Centrifugal concentrator filter (Amicon ultra, 10kDa, Millipore, Billerica, MA, USA) was used to purify FLpepT26-S100A4 (GST) with unmodified S100A4 (GST) from the free, unbound FLpepT26 peptide. As isopeptidase activity prefers slightly acidic pH (Fleckenstein et al., 2002) the buffer was replaced by 20 mM MOPS buffer, pH 6.8 containing 0.5 mM EDTA, 150 mM NaCl, 5 mM DTT, 0.01% Tween 20. The ratio of co-purified FLpepT26-S100A4 (GST) and S100A4 (GST) was calculated based on the total protein concentration (determined by Bio-Rad Protein Assay) and its fluorescein content (absorption at 493 nm) using 79600 M-1 cm-1 as molar extinction coefficient for fluorescein. In optimized conditions the FLpepT26-S100A4 (GST) content was approximately 15 % as an average in the reaction product meaning that 5 μg purified mixture of FLpepT26-S100A4 (GST) and S100A4 (GST) corresponds to 0.5 μM FLpepT26-S100A4 (GST) in 35 μl of the isopeptidase assay.

3.2.6. Activity assays

Transamidase assay

For kinetic amine incorporation assay the reaction mixture contained 50 mM Tris-HCl buffer pH 7.5, 0.5 mM dansyl-cadaverine, 2 mg/ml N,N' dimethylated casein, 3 mM DTT, and 3 mM $CaCl_2$ or 10 mM EDTA. The assay was started with the addition of 20 µl enzyme (100 nM TG2 final concentration) and reaction volume was 100 µl. The linear part of the increasing fluorescence was used to calculate the reaction rate in terms of RFU per minutes (Ex/Em: 360/490 nm; at 37°C).

Crosslinking activity was measured using a real time fluorescence anisotropy assay (Biri et al., 2016). TG2 mediated crosslinking of FL-pepT26 into S100A4 (GST) was monitored by following the increase in the fluorescence anisotropy. The reaction volume was 35 µl and performed at 37ºC for 30 mins with 100 nM FLpepT26, 12.8 µM S100A4 (GST), 10 nM TG2, 3 mM $CaCl_2$ (5 mM EDTA was used as negative control). The reaction buffer contained 20 mM Tris–HCl, pH 7.5, 150 mM NaCl, 5 mM DTT, and 0.01% Tween 20. 384-well untreated Polystyrene Black Microplates (Nunc, Thermo Scientific, Denmark, cat. no. 262260) and Synergy H1 microplate reader (GreenFP filter cube, excitation 485 nm, emission 528 nm; BioTek, Winooski, VT, USA) was used. The initial slopes of the kinetic curves were used to calculate the reaction rates in terms of anisotropy per minute.

End point transamidase activity assay based on labelled glutamine donor peptide incorporation on microtiter plate

Another microtiter plate assay, the commercial CovTest (CovalAb, France) was also used to measure the transamidase activity of TG2. The assay is based on the incorporation of a short biotinylated peptide (biotin-pepT26, acyl-donor) into spermine (acyl-acceptor) which was covalently coupled to the surface of the activated plate. We performed the assay based on the published protocol (Perez et al., 2009). The plates (CovalAb) were washed with solution containing 0.1 % Tween-20 in 0.1M Tris-HCl, pH 8.5 at 37°C for 15 minutes. Then 50 µl of cold reaction mixture containing 20 mM biotin-pepT26 oligopeptide (CovalAb), 30 mM $CaCl_2$, 10 mM DTT in reaction buffer (150 mM NaCl, 40 mM Tris-HCl, pH 8.3) were loaded into the well. The reaction was started by addition of 60 µl sample (0.5 µg protein) and was incubated for 25 minutes at 37°C. After 3 washing steps, ExtrAvidin peroxidase solution (dilution: 1/5000 in washing buffer containing 1 % BSA), 110 µl/well was added and incubated at 37°C for 15 minutes. After three washes, 100 µl 3,3',5,5'-Tetramethylbenzidine (TMB) substrate (freshly prepared from TMB tablet using 200 µl DMSO to resolve the tablet

and then it was completed by 9.8 ml phosphate-citrate buffer, pH 5.0 containing 10 µl 30 % H_2O_2) was added in each well. Around 50 µl 2.5N H_2SO_4 was used to stop the colour reaction and absorbances were read at 450 nm.

Fluorescence isopeptidase assay

The Zedira assay is based on the cleavage of isopeptide bond by transglutaminase in the synthesized substrate releasing the dark quencher (2, 4-dinitrophenyl) linked to the cadaverine spacer followed by the increase of fluorescence from the N-terminally attached fluorophore 2-aminobenzoyl (2-Abz). The reaction mixture contained 50 mM MOPS buffer, pH 6.8, containing 5 mM $CaCl_2$, 100 mM NaCl, 0.1 % (w/v) PEG8000, 100 nM TG2, 50 µM A102 isopeptidase assay substrate (Zedira, Darmstadt, Germany) and 2.8 mM DTT (added with the starting solution which contained the enzyme). The reaction was monitored at 37°C by a BIOTEK Synergy H1 microplate reader (Ex/Em: 318/413 nm) and the initial slopes of the kinetic curves were used to calculate the activities.

Real-time fluorescence method using novel crosslinked protein–peptide substrate was used to measure isopeptidase activity. The TG2 mediated cleavage of the isopeptide bond on FLpepT26–S100A4 (GST) substrate was followed by decrease in fluorescence anisotropy. The reaction buffer contained 20 mM Mops (pH 6.8), 150 mM NaCl, 6 mM glycine methyl ester, 5 mM DTT, and 0.1% Tween 20 with 0.5 µM FLpepT26–S100A4 (GST), 300 nM TG2, 5 mM $CaCl_2$ or EDTA. The volume of the reaction was 35 µl and performed for 60 mins at 37^{o}C in 384-well untreated Polystyrene Black Microplates (Nunc, Thermo Scientific, Denmark, cat. no. 262260) and measured using Synergy H1 microplate reader (GreenFP filter cube, excitation 485 nm, emission 528 nm; BioTek, Winooski, VT, USA). Initial slopes of the kinetic curves were used to calculate the reaction rates in terms of anisotropy per minute.

BODIPY-FL-GTPγS nucleotide binding assay

To compare nucleotide binding 500 nM BODIPY-FL-GTPγS GTP analog (Invitrogen, Carlsbad, CA, United States) was used with increasing amounts of TG2 variants in the presence of 20 mM HEPES pH 7.5, 150 mM NaCl, 0.1 TCEP, 0.05 % Tween-20, 0.1 mM EGTA, and 1 mM $MgCl_2$. When BODIPY FL GTPγS binds to GTP-binding proteins there is an increase in fluorescence providing a non-radioactive alternative tool to analyse protein-nucleotide interactions.

Fibronectin binding assay

The fibronectin-binding property of the variants was tested using a published direct ELISA assay (Király et al., 2013) with modifications. At first fibronectin was coated to the plates and washed off. Then 0.3 μg TG2 variant was incubated in the wells for 1 h at room temperature in Tris-buffered saline (TTBS) containing 2.5 mM $CaCl_2$. The bound TG2 variants were detected using TG100 (1:5,000) or TG2 CUB7402 (1:5,000) monoclonal antibodies and then anti-mouse IgG/ HRP (1:7,500) and TMB substrate at 450 nm.

3.2.7. SDS-PAGE analysis of isopeptidase assay products

The isopeptidase assay reaction was stopped by adding 6x denaturation buffer (375 mM Tris-HCl, pH 6.8, 600 mM DTT, 12% (m/v) SDS, 60% (v/v) glycerol, 0.06% (m/v) bromophenol blue) and the samples were boiled for 10 min. SDS-PAGE was performed using 15% Tris-Glycine gel. The fluorescence was detected immediately by an UV gel documentation system (Protein Simple, AlphaImager, HP system).

3.2.8. Mass spectrometric analysis of the peptide after isopeptide cleavage

Bruker Daltonics Esquire 3000plus (Bremen, Germany) ion trap mass spectrometer with on-line HPLC coupling was used to perform electrospray ionization mass spectrometric measurements. Jasco PU-2085Plus HPLC system using a Supelco Ascentis C18 column (2.1 x 150 mm, 3 μm) was used to perfrom HPLC separation. Linear gradient elution (0 min 2% B; 3 min 2% B; 27 min 60% B) with eluent A (0.1% HCOOH in water) and eluent B (0.1% HCOOH in acetonitrile-water (80:20, V/V)) was used at a flow rate of 0.2 ml/min at ambient temperature. The HPLC was directly coupled to the mass spectrometer. Collision induced dissociation (CID) experiments were used for peptide sequencing. Mass spectrometric analysis was carried out by Dr. Gitta Schlosser, MTA-ELTE Research Group of Peptide Chemistry, Hungarian Academy of Sciences, Eötvös Loránd University, Budapest, Hungary.

3.2.9. Data analysis

Data analysis, curve fitting, kinetic calculations were performed by GraphPad Prism 7 software (Graphpad Software Inc. La Jolla, CA, USA) using the appropriate incorporated equations and tools mentioned where it is appropriate. In case of Ca^{2+}-dependence experiments the free calcium-ion concentrations were calculated using the online version of MaxChelator (WEBMAXC STANDARD; http://www.stanford.edu/~cpatton/maxc.html) due to the EDTA content of the substrate used for the isopeptidase reaction

3.2.10. Production of Lenti-viral TG2 constructs

To express wild-type and mutant transglutaminase 2 proteins in cell systems, lentiviral construct, pLenti CMV blast TG2 vector was used. pLenti CMV blast TG2 lentiviral vectors were produced by recombining the desired transglutaminase mutants containing entry (pENTR 1A) and destination vectors (pLentiCMV Blast DEST, pLentiPGK Blast DEST). (Figure 3). The resulting vectors were transformed into E. coli Stbl3 competent cells after GATEWAY cloning (Invitrogen, Carlsbad, CA, United States).

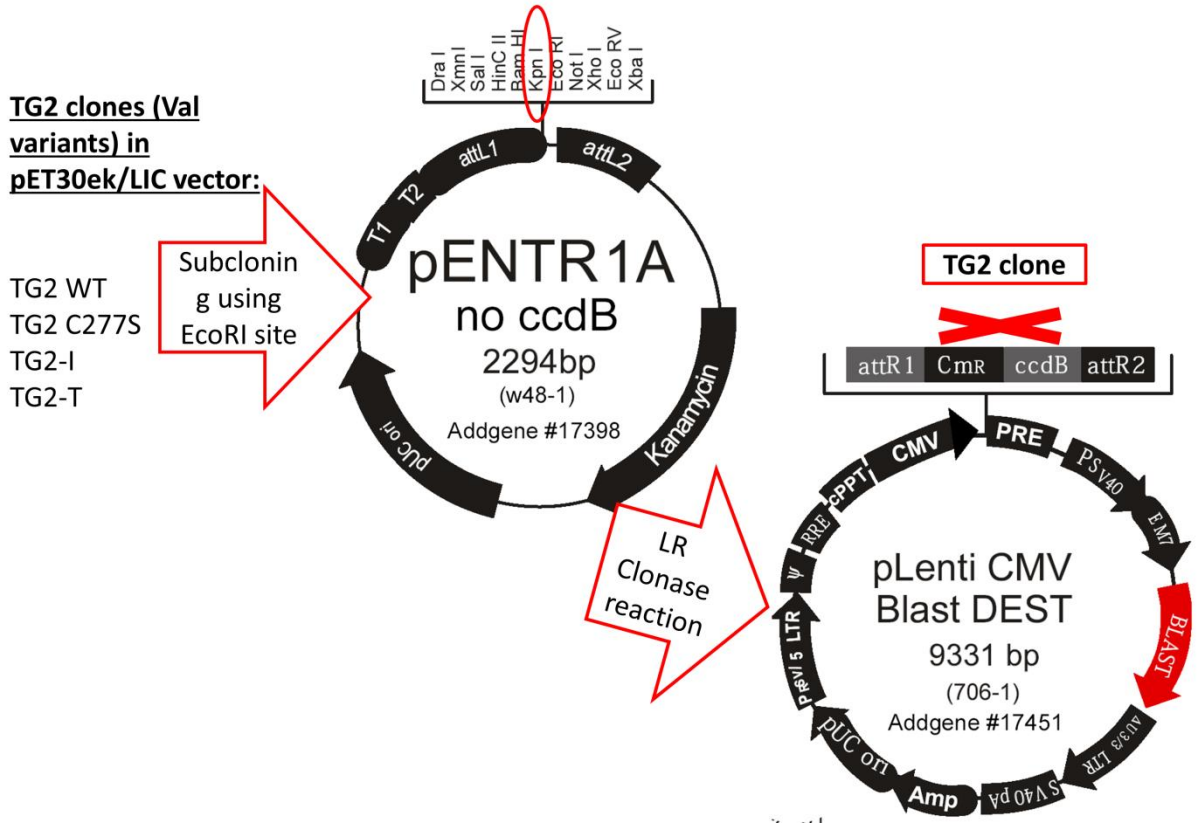


Figure 3. The designed lentiviral vector system WT – wild type; C277S – active site mutant; TG2 I - Isopeptidase active, Transamidase deficient mutant (W332F); TG2 T - Transamidase active, Isopeptidase deficient mutant (W278F).

3.2.11. Mammalian cell culture and transfection

Lentiviral Gateway cloning expression system (Invitrogen, Carlsbad, CA, United States) was used to produce lentiviral plasmid constructs for transfection. 293 FT cells used for the lentivirus production were cultured in tissue culture flask 75 (T75, 75 cm^2, TPP) with DMEM supplemented with 10 % FBS, 100 U/ml penicillin/streptomycin, 5 mM L-glutamine, at 37^oC, 5 % carbon dioxide. Calcium phosphate method of transfection was followed. Two hours

before transfection, fresh medium was replaced. In a 50 ml falcon tube 22.4 µg of vector plasmid (with gene of interest), 7.46 µg of envelop plasmid (pMD2G), 14.5 µg of packaging plasmid (psPAX2) was mixed with Tris-EDTA buffer, water, 2.5 M calcium chloride and mixed briefly. Then Hank's balanced salt solution (2x) was added to this mixture dropwise under agitation by vortexing. After 5 mins at room temperature the mixture was added dropwise on the flask and mixed gently with the culture medium and put at 37ºC, 5 % carbon dioxide. After 14-16 hrs post transfection medium was changed and fresh medium was added. After 24 hrs, supernatant was collected, centrifuged for 5 mins at 1500 rpm and filtered using 0.45 µM filter. The viral supernatants were stored at -80ºC in 1.5 ml aliquots for further use (protocolshttp://tronolab.epfl.ch/webdav/site/tronolab/shared/protocols/LV_production.pdf, tronolab protocols).

PC12 cells (pheochromocytoma) with tetracycline inducible promotor expressing GFP-tagged exon-1 fragment of huntingtin with 23 or 74 glutamine repeats (Q74) (Wyttenbach et al., 2001) were grown in DMEM supplemented with 10 % heat-inactivated horse serum (Gibco), 5 % fetal bovine serum, 100 U/ml penicillin/streptomycin, 5 mM L-glutamine, 5 mM sodium pyruvate and Non-essential aminoacid (5x) at 37ºC, 5 % carbon dioxide. For transduction, cells were grown in T25 (25 cm^2, TPP) with 70 % confluency and viral supernatants were added in 1:1 ratio and left for 24 hrs. After 24 hrs medium was changed and TG2 wildtype or mutants were selected using 3 µg blasticidin (invitrogen) for 10 days.

For treatments, cells were seeded in T75 flasks and were induced with 2 or 4 µg/ml tetracycline for indicated time intervals. Final concentration of autophagy inhibitor, NH_4Cl, 20 mM, proteasome inhibitor, MG132, 5 µM were used for indicated time intervals.

3.2.12. Cell lysate collection and Western blot analysis

The cells were washed twice with ice cold PBS, centrifuged for 10 mins at 1400 rpm. The pellets were lysed in RIPA lysis buffer (Radioimmunoprecipitation assay buffer): 50 mM Tris pH 7.5, 150 mM NACl, 1 mM DTT, 1 mM EDTA, 0.5 % Nonidet P-40, 10 % glycerol, protease inhibitor cocktail (10 µl for 1 ml buffer) and 1 mM PMSF were freshly added. The cell were rotated for 1 hr at 4ºC and sonicated twice on ice. Proteins were quantified by the Bradford assay along with BSA as a standard and denatured for 10 mins at 99ºC. Equal amounts of protein was subjected to SDS-PAGE (7.5 % - 12 % Tris-Glycine gel). The gel was transferred to PVDF membrane using wet blot system (Bio-Rad) at constant voltage of 100 V for 70 mins and blocked with 5 % milk powder for 1 hr. Primary antibodies were used for overnight at 4ºC and secondary antibody for 1 hr in 5 % milk powder. To detect exon-1

fragment of huntingtin with 74 glutamine repeats, GFP antibody (1:5000) and MAB 1574 1C2 (1:7500) against polyglutamine repeats were used. To detect endogenous full length huntingtin MAB 2166 (1:4000) was used. Dilutions of other antibodies: TG2 (TG100, 1:7500), p62 (1:5000), LC3 (1:5000), Atg5 (1:5000), anti-mouse (1:5000-1:8000), anti-rabbit (1:5000-1:15000), anti-rat (1:3000). Actin or tubulin was used as loading control (1:5000).

4. RESULTS

4.1. Comparative genomics reveals novel amino acid clusters with potentially functional effects in transglutaminase 2

For the comparative sequence alignment of human, primate and rodent (mouse) TG2, sequences from UniProt database was used. Chimpanzee and human TG2 differ only in two amino acid residues (D84E and G366E), with 99.7 % sequence identity. Other TG2 primates are also closely related to the human sequence (data not shown). Similarity between human and mouse TG2 sequences is 84.1 %. The similarity in case of F13a (86.7 %), TG6 (85.6 %) and TG5 (81.7 %) is in the same range but in others it is either higher as in TG1 (89.9 %) or lower as in TG3 (76.4 %), TG7 (69 %) and TG4 (53.7 %). Human and mouse TG2 differ in 108 residues, and the differences are distributed in various domains of human TG2: 3.6 % in the β-sandwich, 4.3 % in the catalytic core, and 4 % each in both β-barrel 1 and β-barrel 2 (Figure 4A). Presence of human/primate specific clusters of amino acids is the most striking feature in dissimilarities between human and mouse TG2.

Six novel amino acid clusters containing minimum of three to four amino acids next to one another appeared in human TG2 (Figure. 4A) of which four are triplets and two of them are quartets. Novel clusters are distributed over five exons (exon 2, 7, 10, 12, and 13), with exon 10 containing a triplet and a quartet cluster. The clusters are shown in the crystal structure of human TG2 in opened (PDB ID: 2Q3Z) and closed conformation (PDB ID: 1KV3) in Figure. 4B. β-sandwich and the catalytic core domains contain one novel cluster each and β-barrel 1 and β-barrel 2 contains two clusters each. Chimpanzee, gorilla, northern white-cheeked gibbon have all the six human amino acid clusters and rhesus macaque, sumatran orangutan, olive baboon and green monkey have five clusters (data not shown). Other TG2 species like horse, dog, pig, and zebrafish does not contain these amino acid clusters. As compared to mice, other human transglutaminase family members also contain novel amino acid clusters but they are different from the ones in TG2 and also from one another (data not shown): TG4 has more than ten clusters; TG3 has eight; TG5 has seven; TG7 has six; F13a has four; TG1 has one and TG6 has no cluster. Mostly triplet clusters were observed in F13a and TG7 but the cluster stretches in TG3 and TG5 were longer encoding from three up to nine amino acids.

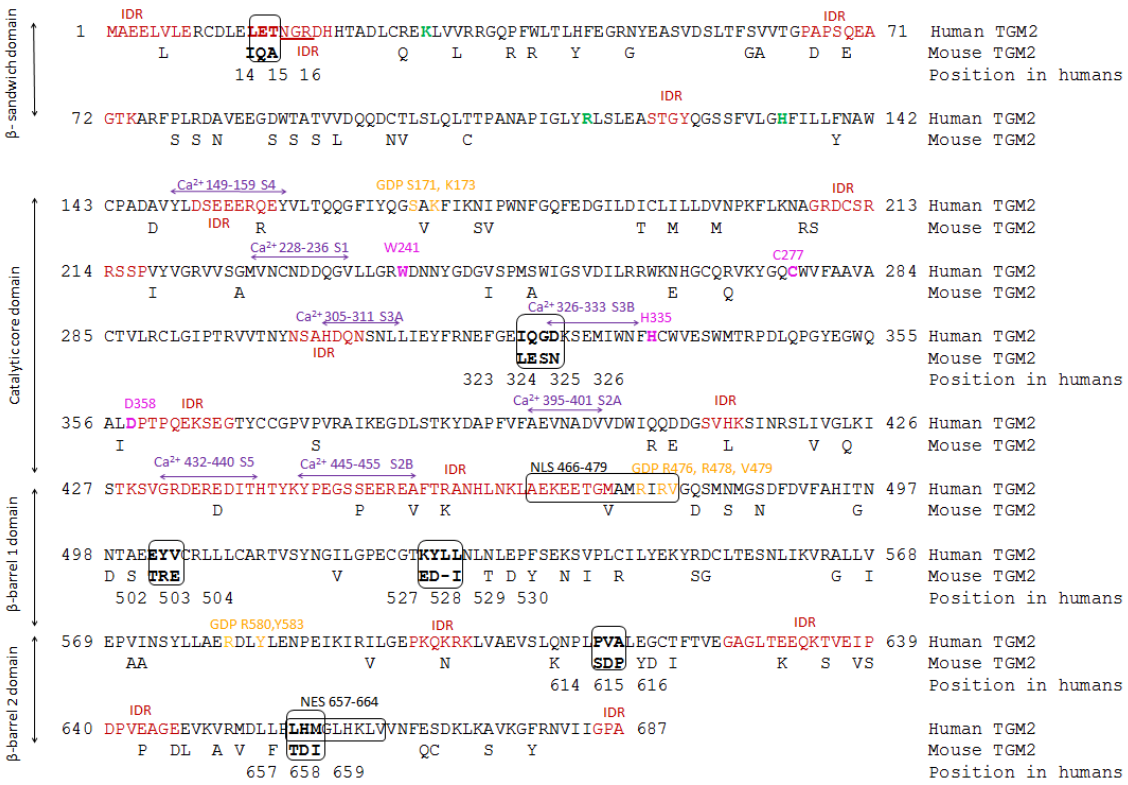


Figure 4A. IDRs and novel amino acid clusters in human transglutaminase 2 Lane1: human TG2 sequence, novel amino acid clusters are highlighted in boxes, IDRs in dark red. Lane2: amino acids different in mouse TG2 as compared to human TG2 sequence. Lane 3: position of amino acid clusters in human TG2. Functional regions of human TG2: intrinsically disordered regions (dark red) (Kanchan et al., 2015), NGR motif residues 17–19 are underlined (Kanchan et al., 2015), fibronectin binding sites (green) K30, R116, H134 (Cardoso et al., 2016), GDP binding residues (orange) (Liu et al., 2002), catalytic residues (pink) (Kanchan et al., 2015), non-canonical Ca^{2+}-binding sites: S4: 149-159, S1: 228-236, S3A: 305-311, S3B: 326-333, S2A: 395-401, S5: 432-440, and S2B: 445-455 (purple) (Király et al., 2009). 14-amino acid nuclear localization signal (NLS) (^{466}AEKEETGMAMRIRV^{479}) and a leucine-rich nuclear export signal (NES) (^{657}LHMGLHKL^{664}) (Shrestha et al., 2015) are boxed. Domains of human TG2 are presented vertically: β-sandwich (1-139), catalytic core (147-460), β-barrel 1 (472-583), and β-barrel 2 (584-687).

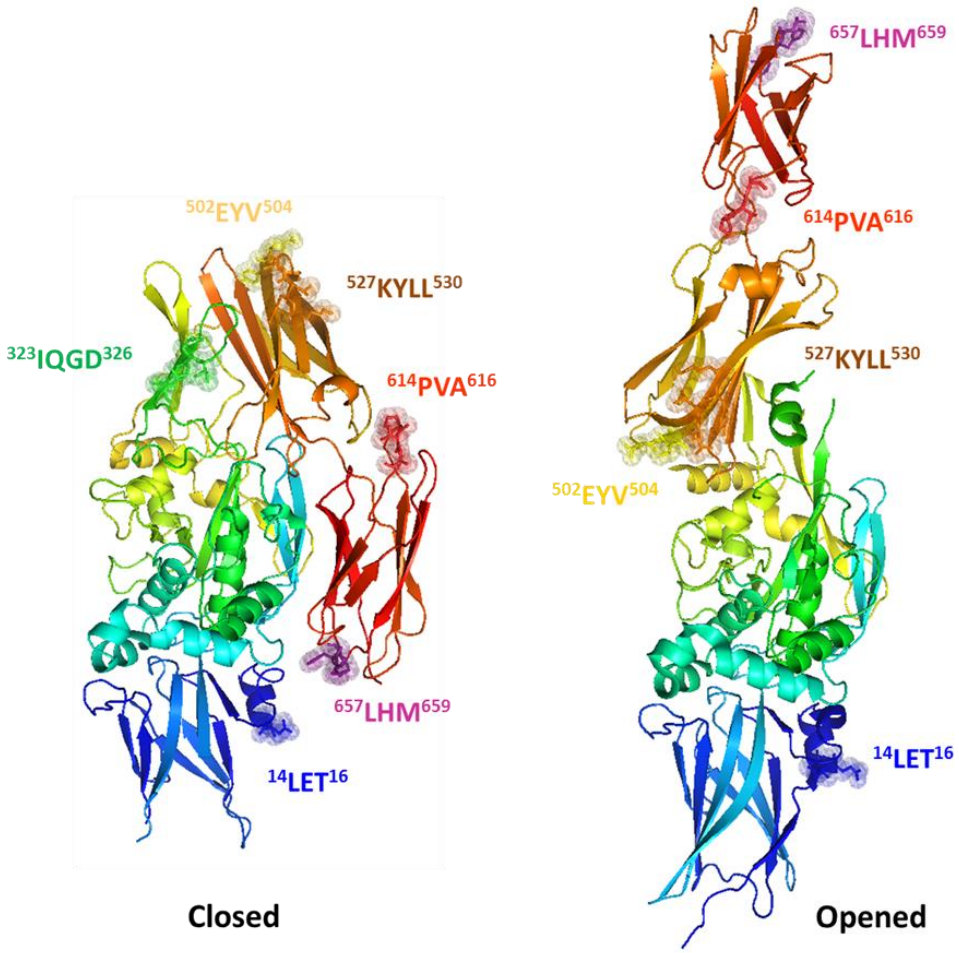


Figure 4B. Representation of human TG2 clusters in opened and closed conformation. Human TG2 closed (PDB ID: 1KV3) and opened conformation (PDB ID: 2Q3Z) is represented. Stick and transparent sphere representation is used to indicate the residues of the clusters. Colour code for different domains of human TG2: blue: β-sandwich; green: catalytic core; orange: β-barrel 1; red: β-barrel 2. L14 residue is not present in the PDB file in closed form, while ^{323}IQGD^{326} cluster residues are not present in the PDB file in the opened form. This figure was prepared by Dr. János András Mótyán.

4.1.1. Impact of novel amino acid clusters on structure and stability of TG2

Secondary structure is an important determinant of Protein folding and stability (Kwok et al., 2002). Conversion from α-helix to β-strand or vice versa can take place when two or more amino acid residues are interchanged. The human TG2 sequence was modified by replacing the clusters with the mouse sequences in order to check the effect of amino acid clusters on secondary structure. By the presence of the mouse residues, conformation of all, but one (502-504) cluster was affected (Table 2). The mouse clusters could destabilize the structure of human TG2 as in most cases regular secondary structural elements were shifted towards coils (α-helices to β-strands).

Although modified amino acid clusters were mostly located in loop regions and were exposed to the solvent, the secondary structure predictions indicate that the novel clusters increased the stability of human TG2. Thus by incorporating the mouse sequences at cluster sites, *stability* changes were computed for both the opened and closed forms of human TG2 (Table 3). Human TG2 was found to be destabilized by all the mouse sequences, especially the 527-530 cluster, where a charge inversion took place (K527E and Y528D). Considerable destabilization was observed in case of clusters 502-504 and 614-616 in both opened and closed conformations (Table 3) but some of the individual changes (e.g. E15Q and A616P) were slightly stabilizing (data not shown). To find an optimal packing for the mouse residues, individual residue contributions to cluster stability might implicate conformational rearrangements in the loop regions. By producing mouse homology models, stability changes for mouse TG2 in both opened and closed conformation were also computed. Except cluster ^{323}IQGD^{326} in the open form, all other human amino acid clusters stabilized the mouse TG2 when mouse sequence was replaced with human amino acid clusters (Figure. 4C). Overall the results indicate a convergence and strong selection pressure on TG2 structure, supported by the increase in stability from mouse to human TG2 during evolution (Tokuriki and Tawfik, 2009).

Table 2. Secondary structure prediction for the human TG2 clusters as compared to the respective mouse sequence

	Cluster in human TG2					**Cluster modified to mouse sequence in human TG2**					
AA position	**wt**	**H**	**E**	**C**	**secondary structure**	**mutant**	**H**	**E**	**C**	**secondary structure**	**secondary structure change**
14	L	**39.3**	34.3	26.4	H	I	**45.9**	37.7	16.5	H	-
15	E	36.4	25.4	**38.3**	C	Q	**53.2**	18.3	28.5	H	C→ H
16	T	21.8	30.8	**47.5**	C	A	35.3	24.8	**39.8**	C	-
323	I	24.9	**45.7**	29.4	E	L	31.7	**37.3**	31.0	E	-
324	Q	23.5	**41.5**	35.0	E	E	32.4	**47.4**	20.2	E	-
325	G	10.6	38.4	**50.9**	C	S	**36.8**	32.5	30.7	E	C → E

326	D	15.8	9.7	**74.5**	C	N	30.6	21.8	**47.6**	C	-
502	E	**48.4**	30.8	20.8	H	T	17.3	**51.2**	31.4	H	-
503	Y	**56.2**	32.6	11.2	H	R	**52.1**	21.6	26.3	H	-
504	V	**47.3**	42.8	9.9	H	E	**60.4**	22.1	17.5	H	-
527	K	29.2	**43.9**	26.9	E	E	25.9	16.0	**58.0**	C	E → C
528	Y	18.6	**51.6**	29.8	E	D	25.9	8.6	**65.5**	C	E → C
529	L	11.3	**56.9**	31.8	E	-	-	-	-	-	-
530	L	11.3	40.4	**48.2**	C	I	18.8	22.4	**58.8**	C	-
614	P	22.2	21.6	**56.2**	C	S	5.4	11.9	**82.7**	C	-
615	V	14.2	**47.2**	38.6	E	D	2.9	4.9	**92.2**	C	E → C
616	A	9.9	**49.7**	40.5	E	P	5.1	3.6	**91.3**	C	E → C
657	L	**43.3**	21.9	34.9	H	T	18.0	15.9	**66.1**	C	H → C
658	H	**50.4**	7.2	42.3	H	D	21.5	5.9	**72.6**	C	H → C
659	M	**50.5**	11.7	37.8	H	I	30.8	14.2	**55.0**	C	H → C

GOR IV method was used to perform the predictions. Predicted probabilities of helices (H), extended strands (E), and coils (C) are displayed (%). AA refers to amino acid. Bold letters indicate the highest probability values. Predicted secondary structure is indicated in column 6. Human TG2 was modified at cluster sites by mouse sequence and values were predicted (column 11). Column 12 indicates the secondary structure change.

Table 3. Stability analysis of TG2 amino acid clusters

Human TG2 (P21980)	Mouse TG2 (P21981)	Human to mouse change	$\Delta\Delta G_{human \rightarrow mouse}$ closed structure [kcal/mol]	$\Delta\Delta G_{human \rightarrow mouse}$ open structure [kcal/mol]
L14 E15 T16	I14 Q15 A16	L14I * E15Q T16A	-0.22	1.82

I323 Q324 G325 D326	L323 E324 S325 N326	I323L** Q324E** G325S** D326N**	0.23	-
E502 Y503 V504	T502 R503 E504	E502T Y503R V504E	0.85	1.56
K527 Y528 L529 L530	E526 D527 -528 I529	K527E Y528D - L530I	4.86	3.95
P614 V615 A616	S613 D614 P615	P614S V615D A616P	2.06	2.19
L657 H658 M659	T656 D657 I658	L657T H658D M659I	2.04	0.44

The analysis was carried out using FoldX software in closed (PDB ID: 1KV3) and opened conformation (PDB ID: 2Q3Z) of TG2. Relative free energy or thermodynamic stability ΔΔG [kcal/mol] was calculated for each cluster after replacing each human TG2 amino acid cluster with rodent sequence. Column 1: human clusters, column 2: mouse sequence at cluster sites, column 3: human to mouse amino acid cluster change. column 4 is the relative free energy in closed form, and column 5 open form. $\Delta\Delta G_{human \rightarrow mouse} = \Delta\Delta G_{mouse} - \Delta\Delta G_{human}$ (mouse relative to human). *Residue not present in the PDB file in closed form. **Residue not present in the PDB file in open form (Clusters involving residues 1-14* and 323-326** for which the coordinates in either or both crystal structures are not available were excluded from the stability analysis).

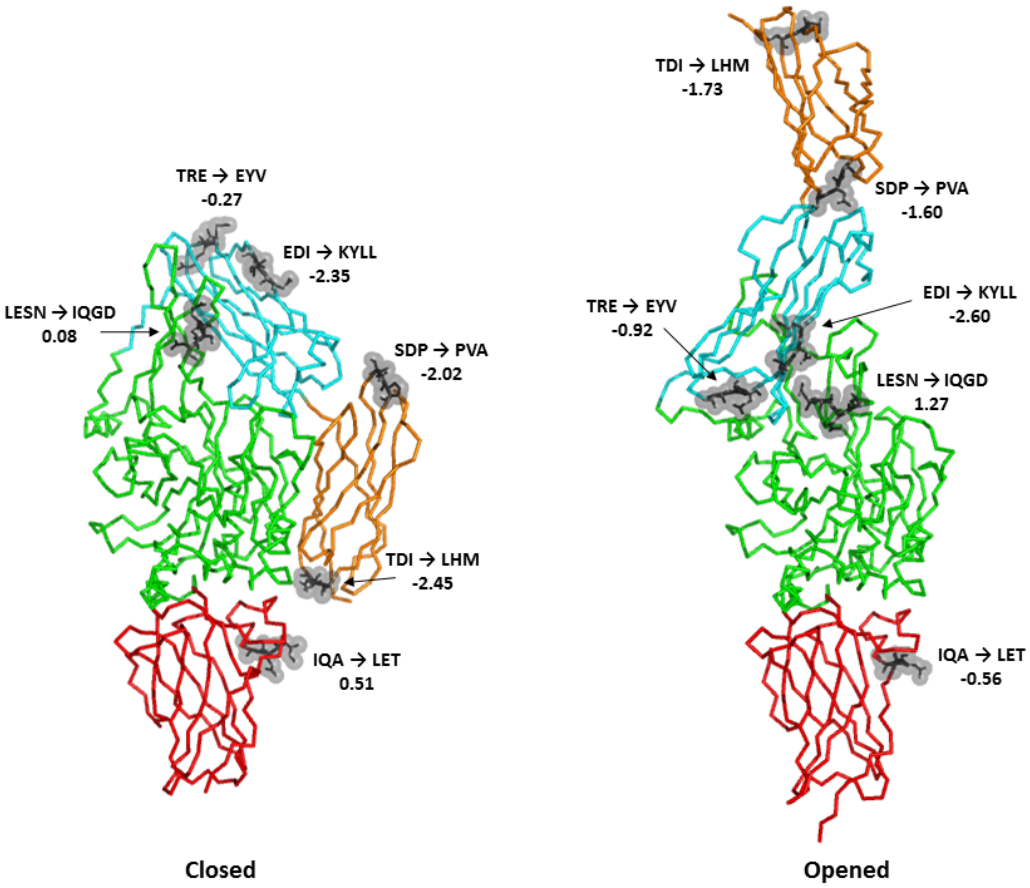


Figure 4C. Influence of human amino acid clusters on stability of mouse TG2. FoldX software was used to do the stability analysis. Relative free energy or thermodynamic stability ΔΔG [kcal/mol] was calculated for each cluster after replacing mouse sequence with human TG2 amino acid clusters. The amino acid cluster changes (mouse to human change) are indicated along with the thermodynamic stability ΔΔG [kcal/mol] values. $\Delta\Delta G_{mouse \rightarrow human} = \Delta\Delta G_{human} - \Delta\Delta G_{mouse}$ (human relative to mouse). Colour code for different domains of mouse TG2: Beta sandwich domain, 1-139 (red) Catalytic core, 140-465 (green) β-barrel 1, 466-583, (cyan) β-barrel 2, 584-686, (orange). This figure was prepared by Dr. János András Mótyán.

4.1.2. Impact of novel amino acid clusters on functional elements

Except for one non-canonical Ca^{2+}-binding site, the novel amino acid clusters are in regions outside the known functional sites in human TG2. S3B Ca^{2+}-binding site 326-333 (Király et al., 2009) present within cluster ^{323}IQGD^{326} might lead to modified Ca^{2+}-regulation of TG2 in humans. Cluster ^{657}LHM^{659} is part of the newly described human specific nuclear export signal motif ^{657}LHMGLHKL^{664} (Shrestha et al., 2015) and also coincides with the celiac epitope residue M659 (Simon-Vecsei et al., 2012) raising the possibility that in pathologic conditions compared to mouse protein, human TG2 is more prone for the development of autoantibodies.

Post-translational modifications and protein-protein interactions are mediated by low complexity short peptide regions (3–20 residues) called SLiMs. Mostly protein segments that lack well-defined tertiary structure termed as IDRs embed SLiMs. Some IDRs and amino acid clusters are located in proximity to each other (Fig. 4A): Cluster ^{14}LET^{16} is part of IDR 14-21, cluster ^{614}PVA^{616} is next to IDR 597–602, and cluster ^{657}LHM^{659} is located near IDR 626-647. The conformational diversity of TG2 can be modulated by novel amino acid clusters via the affected IDRs. Compared to the rodents, human TG2 has gained several new SLiMs (Table 4) since SLiMs were disabled/removed when mouse sequence replaced the human TG2 clusters. This indicates that human TG2 has gained functional motifs during evolution, such as a phosphothreonine binding motif that is found in a subset of FHA domains; an Src homology 2 (SH2) domain binding motif; a major TNF receptor-associated factor 2 (TRAF2)-binding consensus motif; a CK2 phosphorylation site; a PDZ domain recognition motif; a cyclin recognition site; a subtilisin/kexin isozyme-1 cleavage site; anaphase-promoting ubiquitin ligase complex (APCC)-binding destruction motifs; a MAPK docking motif, and a Small Ubiquitin-like Modifier (SUMO) binding site (Table 4).

For both wild-type mouse and wild-type human TG2 phosphorylation and N-glycosylation sites can be predicted. Compared to rodents, human TG2 has gained two phosphorylation sites Y503 and Y528 with the appearance of the related clusters. But humans lost the S325 phosphorylation and N326 N-glycosylation sites as a consequence of the inserted cluster ^{323}IQGD^{326}. Overall, novel amino acid clusters in human TG2 were identified and these peptide sequences contribute to increasing stability of human TG2 and influences secondary structure. Due to the novel clusters, human TG2 have gained functional motifs which could potentially regulate vital functions.

Table 4. Potential SLiMs gained by the appearance of novel clusters in human TG2 during evolution

Amino acid clusters in human TG2	Functional SliMs gained by human TG2	Explanation
^{502}EYV^{504}	*LIG_FHA_2 [497-503]* *LIG_SH2_STAT5 [503-506]* *LIG_TRAF2_1 [499-502]* *MOD_CK2_1 [496-502]* *LIG_PDZ_CLASS_3[501-504]*	- Phosphothreonine motif binding a subset of FHA domain - SH2 domain binding motif - Major TRAF2-binding consensus motif - CK2 phosphorylation site - PDZ domains recognize short sequences at the carboxy terminus of target proteins
^{527}KYLL^{530}	*LIG_SH2_STAT5 [528-531]* *DOC_CYCLIN_1 [527-530]* *CLV_PCSK_SKI1_1 [527-531]* *LIG_FHA_1[524-530]*	- SH2 domain binding motif - Cyclin recognition site - Subtilisin/ kexin isozyme-1 cleavage site - Phosphothreonine motif binding a subset of FHA domain
^{657}LHM^{659}	*DEG_APCC_DBOX_1 [650-658]* *DOC_MAPK_1 [649-657]* *LIG_SUMO_SBM_2 [653-657]*	- Anaphase-promoting ubiquitin ligase complex (APCC) binding destruction motifs - MAPK docking motif - SUMO binding site.

Column 1 shows the amino acid clusters, column 2 shows the potential SLiMs gained (row 1, 2 and 3) by human TG2. SLiM coordinates are given in brackets. Function of the given SLiM is explained in last column. Other clusters not shown here did not affect the linear motifs.

4.2. Genomic polymorphism/variants in genes of the human transglutaminase family

Total number of entries for transglutaminases in ExAC was 5,766 SNVs of which 3,623 SNVs fall under synonymous, non-synonymous or loss-of-function categories in exons (Table 5). In case of each family, non-synonymous SNVs constitute 4.5-6 % of the total entries with TG2, TG4, TG5 having the lowest and TG6 the highest numbers (Table 5). The number of loss-of-function (LOF; including frameshift, splice acceptor and stop gained) variants for TGs are: 22 in F13a, 19 in TG1, 29 in TG2, 20 in TG3, 40 in TG4 and TG5, while 39 in TG6 and 28 in TG7 (Table 5). For each family, the number and percentage of residues polymorphic to nsSNVs were also calculated. Less percent of residues polymorphic to nsSNVs were observed for F13a, TG1, TG2 and TG5 and more for TG3 and TG6 (Figure 5A).

4.2.1. Evolutionary constraint on the transglutaminase genes

For non-synonymous variations of each gene in the ExAc dataset, Z scores were calculated in order to determine the constraint of a particular gene (Lek et al., 2016). The Z values for the transglutaminase nsSNVs are given in Table 5. Positive Z scores indicate increased constraint and negative Z scores indicate decreased constraint. TG1, TG2 and F13a had fewer variants than expected and are under more constraint as specified by highest and positive Z scores. But other members had more variants than expected and are under less constraint with negative Z scores (Table 5).

Recently by gene damage index scores (GDI), load of disease causing mutations and mutational damage on protein-coding human genes was estimated (Itan et al., 2015). The biologically indispensable genes which are under strong purifying selection harbor fewer mutations and had low GDI score. But biologically redundant genes under less purifying selective pressure with more mutations had high GDI score. To estimate the damaged human genes a Phred I-score was also calculated by Ital et al. Phred I-score refers to the ranking of the gene of interest i relative to all other human genes (T = 19,558 genes used in the analyses by Itan et al., 2015). So the least damaged human gene will have lowest Phred-I and GDI score and most damaged human gene will have highest Phred-I score. The GDI and GDI-Phred values are listed for transglutaminases in Table 5 (Itan et al., 2015). Amidst transglutaminases, TG2 and TG1 are under more selective pressure with lowest values and cannot tolerate damaging mutations but other transglutaminases, particularly TG4 and F13a with highest Phred scores tolerate damaging mutations.

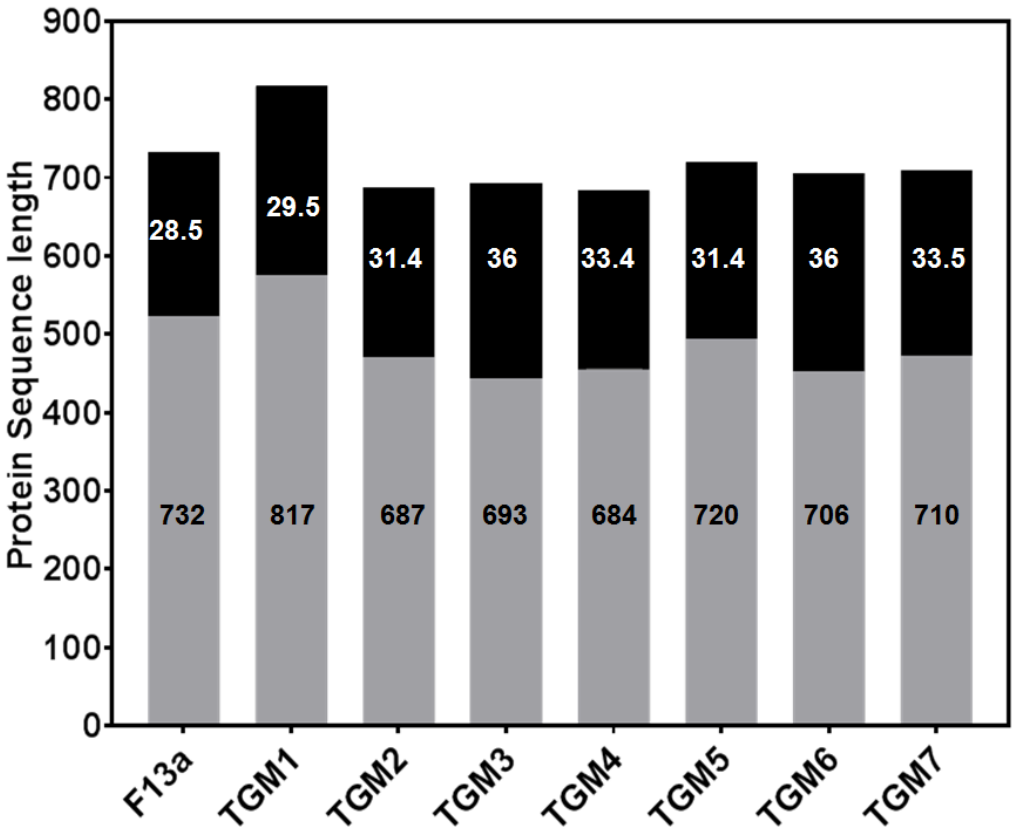
Table 5. Number of different single nucleotide variant and gene constraint in human transglutaminase genes

Gene	Synonymous	Non-synonymous	Loss-of-Function	n_mis	exp_mis	mis_z	GDI	GDI-Phred
F13a	120	283	22	252	269.8	0.53	4254.7	12.9
TG1	163	320	19	314	349.2	0.92	277.2	3.5
TG2	148	272	29	270	290.5	0.59	240.6	3.3
TG3	144	313	20	304	271.9	-0.95	1529.8	7.2
TG4	121	279	39	264	232.6	-1.0	3544.7	11.4
TG5	95	273	40	261	255.9	-0.15	2224.6	8.6
TG6	136	336	39	320	287.7	-0.92	764.5	5.4
TG7	91	293	28	283	263.3	-0.59	698.4	5.2

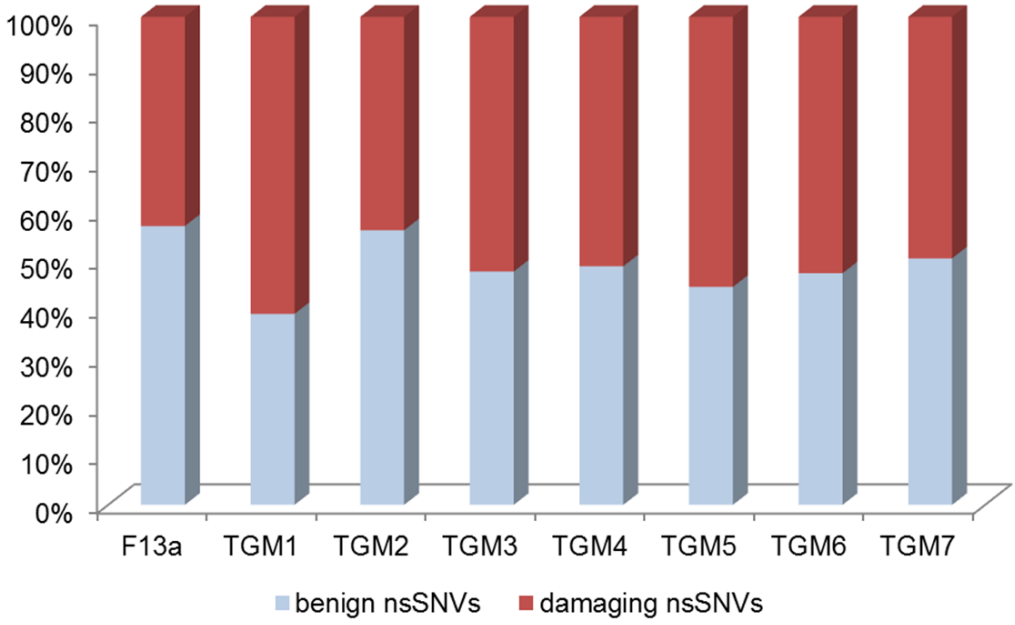
n_mis: number of rare (Minor allele frequency (MAF) < 0.1 %) missense variants found in ExAC. **exp_mis:** depth adjusted number of expected rare (MAF < 0.1 %) missense variants. **mis_z**: corrected missense Z score (taken from Lek et al; 2016). LOF category also includes frameshift mutations. Because for Z score calculations only nsSNVs with MAF < 0.1 % was taken into account (n_mis, column 5) so the number of nsSNVs in columns 3 and 5 differ. The **GDI** and **GDI-Phred** values were taken from (Itan et al., 2015).

The SIFT and polymorphism phenotyping scores provided in the ExAc database were used. According to SIFT analysis F13a and TG6 had lowest and highest ratios of damaging nsSNVs, respectively (data not shown). According to PolyPhen analysis TG2 and F13a had lowest while TG1 and TG5 had the highest ratios of damaging nsSNVs (Figure 5B). The damaging nsSNVs in TG2 is around 45 % and out of these 17 % are concentrated in the β-sandwich domain and 54 % in the catalytic core domain (Figure 5C).

A


Protein Sequence length
900
800
700
600
500
400
300
200
100
0
28.5
29.5
31.4
36
33.4
31.4
36
33.5
732
817
687
693
684
720
706
710
F13a
TGM1
TGM2
TGM3
TGM4
TGM5
TGM6
TGM7

B


100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
F13a
TGM1
TGM2
TGM3
TGM4
TGM5
TGM6
TGM7
benign nsSNVs
damaging nsSNVs

C

Figure 5. Analyses of non-synonymous variants in the transglutaminase family (A). Amino acid residues polymorphic to non-synonymous variants. To get the percentage of polymorphic residues the total number of amino acid residues polymorphic to nsSNVs was divided by the sequence length of the respective protein. The sequence length (shaded grey) and the percentage of polymorphic residues are indicated in the figure (shaded black). **(B). Proportion of damaging and benign non-synonymous variants by the PolyPhen score.** Ratio of damaging nsSNVs: F13a (42.8 %), TG1 (60.8 %), TG2 (43.7 %), TG3 (52.2 %), TG4 (51 %), TG5 (55.3 %), TG6 (52.5 %) and TG7 (49.5 %). **(C). Location of damaging nsSNVs in amino acid sequence of TG2 by PolyPhen/SIFT score.** Lane1: human TG2 sequence, Lane2: damaging nsSNVs in human TG2. Functional regions of human TG2: intrinsically disordered regions (dark red) (Kanchan et al., 2015), amino acid clusters in light blue (described in section 4.1), catalytic residues (pink) (Kanchan et al., 2015), GDP binding residues (orange) (Liu et al., 2002), non-canonical Ca^{2+}-binding sites: S4: 149-159, S1: 228-236, S3A: 305-311, S3B: 326-333, S2A: 395-401, S5: 432-440, and S2B: 445-455 (purple) (Király et al., 2009), fibronectin binding sites (green) K30, R116, and H134 (Cardoso et al., 2016), interaction sequence of other proteins: Syndecan-4: ^{202}KFLKNAGRDCSRRSSPVYVGR^{222}, three SUMO motifs 327-329, 364-366, and 468-470, BAX and BAK: 204-212, α1-adrenoceptor: L547-I561, R564-D581, and Q633-E646, PLCδl: V665-K672 (reviewed in Eckert et al., 2014; Kanchan et al 2015). Domains of human TG2 are presented vertically: β-sandwich (1-139), catalytic core (147-460), β-barrel 1 (472-583), and β-barrel 2 (584-687).

4.2.2. Effect of damaging nsSNVs on TG2 stability and secondary structures

The stability and biochemical functions of the native proteins are affected by the damaging nsSNVs (Han et al., 2015). The influence of PolyPhen or SIFT predicted damaging nsSNVs (indicated in Figure 5C) on protein stability, secondary structure and functional sites including

novel amino acid clusters, IDRs, SLiMs and LC3 (microtubule-associated protein light chain) interacting regions (LIRs) were analyzed.

To understand the sequence–structure relationships and to predict the evolutionary dynamics of proteins, stability effects of mutations are crucial (Tokuriki et al., 2007). FoldX analysis to assess the impact of damaging nsSNVs on TG2 stability was performed using both the opened (PDB ID: 2Q3Z) and closed (PDB ID: 1KV3) conformations. Relative stability (ΔΔG) of the opened and closed conformation was equally influenced by the mutations. Accordingly, 24.4 % of nsSNVs in the opened conformation and 28.6 % in the closed conformation were found to be destabilizing (ΔΔG > 1 kcal/mol), 26.7 % and 23.5 % were highly destabilizing (ΔΔG > 3 kcal/mol) in opened and closed conformations respectively. Catalytic core domain houses more highly destabilizing damaging nsSNVs, in line with the observation that highly destabilizing mutations affect the function of a protein (Tokuriki et al., 2009).

Eight damaging nsSNVs had a minor impact on the secondary structure propensities based on GORIV predictions: at position 12 the insertion of a helix-breaking proline induced helix to coil transition (data not shown). Stability of the human TG2 structure is underscored by the persistence of secondary structures after generation of nsSNVs. Destabilizing ΔΔG values in opened and closed conformations were observed for R214H and W337L nsSNVs and in closed conformation for R377H nsSNV (Data not shown). The V283M nsSNV was highly destabilizing in the opened conformation.

4.2.3. Occurrence of damaging nsSNVs of TG2 at functional and interaction sites

Active site residues (W241, C277, H335, and D358) and novel amino acid clusters (described in section 4.1) does not contain damaging nsSNVs. There is no damaging nsSNVs at residues crucial for transamidase (W278) or isopeptidase (W332) activity (described in section 4.4). There is no nsSNVs at the recently described fibronectin binding residues in TG2 (Cardoso et al., 2016).

GDP binding sites (Liu et al., 2002) are slightly influenced by the damaging nsSNVs S171L, R476Q and R478C but all the non-canonical calcium binding sites (Kiraly et al., 2009) are affected (Figure 5C). Studies indicate that TG2 non-enzymatic protein-protein interactions have important physiological and pathological outcomes (reviewed in Kanchan et al., 2015), so the presence of damaging nsSNVs at various interaction sites were also checked. Six nsSNVs are part of Syndecan-4 interaction sequence of TG2 (Figure 5C) (reviewed in Eckert et al., 2014; Kanchan et al 2015). Similarly, interaction sites of various proteins in TG2 enclose damaging nsSNVs: the α1-adrenoceptor interaction sites and the PLCδl interaction

sequence V665-K672 (reviewed in Eckert et al., 2014; Kanchan et al 2015) embeds eleven and three nsSNVs respectively (Figure 5C). BAX and BAK interaction sequence embeds one, 14-3-3 binding protein interaction sequence 209-223 (Mishra et al., 2006) has seven and two nsSNVs are present in the three SUMO motifs detected on TG2 (Figure 5C) (Luciani et al., 2009). Cellular TG2 functions like transmembrane signaling, cell adhesion, migration, Ca^{2+} regulation of transamidation, cell death induction and protein turnover might be influenced due to the proximity of nsSNVs to the sites targeted by the multiple interacting partners.

In TG2 gene, three heterozygous missense mutations associated with early-onset type 2 diabetes in a small disease cohort were reported (Porzio et al., 2007) and one of those nsSNV, M330R has damaging scores. But in larger diabetes patient cohorts, the association of TG2 mutations and dysfunction has not been confirmed so far and relative to wild-type littermates, TG2 KO mice have no impairment in glucose-stimulated insulin secretion by pancreatic islets (Iismaa et al., 2013).

By generating immunogenic gliadin peptides through deamidation and because of the appearance of disease specific anti-TG2 antibodies, TG2 plays a crucial role in celiac disease pathogenesis. The so far identified celiac epitopes, namely epitope 1 which is composed of Lys30, Arg116, His134 and epitope 2 consisting of Arg 19, Glu 153, Glu 154, Met 659 (Simon-Vecsei et al., 2012; Iversen et al., 2014) does not have any damaging nsSNVs.

4.2.4. Occurrence of damaging nsSNVs of TG2 at intrinsically disordered regions and short linear motifs

Studies show that functionality of IDRs are affected by about 22 % of human disease mutations occurring in intrinsically disordered regions (Davey et al., 2012). We looked for the presence of damaging nsSNVs in recently reported 13 IDRs embedding 39 SLiMs in humans (Kanchan et al., 2015). The damaging nsSNVs located in IDRs embedding SLiMs are given in Table 6A. β-sandwich domain houses a damaging nsSNV located in the IDR 65-74 and a SLiM motif recognized by SH3 domains. Series of six nsSNVs are located in IDR 208-217 and the SLiM motifs involved in proline directed phosphorylation signaling pathways in the catalytic core. Series of SLiMs (TRAF2 and TRAF6 binding motif; CK2 and proline-directed kinase phosphorylation motif) embedded in IDR 358-367 has two nsSNVs. IDR 411–414 embeds SLiMs important for actin binding (WH2 domains), GSK3 and PLK phosphorylation to which four nsSNVs are co-localized. Moreover, four nsSNVs are present in IDR 428–473 with SLiM Ubiquitin-specific-processing protease 7 (USP7), which acts as a deubiquitination enzyme and three nsSNVs in IDRs 597–602, and 626–647 embedding series of SLiMs in the

β-barrel 2 domain. SLiM responsible for MAPK docking within IDR 685–687 has four nsSNVs. Therefore, nsSNV mutations can influence IDRs and their SliMs of TG2 with possible functional consequences.

Autophagy is long been viewed as an intracellular degradation system involving bulk breakdown of cytosolic material. However, role of selective autophagy mediated by the so-called selective autophagy receptors (SARs) is rapidly emerging (Kalvari et al., 2014). Short linear sequence motifs called LIRs mediate interaction between SARs and proteins of the Autophagy-related protein 8 (Atg8) family ensuring the targeting of autophagy receptors to LC3 or other Atg8 family proteins anchored in the phagophore membrane. The short tetrapeptide sequence WxxL (where x could be any residue) present in the canonical LIR motif is necessary for its interaction with two distinct hydrophobic pockets of LC3 (Kalvari et al., 2014). Involvement of TG2 in autophagy (D'Eletto et al., 2009), prompted us to check for LIRs in TG2. TG2 consist of several LIR motifs and 11 damaging nSNVs located within these motifs are shown in Table 6B. The xLIR motif ^{352}EGWQAL^{357} is targeted by two damaging nsSNVs and nine of them target WxxL motifs (Table 6B). Given the role of TG2 in autophagosome maturation and its interaction with autophagy cargo proteins (D'Eletto et al., 2009, 2012), nsSNVs present within LIRs might have functional implications in the autophagic process.

Table 6. Damaging nsSNVs located in short linear motifs

A

IDR	ELM	Explanation	Damaging nsSNVs
65–74	LIG_SH3_3 [59–65]	SH3 binding domain	G64S
208–217	DOC_MAPK_1 [213–220] DOC_WW_Pin1_4 [213–218] MOD_GSK3_1 [209–216] MOD_CK1_1 [212–218] MOD_PKA_2 [212–218] MOD_ProDKin_1 [213–219]	Docking interaction in MAP kinase cascade (exemplified cJun); binds WW domains, involved in proline directed phosphorylation signaling pathways; phosphorylation motifs (CK1, GSK3,PKA, proline-directed kinase)	R209C, R213C, R214H, V218A, V220M
358–367	DOC_WW_Pin1_4 [357–362] LIG_FHA_2 [358–364] LIG_TRAF2_1 [360–363]	Binds WW domains, involved in proline directed phosphorylation signaling pathways; TRAF2 and TRAF6 binding motif; CK2 and proline-directed kinase	T360M, E366K

	LIG_TRAF6 [359–367] MOD_CK2_1 [357–363] MOD_ProDKin_1 [357–363]	phosphorylation motif	
411–414	LIG_ACTIN_WH2_2 [409–427] MOD_GSK3_1 [408–415] MOD_PLK	Actin binding motif (WH2 domains); GSK3 and PLK phosphorylation site;	D409V, D409G, G410E, K414T
428–473	DOC_USP7_1 [446–450] MOD_CK1_1 [427–433] MOD_CK2_1 [446–452]	USP7 binding motif; CK1 and CK2 phosphorylation motif	E447K, G448E, V431L, V431M
597–602	TRG_LysEnd_APsAcLL_1 [599–604] TRG_NLS_MonoExtN_4 [597–604] DOC_CYCLIN_1 [601–604] DOC_MAPK_1 [601–609]	Sorting and directing signal to lysosomal endosomal compartment; NLS; cyclin recognition signal; MAPK docking motif	R601C
626–647	DEG_APCC_DBOX_1 [650–658] DOC_MAPK_1 [649–657] LIG_FHA_1 [633–639]	APCC binding destruction signal; MAPK docking motif; FHA binding motif	P656L, T635M
685–687	DOC_MAPK_1 [674–684]	MAPK docking motif	A675P, V676A, R680P, I684T

B

MOTIF	START	END	LIR sequence	TG2 damaging nsSNV
xLIR	352	357	**EG**WQAL	E352K, G353V
WxxL	38	43	PFWLT**L**	L43R
WxxL	133	138	GHFI**L**L	L137F
WxxL	278	283	WVFA**AV**	A282T, V283M
WxxL	392	397	FVF**A**EV	A395V
WxxL	514	519	VSY**N**GI	N517S
WxxL	677	682	KGF**R**NV	R680W, R680P, R680Q

A Damaging nsSNVs located in ID regions embedding SLiMs in TG2. Position of IDRs are indicated in column 1; names of SLiMs and their coordinates in square brackets in column 2; explanation for the function of the given SLiM in column 3; damaging nsSNVs of TG2 in column 4. **B. Damaging nsSNVs located in potential LC3 interacting regions (LIR)**.

4.3. Heterozygous and homozygous occurrence of TG2 variants

4.3.1. Population occurrence of non-synonymous SNVs in transglutaminases

The population distribution of nsSNVs is crucial to determine the impact of nsSNVs on human health and disease. TG3 and TG4 had the highest (over 0.2 million) nsSNV allele counts in the population covered in the ExAc dataset and TG2 and TG1 had the lowest (2601 and 5174 respectively) (Table 7A). Variants with allele frequency > 5 % are typically defined as common variants and those found with < 0.5 % allele frequency are termed as rare variants. (Greg, 2012). Common nsSNVs were present in F13a, TG3, TG4, and TG6, but TG1 and TG7 had only rare nsSNVs. In case of each family member the allele frequency (%) value for the three most frequent variants is depicted in Figure 6. The nsSNVs in TG2 were rare with allele frequency values less than 0.5 %. The highest allele frequency values in case of TG2 nsSNVs are R76H (0.47 %), V542F (0.38 %), E366K (0.11 %), R433Q (0.10 %) and E469G (0.10 %); of these only V542F has potentially damaging PolyPhen score.

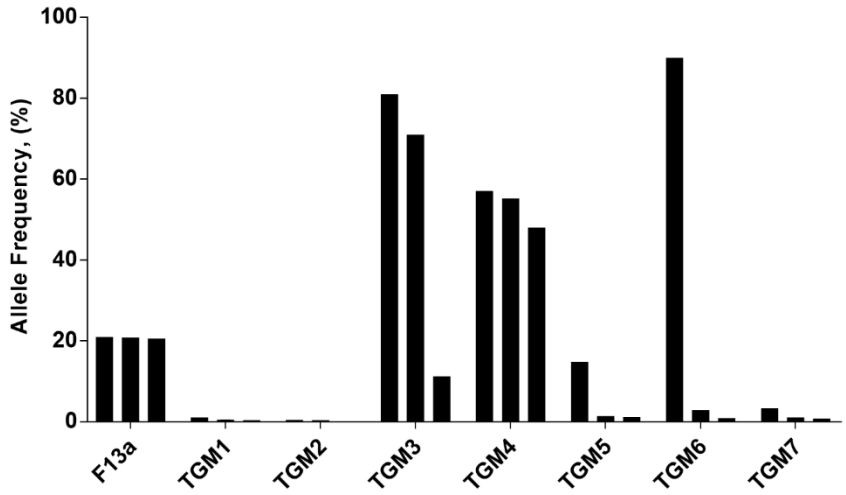


Figure 6. Allele frequency of three most frequent non-synonymous variants in transglutaminase family genes. ExAC database values were used. For each family, nsSNVs with top three allele frequency (%) values are shown. F13a: P565L (21 %), E652Q (20.8 %), V35L (20.6 %); TG1: V518M (1.04 %), E520G (0.56 %), S42Y (0.40 %); TG2: see in text; TG3: T13K (81 %), G654R (71 %), S249N (11.2 %); TG4: R372C (57 %), V409I (55 %), E313K (48 %); TG5: A352G (14.8 %), V504M (1.4 %), Q521R (1.25 %); TG6: M58V (90 %), R448W (2.80 %), A141E (0.9 %); TG7: P564L (3.4 %), V103L (1.07 %), V515L (0.82 %).

4.3.2. Biochemical analysis of homozygous TG2 variants

Occurrence of damaging TG2 nsSNVs at some functional sites of the protein prompted us to screen databases to see whether the homozygous form is tolerated when compared to nsSNV variants of the other transglutaminase family members. Besides, by biochemical assays we wanted to test the function of the protein products of TG2 nsSNV alleles found in homozygotes.

The revealed existing homozygous nsSNVs of transglutaminase family members from ExAc dataset are provided in Table 7A. The number of such homozygous individuals were highest for TG3, TG4, and TG6 (in the range of 47 to 73 thousands), while lowest for TG2, TG1, and TG7 (12, 25 and 170, respectively) (Table 7A). TG2 and TG1 had the lowest number of nsSNV alleles in homozygous form. All together 12 individuals in various populations in the World, contain only six TG2 nsSNVs in homozygous form, by far the lowest number in the transglutaminase family (Table 7B). Amidst these, probably damaging PolyPhen scores were predicted for R222Q, V542F, P612T, and P612T was highly destabilizing in both closed and opened conformations, whereas V542F and D671N mostly affect the closed form (Table 7B).

Table 7. Summary of population frequencies of nsSNV alleles of transglutaminases

A

Genes	Total number of nsSNV types	Allele count	Number associated with homozygotes	Number of homozygote individuals
F13a	283	88654	9	9214
TG1	320	5174	13	25
TG2	272	2601	6	12
TG3	313	205153	12	73495
TG4	279	265931	17	66931
TG5	273	27038	12	2206
TG6	336	111028	23	47462
TG7	293	11585	13	170

B

Position and change	Allele frequency, [%]	Domains	Stability ΔΔG [kcal/mol]		PolyPhen/ SIFT	Number of homozygote individuals	Population
			closed form	open form			
R76His	0.47	β-sandwich	0	0.57	Benign	5	2 East Asian 3 Latino
R222Q	0.048	Catalytic	-0.98	0.71	Probably damaging	1	South Asian
R433Q	0.10	Catalytic	0.67	0.51	Benign	1	South Asian
V542F	0.38	β-barrel 1	2.29	-1.11	Probably damaging	3	2 East Asian 1 African
P612T	0.018	β-barrel 2	2.26	3.83	Probably damaging	1	South Asian
D671N	0.013	β-barrel 2	1.17	0.10	Benign	1	African

A. Number of homozygous nsSNVs and carriers in transglutaminase family. Total number of nsSNVs in column 2; observed allele counts for the nsSNVs in column 3; number of nsSNVs associated with homozygotes in column 4; number of homozygote individuals in column 5. **B. The 12 human TG2 homozygotes related to 6 nsSNVs.** homozygous nsSNVs of TG2 in column 1; allele frequency of the given nsSNV in column 2; domain distribution of TG2 nsSNVs are shown in column 3; stability values in closed and opened conformations are shown in column 4, 5; PolyPhen/SIFT scores in column 6; number of homozygote individuals for each nsSNV in column 7 and population distribution of the 12 homozygote individuals in column 8.

We analyzed the functional impact of TG2 homozygous nsSNVs since they can be associated with various diseases. By site directed mutagenesis all the six homozygous occurring TG2 variants containing the respective nsSNVs were produced and tested in biochemical assays. Using two previously published kinetic assays, transglutaminase activity of the variants was

measured (Lorand et al. 1998; Biri et al., 2016). The transamidase activity of the variants was comparable in both the assays except R222Q, which was completely active in amine incorporation assay but inactive in the protein crosslinking assay (Figure 7). Compared to the wild type enzyme the variant R76H displayed increase in transamidase activity and variant V542F showed 40 % less transamidase activity. P612T variant exhibited 40 % less activity in the amine incorporation assay but only 18 % less activity in the protein crosslinking assay compared to the wild type. Interestingly PolyPhen/SIFT damaging scores were predicted for both V542F and P612T variants (Table 7B). Based on the amine incorporation assay calcium dependence of the transamidase reaction was also measured (Figure 7). At both measured Ca^{2+} concentrations, R76H exhibited high transamidase activity and variants V542F, P612T had less activity than wild type (Figure 7). Compared with wild type at 0.25 mM Ca^{2+} concentration variants R433Q, D671N manifested a several fold increase in the transamidase activity. At 0.25 mM calcium concentration R222Q showed the lowest efficiency, but increasing of which restored activity. Transglutaminase reactions are reciprocally regulated by Ca^{2+} and nucleotides. By using BODIPYFL-GTPγS analog the nucleotide binding of the variants was examined. Compared to the wild type 18 % higher GTP binding was observed for R76H and D671N and 24 % low GTP binding in case of P612T at 250 nM enzyme concentration (Figure 7).

By using a commercially available small chemically produced substrate the isopeptidase activity of the six variants was measured. The R222Q variant was 15 % active and the variants V542F and P612T displayed less isopeptidase activity compared to the wild type (Figure 7). While compared to the wild type, variants R76H, D671N showed increase in isopeptidase activity. The fibronectin-binding property of the variants was tested by an ELISA method. Around 15 % less fibronectin binding was observed for the variants V542F, P612T but other variants bound fibronectin similarly to the wild-type enzyme (Figure 7).

Novel clusters, functional sites, and predicted IDRs were not influenced by homozygous variants R76H and D671N with benign or tolerated scores. R433Q variant is part of the Ca^{2+} binding site S5 (432-440) and IDR (428-473) embedding SLiMs like USP7 binding motif, TRAF2 binding motif, and CK1 and CK2 phosphorylation motif. V542F variant located within the MOD CK1 phosphorylation site and MOD PLK site phosphorylated by polo like kinases has damaging PolyPhen score and exhibits decreased activity. Damaging P612T variant is near to an IDR (597-602) and a novel amino acid cluster PVA (613-615) (described in section 4.1) and displays less activities and GTP binding. The C-terminal class 3 PDZ-binding motif contains the homozygous variant D671N.

The PolyPhen damaging variant R222Q with very low transamidase activity at physiological Ca^{2+} concentration and low isopeptidase activity falls in the catalytic core domain near the Ca^{2+} binding site S1 (228-236) and SLiM STAT5 SH2 domain binding motif. The impact of R222Q allele on the structure was studied by comparing three situations where in a metal occupies none, only site 1 or all the three calcium binding sites (Figure 8). The H-bond is disturbed by the Q222 variant leading to altered conformation of the loop, P359-G372, and consequently to reduced calcium affinities at site 1 and 2. The resulting topology disfavors proper interaction with a protein amine donor.

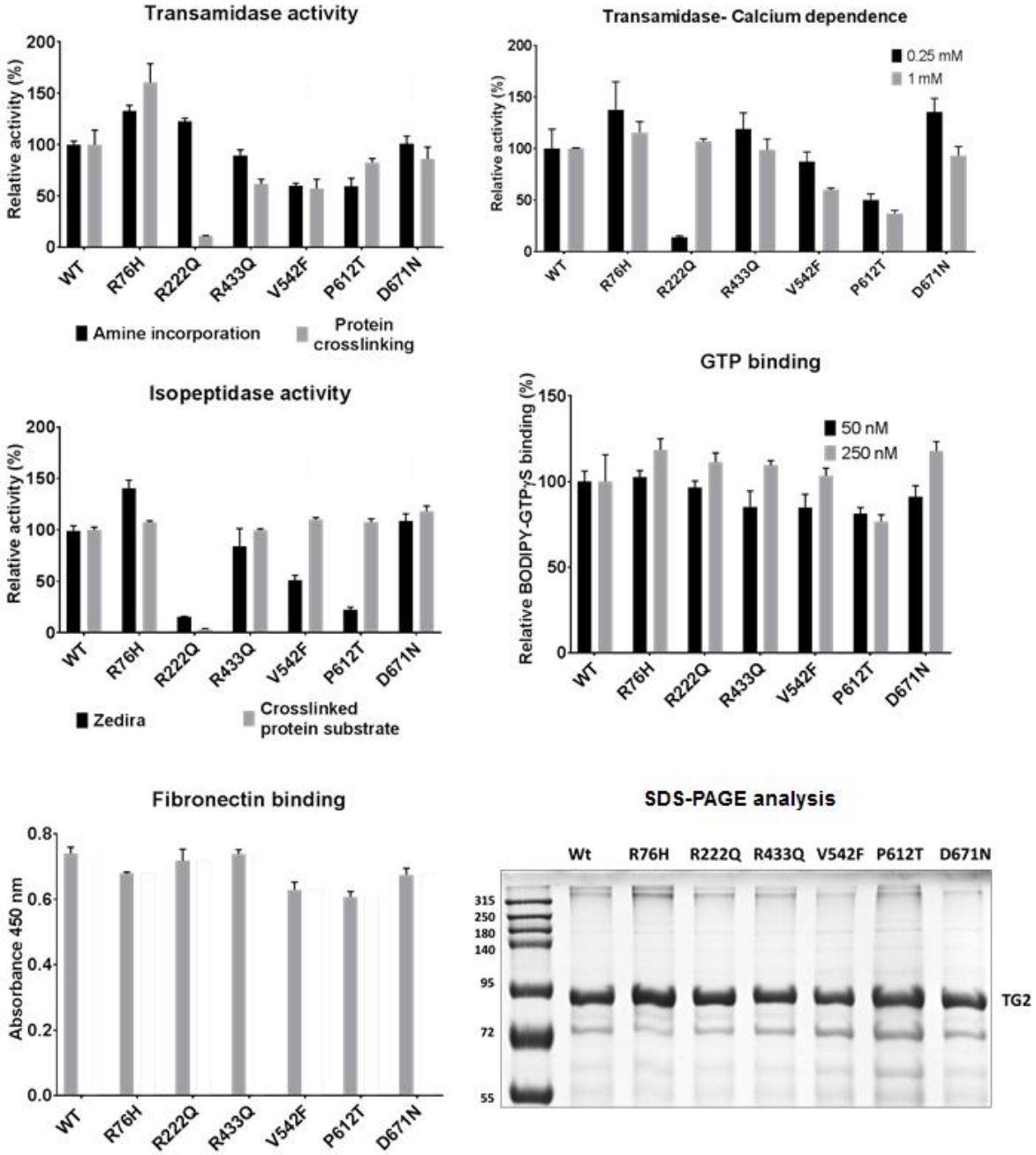


Figure 7. Biochemical characterization of the six homozygous nsSNVs of TG2. Based on kinetic amine incorporation assay and protein crosslinking assay at 3 mM Ca^{2+} concentration, transamidase activity of TG2 wild-type and variants was characterized (Lorand et al., 1998; Biri et al., 2016). Amine incorporation assay with 0.25, and 1 mM calcium concentrations was used to measure calcium dependence of transamidase reaction. Kinetic Zedira assay was used for measuring the isopeptidase activity of TG2 variants at 5 mM Ca^{2+} concentration. The relative activities were calculated as a percentage of the activity values of the wild type TG2. Different concentrations of TG2 (50, and 250 nM) was used to compare the BODIPY-FL-GTPγS binding of variants and wild type TG2 proteins. After 15 minutes of incubation, the change in the fluorescence intensity (Ex/Em: 485/520 nm) was determined. Binding is shown as a percentage of maximum binding in case of wild type TG2. By using previously published direct ELISA assay (Kiraly et al., 2013) fibronectin binding of TG2 variants was performed. Data are presented as means with ± standard deviations from three separate experiments done in triplicate. All the data was analyzed by GraphPad Prism 7. Equal amount of purified proteins were separated by 10 % SDS-PAGE (polyacrylamide gel electrophoresis) and stained using Coomassie Brilliant Blue R250 dye (Thermo Fisher Scientific, Waltham, Massachusetts, USA).

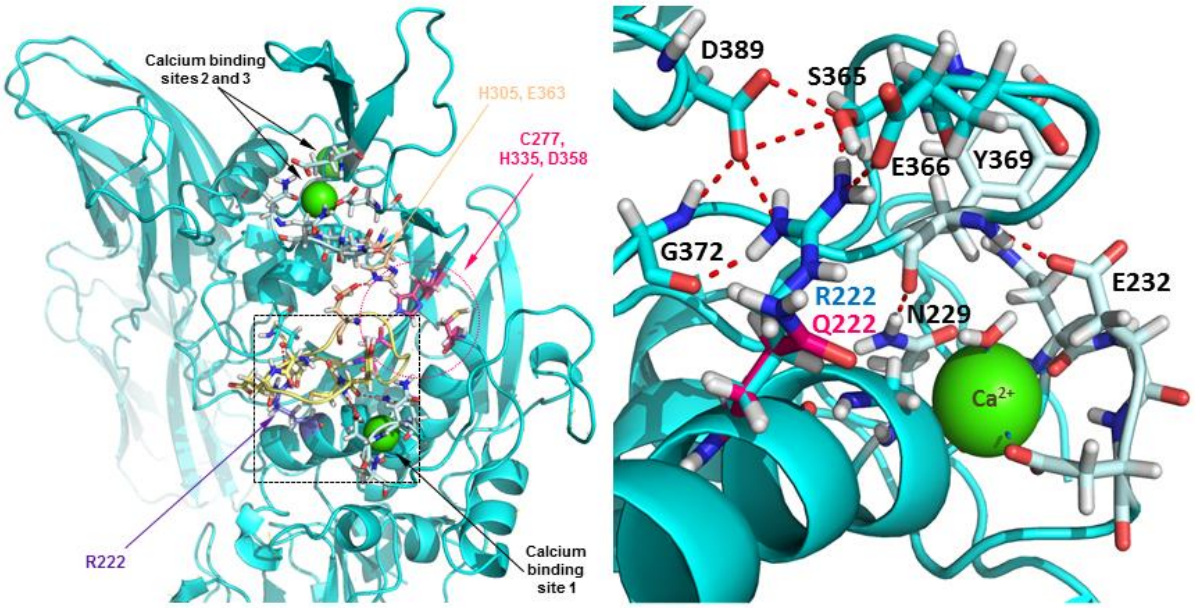


Figure 8. Structural interpretation of the effect of the R222Q variation on enzyme activity. The R222 interactions were analyzed using a homology model of TG2 containing three bound calcium ions (green spheres), which supposedly corresponds to the active form. The squared area in the left panel is magnified and shown from an angle rotated by 90° on the right. R222 is located in the middle of the solvent accessible surface of the α-helix leading up to calcium binding site S1 (226-233). R222 is at the core of an H-bond network (red dashed lines) that serves to bundle neighbouring structural elements of TG2 together. Upon binding of calcium to site 1, H-bonds between E232, N229 and backbone atoms of Y369, and H-bonds of R222 to S365, E366, G372 and D389 cooperatively tether the flexible loop, P359-G372 (yellow). The changing conformation of this loop leads to reorganization of another non-covalent interaction network near calcium binding site 2, including directly a calcium binding residue, N306 for metal binding, and to honing of the charge relay duad, E305-E363

that has recently identified importance for catalysis (Stieler et al., 2013). The Q222 variant fails to establish the critical H-bonds with S365, E366, and E389, thus the calcium binding of both sites are impaired and the charge relay system is also negatively affected. The same loop, most probably, also contributes residues to the amine substrate binding surface and controls access to the active site (C277/H335/D358), likely explaining that the Q222 enzyme has conserved transamidase activity for a small molecular amine, but compromised cross-linking activity towards a protein amine donor and lost isopeptidase activity for a protein-peptide conjugate. This figure was prepared by Dr. Máté Demény.

4.3.3. Heterozygous and homozygous LOF variants of transglutaminases

The damaging nsSNVs of TG2 were rare and those found in homozygous individuals did not lose the biochemical function completely (one displayed reduced activity). This prompted us to investigate the population frequency of TG2 LOF variants to explore the relationship between TG2 and human diseases. So far, TG2 disease causing mutants have not been found and no TGs are among the 3230 haploinsuffiency genes, defined as genes with knockout variants at least in heterozygous state (Lek et al., 2016).

Based on the available data we compiled the LOF variants of transglutaminase family members from different populations. As part of the pilot phase of the 1000 Genomes Project exon sequencing data from 185 individuals was analyzed and 2951 LOF variants, rare and likely deleterious LOF alleles, including 26 known and 21 predicted severe disease-causing variants were reported (MacArthur et al., 2012). A homozygous TG6 LOF variant is present in this dataset. Recently, in the Icelandic population a list of autosomal genes with rare LOF mutations have been published (Sulem et al., 2015). The sequence variants identified in whole genome sequencing of 2,636 Icelanders were imputed into a set of 101,584 additional chip-genotyped Icelanders. Because of rare loss-of-function variants, 1,171 genes in 8,041 individuals (7.7 %) are completely knocked out. Icelandic dataset has homozygous LOF variants belonging to transglutaminase family members TG1, TG4, and TG5 (Sulem et al., 2015). Recently, 1111 rare gene knockouts in 781 genes were identified in 3222 exomes of consanguineous British population of Pakistani heritage. Single homozygous TG4 LOF variant was identified in this dataset (Narasimhan et al., 2016). In an aim to improve the understanding and research of autism spectrum disorders a project called Simons Simplex Collection was established. A homozygous TG4 LOF variant was identified in this dataset (Lim et al., 2013). In Atherosclerosis Risk in Communities cohort study samples, heterozygous LOF variants for all the transglutaminases were identified (Li et al., 2015). Loss-of-function heterozygotes were identified for all the members of the family in ExAC dataset but in homozygous form only for TG4 and TG6 (Lek et al., 2016). Altogether, human

TG2 has low intolerance to genomic variations compared to other members of transglutaminase family. Variants influence the protein stability and functional sites but do not influence active site and some critical residues. Homozygous variants are very rare and even those present have functional activities.

4.4. Separation and comparison of transamidase and isopeptidase activity of transglutaminase 2 and their biological relevance

4.4.1. In silico considerations for the separation of transamidase and isopeptidase activity of TG2

Crystal structures of human transglutaminases compared with corresponding papain structures reveals that the active site of both are surrounded by bulky hydrophobic residues, and by being buried in the core domain they are isolated from ambient water (Pinkas et al., 2007, Chica et al., 2004, Iismaa et al., 2003, Murthy et al., 2002). The active site is quite exposed to water in the presence of substrate (inhibitor) as indicated by space-filling models. But in substrates absence the hydrophobic tunnels isolated by annealing hydrophobic side chains, such as W241-W332, W278-F334 repell water from the catalytic cavity. Active site is also shielded from the intrusion of water by W180, W337 and Y510 residues. Reaction catalysis of the human TG2 not only depends on the hydrophobicity of the amino acids around the active site and in the substrate binding area. It also depends on the complexity of the surrounding Tryptophan residues (space filling property, aromatic carbon atom ring containing nitrogen), which have important effect on stabilisation of the transition states of transglutaminase activities (reviewed in Keillor et al., 2014). Study by Pinkas et al. (2007) showed W332F mutation in rat TG2 lead to loss of transamidase activity but isopeptidase activity was not reported. In relation to this, Keillor et al. (2008) targeted selected amino acids around the active site in guinea pig TG2 and identified Trp332 which when replaced by Phe or Tyr conferred very low transamidation activity on the enzyme but this study did not target Trp278 (W278F).

Replacement of one or two of the seven residues (W241-W332, W278-F334, W180, W337 and Y510) by in silico modeling suggested that these residues would make the catalytic core "leaky" to the solvent water molecules without disrupting secondary domain structures. The catalytic site exposed to ambient water is expected to divert transglutaminase activity towards hydrolysis of the transglutaminase reaction products. Amino acids with shorter side chains were designed to prevent the disruption of the active site and to save the hydrophobic character. Based on this consideration we designed 12 single mutants. But only mutants W278F, W332F, inactive conserved Trp mutant W241F (control) and active site mutant C277S (control) relative to this dissertation are discussed here. In order to characterize the isopeptidase property of human TG2 these mutants have not yet been purposefully created in other studies (Murthy et al., 2002, Iismaa et al., 2003).

4.4.2. Characterisation of transglutaminase 2 mutants with dominant isopeptidase or transamidase activity

By site directed mutagenesis two special mutants, one deficient in transamidase activity with higher isopeptidase activity (W332F) and another with opposite properties (W278F) were found and characterised. Based on the genetic studies (e. g. 1000 Genomes) 224 position in TG2 contains valine but TG2 clone used by laboratories in the transglutaminase research field had Gly amino acid at 224 position. But in 2013, we reported structural significance of 224 position in TG2 and biochemical differences between Val^{224} and Gly^{224} containing variants. So we produced the Val^{224} containing wildtype and mutants and compared the transamidase and isopeptidase activities with the Gly^{224} containing variant. Compared to the Gly^{224} variant, Val^{224} wildtype TG2 displayed 33 % higher transamidase activity (Figure 9). As expected the Gly^{224} and Val^{224} W332F mutants (TG2-I) were deficient in transamidase activity. The Gly^{224} and Val^{224} W278F mutants (TG2-T) had similar transamidase activity compared to Gly^{224} variant and Val^{224} wildtype respectively (Figure 9). Regarding isopeptidase activity, Val^{224} wildtype TG2 showed 45 % higher activity than Gly^{224} variant. The Val^{224} W332F and W278F mutants showed two fold higher isopeptidase activity than Gly^{224} variant mutants respectively. Compared to Val and Gly 224 variants, the corresponding W332F TG2 mutants presented 1.5 and 1.3 times higher isopeptidase activity respectively. But Gly and Val 224 W278F mutant showed very less isopeptidase activity compared to the wildtype and W332F mutant. The C277S active site mutant was completely inactive in all measurements but Val^{224} W241F mutant showed detectable transamidase but no isopeptidase activity (Figure 9).

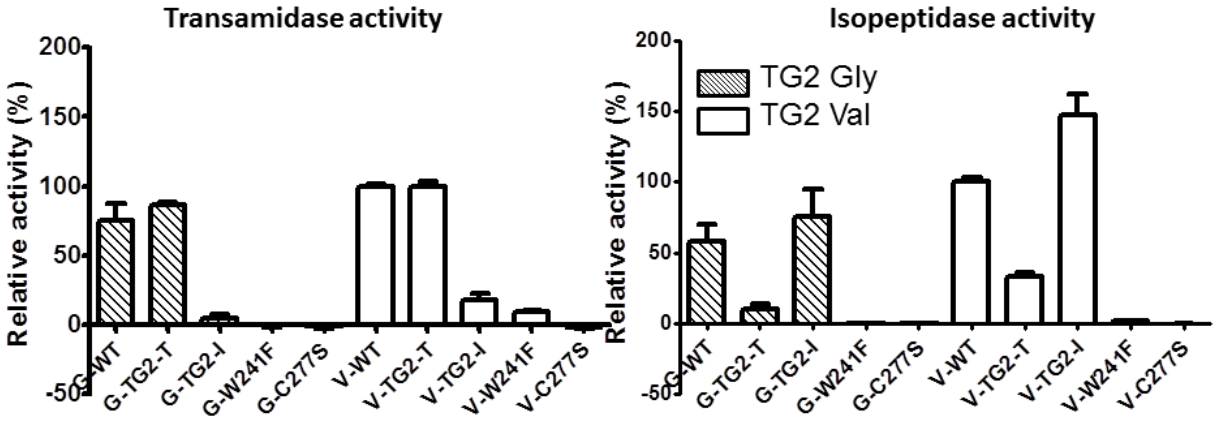


Figure 9. Kinetic characterisation of the transamidase and isopeptidase reaction of TG2-I and TG2-T and wild-type TG2. Comparison of the transamidase and isopeptidase activities of mutant TG2 proteins containing Gly variant or Val wild-type at 224 amino acid position. G and V indicates glycine or valine at 224 position and TG2-T refers to W278F

mutant, while TG2-I refers to W332F mutant. Transamidase activity was tested using, endpoint microtiter plate transamidase method (biotin-T26 peptide incorporation into immobilized spermine) (Perez et al., 2009) in the presence of 0.5 µg enzyme and isopeptidase reaction using A102 substrate; (Zedira) with 10 µg enzyme. The relative activities were calculated as a percentage of the activity values of the human Val^{224} containing wild type TG2. Data are presented as means with ±SD from two or three separate experiments done in triplicate. This figure was prepared by Dr. Róbert Király.

4.4.3. Role of transamidase and isopeptidase activities on cell aggresome system in rat huntington cell model

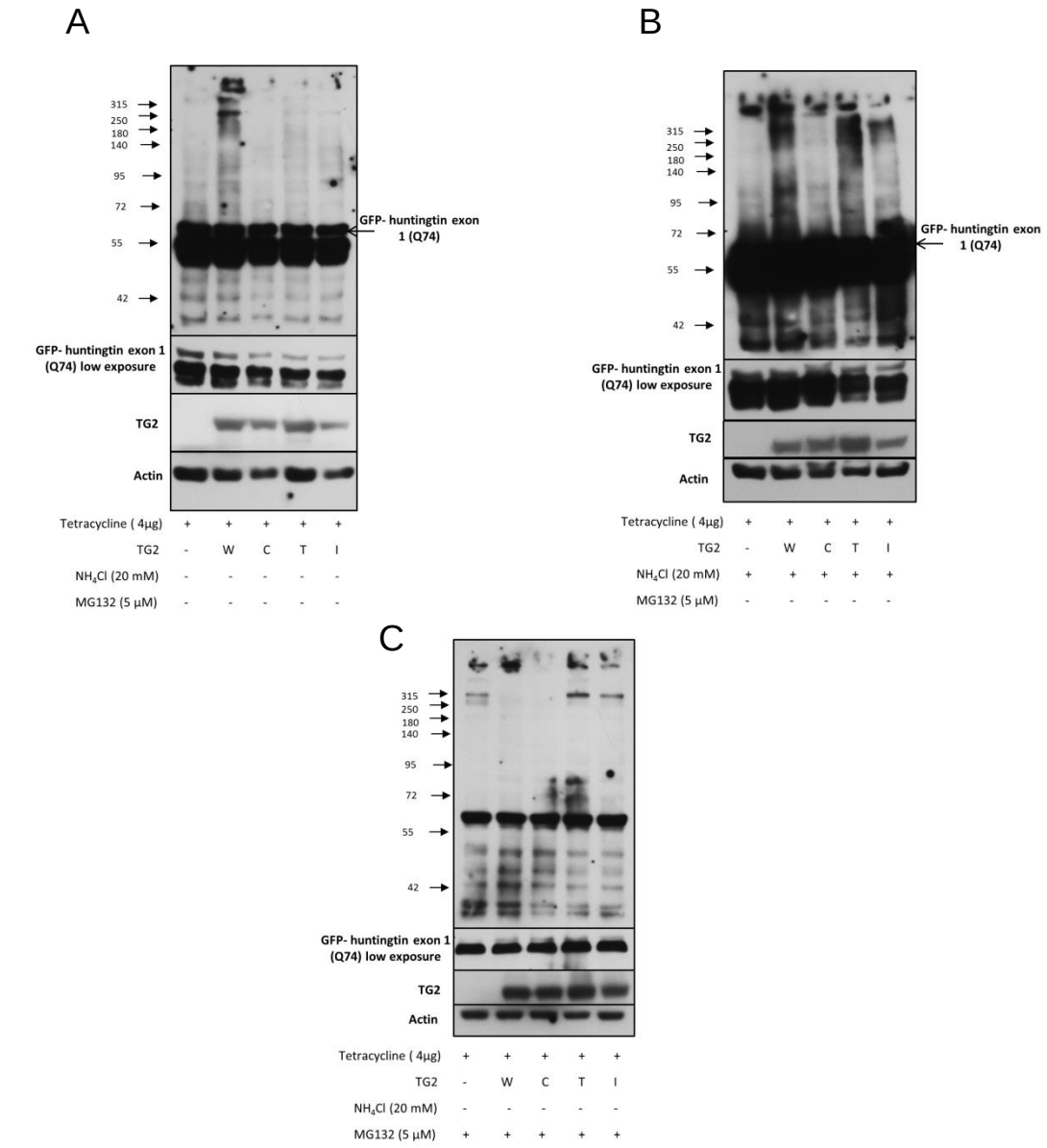
The above characterized transamidase (W278F, TG2 T) and isopeptidase mutants (W332F, TG2 I) along with wild-type and C277S active site mutant was used to determine the putative role of TG2 in aggregate formation and mutant huntingtin crosslinking. The *in situ* effect of TG2 on aggregate formation was studied in PC12 neuronal rat cells, which express tetracycline inducible GFP-tagged exon 1 fragment of htt gene with 23 (control) or 74 (prone to aggregate formation) glutamine repeats and stably transfected with human wild type TG2 and TG2 mutants described above. Before transfection the PC12 cells were checked for endogenous TG2 activity by radiolabelled putrescine incorporation assay and westernblot. The results showed that there is no endogenous TG2 activity in PC12 cells (data not shown).

In first set of experiments, cells were induced with 1, 2 or 4 µg/ml of tetracycline for 24 hrs, 48 hrs and 66-72 hrs. The aggregates were visualized in microscopy and exon 1 fragment of huntingtin was confirmed by westernblot. As 4 µg/ml tetracycline induction for 66 hrs displayed more aggregates, this setting was used for further experiments. In a next set of experiments, we analyzed whether TG2 can modify exon 1 fragment of huntingting with 74 glutmine repeats. The antibody aganist GFP identified few bands around 55 kDa, but the band slightly above 55-kDa is likely to be the exon-1 fragment of mutant huntingtin and it was confirmed by 1C2 antibody aganist polyglutamine repeats (data not shown). As reported in Figure 10 A there was no significant difference in the GFP tagged exon1 fragment of huntingtin (band around 57 kDa) in the presence or absence of TG2. But high molecular weight polymers were observed only in PC12 Q74 cells transfected with wild type TG2 (Figure 10 A). Without tetracycline addition these crosslink polymers were not observed (data not shown).

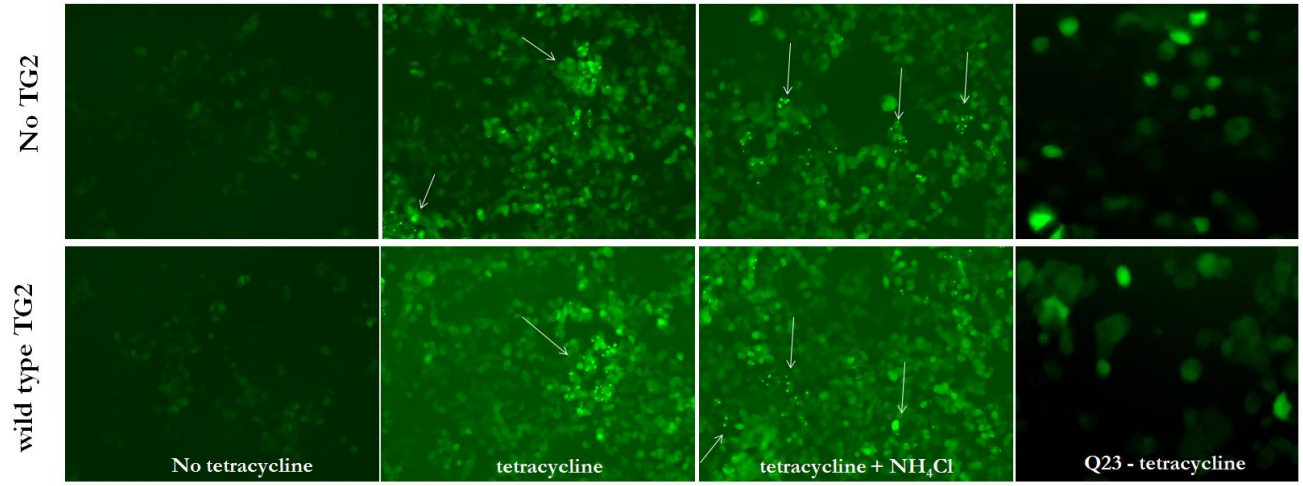
It is well known that aggregates are cleared either by ubiquitin-proteasome or autophagy lysosomal system. In line with this, we performed autophagy and proteasome inhibition experiments to study the effect of TG2 on huntingtin aggregates. For inhibiting autophagy we used NH_4Cl, which hampers the activation of the lysosomal enzymes, hence blocking the

degradation process. After 48 hrs of tetracycline addition, NH_4Cl was added for 18 hrs with fresh addition of tetracycline. Pathological huntingtin polymers were more pronounced in cells expressing Wt TG2 and TG2 T mutant compared to cells without TG2 or expressing either TG2 I or TG2 C277S mutant (Figure 10 B). There was no drastic difference in accmulation of autophay markers LC3 I, II, an ubiquitin binding protein p62/SQSTM1 and E3 ubiquitin ligase, Atg 5 before and after autophagy inhibition (11 A, B). There was an increase in aggregates after autophagy inhibition in the presence and absence of TG2 (Figure 10 D).

In next experiments the cells were subjected to proteasome inhibition using commercially avaliable proteasome inhibitor MG132. After 48 hrs of tetracycline addition, MG132 was added for 18 hrs with fresh tetracycline. Compared to autophagy inhibition, less crosslink polymers were observed when proteasome was blocked (Figure 10 C). After proteasome inhibition only one band was observed around 55 kDa compared to three in absence of the same or in the presence of autophagy inhibitor. After proteasome inhibition, there was no remarkable difference in the accmulation of LC3 I, II and Atg 5 and p62/SQSTM1 levels (Figure 11 C). Moreover the control cells with Q23 repeats induced for exon 1 fragment of huntingtin and transfected with Wt TG2 did not show any crosslink polymers (Figure 12 A). The was also no difference in the endogenous full length huntingtin levels in Q74 cells transfected with TG2 Wt and I mutant (Figure 12 B).


A
315
250
180
140
95
72
55
42
GFP- huntingtin exon 1 (Q74)
GFP- huntingtin exon 1 (Q74) low exposure
TG2
Actin
Tetracycline (4µg) + + + + +
TG2 - W C T I
NH4Cl (20 mM) - - - - -
MG132 (5 µM) - - - - -
B
315
250
180
140
95
72
55
42
GFP- huntingtin exon 1 (Q74)
GFP- huntingtin exon 1 (Q74) low exposure
TG2
Actin
Tetracycline (4µg) + + + + +
TG2 - W C T I
NH4Cl (20 mM) + + + + +
MG132 (5 µM) - - - - -
C
315
250
180
140
95
72
55
42
GFP- huntingtin exon 1 (Q74) low exposure
TG2
Actin
Tetracycline (4µg) + + + + +
TG2 - W C T I
NH4Cl (20 mM) - - - - -
MG132 (5 µM) + + + + +

D


No TG2
wild type TG2
No tetracycline
tetracycline
tetracycline + NH4Cl
Q23 - tetracycline

E

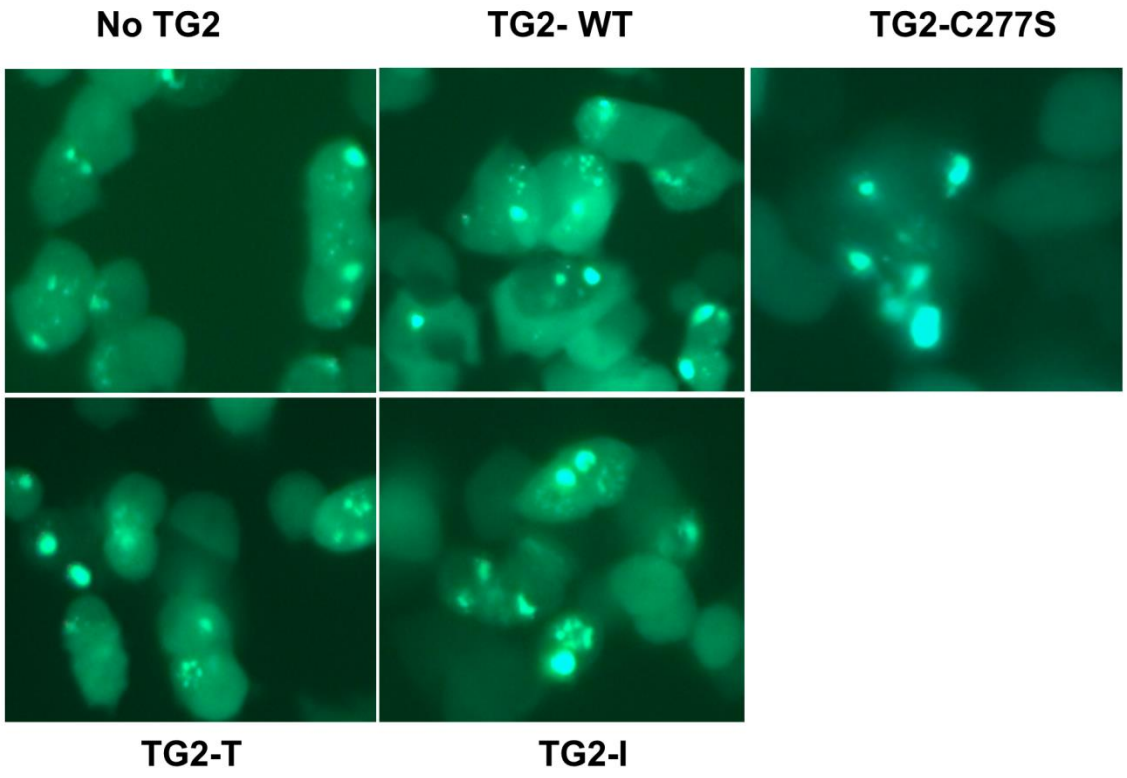


Figure 10. Influence of TG2 activities on exon 1 fragment of huntingtin with Q74 repeats. 50 µg of proteins were separated by 10 % SDS-PAGE and analysed by western blotting with anti-GFP, TG2 and actin antibodies (see materials and methods for details). **(A)** Exon 1 fragment of mutant huntingtin was induced for 66 hours with 4 µg/ml tetracycline and after 48 hours **(B)** cells were either blocked for autophagy (18 hours) with NH_4Cl or **(C)** for proteasome (18 hours) with MG132. W- Wild type TG2, C- C277S TG2 mutant, T- W278F TG2 mutant, I- W332F TG2 mutant. First panel in part A, B, C shows the whole blot developed for anti-GFP and the second panel shows only the exon-1 fragment of huntingtin with low exposure. **(D)** Microscopic images taken after 66 hrs of tetracycline addition and 18 hrs of autophagy inhibition. Aggregates are marked with arrows. Last panel shows the PC12 cells with Q23 glutamine repeats induced for exon-1 fragment of human huntingtin. **(E)** Magnified view of aggregates in PC12 Q74 cells transfected with different TG2 and after 48 hrs of tetracycline addition.

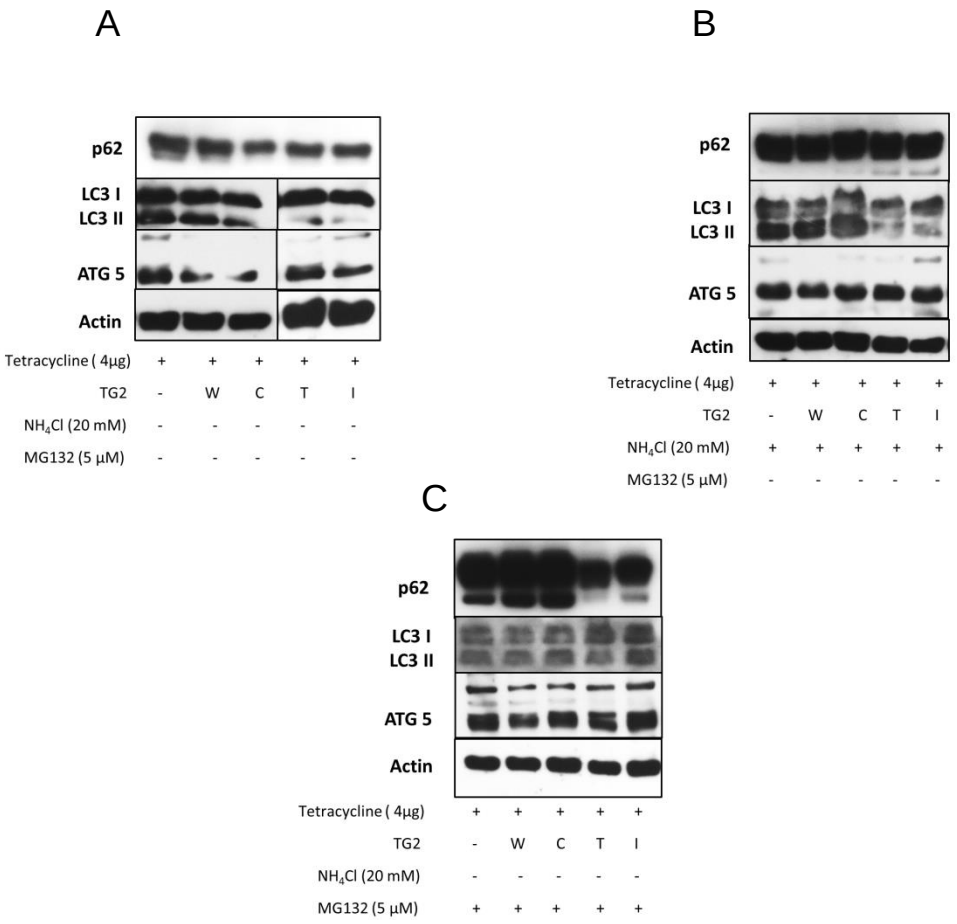


Figure 11. Expression of autophagy associated proteins. 50 µg of proteins were separated by 10 % SDS-PAGE and analysed by western blotting with p62, LC3 I, II, ATG 5, and actin antibodies (see materials and methods for details). Exon 1 fragment of mutant huntingtin was induced for 66 hours with 4 µg/ml tetracycline **(A)** and after 48 hours cells were either blocked for autophagy (18 hours) with NH_4Cl (**B)** or for proteasome (18 hours) with MG132 **(C)** W- Wild type TG2, C- C277S TG2 mutant, T- W278F TG2 mutant, I- W332F TG2 mutant.

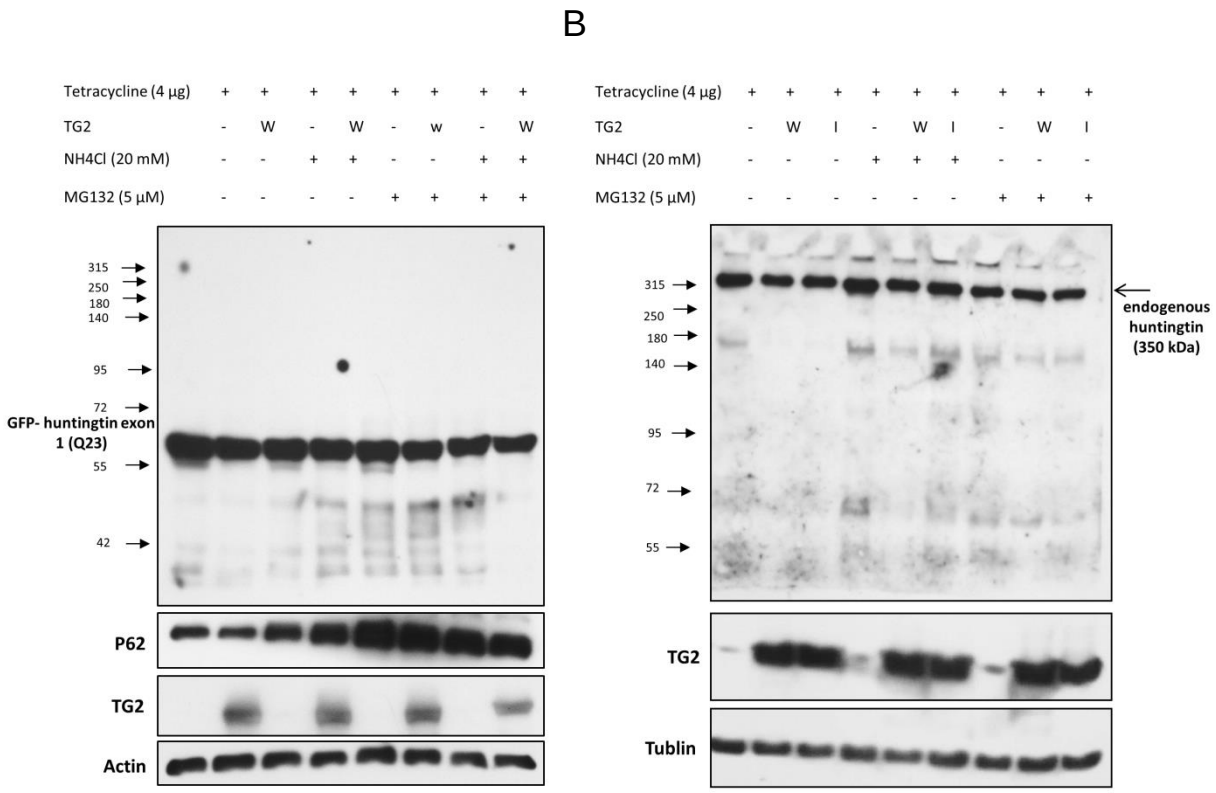


Figure 12. Influence of TG2 on exon 1 fragment of huntingtin with Q23 repeats and endogenous full length huntingtin. (A) 50 µg of proteins were separated by 10 % SDS-PAGE and analysed by western blotting with TG2, p62, and actin antibodies (see materials and methods for details). Exon 1 fragment of mutant huntingtin was induced for 66 hours with 4 µg/ml tetracycline (all lanes) and after 48 hours cells were either blocked for autophagy (18 hours) with NH_4Cl (lanes 3, 4) or for proteasome (18 hours) with MG132 (lanes 5, 6) or for both autophagy and proteasome (lanes 7, 8). W- Wild type. The Q23 blot was exposed for a shorter time compared to Q74 blots in Figure 10. **(B)** 50 µg of proteins were separated by 7.5 % SDS-PAGE and analysed by western blotting with MAB 2166 huntingtin, TG2, and tubulin antibodies (see materials and methods for details). PC12 cells expressing exon 1 fragment of huntingtin with Q74 repeats was used for this experiment. Exon 1 fragment of mutant huntingtin was induced for 66 hours with 4 µg/ml tetracycline (all lanes). After 48 hrs the induced cells were blocked for autophagy (18 hours) with NH_4Cl (lanes 4, 5, 6) or for proteasome (18 hours) with MG132 (lanes 7, 8, 9). W- Wild type, I- W332F TG2 mutant.

Taken together, the transamidase and isopeptidase activities of TG2 were successfully separated by site-directed mutagenesis. TG2 transamidase activity was shown to be involved in the formation of covalently cross-linked protein polymers and the potential role of isopeptidase activity in reversing the protein crosslinks was also demonstrated.

4.5. Real time kinetic assay to study the isopeptidase activity of transglutaminase 2

4.5.1. Design and implementation of the fluorescence anisotropy assay

In a recent study, S100A4 turned out to be a good amine donor substrate of TG2 (Biri et al., 2016). The product formed by crosslinking the fluorescently labelled FLpepT26 dodecapeptide and S100A4 was purified and used as a substrate in the isopeptidase assay. For isopeptidase activity measurements, both FLpepT26-BSA and FLpepT26-S100A4 cross-linked products were considered as substrates. S100A4 has only one TG2 reactive Lys residue (Lys100) and does not contain transglutaminase reactive glutamine residues unlike BSA. But BSA contains several surface exposed Lys residues. In line with earlier observations, the enzymatic dissociation constant (Km) value of both the substrates were comparable for transamidation reaction but 10-fold lower Vmax was determined for BSA compared to S100A4. This suggest that to produce an isopeptidase protein substrate with a single isopeptide bond, S100A4 (GST) could be a very potent Lys donor. S100A4 was expressed with GST-tag in order to get a higher anisotropy change, better fluorescence polarization signal and to compensate for the size difference between S100A4 (11.5 kDa) and BSA. We confirmed that GST did not effect TG2-catalyzed transamidation reaction as TG2 did not crosslink GST alone into FLPepT26 (Figure 13 A and B).

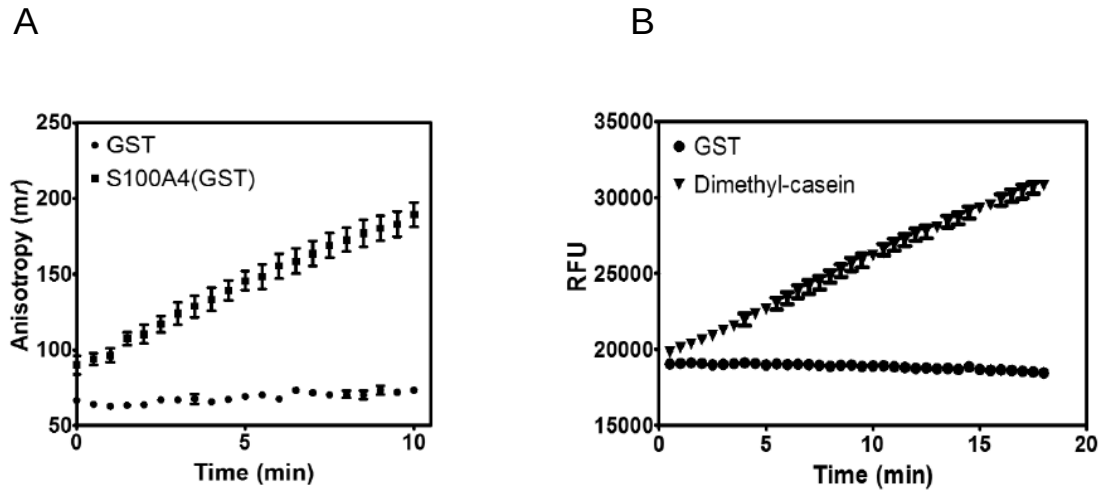


Figure 13. Analysis of the effect of GST on the crosslinking activity. Effect of GST was checked by crosslinking anisotropy assay with FLpepT26 peptide **(A)** and dansyl-cadaverine incorporation assay **(B)**. Results of the anisotropy assay and amine incorporation assay show that GST was not utilised neither as an amine donor nor as an amine acceptor respectively. Representative data from two independent experiments done in triplicate. Data are presented as means with ±SEM, standard error of the mean from two separate experiments done in triplicate.

The substrate for the isopeptidase reaction, FLpepT26-S100A4 (GST) containing the Nε-(γ-glutamyl) lysine bond was produced in larger scale by carrying out crosslinking reaction as described in Materials and Methods. In order to avert unwanted further modification of the crosslinked molecules during their separation from free FLpepT26 peptide the reaction was stopped by the addition of 10 mM EDTA (final concentration) and for optimal isopeptidase activity the buffer was replaced to 3-(N-morpholino)propanesulfonic acid (MOPS) buffer, pH 6.8 (Fleckenstein et al., 2002) (Scheme on Figure 14 A). SDS-PAGE was used to confirm the efficiency of the separation and at ~42 kDa the presence of the crosslinked, fluorescently labelled FLpepT26-S100A4 (GST) molecule was confirmed. Furthermore, disappearance of the 1885 Da FLpepT26 peptide which normally appears in the gel as a 10 kDa diffuse band was also confirmed (Figure 14 B).

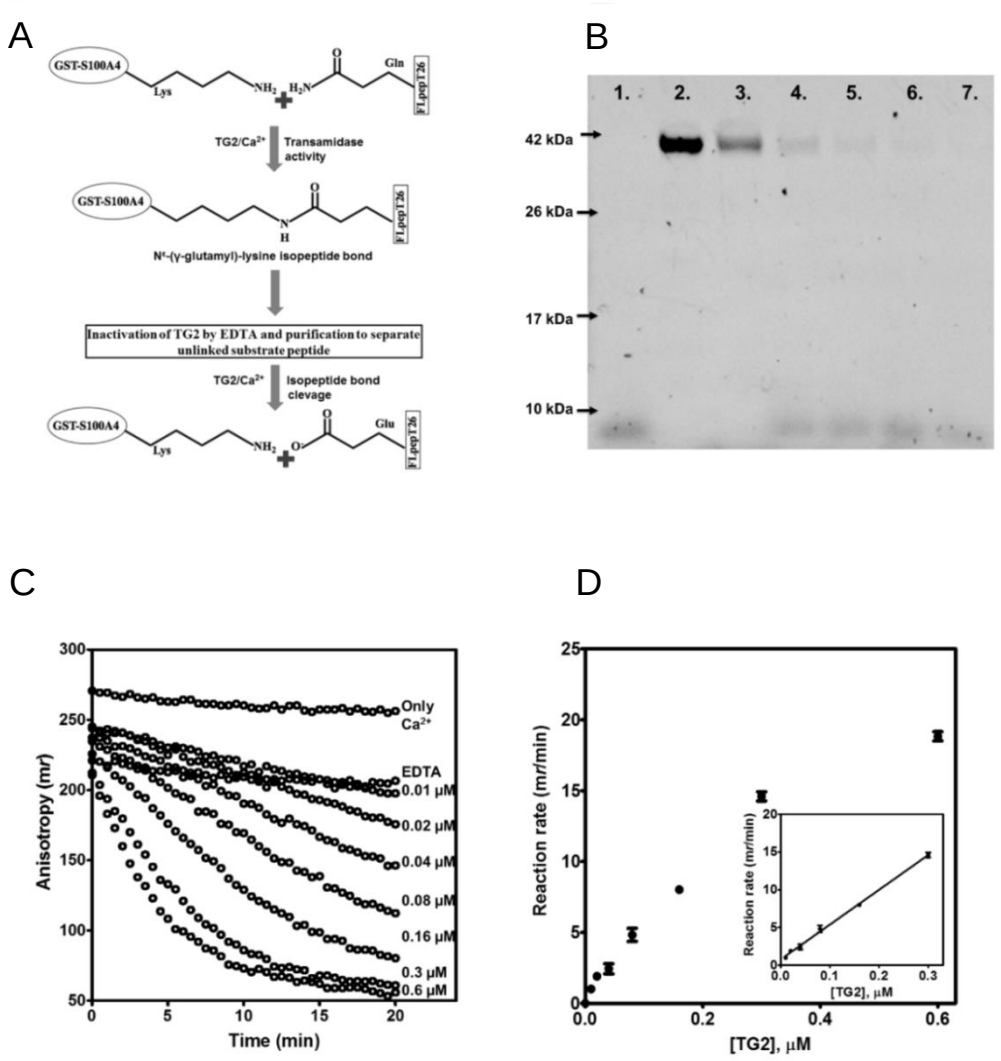


Figure 14. Design and implementation of the fluorescence anisotropy assay that monitors isopeptidase activity of TG2. (A) Work scheme: TG2 mediated crosslinking of

FLpepT26 peptide into S100A4 (GST) and the isopeptide containing FLpepT26-S100A4 (GST) serves as isopeptidase assay substrate after separation from the free peptide. **(B)** Visualization of the isopeptide cleavage on FLpepT26-S100A4 (GST) by 15 % SDS-PAGE; the kinetic reactions were performed for 90 mins at 37°C. Representative gel image of FLpepT26 peptide (11.3 ng peptide, lane 1), separated FLpepT26-S100A4 (GST) isopeptidase substrate (8.6 µg protein, lane 2), product of isopeptidase reactions on the substrate in the presence of 300 nM TG2 and 5 mM EDTA (lane 3) or in the presence of 5 mM Ca^{2+} and different TG2 concentrations (0.06, 0.3, 0.6, 1.2 µM, lane 4-7, respectively; 2.8 µg substrate protein was used). **(C)** The isopeptide cleavage was monitored in real-time by measuring the anisotropy change. The 35 µl reaction mix contained 5 µg substrate protein (0.5 µM FLpepT26-S100A4 (GST), 5 mM Ca^{2+} or EDTA and various concentrations of TG2. **(D)** The rate of anisotropy change shows linear correlation between 10-300 nM TG2 concentration. The linear range analysed using Pearson correlation analysis ($P<0.0001$, $r=0.990$) is shown in the box. Data are presented as means with ±SD from three separate experiments done in triplicate.

Before it can be used as a substrate for TG2 isopeptidase activity measurement, the purified crosslinked product was validated. SDS-PAGE clearly displayed the cleavage of isopeptide bond in FLpepT26-S100A4 (GST) by TG2. During the enzymatic reaction there was a decrease in amount of substrate while the amount of released fluorescein-conjugated peptide showed a marked increase (Figure 14 B). The isopeptidase activity was validated by applying the product of the isopeptidase reaction to a Jupiter 300 C5 RP-HPLC column (Phenomenex) and HPLC-ESI-MS/MS (tandem mass spectrometry) was further used to analyse the peak corresponding to the deamidated FLpepT26 peptide. Compared to FLpepT26 (1886.1 Da) the released peptide was a single compound with a molecular mass of 1884.9 Da with slightly higher retention time (Thangaraju et al., 2016). To confirm the location of the mass difference both peptides were subjected to collision induced dissociation (Thangaraju et al., 2016).

Complete set of mainly *b* and *y* type fragment ions were the result of MS/MS sequencing of the triply protonated parent ions (m/z 629.3 and 629.7) (Thangaraju et al., 2016). Sequence of FLpepT26 correlate with the identified peptide fragments and the Gln to Glu transformation in the released peptide was unequivocally confirmed based on the mass differences of the *N*-terminal peptide fragments (Theoretical molecular mass: FLpepT26: 1885.0; Cleaved FLpepT26: 1886.0; Measured molecular mass: FLpepT26: 1884.9; Cleaved FLpepT26: 1886.1).

4.5.2. Characterisation of the developed protein based isopeptidase assay

After successfully demonstrating the TG2 mediated hydrolysis of FLpepT26-S100A4 we checked its connection with decreased anisotropy. Anisotropy is expected to decrease over time during the isopeptidase reaction as opposed to crosslinking reaction were there is an increase of anisotropy (Biri et al., 2016). Indeed, there was a decrease in anisotropy signal over time with increasing TG2 concentration. Reaction rates were calculated in the range of 10-600 nM TG2 concentration by using linear regression based on the first phase (9-10 minutes) of the decrease (Figure 14 C). The isopeptidase activity of the remaining enzyme used for isopeptide substrate generation was checked in the presence of 5 mM calcium or EDTA without adding any further TG2. The comparison of the reaction rates revealed that there was no significant difference between the slope of the curves (Figure 14 C). The isopeptidase activity could be detected at 10 nM TG2 concentration, using 5 µg substrate (which corresponds to 0.5 µM concentration) while the reaction rate was linearly dependent on TG2 concentration in the 10-300 nM range (Figure 14 D).

We further characterized the kinetic parameters of isopeptidase reaction. At fixed TG2 level, (0.3 µM TG2, 5 mM Ca^{2+}) the effect of increasing FLpepT26-S100A4 (GST) substrate concentration on the reaction rate was measured. The curve fitted based on the Michaelis-Menten equation (Figure 15 A coefficient of determination: R^2=0.9871) followed saturation kinetics. Activity reached plateau phase here and by using Graphpad Prism 7, Vmax was calculated to be 57.94+0.92 mr/min/µM TG2 and Km to be 53.91+4.4 nM for FLpepT26-S100A4 (GST). The isopeptidase assay substrate, isopeptide bond containing FLpepT26-S100A4 (GST) might also contain significant amount of free S100A4 (GST) that can influence the isopeptidase reaction. Using the previously mentioned Zedira assay and our protein substrate based assay we addressed the effect of free S100A4 (GST) on the isopeptidase activity (Figure 16 A and B). In the Zedira assay there is a dose dependent increase of isopeptidase activity when the S100A4 (GST) concentration is increased. This may be attributed to the interaction between S100A4 and TG2 (Biri et al., 2016), which leads to potential stabilization of the active conformation of the enzyme. Our study has shown that isopeptidase activity is not influenced by amine donor substrates (Kiraly et al., 2016). But in our protein-based assay, there was no change in the isopeptidase activity when the S100A4 (GST) concentration was increased. The high ratio of S100A4 (GST) concentration in the substrate could explain the attained saturation limit without addition of S100A4. Glycine methyl ester was present in the assay solutions as it was shown to facilitate isopeptidase reaction by aiding aminolysis (Adamczyk et al., 2013). The fluorescence polarization was

comparable in the presence or absence of glycine methyl ester (data not shown) and the transamidated peptide product was not identified by MS analysis (Thangaraju et al., 2016).

TG2 undergoes a large conformational change during the activation by Ca^{2+} and exhibits a so-called open conformation (Pinkas et al., 2007). Influence of calcium on isopeptidase activity of TG2 was checked at different Ca^{2+} concentrations (Figure 15 B). Generally, in the presence of 1.5 mM Ca^{2+} concentration maximum reaction rate is reached during crosslinking and the deaminylation activity (Kanchan et al., 2013). But for isopeptidase bond cleavage, 4-5 mM Ca^{2+} concentration was required to attain saturation with EC50 value of 0.96±0.17 mM (calculated by fitting a dose-response curve). In intracellular environment Ca^{2+} and nucleotides reciprocally regulate the TG2 activity (reviewed in Eckert et al., 2014). By a stable GTP compound, GTPγS the inhibitory effect was tested on TG2 isopeptidase activity (Figure 15 D). The inhibitory effect of GTPγS on isopeptidase activity of TG2 was similar in the protein-based and Zedira A102 substrate assay (described in section 4.4). Due to TG2 involvement in various pathologies (reviewed in Eckert et al., 2014), pharmacological intervention to regulate transglutaminase activity by inhibitors is gaining ground. Effect of a cell permeable, irreversible active site directed inhibitor of TG2, Z-DON (Zedira) (McConoughey et al., 2010) was also tested on the newly developed method (Figure 15 C). Around 0.3 μM Z-DON was enough to inhibit the isopeptidase activity completely with an IC50 value of 30.7 nM.

We also characterised the effect of earlier described transamidase mutant (TG2 T, W278F) and isopeptidase mutant (TG2 I, W332F) (described in section 4.4) using the new method. In the protein-based assay, TG2 T mutant showed 46.6 % and TG2 I mutant showed 78.6 % lower activity than wild type TG2 respectively (Figure 15 E). Compared to the previously discussed Zedira deaminylation assay (described in section 4.4) W278F mutant showed higher while W332F mutant showed lower activity compared to the wild type. BSA-FLpepT26 crosslinked substrate was also produced and tested in the isopeptidase assay, to see the importance of substrate quality. No isopeptidase reaction could be detected either by monitoring polarization change or by SDS-PAGE using the BSA-FLpepT26 crosslinked substrate (Thangaraju et al., 2016).

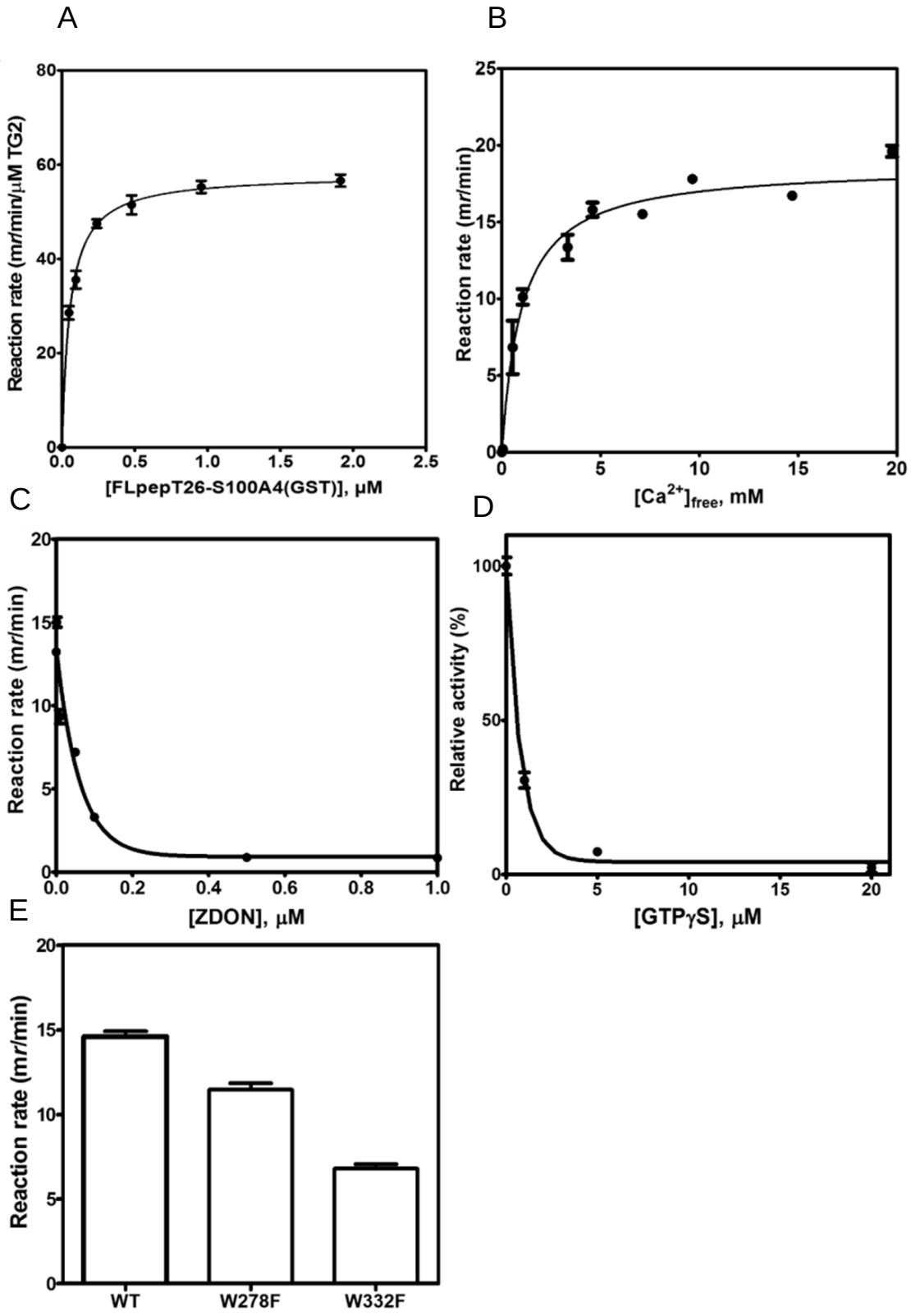


Figure 15. Characterisation of the newly developed protein based isopeptidase assay. Determination of the Michaelis constant **(A)**, the Ca^{2+}-dependence **(B)**, the inhibitory effect of Z-DON **(C)** and GTPγS **(D)**. In the presence of 5 mM Ca^{2+}, 0.3 µM TG2 and 5 µg of substrate, Z-DON inhibitor was tested after 5 minutes preincubation while GTPγS-dependence with 2 mM Ca^{2+}. 300 nM TG2, and 5 mM Ca^{2+} were used to test the mutants **(E)**. Instead of Ca^{2+}, 5 mM EDTA was present in all negative control experiments. Data are presented as means with ±SD from two separate experiments done in triplicate.

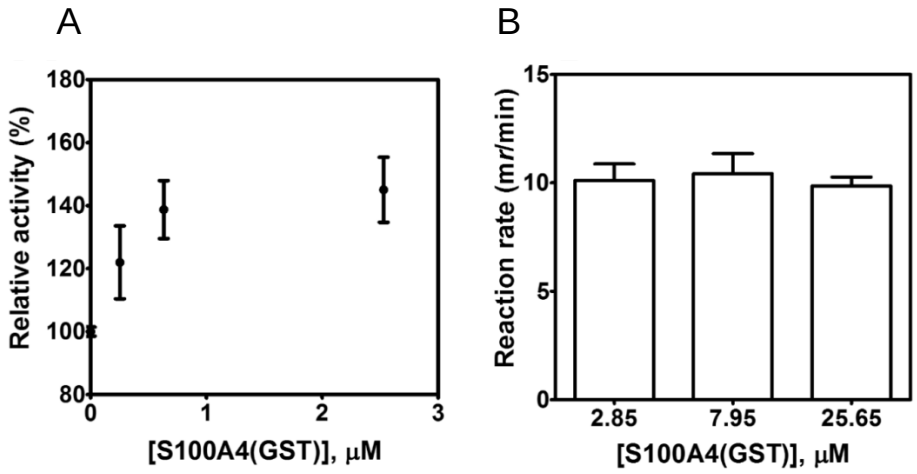


Figure 16. Effect of free S100A4 (GST) on the isopeptidase activity. Using A102 Zedira substrate **(A)** or FLpepT26-S100A4 (GST) peptide based substrate **(B)** the effect of S100A4 (GST) on isopeptidase activity was studied. Data are presented as means with ±SEM from two separate experiments done in triplicate or duplicate.

4.5.3. Analysis of homozygous TG2 variants by the newly developed assay

The isopeptidase activity of six homozygous nsSNVs of TG2 (Discussed in section 4.3.2) was also determined using the newly developed real-time kinetic assay (Figure 17). The relative activity was calculated for all the variants compared to the wild type TG2. Except R222Q, all the other variants showed similar activity like the wild type TG2. Variant D671N was 13 % more active compared to the wild type but variant R222Q showed less activity. The result of the R222Q variant is similar to the previously described Zedira assay (Figure 7 isopeptidase activity) but V542F and P612T are more active in the protein based kinetic assay compared to the Zedira assay (Figure 7 isopeptidase activity). This difference may be owed to the sensitivity of the enzyme to the substrate, enzyme concentration, nature and concentration of the substrate used in both the assays. Zedira assay uses 50 μM peptide as a substrate but in protein based assay it is 0.5 μM. But the concentration of the enzyme is 100 nM in the Zedira assay and 300 nM in the protein based assay. Overall, we were able to demonstarte the cleavage of the isopeptide bond by measuring the decrease in fluorescence polarization in real time and also characterized the kinetic parameters of the isopeptidase method.

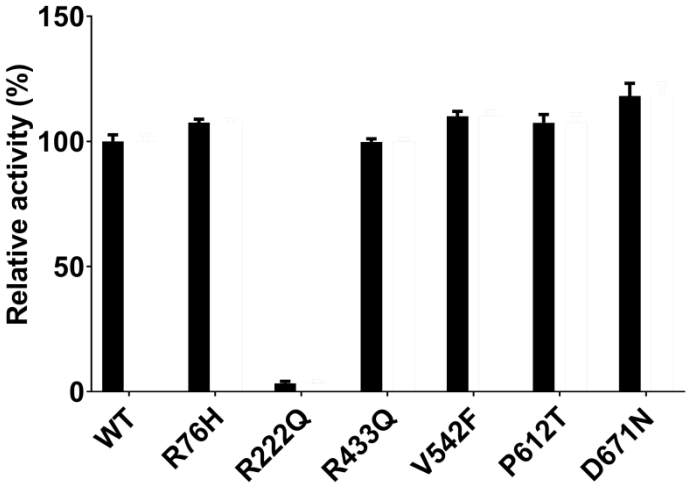


Figure 17. **Characterisation of isopeptidase activity of homozygous nsSNVs of TG2**
The isopeptidase activity of variants were measured using the newly developed protein based method. The reaction was measured at 300 nM TG2, 5 mM calcium and 0.5 μM substrate. The reaction rate was measured for the first 15 mins and relative activity was calculated compared to the wild type TG2.

5. DISCUSSION

5.1. Non-synonymous SNVs of TG2 are rare, influence protein stability and some functional sites but are not associated with diseases

To understand the relationship between genetic and phenotypic variations, structural analysis of non-synonymous mutations is essential (Sunyaeva et al., 2000). The TG2 damaging nsSNVs concentrated in the catalytic core domain could destabilize the protein. Mutation of a hydrophobic residue to a charged or polar residue could destabilize the protein as core of the protein is hydrophobic in nature. Based on literature data, generally, core of a protein is concentrated with disease causing SNVs (Yates et al., 2013) and three quarters of nsSNVs associated with the Mendelian diseases affect protein stability (Katsonis et al., 2014). Destabilizing mutations are associated with diseases like Cystic fibrosis, Parkinson's, Alzheimer's and certain types of cancers and some enhance the protein stability and linked to phenotypes like mental retardation (reviewed in Stefl et al., 2013). In transglutaminases, mutation R252L in F13a destabilizes the folded structure and results in poor clot formation (Mikkola and Palotie, 1996). So the structural perturbation caused by the destabilizing TG2 nsSNVs to the protein should be looked in detail.

The absence of damaging nsSNVs at the active site residues, and the residues critical for transamidase (W278) and isopeptidase (W332) activities underlines the importance of these residues in regulating TG2 activities. But the presence of damaging nsSNVs in the calcium, GDP, and many interaction sites should be evaluated in detail as it might have structural and functional significance. Also, occurrence of damaging nsSNVs in the heterozygous state within IDRs embedding SLiMs indicates towards a potential functional significance due to sequence variability at these sites. SLiMs can potentially serve as sites for post-translational modification, proteolytic cleavage, ligand docking or binding (Uyar et al., 2014). SLiMs in IDRs mediate many protein-protein interactions and nsSNVs presence could influence these interactions. One of our previous study revealed that substrate proteins prefer glutamine and lysine residues in IDRs of TG2 for interactions (Csosz et al., 2008). Functionally important residues of SLiMs embedding IDRs are enriched in disease-related mutations (Gfeller et al., 2014) and nsSNVs (Uyar et al., 2014). Intrinsically disordered proteins have been implicated in many human diseases including neurodegenerative disorders (Uversky et al. 2008). Disordered regions display low sequence conservation and have no evident structure, so functional consequence of nsSNVs in these regions are hard to predict.

Rare variants are more likely to be functional and tend to have larger effect size compared to common variants (Gorlov et al., 2011). Function of the rare variants can offer clues about common variants mechanism but large sample sizes and high-throughput sequencing are required to identify them. It is important to note that all the nsSNVs in TG2 are rare with minor allele frequency below 0.5 % but none so far is associated with any phenotype. Weak purifying selection leads to an abundance of rare variants and they give obvious insight into disease mechanism (Gorlov et al., 2011). Deleterious rare, non-synonymous variants with extremely diverse effects on molecular function and phenotype are connected to human diseases (Kryukov et al., 2007). According to recent studies, genomes of healthy individuals contain rare, homozygous LOF variants or knockouts (MacArthur et al., 2010). Currently, 1,717 genes have been reported to be linked to various Mendelian recessive disorders (the most common is cystic fibrosis), although most of them are compatible with life. Clinical and genetic studies reveal that F13a (bleeding disorder, prevalence 1 in 2 millions) (Katona et al., 2014), TG1 (lamellar ichthyosis, prevalence 1 in 150 thousands) (Huber et al., 1995), and TG5 (acral peeling skin, prevalence <1 in 1 million) (Cassidy et al., 2005) contain humans knock out, homozygous LOF variants associated with disease phenotypes. Until now, except TG2 heterozygous mutations associated with early-onset type 2 diabetes (Iismaa et al 2013; Bernassola et al., 2002; Porzio et al., 2007), no homozygous LOF or disease associated nsSNV has been reported for TG2 in any population.

As discussed earlier, multifunctional nature of TG2 is the most plausible explanation for this remarkable purifying selection. With broad interaction potential and its ubiquitous presence TG2 interacts with a large number of ligands and proteins (reviewed in Eckert et al., 2014; Kanchan et al., 2015). This underscores the likely structural and functional importance of almost all amino acids in the TG2 sequence and probably there would be deleterious consequences if amino acids manifested in homozygotes are lost. Adding to this, fertile TG2 knock-out mice without any distinct developmental phenotype, raises the possibility that human TG2 gained vital functions compared to rodents that has not been fully revealed, yet. This suggests that during evolution human TG2 might have gained novel functions of critical significance essential for life. So we carried out comparative genomics study between human, primate, and rodent TG2 sequences to elucidate the evolutionary and structural background for gain-of-function changes in TG2.

5.2. Novel clusters gained by human TG2 stabilizes the protein and could potentially regulate vital functions

Compared to mouse during evolution, human TG2 has gained novel functions by the appearance of amino acid clusters and IDRs. It can be drawn to conclusion that humans and other primates share the potential new biochemical and physiological functions mediated by these amino acid clusters. Evolution of a protein depends on protein stability and aggregation and both are connected to protein folding and function (Somero et al., 1995; Fields, 2001; Závodszky et al., 1998). To accommodate bulky side chains and due to regular secondary structure rearrangements of human TG2, the amino acid clusters gained during evolution from mouse to human TG2 increase the stability of the protein structure. The cluster sequences gained by human/primates probably provided stability to efficiently execute catalytic and scaffolding functions compared to other species. Apart from the clusters the single amino acid changes between human and mouse could have potential impact on stability. The stabilizing contacts like electrostatic or hydrogen-bonding interactions could be influenced by changes of polar and charged side-chains residues to hydrophobic ones. Increase in stability might result in a rigid molecule with reduced enzymatic activity so the biochemical and binding properties of these clusters should be studied in future.

The evidence for the influence of amino acid clusters beyond stability comes from a recent study reporting a novel 14-amino acid nuclear localization signal (NLS) (**^{466}**AEKEETGMAMRIRV**^{479}**) and a leucine-rich nuclear export signal (NES) (**^{657}**LHMGLHKL**^{664}**) which promote human TG2 shuttling between the nuclear and cytosolic milieu (Shrestha et al., 2015). One of the clusters ^{657}LHM^{659} is part of the first three amino acids of the nuclear export signal sequence and TG2 exporting activity is reduced with a mutation of L657 to glutamine residue. But the consequence of this TG2 exporting mechanism in cellular physiology remains to be elucidated. Nuclear TG2 interact with transcriptional factors like E2F1, hypoxia inducible factor 1 and related proteins like Sp1 and histones and/or involved in regulating gene expression via post-translational modification (Kuo et al., 2011). By doing so, TG2 controls processes like cell growth and survival, differentiation and apoptosis.

Some of the novel clusters proximity to functional sites of TG2 may increase their regional stability. Integrin β1 and β3 contacts in the N-terminal domain IDR (14-21) of TG2 which contains ^{17}NGR^{19} motif (Asp-Gly-Arg), thus amino acid cluster ^{14}LET^{16} next to this motif might be vital in humans. Isoaspartyl residues (isoAsp) could be formed due to asparagine deamidation at the NGR motif which in turn results in an iso-DGR motif (isoAsp-Gly-Arg).

Iso-DGR motif mimics a well-known integrin binding site, RGD motif (Arg-Gly-Asp) (Corti and Curnis, 2011; Kanchan et al., 2015). IDRs rich in putative phosphorylation sites in TG2 contains SUMO interacting sites (Kanchan et al., 2015) and due to cluster ^{657}LHM^{659} new SUMO binding motif (653-657) may have been generated in its proximity. In nasal epithelium, SUMO 1 interaction with TG2 regulates oxidative stress and inflammation (Luciani et al., 2009). Phosphorylation sites lost or gained by human TG2 because of the novel clusters could provide new regulatory connection to signaling pathways mediated by protein phosphorylation. Phosphorylation plays an important role in cell cycle, developmental and cancer pathways (Reimand et al., 2013; Pawson et al., 2005). As many proteins target phosphorylation and glycosylation for binding, their loss could impact protein-protein interactions mediated by human TG2.

Apart from providing conformational flexibility, IDRs and SLiMs play a vital role in protein function (Van der Lee et al., 2014). Given the role of SLiMs in protein-protein interactions and non-enzymatic interactions of TG2, the new SLiMs (interactions sites, binding motifs and cleavage sites) generated by the appearance of amino acid clusters hold high significance (Kanchan et al., 2015). Frequency of IDRs in eukaryotes compared to bacteria and archaea suggest their role in eukaryotic evolution (Bogatyreva et al., 2006). Besides, disordered proteins have fewer evolutionary constraints and evolve faster than ordered proteins (Brown et al., 2009). These clusters contributes to new functional sites for interactions, posttranslational modifications and also seems to stabilize human TG2. Also the absence of damaging nSNVs in the cluster sites shows the importance of these sites gained by human TG2. Further comparing mouse and human TG2 systematically by biochemical and cellular studies would completely reveal the importance of these amino acid segments in human physiology and pathology.

5.3. Human TG2 has low intolerance to genomic variations

Transglutaminase 2 research is in continuous expansion as demonstrated by the large number of studies however, the evolutionary significance and effect of single nucleotide variations of TG2 is unexplored so far. The genomic variations in the TG2 gene should be limited based on the unusual multiple roles encoded into one polypeptide chain. The molecular background of polymorphism, common and rare variants can be extensively studied by recent advances in large scale exome sequencing technologies. This helps to understand unsolved Mendelian and non-Mendelian genetic disorders (Ng et al., 2010; O'Roak et al., 2011; Tennessen et al., 2010; Yi et al., 2010). It has been predicted that each individual carries total of new 13-14,000

SNVs and 2-3 % of which influence protein function and 97 % of functionally important SNVs are rare (Tennessen et al., 2012; Nelson et al., 2012). The available information in databases shows that there are no common single nucleotide variations in exons of human TG2 unlike in the case of most other transglutaminases. The rate of mutations in humans for single-base substitutions is estimated to be ~1–1.5 × 10^{-8} per site per generation. In accord with this rate the human transglutaminase genes harbor the same number of nsSNV types, but differ very much in their response to purifying selection.

Our survey indicates that TG2 is under higher gene constraint and can tolerate few exonic genomic alterations, particularly in homozygous form, and none in homozygous loss of function state. We presume that high selective pressure acting on human TG2 does not allow generation of even heterozygous common variants. Human TG2 nsSNVs occurring in homozygous state tested by us show only minor changes in biochemical activities (except R222Q) and did not affect any structural and functional features.

Regarding the R22Q variant, the only homozygous one with decreased activity, when amine donor was a protein (S100A4) it showed reduced transamidase and isopeptidase activities. At physiologically relevant range of Ca^{2+} concentration this variant also showed diminished responsiveness in the amine incorporation assay. This behavior can be attributed to the importance of the region surrounding the R222 residue. The position of the loop, P359-G372 is determined by the non-covalent interaction network around R222 and calcium binding site 1. Amine substrate binding is achieved by the interaction between the residues in the loop and calcium binding site 2. Finally, through a charge-relay amino acid diad, E363, H305, the loop catalysis itself by soaking up a proton when the amine attacks the acyl-enzyme intermediate. Most of this loop is intrinsically disordered and highly dynamic without cooperative stabilizing interactions with site 1 and 2 and R222. The calcium affinities at site 1 and 2 could be probably reduced because the conformation of the loop is altered by the disrupted H-bond network in the Q222 variant resulting in a topology which disfavors proper interaction with the protein amine donor. Besides, the absence of homozygous TG2 nsSNVs in the conserved sites shows the importance of these variants.

5.4. Potential application of separated TG2 activities and their role in formation of covalently crosslinked polymers

To study the importance of TG2 activities we separated transamidase and isopeptidase activities by site directed mutagenesis. By replacing the amino acid residues around the active site we could achieve separation of different catalytic activities of transglutaminase 2. The

characterised mutants can serve as tools to develop cellular models for better examination of biological functions and applications of transglutaminases. For producing covalently crosslinked polymers and to better understand the role of amine incorportaion in cellular context the W278F (TG2-T) mutant with elevated transamidase and deficient isopeptidase activity will be helpful. The W332F (TG2-I) mutant with its high isopeptidase activity and without transamidation could be potentially applied to reverse protein crosslinks in cells and body fluids. Deubiquitinases reverse isopeptide bonds formed between proteins and ubiquitin by ubiquitin ligases and regulate fates of many proteins in different cell compartments. We hypothesize that transglutaminases with both the activities can form and break the protein-bound γ-glutamyl-linked crosslinks and also regulate critical cellular functions. It could be potential therapeutic tool in neurodegenerative disorders manifested by increase in protein crosslinks.

We applied the characterized mutants to study the formation and reversal of protein crosslinks in neurodegenerative disorders using a rat cell model of huntington. As previously shown by other studies the huntingtin with normal repeats are neither good substrates or nor crosslinked by TG2. The cells expressing wild-type TG2, displayed more crosslinked polymers and it increased with autophagy inhibition. The absence of crosslinked polymers in the TG2 C277S transamidating mutant also confirms that TG2 transamidation activity is indispensable for crosslinked polymer formation. Less crosslinked polymers present after proteasome inhibition in our study reveals that they are rerouted for autophagy degradation. Misfolded proteins once aggregated cannot pass through the proteasome narrow structures and hence not degraded by them. A recent study disclosed that mutant huntingtin was recruited to exosomes by TG2 mainly upon proteaseome inhibition (Hidalgo et al., 2016). TG2 is reported to play a role in autophagosome maturation (D'Eletto et al., 2009) and in autophagy mediated removal of ubiquitinated proteins (D'Eletto et al., 2012). In the later study it was shown that TG2 transamidation activity is essential for clearance of protein aggregates but not required for the binding of ubiquitinated proteins. Studies have also shown that during early stages of autophagy free calcium ions are accumulated in cytoplasm and this could potentially activate TG2 transamidating activity.

The less crosslinked polymers present in the cells with W332F mutant shows that reduction in the transamidation activity can reduce the polymer formation and also a potential role of isopeptidase activity in reversing the crosslinks. It can be used as a potential therapeutic tool in combination with inhibitors. As human brain contains various other TGs, the application of a non-selective inhibitor could be fatal, for example it can inhibit Factor XIII and cause

bleeding disorder. Our study has explored a possible involvement of TG2 activities in cell aggresome system but further experiments for example, morphology and quantification of aggregates are required to completely confirm these observations. Besides, the LIR motifs of TG2 should be investigated along with the heterozygous damaging nsSNVs within these motifs to study whether they influence the autophagic process.

5.5. Sensitivity and application of the newly developed isopeptidase method

The isopeptidase activity of TG2 is not studied due to the lack of proper protein based assay. We are successful in developing a protein based, fluorescence polarization assay to measure the isopeptidase activity of TG2 in real time. Our study also shows that the quality of the crosslinked substrate applied to measure the isopeptidase reaction is vital. Compared to BSA, the presence of only one TG2-reactive lysine residue, Lys100 in the C-terminal tail of S100A4 makes it easily accessible for the active site of TG2. The detection limit of newly developed assay (10 nM TG2) is similar to deaminylation assay using small non-specific peptides (Zedira). The sensitivity and detection limit of 13 nM TG2 noticed in commercial non protein based kinetic assays are similar but they use 50 μM substrate concentration (Adamczyk et al., 2013). But the amount of substrate used in our assay is only 0.5 μM compared to 50 μM in the above mentioned assay. While studying the effect of increasing FLpepT26-S100A4 (GST) substrate concentration on the reaction rate, we observed a very low enzymatic dissociation constant (Km) compared to previously published values for human TG2. The Km values were in micromolar range for Zedira assay measuring deaminylation activity with peptide substrate (peptides A101 and A102 with 13.3 μM and 54.3 μM Km values respectively) and in millimolar range for kinetic spectrophotometric assay detecting deamidation using Cbz-Gln-Gly as substrate (Yi et al., 2015). The Km values were in micromolar range for assay measuring cleavage of isopeptide bond between crosslinked peptides (Stamnaes et al., 2008). The lower Km value in nanomolar range demonstarte high sensitivity and reaction specificity of our newly developed method. Isopeptidase activity might have physiological relevance given the high reactivity of TG2 with isopeptide bonds.

Generally, maximum reaction rate is attained with 1.5 mM Ca^{2+} in case of crosslinking and deaminylation activity. But for isopeptidase activity, 4.5 mM Ca^{2+} was required to attain maximum reaction rate suggesting a potential unexplored role for isopeptidase activity in Ca^{2+} rich extracellular matrix compared to cytosol. Compared to Wt TG2, both W278F and W332F mutants showed less isopeptidase activity in the protein based assay than other assays. Earlier

observed differential effect of these mutations might have been masked by the larger substrate size used in this assay suggesting the importance of steric features on the TG2 isopeptidase activity. The potential role of isopeptidase activity in cystic fibrosis or in neurodegenerative disorders to remove pathologically accumulated, Nε-(g-glutamyl) lysine crosslinks should be investigated in detail. Production of TG2-specific substrates could provide an avenue to study the biological and regulatory significance and reveal new mechanistic details of the isopeptidase reaction.

6. SUMMARY

- Novel amino acid clusters in human TG2 were identified and computational predictions revealed that these peptide sequences contribute to increasing stability of human TG2 and could potentially regulate vital functions.

- Based on the information from exome databases, TG2 non-synonymous single nucleotide variants were rare and under selective evolutionary constraint compared to other members of transglutaminase family.

- The damaging non-synonymous single nucleotide variants destabilize the protein structure and can influence vital functions. Six existing homozygous non-synonymous single nucleotide variants of TG2 were experimentally characterized and only one, the R222Q variant was shown to influence TG2 transmidase and isopeptidase activities.

- The transamidase and isopeptidase activities of TG2 were successfully separated by site-directed mutagenesis. TG2 transamidase activity was shown to be involved in the formation of covalently cross-linked protein polymers and the potential role of isopeptidase activity in reversing the protein crosslinks was also demonstrated.

- A kinetic real-time protein based method to monitor the isopeptidase activity of TG2 was successfully developed.

7. REFERENCES

1. Adamczyk, M. Heil, A. and Aeschlimann, D., 2013. Real-time fluorescence assay for monitoring transglutaminase activity. BMG LABTECH, Application Note 234, Rev. 04/2013.

2. Ahvazi B, Kim HC, Kee SH, Nemes Z, Steinert PM. 2002. Three-dimensional structure of the human transglutaminase 3 enzyme: binding of calcium ions changes structure for activation. EMBO J, 21, pp.2055-2067

3. Alkuraya, F.S. 2015. Human knockout research: new horizons and opportunities. *Trends in Genetics*, 31(2), pp.108-115.

4. Arnold K, Bordoli L, Kopp J, and Schwede T. 2006. The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling. Bioinformatics, 22, pp.195-201.

5. Balajthy, Z. et al., 2006. Tissue-transglutaminase contributes to neutrophil granulocyte differentiation and functions. *Blood*, 108, pp.2045–2054.

6. Bernassola, F. et al., 2002. Role of transglutaminase 2 in glucose tolerance: knockout mice studies and a putative mutation in a MODY patient. FASEB J, 16(11), pp.1371-1378.

7. Biri, B. Kiss, B. Király, R. Schlosser, G. Láng, O. Kőhidai, L. Fésüs, L. Nyitray, L., 2016. Metastasis-associated S100A4 is a specific amine donor and an activity-independent binding partner of transglutaminase-2. Biochem. J, 473, pp.31-42.

8. Blom, N. Gammeltoft, S. Brunak, S., 1999. Sequence- and structure-based prediction of eukaryotic protein phosphorylation sites. J Mol Biol, 294(5), pp.1351-1362.

9. Bogatyreva NS, Finkelstein AV, Galzitskaya OV. 2006. Trend of amino acid composition of proteins of different taxa. J Bioinformatics Comput Biol, 4(2), pp.597-608.

10. Bolger, R. Checovich, W. 1994. A new protease activity assay using fluorescence polarization. BioTechniques, 17, pp.585-589.

11. Boros, S. Ahrman, E. Wunderink, L. Kamps, B. de Jong, W.W. Boelens, W.C. Emanuelsson, C.S., 2006. Site-specific transamidation and deamidation of the small heat-shock protein Hsp20 by tissue transglutaminase. Proteins, 62(4), pp.1044-52.

12. Boros, S. Wilmarth, P.A. Kamps, B. de Jong, W.W. Bloemendal, H. Lampi, K. Boelens, W.C., 2008. Tissue transglutaminase catalyzes the deamidation of glutamines in lens betaB(2)- and betaB(3)-crystallins. Exp Eye Res, 86(2), pp.383-93.

13. Brown CJ, Johnson AK, Daughdrill GW. 2009. Comparing Models of Evolution for Ordered and Disordered Proteins. Mol. Biol. Evol, 27(3), pp.609–621

14. Cardoso, I, Østerlund, E.C. Stamnaes, J. Iversen, R. Andersen, J.T. Jørgensen, T.J.D. Sollid, L.M., 2016. Dissecting the interaction between transglutaminase 2 and fibronectin. Amino Acids, DOI 10.1007/s00726-016-2296-y.

15. Cassidy, A.J. van Steensel, M.A. Steijlen, P.M. van Geel, M. van der Velden, J. Morley, S.M. et al., 2005. A homozygous missense mutation in TGM5 abolishes epidermal transglutaminase 5 activity and causes acral peeling skin syndrome. Am J Hum Genet, 77(6), pp.909-17.

16. Chica, R.A. Gagnon, P. Keillor, J.W. Pelletier, J.N. 2004. Tissue transglutaminase acylation: Proposed role of conserved active site Tyr and Trp residues revealed by molecular modeling of peptide substrate binding. Protein Sci, 13, pp.979-991.

17. Chimpanzee Sequencing and Analysis Consortium, 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature*, 437, pp.69–87.

18. Chinwalla A.T. et al., 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature*, 5;420(6915), pp.520-62.

19. Chun, W. Lesort, M. Tucholski, J. Ross, C.A. Johnson, G.V.W., 2001. Tissue Transglutaminase Does Not Contribute to the Formation of Mutant Huntingtin Aggregates. Cell Biol, 153(1) pp.25–34.

20. Cortes, C.J. and La Spada, A.R., 2014. The many faces of autophagy dysfunction in Huntington's disease: from mechanistic pathways to therapeutic opportunities. Drug Discov Today, 19(7), pp.963–971.

21. Corti A, Curnis F (2011) Isoaspartate-dependent molecular switches for integrin-ligand recognition. J Cell Sci 124(4):515-522.

22. Csomos, K. et al., 2010. Tissue transglutaminase contributes to the all-trans retinoic acid induced differentiation syndrome phenotype in the NB4 model of acute promyelocytic leukemia. *Blood*, 116, pp.3933 3943.

23. Csosz E et al. 2008. Substrate Preference of Transglutaminase 2 Revealed by Logistic Regression Analysis and Intrinsic Disorder Examination. J. Mol. Biol, 383(2), pp.390-402.

24. Davey, N.E. Van Roey, K. Weatheritt, R.J. Toedt, G. Uyar, B. Altenberg, B. Budd, A. Diella, F. Dinkel, H. Gibson, T.J. et al., 2012. Attributes of short linear motifs. Mol Biosyst 8(1), pp.268-281.

25. D'Eletto, M. Farrace, M.G. Falasca, L. Reali, V. Oliverio, S. Melino, G. et al., 2009. Transglutaminase 2 is involved in autophagosome maturation. Autophagy, 5(8), pp.1145-54.

26. D'Eletto, M. Farrace, M.G. Rossin, F. Strappazzon, F. Di Giacomo, G. Cecconi, F. et al., 2012. Type 2 transglutaminase is involved in the autophagy-dependent clearance of ubiquitinated proteins. Cell Death Differ, 19(7), pp.1228–1238.

27. Demény, M.A et al., 2015. Transglutaminases Multiple Functional Modifiers and Targets for New Drug Discovery: Book Chapter 1: structure of transglutaminases. unique features serve diverse functions. *Springer*, pp.1-45.

28. Dinkel, H. et al., 2014. The eukaryotic linear motif resource ELM: 10 years and counting. Nucleic Acids Res 42 (database issue), pp.D259-266.

29. Dosztányi, Z. Csizmok, V. Tompa, P. Simon, I., 2005. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. Bioinformatics, 21(16), pp.3433-3434.

30. D'Souza, D.R. Wei, J. Shao, Q. Hebert, M.D. Subramony, S.H. Vig, P.J., 2006. Tissue transglutaminase crosslinks ataxin-1: Possible role in SCA1 pathogenesis. Neurosci Lett, 409(1), pp.5–9.

31. Eckert, R.L. et al., 2014. Transglutaminase regulation of cell function. *Physiol Rev*, 94, pp.383-417.

32. Fesus, L. et al., 1987. Induction and activation of tissue transglutaminase during programmed cell death. *FEBS Lett*, 224, pp.104–108.

33. Fesus, L. et al., 1989. Apoptotic hepatocytes become insoluble in detergents and chaotropic agents as a result of transglutaminase action. *FEBS Lett*, 245, pp.150–154.

34. Fesus, L and Piacentini, M., 2002. Transglutaminase 2: an enigmatic enzyme with diverse functions, *Trends in Biochemical Sciences*, 27(10), pp.534–539.

35. Fesus, L & Szondy, Z., 2005. Transglutaminase 2 in the balance of cell death and survival. *FEBS Lett*, 579, pp.3297–3302.

36. Fields, PA., 2001. Protein function at thermal extremes: balancing stability and flexibility. Comp Biochem Physiol A Mol Integr Physiol, 129(2-3), pp.417-431.

37. Fleckenstein, B. Molberg, Ø. Qiao, S.W. Schmid, D.G. von der Mülbe, F. Elgstøen, K. Jung, G. Sollid, M., 2002. Gliadin T cell epitope selection by tissue transglutaminase in celiac disease. Role of enzyme specificity and pH influence on the transamidation versus deamidation process. J. Biol. Chem, 277, pp.34109-34116.

38. Folk, J.E. Cole, P.W. Mullooly, J.P., 1967. Mechanism of action of guinea pig liver transglutaminase III. The metal-dependent hydrolysis of p-nitrophenyl acetate; further observations on the role of metal in enzyme activation. J Biol Chem, 242,pp.2615-2621.

39. Gentile, V. Saydak, M. Chiocca, E.A. Akande, O. Birckbichler, P.J. Lee, K.N. et al., 1991. Isolation and characterization of cDNA clones to mouse macrophage and human endothelial cell tissue transglutaminases. J Biol Chem, 266(1), pp.478-483.

40. Gentile, V. et al., 1994. The human tissue transglutaminase gene maps on chromosome 20q12 by in situ fluorescence hybridization. *Genomics*, 20(2), pp.295– 297.

41. Gfeller D, Ernst A, Jarvik N, Sidhu SS, Bader GD. 2014. Prediction and experimental characterization of nsSNPs altering human PDZ-binding motifs. PLoS One 10, 9(4):e94507.

42. Gibbs, R. A. et al., 2007. Evolutionary and biomedical insights from the rhesus macaque genome. *Science*, 316, pp.222–234.

43. Gorlov, I.P. et al., 2011. Evolutionary evidence of the effect of rare variants on disease etiology. *Clin Genet*. 79(3), pp.199–206.

44. Greg, G., 2012. Rare and common variants: twenty arguments. Nat Rev Genet, 13, pp.135-145.

45. Grenard, P et al., 2001. Evolution of Transglutaminase Genes: Identification of a Transglutaminase Gene Cluster on Human Chromosome 15q15. *The Journal of biol. chemistry*, 2001, 276(35),pp.33066–33078.

46. Grosso, H. and Mouradian, M.M., 2012. Transglutaminase 2: Biology, Relevance to Neurodegenerative Diseases and Therapeutic Implications. Pharmacology & Therapeutics, 133, pp.392–410.

47. Guerois, R. Nielsen, J.E. Serrano, L., 2002. Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. J Mol Biol, 320(2), pp.369-387.

48. Guilluy, C. Rolli-Derkinderen, M. Tharaux, P.L. Melino, G. Pacaud, P. Lorand, G., 2007. Transglutaminase-dependent RhoA activation and depletion by serotonin in vascular smooth muscle cells. J Biol Chem, 282, p.2918-2928.

49. Gundemir, S & Johnson, G.V., 2009. Intracellular localization and conformational state of transglutaminase 2: implications for cell death. *PLoS ONE*, 4, e6123.

50. Gundemir, S, Colak, G, Tucholski, J, Johnson, G.V. 2012. Transglutaminase 2: a molecular Swiss army knife. Biochim Biophys Acta, 1823(2), pp.406-419.

51. Gupta, R. and Brunak, S., 2002. Prediction of glycosylation across the human proteome and the correlation to protein function. Pac Symp Biocomputing, pp.310-322.

52. Han, Y. Tung, T.Q. Ann, I., 2015. Computational approaches for identifying deleterious nonsynonymous variants from human genomes. Advanced Science and Technology Letters, 105, pp.44-48.

53. Hasegawa, G. et al., 2003. A novel function of tissue-type transglutaminase: protein disulphide isomerase. *Biochem. J*, 373, pp.793–803.

54. Hidalgo, L.D. et al 2016. Transglutaminase type 2-dependent selective recruitment of proteins into exosomes under stressful cellular conditions. Biochim Biophys Acta, 1863(8), pp.2084-2092

55. Huber, M. Rettler, I. Bernasconi, K. Frenk, E. Lavrijsen, S.P. Ponec, M. et al., 1995. Mutations of keratinocyte transglutaminase in lamellar ichthyosis. Science, 267(5197), pp.525-8.

56. Ichinose, A. and Aoki, N., 1982. Reversible cross-linking of alpha 2-plasmin inhibitor to fibrinogen by fibrin-stabilizing factor. Biochim Biophys Acta, 706, pp.158-164.

57. Iismaa, S.E. et al., 2003. Evolutionary specialization of a tryptophan indole group for transition-state stabilization by eukaryotic transglutaminases. Proc Natl Acad Sci U S A, 100, pp.12636-12641.

58. Iismaa SE, Mearns BM, Lorand L, Graham RM. 2009. Transglutaminases and Disease: Lessons From Genetically Engineered Mouse Models and Inherited Disorders. Physiol Rev 89(3):991-1023.

59. Iismaa, S.E. et al., 2013. Glucose Homeostasis in Mice Is Transglutaminase 2 Independent. PLoS ONE, 8(5), e63346. doi:10.1371/journal.pone.0063346.

60. International Human Genome Sequencing Consortium., 2004. Finishing the euchromatic sequence of the human genome. *Nature*, 431, pp.931-945.

61. Itan, Y. Shang, L. Boisson, B. Patin, E. Bolze, A. et al., 2015. The human gene damage index as a gene-level approach to prioritizing exome variants. Proc Natl Acad Sci U S A, 112(44), pp.13615–13620.

62. Iversen, R. Mysling, S. Hnida, K. Jørgensen, T.J.D. Sollid, L.M., 2014. Activity-regulating structural changes and autoantibody epitopes in transglutaminase 2 assessed by hydrogen/deuterium exchange. Proc Natl Acad Sci U S A, 111(48), pp.17146–17151.

63. Jabri, B. and Sollid, L.M., 2006. Mechanisms of Disease: immunopathogenesis of celiac disease. Nature Reviews Gastroenterology and Hepatology, 3, pp.516-525.

64. Jacob, H.J. and Kwitek, A.E., 2002. Rat genetics: Attaching physiology and pharmacology to the genome. *Nat. Rev. Genet*, 3(1), pp.33–42.

65. Jang, G.Y. et al., 2010. Transglutaminase 2 suppresses apoptosis by modulating caspase 3 and NF-κB activity in hypoxic tumor cells. *Oncogene*, 29, pp.356–367.

66. Johnson, T.S. et al., 2007. Transglutaminase inhibition reduces fibrosis and preserves function in experimental chronic kidney disease. J Am Soc Nephrol, 18, pp.3078-3088.

67. Kahlem, P. Terre, C. Green, H. Djian, P., 1996. Peptides containing glutamine repeats as substrates for transglutaminase-catalyzed cross-linking: relevance to diseases of the nervous system. Proc Natl Acad Sci U S A, 93(25), pp.14580–14585.

68. Kahlem, P. Green, H. Djian, P., 1998. Transglutaminase action imitates Huntington's disease: selective polymerization of Huntingtin containing expanded polyglutamine. Mol Cell, 1(4), pp.595–601.

69. Kalvari, I. Tsompanis, S. Mulakkal, N.C. Osgood, R. Johansen, T. Nezis, I.P. Promponas, V.J., 2014. iLIR: A web resource for prediction of Atg8-family interacting proteins. Autophagy, 10(5), pp.913-925.

70. Kanchan, K. Ergülen, E. Király, R. Simon-Vecsei, Z. Fuxreiter, M. Fésüs, L. 2013. Identification of a specific one amino acid change in recombinant human transglutaminase 2 that regulates its activity and calcium sensitivity. Biochem J, 455, pp.261-272.

71. Kanchan, K. et al., 2015. Physiological, pathological, and structural implications of nonenzymatic protein–protein interactions of the multifunctional human transglutaminase 2. *Cell Mol Life Sci*, 72(16), pp.3009-35.

72. Katona, É. Muszbek, L. Devreese, K. Kovács, K.B. Bereczky, Z. Jonkers, M. et al., 2014. Factor XIII deficiency: complete phenotypic characterization of two cases with novel causative mutations. Haemophilia, 20(1), pp.114-120.

73. Katsonis P et al. 2014. Single nucleotide variations: Biological impact and theoretical interpretation. Protein science. 23(12), pp.1650-1666.

74. Keillor, J.W. 2008. The bioorganic chemistry of transglutaminase: From mechanism to inhibition and engineering. Can J Chem, 86, pp.271-276.

75. Keillor, J.W. et al., 2014. Acyl transfer mechanisms of tissue transglutaminase. Bioorg Chem, 57, pp.186-197.

76. Keinan, A. and Clark, A.G. 2012. Recent Explosive Human Population Growth Has Resulted in an Excess of Rare Genetic Variants. *Science*, 336, 740. DOI: 10.1126/science.1217283

77. Kenniston, J.A. et al., 2013. A homogeneous fluorescence anisotropy assay for measuring transglutaminase 2 activity. Anal. Biochem, 436, pp.13-15.

78. Király, R. Csosz, E. Kurtán, T. Antus, S. Szigeti, K. Simon-Vecsei, Z. Korponay-Szabó, I.R. Keresztessy, Z. Fésüs, L., 2009. Functional significance of five noncanonical Ca2+-binding sites of human transglutaminase 2 characterized by site-directed mutagenesis. FEBS J, 276(23), pp.7083-7096.

79. Király, R, Demény, M, Fésüs, L. 2011. Protein transamidation by transglutaminase 2 in cells: a disputed Ca^{2+}-dependent action of a multifunctional protein. FEBS J, 278(24), pp.4717-39.

80. Király, R. Barta, E. Fésüs, L., 2013. Polymorphism of transglutaminase 2: unusually low frequency of genomic variants with deficient functions. Amino Acids, 44, pp.215–225.

81. Korstanje, R. and DiPetrillo, K., 2004. Unraveling the genetics of chronic kidney disease using animal models. Am. J. Physiol. Renal Physiol, 287(3), pp.F347–F352.

82. Krieger E, Vriend G. 2002. Models@Home: distributed computing in bioinformatics using a screensaver based approach. Bioinformatics, 18(2), pp.315-318.

83. Kryukov, G.V. et al., 2007. Most Rare Missense Alleles Are Deleterious in Humans: Implications for Complex Disease and Association Studies. *The American Journal of Human Genetics.* 80, pp.727-739.

84. Kuo TF, Tatsukawa H, Kojima S (2011) New insights into the functions and localization of nuclear transglutaminase 2. FEBS J, 278(24), pp.4756-4767.

85. Kwok, S.C. Mant, C.T. Hodges, R.S., 2002. Importance of secondary structural specificity determinants in protein folding: Insertion of a native β-sheet sequence into an α-helical coiled-coil. Protein Sci, 11(6), pp.1519-1531.

86. Lai, T.-S and Greenberg, C.S., 2013. TGM2 and implications for human disease: role of alternative splicing. *Frontiers in Bioscience*, 18, pp.504–519.

87. Lai, T.S, Lin, C.J, Greenberg, C.S. 2016. Role of tissue transglutaminase‐2 (TG2)‐mediated aminylation in biological processes. Amino acids, DOI 10.1007/s00726-016-2270-8.

88. Lawrie, D.S. and Petrov, D.A., 2014. Comparative population genomics: power and principles for the inference of functionality. *Trends in Genetics*, 30(4), pp.133-139.

89. Lek, M. et al. 2016. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*, 536, pp.285–291.

90. Li, A.H. 2015. Analysis of loss-of-function variants and 20 risk factor phenotypes in 8,554 individuals identifies loci influencing chronic disease. Nat Genet, 47(6), pp.640-642.

91. Lim, E.T. et al., 2013. Rare complete knockouts in humans: population distribution and significant role in autism spectrum disorders. Neuron, 77(2), pp.235–242.

92. Liu, S. Cerione, R.A. Clardy, J., 2002. Structural basis for the guanine nucleotide-binding activity of tissue transglutaminase and its regulation of transamidation activity. Proc Natl Acad Sci USA, 99(5), pp.2743-2747.

93. Lorand, L. et al., 1988. Fibronectin as a carrier for the transglutaminase from human erythrocytes. *Proc Natl Acad Sci USA*, 85(4), pp.1057–1059.

94. Lorand, L Stern, A.M. Velasco, P.T., 1998. Novel inhibitors against the transglutaminase-catalysed crosslinking of lens proteins. Exp Eye Res, 66, pp.531-536.

95. Lorand, L and Graham, R.M. 2003. Transglutaminases: crosslinking enzymes with pleiotropic functions. Nature Reviews Molecular Cell Biology, 4, pp.140-156.

96. Luciani, A. et al., 2009. SUMOylation of tissue transglutaminase as link between oxidative stress and inflammation. J Immunol, 183(4), pp.2775-2784.

97. MacArthur, D.G. and Tyler-Smith, C., 2010. Loss-of-function variants in the genomes of healthy humans. *Hum Mol Genet*, 19(R2), R125–R130.

98. MacArthur, D.G. et al., 2012. A systematic survey of loss-of-function variants in human protein-coding genes. *Science*, 335(6070):pp.823-828.

99. Martin, A. et al., 2013. Possible involvement of transglutaminase-catalyzed reactions in the physiopathology of neurodegenerative diseases. Amino Acids, 44, pp.111-118.

100. Mastroberardino, P.G. et al., 2002. 'Tissue' transglutaminase ablation reduces neuronal death and prolongs survival in a mouse model of Huntington's disease. Cell Death Differ. 9(9), pp.873-880.

101. Mastroberardino P.G. et al., 2006. "Tissue" transglutaminase contributes to the formation of disulphide bridges in proteins of mitochondrial respiratory complexes. *Biochim. Biophys. Acta*, 1757, pp.1357–1365.

102. Mastroberardino, P.G. & Piacentini, M., 2010. Type 2 transglutaminase in Huntington's disease: a double-edged sword with clinical potential. J Intern Med, 268(5), pp.419–431.

103. McConoughey, S.J. 2010. Inhibition of transglutaminase 2 mitigates transcriptional dysregulation in models of Huntington disease, EMBO. Mol. Med, 2, pp.349-370.

104. Mikkola H, Palotie A. 1996. Gene defects in congenital factor XIII deficiency. Semin Thromb Hemost, 22(5), pp.393-398.

105. Mimuro, J. Kimura, S. and Aoki, N., 1986. Release of alpha 2-plasmin inhibitor from plasma fibrin clots by activated coagulation factor XIII. Its effect on fibrinolysis. J Clin Invest, 77, pp.1006-1013.

106. Mishra, S. and Murphy, L.J. 2004. Tissue transglutaminase has intrinsic kinase activity: identification of transglutaminase 2 as an insulin-like growth factor-binding protein-3 kinase. *J. Biol. Chem*, 279, pp.23863–23868.

107. Murthy, S.N. et al., 1999. Interaction of G(h)/transglutaminase with phospholipase Cδ1 and with GTP. *Proc. Natl. Acad. Sci. U. S. A.* 96, pp.11815–11819.

108. Murthy, S.N. 2002. Conserved tryptophan in the core domain of transglutaminase is essential for catalytic activity. Proc Natl Acad Sci U S A, 99, pp.2738-2742.

109. Narasimhan, V.M. et al., 2016. Health and population effects of rare gene knockouts in adult humans with related parents. Science, 352 (6284), pp.474-477.

110. Nelson MR, Wegmann D, Ehm MG, Kessner D, Jean PS, Verzilli C, et al. 2012. An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. Science, 337(6090), pp.100-104.

111. Nemes, Z. Petrovski, G. Aerts, M. Sergeant, K. Devreese, B. Fesus, L., 2009. Transglutaminase-mediated intramolecular cross-linking of membrane-bound alphasynuclein promotes amyloid formation in Lewy bodies. J Biol Chem, 284(40), pp.27252–27264.

112. Ng, S.B. et al., 2010. Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet*, 42(1):pp.30-35.

113. Nielsen, R. et al., 2007. Recent and ongoing selection in the human genome. *Nat Rev Genet*, 8(11), pp.857–868.

114. Nurminskaya, M.V. and Belkin, A.M. 2012. Cellular functions of tissue transglutaminase.Int Rev Cell Mol Biol, 294, pp.1-97.

115. O'Roak BJ, Deriziotis P, Lee C, Vives L, Schwartz JJ, Girirajan S, et al. 2011. Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. Nat Genet, 43(6), pp.585-589.

116. Ortega, Z. and Lucas. J.J., 2014. Ubiquitin-proteasome system involvement in Huntington's disease. Front Mol Neurosci, 7(77), pp.1-11.

117. Parameswaran, K.N. Cheng, X.F. Chen, E.C. Velasco, P.T. Wilson, J.H. Lorand, L., 1997. Hydrolysis of gamma:epsilon isopeptides by cytosolic transglutaminases and by coagulation factor XIIIa. J Biol Chem, 272, pp.10311-10317.

118. Pawson T, Scott JD (2005) Protein phosphorylation in signaling- 50 years and counting. Trends Biochem Sci, 30(6), pp.286-290.

119. Peitsch, M.C. et al., 1993. Characterization of the endogenous deoxyribonuclease involved in nuclear DNA degradation during apoptosis (programmed cell death). *EMBO J*, 12, pp.371–377.

120. Pennacchio, L.A. and Rubin, E.M., 2003. Comparative genomic tools and databases: providing insights into the human genome. *The Journal of Clinical Investigation, 111(8),*pp.1099-1106.

121. Perez Alea, M. Kitamura, M. Martin, G. Thomas, V. Hitomi, K. El Alaoui, S., 2009. Development of an isoenzyme-specific colorimetric assay for tissue transglutaminase 2 cross-linking activity. Anal Biochem, 389, pp.150-156.

122. Pinkas, D.M. Strop, P. Brunger, A.T. Khosla, C. 2007. Transglutaminase 2 Undergoes a Large Conformational Change upon Activation. Plos Biol, 5(12), pp.2788-2796.

123. Porzio O, Massa O, Cunsolo V, Colombo C, Malaponti M, Bertuzzi F, et al. 2007. Missense mutations in the TGM2 gene encoding Transglutaminase 2 are found in patients with early-onset Type 2 diabetes. Hum Mutat 28:11(1150).

124. Qiao, S.W. Piper, J. Haraldsen, G. Oynebråten, I. Fleckenstein, B. Molberg, O. Khosla, C. Sollid, L.M., 2005. Tissue transglutaminase-mediated formation and cleavage of histamine-gliadin complexes: biological effects and implications for celiac disease. J Immunol, 174, pp.1657-1663.

125. Raczyński, G. Snochowski, M. Buraczewski, S., 1975. Metabolism of epsilon-(gamma-L-glutamyl)-L-lysine in the rat. Br J Nutr, 34, pp.291-296.

126. Ramensky, V et al., 2002. Human non-synonymous SNPs: server and survey. *Nucleic Acids Res*. 30(17), pp.3894–3900.

127. Recktenwald, C.V. and Hansson, G.C., 2016. The Reduction-insensitive Bonds of the MUC2 Mucin are Isopeptide Bonds. J Biol Chem, 291(26), pp.13580-13590.

128. Reimand J, Wagih O, Bader GD (2013) The mutational landscape of phosphorylation signaling in cancer. Sci Rep, 3:2651.

129. Reyes-Turcu, F.E. Ventii, K.H. Wilkinson, K.D. 2009. Regulation and cellular roles of ubiquitin-specific deubiquitinating enzymes. Annu. Rev. Biochem, 78, pp.363-397.

130. Rollins, J. et al., 2006. In search of new targets for plasma highdensity lipoprotein cholesterol levels: Promise of human-mouse comparative genomics. *Trends Cardiovasc. Med*, 16(7), pp.220–234.

131. Romero, P. Obradovic, Z. Li, X. Garner, E.C. Brown, C.J. Dunker, A.K., 2001. Sequence complexity of disordered protein. Proteins: Struct Funct Gen, 42, pp.38-48.

132. Ross, C.A, Tabrizi, S.J. 2011. Huntington's disease: from molecular pathogenesis to clinical treatment. Lancet Neurol, 10(1), pp.83-98.

133. Saitou, N and Nei, M. et al., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol*, 4(4), pp.406-25.

134. Sali A, Blundell TL. 1993. Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 234(3):779-815.

135. Sarkar N.K. et al., 1957. An enzymically catalyzed incorporation of amines into proteins. *Biochim Biophys Acta*, 25(2), pp.451-2.

136. Schmid, A.W. Condemi, E. Tuchscherer, G. Chiappe, D. Mutter, M. Vogel, H. et al., 2011. Tissue transglutaminase-mediated glutamine deamidation of beta-amyloid peptide increases peptide solubility, whereas enzymatic cross-linking and peptide fragmentation may serve as molecular triggers for rapid peptide aggregation. J. Biol. Chem, 286, pp.12172–12188.

137. Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, Serrano L. 2005. The FoldX web server: an online force field. Nucleic Acids Res, 33, W382-388.

138. Shan, L. Molberg, O. Parrot, I. et al., 2002. Structural basis of gluten intolerance in celiac sprue. Science, 297(5590), pp.2275-2279.

139. Shrestha, R. Tatsukawa, H. Shrestha, R. Ishibashi, N. Matsuura, T. Kagechika, H. Kose, S. Hitomi, K. Imamoto, N. Kojima, S., 2015. Molecular mechanism by which acyclic retinoid induces nuclear localization of transglutaminase 2 in human hepatocellular carcinoma cells. Cell Death Dis, 6, e2002, doi:10.1038/cddis.2015.339.

140. Signorini, M. et al., 1988. Human erythrocyte transglutaminase: purification and preliminary characterisation. *Biol Chem Hoppe Seyler*, 369(4), pp.275–281.

141. Simon-Vecsei, Z. et al., 2012. A single conformational transglutaminase 2 epitope contributed by three domains is critical for celiac antibody binding and effects. Proc Natl Acad Sci U S A, 109(2), pp.431-436.

142. Somero, G.N., 1995. Proteins and temperature. Annu Rev Physiol, 57, pp.43-68.

143. Stamnaes, J. Fleckenstein, B. and Sollid, L.M., 2008. The propensity for deamidation and transamidation of peptides by transglutaminase 2 is dependent on substrate affinity and reaction conditions. Biochim Biophys Acta, 1784(11), pp.1804–1811.

144. Stefl S et al. 2013. Molecular mechanisms of disease-causing missense mutations. J Mol Biol, 425(21), pp.3919-3936

145. Stieler M, Weber J, Hils M, Kolb P, Heine A, Buchold C, et al. 2013. Structure of Active Coagulation Factor XIII Triggered by Calcium Binding: Basis for the Design of Next-Generation Anticoagulants. Angewandte Chemie. International Edition English. 52, pp. 11930-11934.

146. Stoll, M. et al., 2000. New target regions for human hypertension via comparative genomics. *Genome Res*, 10(4), pp.473–482.

147. Sugiyama, F. et al., 2001. Concordance of murine quantitative trait loci for salt-induced hypertension with rat and human loci. *Genomics*, 71(1), pp.70–77.

148. Sulem, P. et al., 2015. Identification of a large set of rare complete human knockouts. *Nat Genet*, 47(5), pp.448–452.

149. Sunyaeva S, Ramensky V, Borka P. 2000. Towards a structural basis of human non-synonymous single nucleotide polymorphisms. Trends Genet, 16(5), pp.198-200.

150. Szondy, Z. et al., 2003. Transglutaminase 2 -/- mice reveal a phagocytosis-associated crosstalk between macrophages and apoptotic cells. *Proc Natl Acad Sci USA*, 100, pp.7812–7817.

151. Takada, Y et al., 2007. The integrins. *Genome Biol*, 8(5), pp.215. doi:10.1186/gb-2007-8-5-215

152. Takeuchi, Y et al., 1998. Nuclear translocation of tissue type transglutaminase during sphingosine-induced cell death: a novel aspect of the enzyme with DNA hydrolytic activity. Z *Naturforsch C*, 53(5-6), pp.352-8.

153. Tee, A.E. Marshall, G.M. Liu, P.Y. Xu, N. Haber, M. Norris, M.D., et al. 2010. Opposing effects of two tissue transglutaminase protein isoforms in neuroblastoma cell differentiation. J. Biol. Chem., 285, pp.3561–3567.

154. Tennessen JA, Madeoy J, Akey JM. 2010. Signatures of positive selection apparent in a small sample of human exomes, Genome Res, 20(10), pp.1327-34.

155. Tennessen, J.A. et al., 2012. Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science*. 6;337(6090), pp.64-9.

156. Thangaraju K, Biri B, Schlosser G, Kiss B, Nyitray L, Fesüs L, et al., 2016. Real-time kinetic method to monitor isopeptidase activity of transglutaminase 2 on protein substrate. Anal. Biochem, 505, pp.36-42.

157. Thomas, P.D. Campbell, M.J. Kejariwal, A. Mi, H. Karlak, B. Daverman, R. et al., 2003. PANTHER: a library of protein families and subfamilies indexed by function. Genome Res, pp.2129–2141.

158. Tokuriki, N. Stricher, F., Schymkowitz, J. Serrano, L. Taw, D.S., 2007. The Stability Effects of Protein Mutations Appear to be Universally Distributed. J Mol Biol, 369, pp.1318–1332.

159. Tokuriki, N. Tawfik, D.S., 2009. Stability effects of mutations and protein evolvability. Curr Opin Struct Biol, 19(5), pp.596-604.

160. Turner, P.M and Lorand, L 1989. Complexation of fibronectin with tissue transglutaminase. *Biochemistry*, 28(2), pp.628–635.

161. UniProt Consortium., 2015. UniProt: a hub for protein information. Nucleic Acids Res. 43(Database issue), pp.D204-212.

162. Uversky VN, Oldfield CJ, Dunker AK. 2008. Intrinsically disordered proteins in human diseases: introducing the D2 concept. Annu Rev Biophys, 37, pp.215-246

163. Uyar B, Weatheritt RJ, Dinkel H, Davey NE, Gibson TJ. 2014. Proteome-wide analysis of human disease mutations in short linear motifs: neglected players in cancer? Mol Biosyst, 10(10), pp.2626-2642.

164. Vallender, E.J. and Lahn, B.T., 2004. Positive selection on the human genome. *Human Molecular Genetics*, 13 (issue 2), pp.R245–R254.

165. Van de Peer Y and Wachter D.R. 1994. TREECON: a software package for the construction and drawing of evolutionary trees for the Microsoft Windows environment. *Comput Appl Biosci*, 10(5), pp.569-570.

166. Van der Lee R, Buljan M, Lang B, Weatheritt RJ, Daughdrill GW, Dunker AK, et al. 2014. Classification of intrinsically disordered regions and proteins. Chem Rev, 114 (13), pp.6589–6631.

167. Verhaar, R. Drukarch, B. Bol, J.G. Jongenelen, C.A. Wilhelmus, M.M., 2013. Tissue transglutaminase cross-links beclin 1 and regulates autophagy in MPP+-treated human SH-SY5Y cells. Neurochemistry International, 62, pp.486–491.

168. Wang, X. and Paigen, B., 2005. Genome-wide search for new genes controlling plasma lipid concentrations in mice and humans. *Curr. Opin. Lipidol*, 16(2), pp.127–137.

169. Waterston et al., 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature*, 420(6915), pp.520-62.

170. Wei, L. Liu, Y. Dubchak, I. Shon, J. Park J., 2002. Comparative genomics approaches to study organism similarities and differences. Journal of Biomedical Informatics, 35(2), pp.142-150.

171. Wyttenbach, A et al., 2001. Polyglutamine expansions cause decreased CRE- mediated transcription and early gene expression changes prior to cell death in an inducible cell model of Huntington's disease. *Human Molecular Genetics*, 10(17), pp.1829-1845.

172. Yates C.M, Sternberg M.J.E. 2013. The Effects of Non-Synonymous Single Nucleotide Polymorphisms (nsSNPs) on Protein–Protein Interactions. J Mol Biol, 425(21), pp.3949–3963.

173. Yi, M.C. et al., 2015. An unprecedented dual antagonist and agonist of human transglutaminase 2. Bioorg. Med. Chem. Lett, 25, pp.4922-4926.

174. Yi, X. Liang, Y. Huerta-Sanchez, E. Jin, X. Cuo, Z.X. Pool, J.E. et al., 2010. Sequencing of 50 human exomes reveals adaptation to high altitude. Science, 329(5987), pp.75-78.

175. Zainelli, G.M. Ross, C.A. Troncoso, J.C. Muma, N.A., 2003. Transglutaminase cross-links in intranuclear inclusions in Huntington disease. J Neuropathol Exp Neurol, 62(1), pp.14-24.

176. Zainelli, G.M. Dudek, N.L. Ross, C.A. Kim, S.Y. Muma, N.A., 2005. Mutant huntingtin protein: a substrate for transglutaminase 1, 2, and 3. J Neuropathol Exp Neurol, 64(1), pp.58-65.

177. Závodszky P, Kardos J, Svingor A, Petsko GA (1998) Adjustment of conformational flexibility is a key event in the thermal adaptation of proteins. Proc Natl Acad Sci U S A, 95(13), pp.7406-7411.

178. Zemaitaitis, M.O. Kim, S.Y. Halverson, R.A. Troncoso, J.C. Lee, J.M. Muma, N.A., 2003. Transglutaminase activity, protein, and mRNA expression are increased in progressive supranuclear palsy. J Neuropathol Exp Neurol, 62(2), pp.173–184.

8. KEYWORDS

transglutaminase 2; transamidase activity; isopeptidase activity; mutants; non-synonymous single nucleotide variants; novel amino acid clusters; protein stability; real-time kinetic method; huntington's disease; crosslinking.

9. ACKNOWLEDGEMENTS

First and foremost I would like to express my gratitude to my supervisor Prof. Dr. László Fésüs for providing me an opportunity to work in his research group and for the continuous support, motivation and guidance. I am deeply indebted to him for the confidence he bestowed upon me.

I am very grateful to Dr. Róbert Király for guiding me through the initial phase of my PhD and for teaching me experiments, presentation skills and for valuable suggestions.

My sincere gratitude goes to all my colleagues and friends at the Stem cell, Apoptosis and Genomics research laboratory, coeliac research laboratory and laboratory of Protein Dynamics for their consistent help and co-operation during my PhD work. I also thank them for helping me in any condition when I needed.

I would like to thank Jennifer Nagy, Klem Attiláné and Szilvia Szalóki for their excellent technical assistance. Special thanks to Dr. Máté A. Demény and Dr. Kajal Kanchan for their valuable inputs and feedback regarding my project.

I would also like to thank all the co-authors in the publications for their contributions to my project.

I acknowledge the financial support of Hungarian Scientific Research Fund (OTKA NK 105046) and the European Union Marie Curie Actions, Framework Programme 7 TRANSPATH ITN 289964, which provided me an opportunity to participate in conferences and to visit pioneer institutes for training as part of the PhD program.

I would like to thank my friends in India, United States and Germany for their encouragement and support during difficult times. Finally, I express my sincere thanks to my parents, wife and sister for their unconditional love, sacrifice and emotional support.

10. LIST OF PUBLICATIONS

Registry number: DEENK/264/2016.PL
Subject: PhD Publication List

Candidate: Kiruphagaran Thangaraju
Neptun ID: ND9YVQ
Doctoral School: Doctoral School of Molecular Cellular and Immune Biology

List of publications related to the dissertation

1. **Thangaraju, K.**, Király, R., Mótyán, J. A., Ambrus, V. A., Fuxreiter, M., Fésüs, L.: Computational analyses of the effect of novel amino acid clusters of human transglutaminase 2 on its structure and function.
   Amino Acids. [Epub ahead of print], 2016.
   DOI: http://dx.doi.org/10.1007/s00726-016-2330-0
   IF: 3.196 (2015)

2. **Thangaraju, K.**, Biri, B., Schlosser, G., Kiss, B., Nyitray, L., Fésüs L., Király R.: Real-time kinetic method to monitor isopeptidase activity of transglutaminase 2 on protein substrate.
   Anal Biochem. 505, 36-42, 2016.
   DOI: http://dx.doi.org/10.1016/j.ab.2016.04.012
   IF: 2.243 (2015)

3. Király, R., **Thangaraju, K.**, Nagy, Z., Collighan, R., Nemes, Z., Griffin, M., Fésüs, L.: Isopeptidase activity of human transglutaminase 2: disconnection from transamidation and characterization by kinetic parameters.
   Amino Acids. 48 (1), 31-40, 2016.
   DOI: http://dx.doi.org/10.1007/s00726-015-2063-5
   IF: 3.196 (2015)

Address: 1 Egyetem tér, Debrecen 4032, Hungary Postal address: Pf. 39. Debrecen 4010, Hungary
Tel.: +36 52 410 443 Fax: +36 52 512 900/63847 E-mail: publikaciok@lib.unideb.hu, ¤ Web: www.lib.unideb.hu

Total IF of journals (all publications): 8,635
Total IF of journals (publications related to the dissertation): 8,635

The Candidate's publication data submitted to the iDEa Tudóstér have been validated by DEENK on the basis of Web of Science, Scopus and Journal Citation Report (Impact Factor) databases.

07 October, 2016

Address: 1 Egyetem tér, Debrecen 4032, Hungary Postal address: Pf. 39. Debrecen 4010, Hungary
Tel.: +36 52 410 443 Fax: +36 52 512 900/63847 E-mail: publikaciok@lib.unideb.hu, ¤ Web: www.lib.unideb.hu

SUBMITTED MANUSCRIPT'S

1. Kiruphagaran Thangaraju, Róbert Király, Máté A. Demény, János András Mótyán, Mónika Fuxreiter, László Fésüs. Genomic variants reveal differential evolutionary constraints on human transglutaminases and point towards unrecognized significance of transglutaminase 2. PLoS One. Under review.