Unravelling the mechanisms by which mutations in TG6 cause ataxia

Shannon Turberville

PhD Thesis

2019

Acknowledgments

I would like to say thank you to my supervisors, Prof Daniel Aeschlimann and Dr Konrad Beck and Zedira for their support during my PhD and Dr Michail Nomikos and Prof Tony Lai for the help they provided in setting up the *E.coli* expression system. In addition, I would like to thank both Dr Vera Knauper and Prof Jane Endicott for their support and feedback during the write up period of my PhD. A special thank you to my friends and family, specifically Rhiannon and Helen, I 100% could not have done this without you.

Abstract

Transglutaminase 6 (TG6) is an enzyme involved in the transamidation, deamidation and isopeptide cleavage of target substrates. TG6 is predominantly expressed in the brain and has been linked to progressive neurological dysfunction, most commonly ataxia. Spinocerebellar ataxia (SCA) is a group of neurological diseases characterised by progressive imbalance and cognitive impairment. Mutations in TG6 have been specifically linked to SCA35 using exome sequencing and genetic linkage analysis. Recent work investigating the TG6 mutants in a cellular context has identified that some mutants initiate the unfolded protein response and are targeted for proteolytic degradation by the proteasome, whilst other are not (Tripathy *et al.* 2017). In order to enable future intervention in the disease process, it is important to establish the underlying mechanism by which the pathogenic TG6 mutants cause neurodegeneration.

Here, the structure and function of wild-type TG6 and the pathogenic TG6 mutants were first analysed using a novel bioinformatic method to predict the impact of TG6 mutations. The biochemical methods optimised using TG2 and TG2 GTP binding mutants showed high sensitivity whereby the point mutations which affect TG regulation, activation and structure could be characterised. These methods in turn were used to distinguish the various molecular mechanisms the TG6 mutants undertake to cause neuronal cell death, work which has not been previously been carried out. Cellular studies provided further evidence for the diverse pathogenic mechanisms involved in the pathogenesis of SCA35 and highlighted the involvement of the unfolded protein response in those TG6 mutants impacting TG6 structure. Ongoing research will provide valuable insight into the mechanisms of neuronal cell death and will aid in the validation of the bioinformatic method. Consequently, a treatment for SCA35 can be developed.

List of Abbreviations

A3v	Aggregation propensity value
A4v	Average aggregation propensity value
AATr	Area of the aggregation profile above the threshold per residue
AEBSF	4-(2-aminoethyl) benzenesulfonyl fluoride
AGA	Anti-Gliadin antibodies
AML	Acute myeloid leukaemia
ANOVA	Analysis of variance
ATP	Adenosine Trisphosphate
ATPγS	Adenosine 5'-O-(3- thiotriphosphate)
BCA	Bicinchoninic
BODIPY	Boron-dipyrromethene
BSA	Bovine serum albumin
BzATP	2' (3')-*O*-(4-benzoylbenzoyl) adenosine triphosphate
C/EBP	CCAAT/enhancer-binding protein
CADD	Combined annotation dependent depletion
CD	Circular dichroism
CHOP	C/EBP homologous protein
CHX	cycloheximide
CNS	Central nervous system
CSF	Cerebral spinal fluid
DMEM	Dulbecco's modified eagle media
DMSO	Dimethyl sulfoxide
DON	6-diazo-5-oxo-L-norleucine
DSSP	Database of secondary structure
DTT	Dithiothreitol
EAE	Experimental autoimmune encephalomyelitis
ECM	Extracellular matrix
EDTA	Ethylenediaminetetraacetic acid
EGTA	Ethylene glycol-bis (β-aminoethylether)-N, N, N', N'-tetraaccetic acid
ELISA	Enzyme-linked immunosorbent assay
ER	Endoplasmic reticulum

ExAC	Exome aggregation consortium
FL	Full length
FXIIIA	Factor 13 chain A
GA	Gluten ataxia
GAD67	Glutamate decarboxylase 67
GAE	Gluten ataxia with enteropathy
GAo	Gluten ataxia without enteropathy
GAP	GTPase activating protein
GDP	Guanosine diphosphate
GEF	Guanine exchange factor
GFAP	Glial fibrillary acidic protein
GFP	Green fluorescent protein
GMP	Guanosine monophosphate
GnomAD	Genome aggregation database
GST	Glutathione S-transferase
GTP	Guanosine triphosphate
GTPγSS	Guanosine 5'-O-(3- thiotriphosphate)
HEK	Human embryonic kidney
hiFBS	Heat inactivated foetal bovine serum
HLA	Human leukocyte antigen
HMM	Hidden markov models
HRP	Horseradish peroxidase
HSIE	His-SUMO-Intein
Hsp 60	Heat shock protein 60
HST	Hotspot threshold
IMAC	Immobilised metal affinity chromatography
IP_3	Inositol 1, 4, 5-trisphosphate
IPTG	Isopropyl-β-D-thiogalactopyranoside
ITC	Isothermal calorimetry
LMW	Low molecular weight
MBP	Maltose binding protein
MDC	Mono-dansyl cadavarine
MOI	Multiplicity of infection

MRI	Magnetic resonance imaging
MS	Multiple Sclerosis
MS/MS	Multiple steps of mass spectrometry
MSA	Multiple sequence alignment
NavSS	Sum of the average aggregation propensity values across the whole polypeptide sequence
NCBI	National centre for biotechnology information
NEM	N-ethylmaleimide
NeuN	Neuronal Nuclei
NT	Nucleotide
NusA	N-utilisation substance A
OD_{600}	Optical density at 600nm
P2X7R	P2X purinoceptor 7
PBS	Phosphate buffered saline
PBS-T	Phosphate buffered saline Tween 20
PD	Parkinsons Disease
PDB	Protein database
PDON	Acetyl-P (DON) LPF-NH2
PHQ	Benzyl chlorocarbonate
PMSF	Phenylmethylsulfonyl fluoride
PolyPhen-2	Polymorphism Phenotyping
PVDF	Polyvinylidene fluoride
RIPA	Radioimmunoprecipitation assay
RT	Room temperature
SCA	Spinocerebellar ataxia
SD	Standard deviation
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SEM	Standard error of the mean
sf	*Spodoptera frugiperda*
SFP	Stansted fluid power
SIFT	Sorting Intolerant From Tolerant
SPR	Surface plasmon resonance

STS	Staurosporine
sXBP1	Spliced X-box binding protein 1
TBS	Tris buffered saline
TBS-T	Tris buffered saline tween 20
TG	Transglutaminase
TG6-S	Transglutaminase 6 – short
THSA	Total hotspot area
TLR4	Toll like Receptor 4
Trx	Thioredoxin
UHS	Uncombable hair syndrome
UPR	Unfolded protein response
WT	Wild type
XW1	(2S/2R)-2-amino-7-ethoxy-7-oxoheptanoic acid
βcore	β sandwich and catalytic core
ΔG	Change in Gibbs free energy
ΔΔG	Change in the change of Gibbs free energy

Table of contents

List of Figures

List of Tables

CHAPTER 1

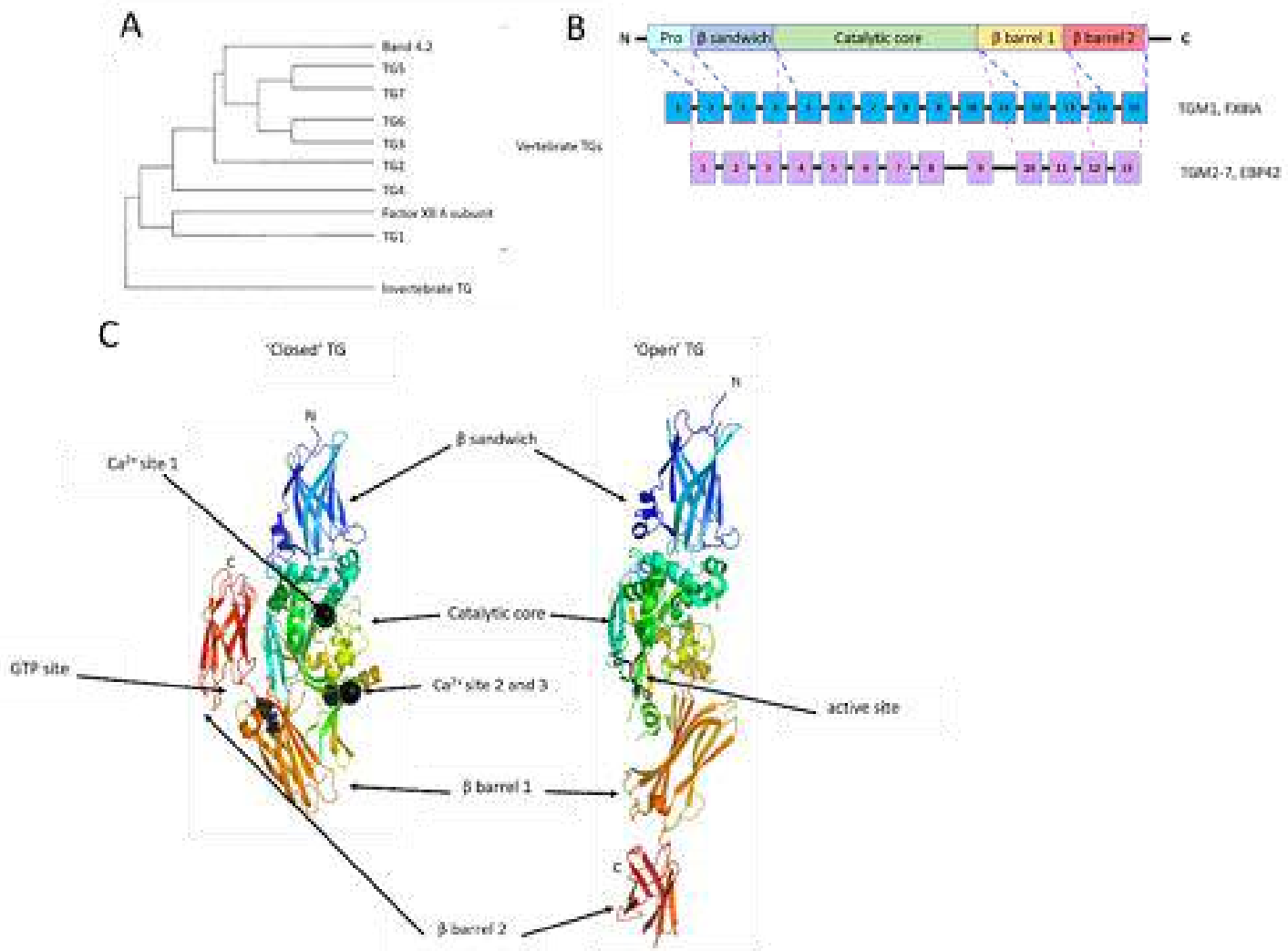
1. INTRODUCTION

Transglutaminases (TGs) are calcium-dependent enzymes responsible for the post-translational modification and crosslinking of target substrates. TGs are expressed in both vertebrates and invertebrates but they are most well characterised in mammals.

1.1 The Transglutaminase family.

The TG family consists of nine members in humans: TG1-7, Factor XIIIA (FXIIIA) and band 4.2 protein (Fig 1.1A). They are closely related to the cysteine protease family, which make up part of the papain-cysteine protease superfamily (Lorand *et al.* 1992). They have a well-preserved gene organisation and can be split into two subclasses based on their number of exons (Aeschlimann and Paulsson 1994) (Fig. 1.1B). Seven members (TG2 (Fraij and Gonzales 1997), TG3 (Kim *et al.* 1994), TG4 (Dubbink *et al.* 1998), TG5, TG6, TG7 (Grenard *et al.* 2001) and band 4.2 (Korsgren and Cohen 1991)) are coded by genes with 13 exons, whilst two members (TG1 (Kim *et al.* 1992) and Factor XIIIA (Ichinose and Davie 1988) are coded by genes containing 15 exons (Fig. 1.1B). TG1 and FXIIIA have an additional *N*-terminal extension, which is coded for by exon 1 and is not present in the other TGs. Exon 9 splits to form two exons, giving rise to 15 exons. Furthermore they all have similar protein domain structure (β sandwich, catalytic core and two β barrel domains) that can adopt two main conformations (Fig. 1.1C) which is discussed further in Section 1.5

TG2 is the most extensively researched of the TG isoforms and has been implicated in a number of diseases such as cancer, osteoarthritis and gluten related disorders (Iismaa *et al.* 2009; Eckert *et al.* 2014). TG3 and TG6 have also been implicated in gluten related disorders such as dermatitis herpetiformis (DH) and gluten ataxia (GA) (Sardy *et al.* 2002; Hadjivassiliou *et al.* 2008a). Whilst TG3 has been comprehensively researched, not much is known about the biochemical characteristics of TG6. TG3 and TG6 have a sequence identity of 50% and are present in the same gene cluster and are predicted to have similar structure and function (Fig. 1.1A and Fig. 1.2). Consequently, the work carried out on TG3 can be used to inform on the expected parametres of TG6. Furthermore, the data obtained on TG2 can also be used to further research in TG6. TG2 and TG6 share sequence identity of 41%, so data collected on TG2 could be relevant for TG6 (Fig. 1.2).


A
Band 4.2
TG5
TG7
TG6
TG3
TG2
TG4
Factor XIII A subunit
TG1
Invertebrate TG
Vertebrate TGs
B
N
Pro
β sandwich
Catalytic core
β barrel 1
β barrel 2
C
TGM1, FXIIIA
TGM2-7, EBP42
C
'Closed' TG
'Open' TG
N
β sandwich
Ca2+ site 1
Catalytic core
GTP site
Ca2+ site 2 and 3
active site
β barrel 1
β barrel 2
C

1.1.1 Transglutaminase 2

The human *TGM2* gene is located on chromosome 20q11-12 and codes for TG2, a protein with a molecular weight of 78 kDa. TG2 is constitutively expressed in a wide variety of cells e.g. epithelial cells, smooth muscle cells and fibroblasts (Iismaa *et al.* 2009). Intracellular TG2 is mostly cytosolic, but can also be localised to the inner side of the plasma membrane or the nuclear membrane (Iismaa *et al.* 2009). Intracellular TG2 contributes to crosslinking proteins during apoptosis, when there is sufficient Ca^{2+} for TG2 activation (Iismaa *et al.* 2009). Independently of transamidase activity, TG2 can act as a GTPase and mediate G-protein signalling via the α1B/ α1D adrenergic receptor (Iismaa *et al.* 2000). The biological roles of extracellular TG2 include wound healing, angiogenesis and bone remodelling, via stabilisation of the extracellular matrix (ECM) (Aeschlimann and Thomazy 2000).

1.1.2 Transglutaminase 3

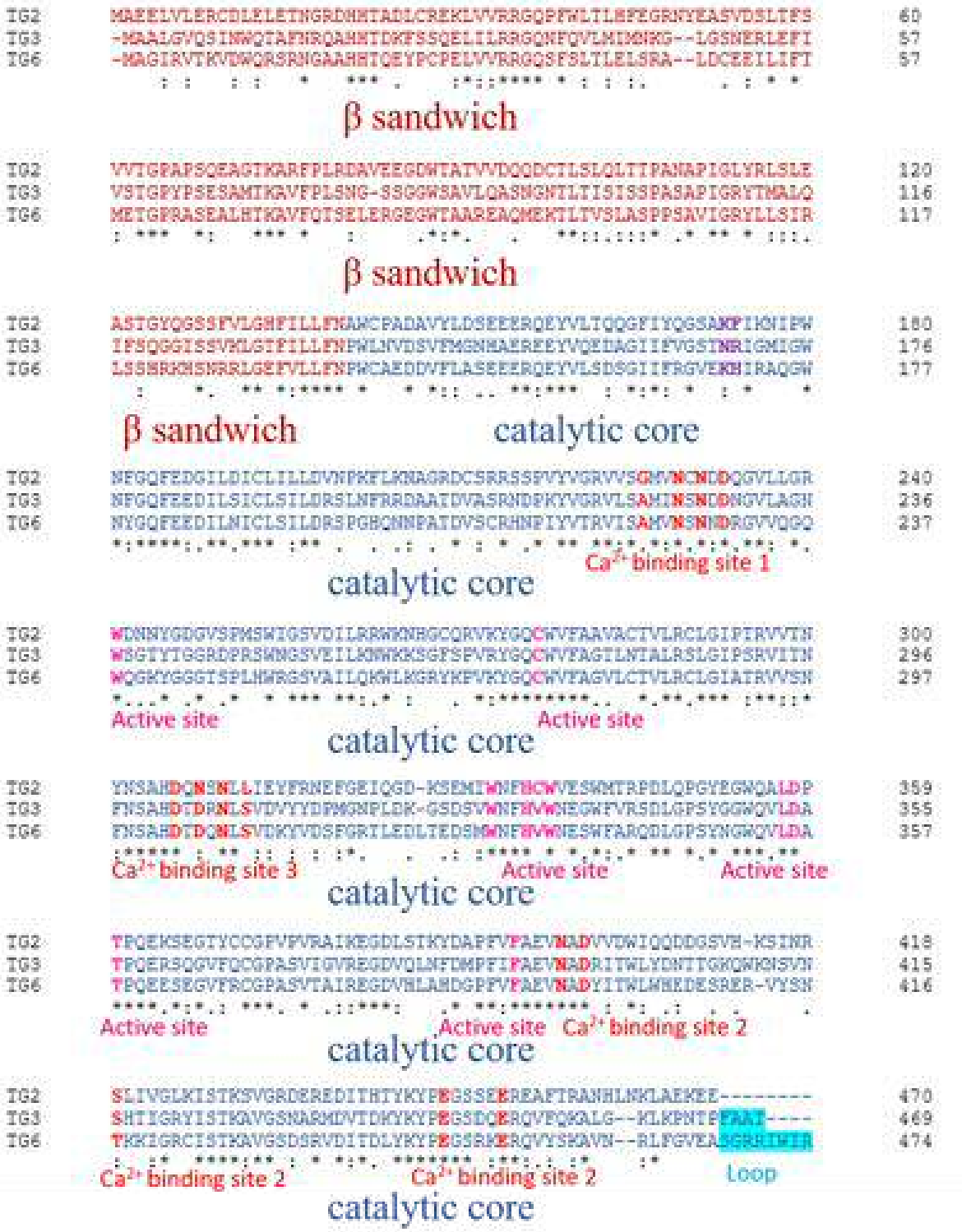
TGM3 is present on chromosome 20p13 and codes for the inactive zymogen, TG3, which is approximately 77 kDa (Kim *et al.* 1990). TG3 is activated by proteolytic cleavage, giving rise to a 50 kDa β sandwich and catalytic core domain, and a 25 kDa fragment that encodes 2 β barrel domains (Kim *et al.* 1990; Kim *et al.* 1993). It is most abundantly expressed in the skin, brain and placenta (Kim *et al.* 1990; Greenberg *et al.* 1991; Kim *et al.* 1999). In the epidermis, TG3 crosslinks various structural proteins that are important in the formation of the cell envelope barrier, for example lokrin, small proline-rich proteins, involucrin, keratins and filaggrin (Steinert and Marekov 1995). TG3 also has a role in the crosslinking of trichohyalin in the hair follicle and modification of keratin intermediate filament in hair fibre morphogenesis (Steinert *et al.* 2003). These activities contribute to the appearance of $TG3^{-/-}$ mice, which have curled whiskers and fur (John *et al.* 2012). Mutations in *TGM3* have been linked to uncombable hair syndrome (UHS) (Sardy *et al.* 2002; Basmanav *et al.* 2016). Also, TG3 autoantibodies are associated with the immune-related disease, Dermatitis herpetiformis (DH), which presents in the skin (Sardy *et al.* 2002).

1.1.3 Transglutaminase 6

TGM6, the gene encoding TG6 is located on chromosome 20p13 (Grenard *et al.* 2001; Wang *et al.* 2010). It is expressed predominantly in the central nervous system (CNS) and is therefore known as the neuronal transglutaminase (Thomas *et al.* 2013). However, its expression is also prominent in the testis and skin (Stamnaes *et al.* 2010; Thomas *et al.* 2013). TG6 is also expressed in the lung small cell carcinoma cell line (H69, cell line with neuronal characteristics) which was used to isolate the full-length (FL) cDNA (Thomas *et al.* 2013). The open reading frame was predicted to be 2120 nucleotides (nt) coding for 706 amino acids, with a calculated molecular mass of 79276 Da.

Alternative splicing at exon 12 of *TGM6*, was also identified in H69 cells and resulted in early termination within exon 13. The resulting TG6 (called TG6-S) is truncated after β barrel 1, is composed of 625 amino acids (Mr =70480 Da) and has a unique *C*-terminus. Additional *TGM6* splice variants have been identified in the mouse brain: one lacking exon 3 and the other lacking both exon 3 and 4 (Thomas *et al.* 2013). Both splicing events result in TG6 with an altered β sandwich domain, which is also seen in other TGs (TG2 (Aeschlimann *et al.* 1998), TG4 (Cho *et al.* 2010) and TG5 (Candi *et al.* 2001)).

TG6 has been associated in a number of neurological diseases such as GA, multiple sclerosis (MS), spinocerebellar ataxia (SCA) and Parkinsons disease (PD), which are discussed further in Section 1.10 and 1.13). As TG6 is prominently expressed in the brain, regions of the CNS expressing TG6 were identified (Liu *et al.* 2013). High TG6 expression was detected in the cell layers of the cerebral cortex, olfactory bulb and the cerebellum, which contain the neuronal cell bodies (Thomas *et al.* 2013). In the cerebellar cortex Purkinje cells showed weak TG6 staining, whilst the neurones deep in the cerebellar nucleus were positive for TG6 (Liu *et al.* 2013). TG6 immunoreactivity was observed in the ventral horn and the substantia nigra, most predominantly in the reticular part (Liu *et al.* 2013).


β sandwich
β sandwich
β sandwich
catalytic core
Ca2+ binding site 1
catalytic core
Active site
Active site
catalytic core
Ca2+ binding site 3
Active site
Active site
catalytic core
Active site
Active site
Ca2+ binding site 2
catalytic core
Ca2+ binding site 2
Ca2+ binding site 2
Loop
catalytic core

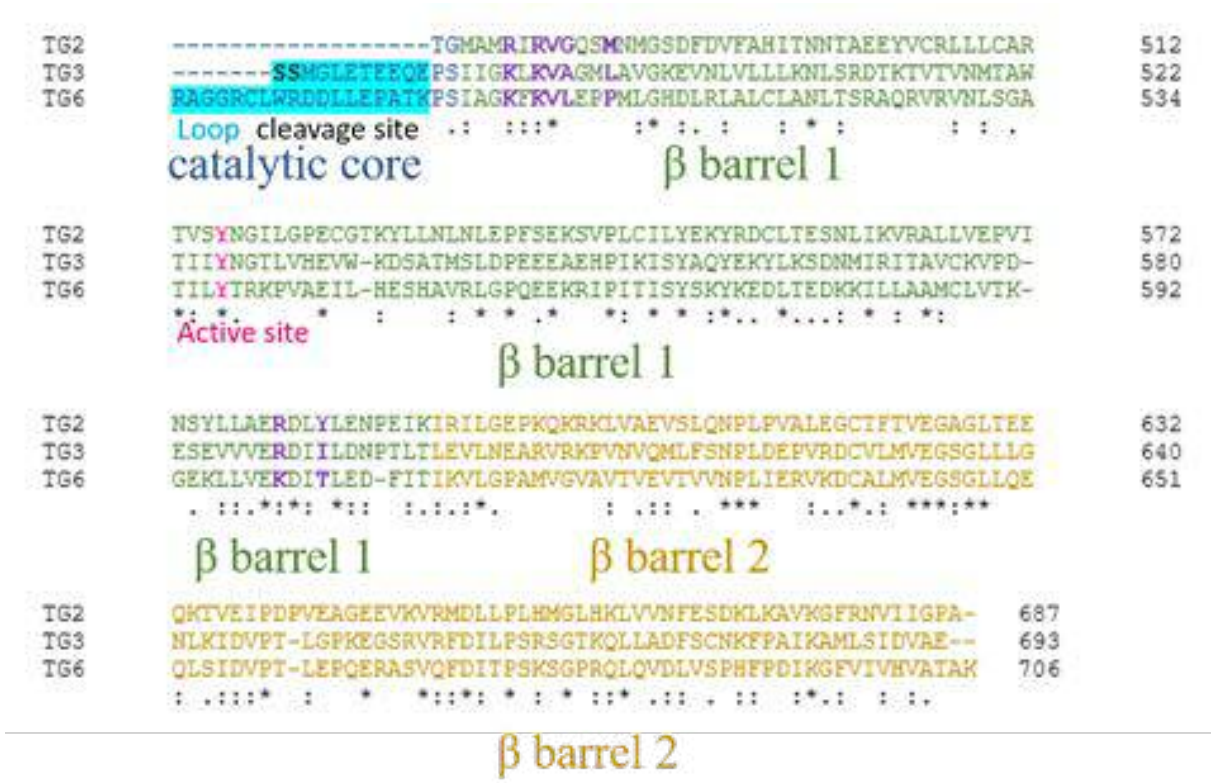


Figure 1.2: Sequence alignment of TG2, TG3 and TG6.
Multiple sequence alignment (MSA) for TG2, TG3 and TG6. The four domains and are highlighted (β sandwich (dark red, res.3-137), catalytic core (dark blue, res. 138-494), β barrel 1 (green, res. 495-606) and β barrel 2 (dark yellow, res 607-706). The active site residues (pink), Ca^{2+} binding sites (red) and GTP binding pocket (purple) residues are highlighted in bold. The cleaved loop (cleavage site in black) in TG3 and the extended loop domain of TG6 is highlighted in cyan. Conservation is denoted as *, fully conserved residue, :, strongly conserved properties of residues and ., weakly conserved properties of residue

In all regions, TG6 expression was restricted to neurons (co-staining with a neuron specific nuclear protein, neuronal nuclei N (NeuN)) and TG6 was localised to the cytoplasm of the neuron cell bodies. GABAergic inhibitory neurons were investigated for TG6 expression using glutamate decarboxylase 67 (GAD67, a marker for GABAergic neurones)-GFP knock in mice. This study showed that the majority of TG6-expressing cells were GFP positive and TG6 was therefore expressed in inhibitory GABAergic neurones. Together, these results show the regions of the brain important for locomotion are highly TG6 positive and expression is localised to GABAergic neurons.

To identify the role of TG6 in the CNS, TG6 expression was monitored throughout embryogenesis. At embryonic day 11, when the generation of neurons starts, a small amount of TG6 can be detected in the mouse brain, and there is greater TG6 reactivity in the developing spinal cord (Liu *et al.* 2013). At embryonic day 13, TG6 expression can be detected throughout the brain and expression is prominent in areas undergoing neuronal differentiation e.g. the mesencephalon and layer of the cortex thought to be the progenitor cells. By embryonic day 16 TG6 expression is observed in the outer layers of the neocortex. TG6 mRNA is detected in layers I-IV and the highest TG6 expression is detectable in layers V/VI, indicating that TG6 expression correlates with neurogenesis and neuronal differentiation (Liu *et al.* 2013).

1.2 TG enzymatic function.

TGs are involved in many different types of reaction, which can be grouped into three main reaction types: transamidation, hydrolysis and esterification (Fig. 1.3) (Lorand and Graham 2003). In the transamidation reaction, glutamine-rich substrates act as the acceptor substrate which are covalently crosslinked via the formation of a γ-glutamyl-ε-lysine linkage to a lysine donor substrate (Fig. 1.3A). This reaction often leads to the formation of insoluble supramolecular structures (Aeschlimann and Thomazy 2000). A primary amine can replace lysine and be incorporated in the same manner onto the glutamine side chain of the TG substrate (Fig. 1.3B, (Lorand and Graham 2003)).
Acylation is carried out in the formation of the acylenzyme intermediate (Fig. 1.4, step C) and the acylation of the lysine substrate onto the donor protein, this results in the release of ammonia (Fig. 1.3C). As a result acylation and amine incorporation are competing reactions. Glutamine-containing peptides can undergo hydrolysis via two mechanisms: site specific deamidation and isopeptide cleavage Site-specific deamidation results in side chain conversion from glutamine to glutamate (Fig. 1.3E). TG's isopeptidase activity induces the cleavage of the γ-glutamyl-ε-lysine linkage formed by crosslinking (Fig. 1.3F). Furthermore glutamine substrates can undergo esterification, whereby a fatty acid molecule is added to the glutamine substrate (Fig. 1.3D).

IMAGE REMOVED DUE TO COPYRIGHT.

Figure 1.3: Post-translational reactions carried out by TGs.
TG reactions can be split into three main groups: transamidation, esterification and hydrolysis. A: Protein crosslinking involves the formation of a γ-glutamyl-ε-lysine linkage between lysine donor (purple ellipse) and acceptor glutamine (blue square), B: Amine incorporation involves the incorporation of a primary amine into the glutamine acceptor substrate. C: Acylation of donor lysine. D: Esterification of glutamine acceptor substrate. E: Deamidation involves the conversion of glutamine to glutamate on glutamine substrate. F: Isopeptide cleavage of γ-glutamyl-ε-lysine linkage. Figure adapted from (Lorand and Graham 2003).

results in the release of NH_3 and formation of the acylenzyme intermediate via a thioester bond (Fig 1.4C). Upon approach of the lysine donor substrate (substrate 2) a complex forms) and the lysine donor attacks thio-acyl-bond between TG and the glutamine substrate (Fig. 1.4E, oxyanion intermediate 2). The cross-linked product is then released by a deacylation reaction and the TG2 active site is regenerated (Fig. 1.4F).

IMAGE REMOVED DUE TO COPYRIGHT.

Figure 1.4: TG Transamidation mechanism.
Transamidation is one of the major reactions carried out by mammalian transglutaminases. This involves the binding of the glutamine substrate via its γ carboxyamide group to the thiolate group of cysteine in the catalytic triad which forms the oxyanion intermediate. The subsequent acylation reaction causes the release of NH_3 and formation of the acylenzyme intermediate. The second oxyanion intermediate is formed by the nucleophilic attack of the lysine donor substrate (substrate 2). The crosslinked product is released by a deacylation reaction (Figure adapted from (Iismaa *et al.* 2009)).

1.4 TG Structure.

To understand the structural functional relationship of TGases, high-resolution structures of TG2 (Liu *et al.* 2002; Pinkas *et al.* 2007), TG3 (Ahvazi *et al.* 2002) and FXIIIA (Stieler *et al.* 2013) have been solved. These structures revealed that mammalian TGs consist of four domains, an *N*-terminal β sandwich, followed by an α/β catalytic core and then two β barrel domains. As discussed in greater below, TG activation is accompanied by a closed to open conformational change whereby closed conformation is adopted upon binding of GTP. The binding of three Ca^{2+} and the TG substrate drives a conformational change resulting in an open structure (Fig. 1.5) (discussed in Section 1.4.1-2) (Ahvazi and Steinert 2003; Pinkas *et al.* 2007). However, various intermediate conformations have been captured where by all three Ca^{2+} sites are occupied by Ca^{2+} but the open extended conformation has not been adopted. Also, TG3 has been captured where the closed conformation is adopted and with Ca^{2+} occupying site 1 and 2 and Mg^{2+} in site 3 in the presence of nucleotide (discussed in Section 1.5.2). Furthermore, TG3 specifically has been captured as an inactive zymogen binding Ca^{2+} in site 1 and requires cleavage to allow binding of the two additional Ca^{2+}. This cleavage event has so far not been observed in TG2 and TG6 and is unlikely due to the high transamidation activity of the intact protein (Thomas *et al.* 2013).

1.4.1 TG2 structure.

TG2 in closed compact (Fig. 1.5, left) has been determined on multiple occasions bound to various small molecule ligands that occupy the nucleotide binding pocket (Table 1.1, Fig 1.6A). The overall architecture is similar for these structures and the structures gave an insight into the nucleotide binding pocket (Section 1.6.1). A comparative structural analysis of TG2 revealed that in an intracellular environment with low concentrations of Ca^{2+} and high concentrations of GTP, it assumes a closed compact conformation (Jang *et al.* 2014).

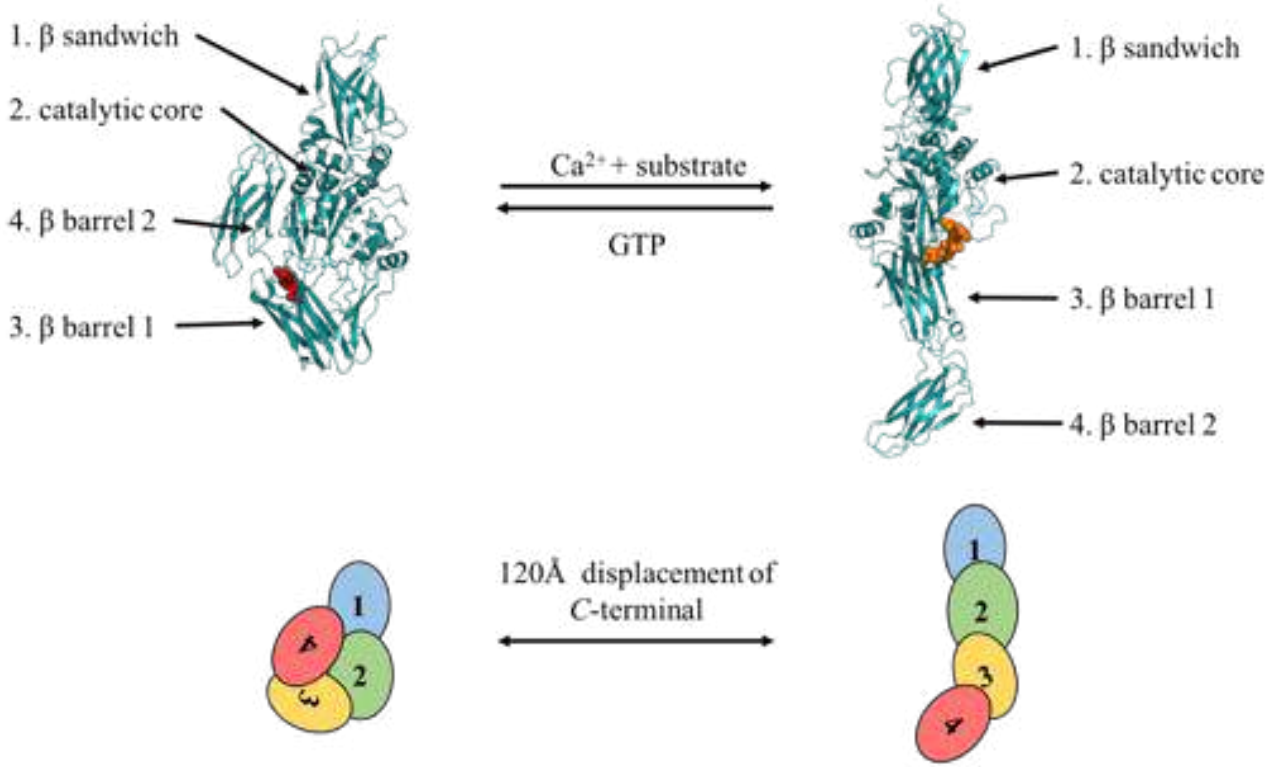


Figure 1.5: Overall architecture of TG2 structures.

TG2 has been captured in two main conformations, the closed compact conformation (left) and the extend environment with low concentrations of Ca^{2+} and high concentrations of GTP (intracellular), it assumes a cl *al.* 2014) . Upon Ca^{2+} and substrate binding, TG2 undergoes a 120 Å shift in the *C*-terminal region and ado 2007). This commonly occurs in an environment with high concentrations of Ca^{2+} (in ECM). Structures are and the inhibitor, PDON, (orange) are shown as spheres. The below schematic is representative of the confor

Based on sequence conservation C336, W337, L357, T360, F392 and Y516 were proposed to make up the TG2 active site (Ahvazi and Steinert 2003) (Fig. 1.6A). T360 is located at the mouth of the acyl acceptor tunnel and is conserved across all catalytically active members of the TG family (Liu *et al.* 2002). The pocket is bridged by two tryptophan residues (W241 and W332). W241 is highly conserved and is predicted to be important for the stabilisation of the transition state during transamidation catalysis (Figure 1.4B and E) (Murthy *et al.* 2002). This stabilisation is via the positioning of the indole nitrogen towards the oxyanion intermediate and contributes van der Waals or quadrupole-ion interactions. In addition, Y516 from the first loop forms a hydrogen bond with the catalytic C277, which must be broken to allow the first substrate access to the catalytic triad.

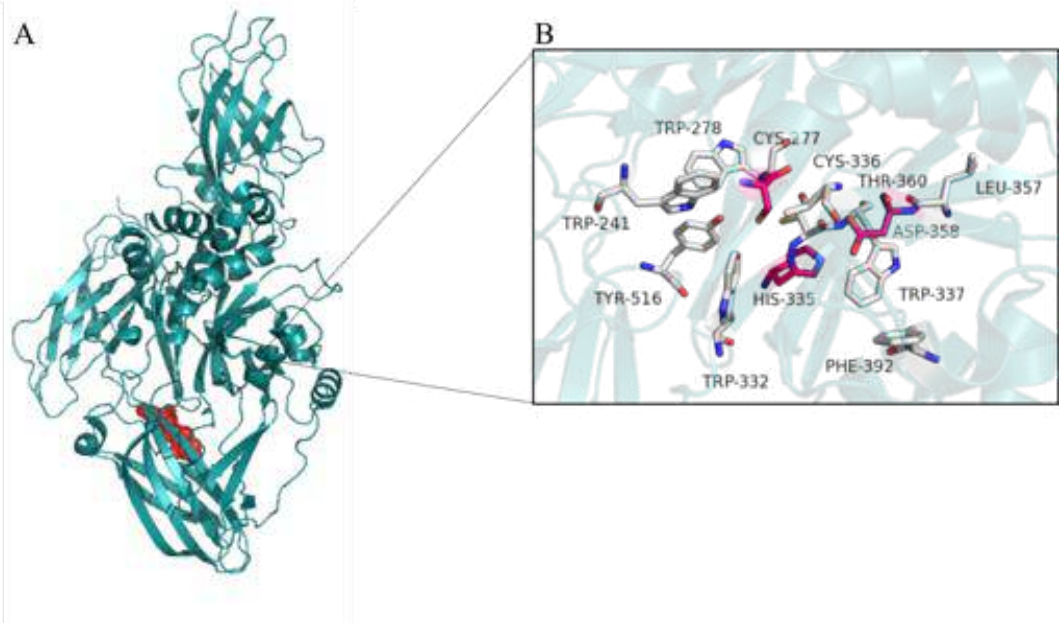


Figure 1.6: Closed TG2 and active site.
TG2 in closed GTP bound conformation. A: Overall structure of TG2 (PDB: 4PYG, (Jang *et al.* 2014). B: TG2 active site residues (white) and catalytic triad residues (pink) of TG2.

In higher concentrations of Ca^{2+} as for example experienced in the ECM, it predominantly assumes an extended open and active conformation upon substrate binding (Pinkas *et al.* 2007). When the *N*-terminal β sandwich and catalytic core domains of the extended and compact TG2 conformations are superimposed, the *C*-terminal residues are displaced by up to ~120 Å (Fig 1.5). Furthermore, the β strand, L312 to R317 (hinge region), becomes α helical upon transition to the open conformation causing changes which extend to the active site.

Table 1.1: TG2 X-ray crystal structures.

TG2 conformation	Bound molecule	PDB	Ref
Closed	ATP	3LY6	Han *et al.* 2010
Closed	GDP	1KV3	Liu *et al.* 2002
Closed	GTP	4PYG	Jang *et al.* 2014
Closed (G224V)[1]	GTP	6A8P	Ha *et al.* 2018
Open	PDON	2Q3Z	Pinkas *et al.* 2007
Open	N-[(benzyloxy)carbonyl]-6-diazonio-5-oxo-L-norleucyl-L-valyl-L-prolyl-L-leucine.	3S3J	Lindemann *et al.* 2012
Open	N~2~-[(2S)-2-[(benzyloxy)carbonyl]amino-7-ethoxy-7-oxoheptanoyl]-L-glutaminyl-L-prolyl-L-leucine	3S3P	Lindemann *et al.* 2012
Open	N-[(2R)-2-[(benzyloxy)carbonyl]amino-7-ethoxy-7-oxoheptanoyl]-L-valyl-L-prolyl-L-leucine	3S3S	Lindemann *et al.* 2012

[1] TG2•GTP containing point mutation G224V

To investigate this open conformation of TG2, the irreversible pentapeptide inhibitor, PDON (Acetyl-P(DON (6-diazo-5-oxo-L-norleucine))LPF-NH_2, single letter amino acid code) was used in the presence of Ca^{2+} (Fig. 1.7A) (Pinkas *et al.* 2007). However, no Ca^{2+} occupied the Ca^{2+} binding sites of this structure. PDON mimics the acyl donor substrate (substrate 1, Fig 1.4) and binds TG2 with a K_i of 60 nM. The hydrogen bonding and hydrophobic interactions between TG2 and PDON are extensive (Fig. 1.7A). The PDON ketone interacts with W241 and the backbone amide of C277 which reflects the oxyanion intermediate observed in acylation and deacylation transition states (Figure 1.4B and E). Hydrophobic stacking interactions can occur between the phenylalanine moiety in PDON and TG2 active site residues further stabilising the

inhibitor. Upon TG-inhibitor complex formation the peptide is buried by more than 70%, which is facilitated by the proline of the inhibitor assuming the *trans* configuration and causing the inhibitor phenylalanine to bind deeper within the hydrophobic pocket. PDON occupies one side of the hydrophobic tunnel. A second binding pocket within the hydrophobic tunnel has been predicted for the lysine donor substrate (substrate 2) which participates in the transamidation reaction (Ahvazi and Steinert 2003).

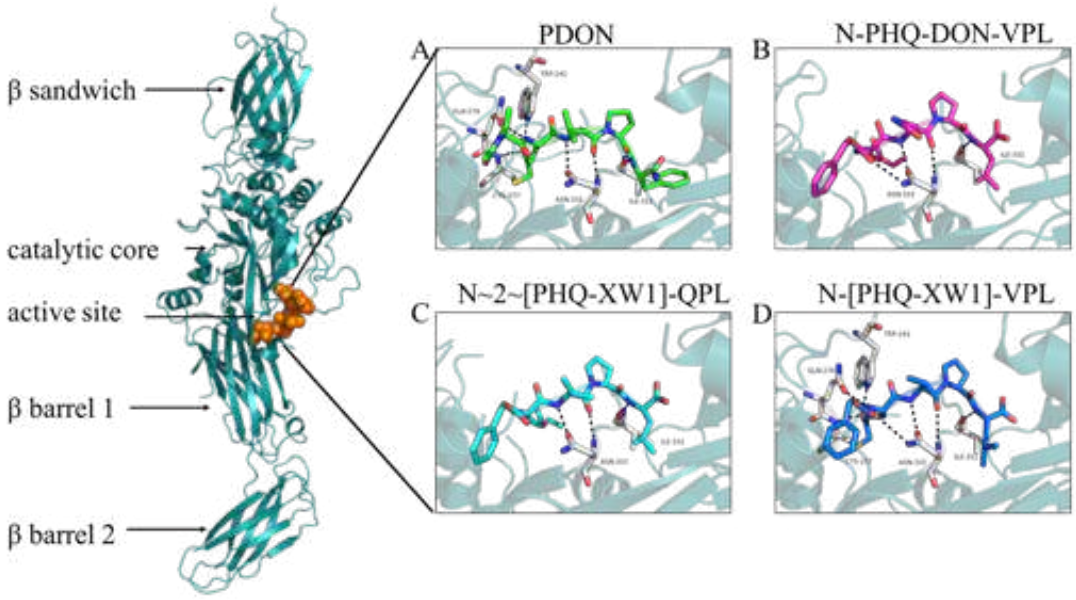


Figure 1.7: Binding TG2 gluten peptidomimetic inhibitors.
TG2 has been captured in open and extended conformation (cartoon representation). The inhibitors mimic the glutamine substrate required for transamidation (shown as sticks. Interacting residues (white) are highlighted and polar interactions <3.5Å are shown (black dotted lines) A: PDON (green) irreversibly binds to the active site cysteine of TG2 (PDB: 2Q3Z, (Pinkas *et al.* 2007). B: Reversible binding of N-PHQ-DON-VPL (magenta), peptide inhibitor (PDB: 3S3J, (Lindemann *et al.* 2012)). C: Reversible binding of N~2~ [PHQ-XW1]-QPL (cyan), peptide inhibitor (PDB: 3S3P, Lindemann *et al.* 2012)) D: Irreversible binding of peptide inhibitor: N-[PHQ-XW1]-VPL (blue) via active site cysteine (PDB: 3S3S, Lindemann *et al.* 2012).

Additional inhibitors have since been identified (PDB: 3S3J, 3S3P and 3SPS, Table 1.1 and Fig.1.7). For example, N-PHQ-DON-VPL (Fig. 1.7B), N~2~ [PHQ-XW1]-QPL (Fig. 1.7C) and N-[PHQ-XW1]-VPL (Fig. 1.7D), where PHQ represents benzyl chlorocarbonate and XW1 is (2S/2R)-2-amino-7-ethoxy-7-oxoheptanoic acid (full names listed in Table 1.1). The overall architecture of the four inhibitor-bound structures (PDB: 2Q3Z, 3S3J, 3S3P and 3S3S) are superimposable and allowed investigation of the peptide binding of TG2 (Pinkas *et al.* 2007). Despite successful

attempts of trapping TG2 in the open, 'active' conformation, no structures have been determined where the three Ca^{2+} sites are occupied with Ca^{2+} (with or without the presence of the substrate).

1.4.2 Structure of TG3.

TG3 has been structurally characterised in the inactive zymogen (contains Ca^{2+} at site 1 and no loop cleavage (S469-S470)), activated and nucleotide bound (guanosine 5'-O-(3-thiotriphosphate (GTPγS), guanosine diphosphate (GDP) and guanosine monophosphate (GMP)) conformations (detailed in Table 1.2, (Ahvazi *et al.* 2002; Ahvazi *et al.* 2003; Ahvazi *et al.* 2004a; Ahvazi *et al.* 2004b)).

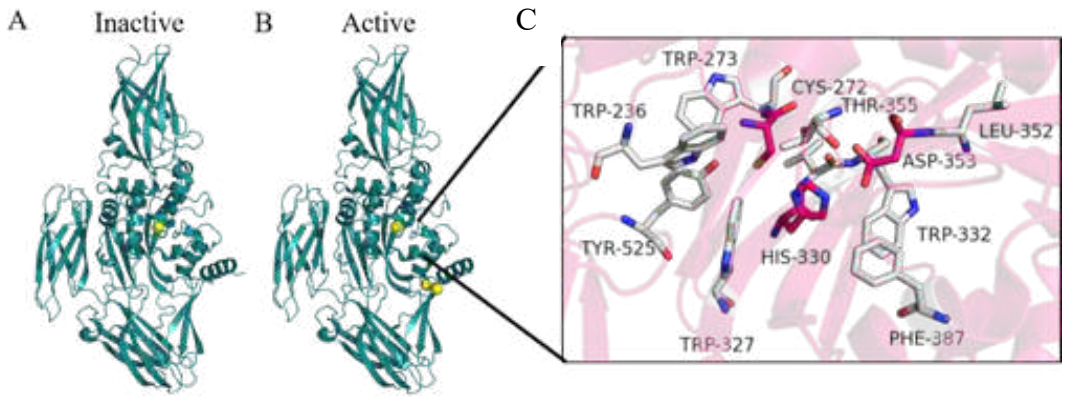


Figure 1.8: Overall architecture of TG3 structures.
TG3 has been crystallised in the inactive and active form. Structures are shown in cartoon representation, Ca^{2+} shown as spheres (yellow). A: TG3 inactive zymogen, showing one Ca^{2+} bound (PDB: 1L9M, (Ahvazi *et al.* 2002)) B: Dispase cleavage is required for binding of additional Ca^{2+} and activation (3x Ca^{2+}, PDB: 1L9N, (Ahvazi *et al.* 2002)). C: The active site residues (white) and catalytic triad residues (pink) of TG3.

The overall architecture of these TG3 structures are similar (Fig. 1.8). A flexible loop including residues 462-478 (subunit A) and 460-472 (subunit B) structure could not be determined, in all cases. An open substrate-bound conformation of TG3 is yet to been identified. Furthermore, there are differences in a loop domain (K321-S325) between the inactive zymogen, active TG3 (3x Ca^{2+}) and nucleotide bound TG3 (2x Ca^{2+} and 1x Mg^{2+}). These differences are explained in more detail in Section 1.5.2.

							al. 20036
Active	Dispase	Ca^{2+}	Ca^{2+}	Mg^{2+}/ Ca^{2+}	GDP	1VJJ	Ahvazi *et al.* 2004a
Active	Dispase	Ca^{2+}	Ca^{2+}	Mg^{2+}	GTPγS	1RLE	Ahvazi *et al.* 2004a
Active	Dispase	Ca^{2+}	Ca^{2+}	Mg^{2+}	GMP	1SGX	Ahvazi *et al.* 2004c

[1]NB: Not bound

The TG3 catalytic triad is made up of C272, H330 and D353 (Fig 1.8C) (Ahvazi *et al.* 2002; Ahvazi and Steinert 2003; Ahvazi *et al.* 2004a). In the inactive zymogen the active site is shielded from contact and buried within a narrow cleft. This is formed by two β strands contributed by the catalytic core domain and carboxyl terminal region, respectively. The conserved residues that compose the active site are W236, W273, F275,W327, F329, V331, W332, L352, T355, F387 and Y525 (Fig. 1.8C) (Ahvazi and Steinert 2003). These residues were identified using molecular docking software and sequence alignments with the other human TGs and are identical to those present in TG2 active site, with the exception of V331 in place of C336 in TG2.

1.4.3 Molecular modelling of TG6.

No high-resolution structure for TG6 is available, so molecular modelling of TG6 has been carried out to predict its structure (Fig. 1.9). TG6 has high sequence similarity to TG3 (50% overall and 59% in the catalytic core domain, Figure 1.2) and therefore Ca^{2+} bound TG3 was used as the template (PDB: 1L9N (Ahvazi *et al.* 2002). In both TG3 and TG6 there is a large flexible loop (residues 481-563 and 463-492 in TG3 and TG6 respectively) that lies between the catalytic core domain and the first β barrel that is not resolved in the TG3 crystal structure (Thomas *et al.* 2013). This loop is more

extended in TG6 compared with TG3 and is predicted to occlude the active site when in closed conformation. In TG3 this loop is cleaved during activation (dispase activated *in vitro*) to allow substrate access (Ahvazi *et al.* 2002; Thomas *et al.* 2013). A cleavage event is unlikely for TG6 as high specific activity of intact TG6 can be measured and therefore, it has been proposed the extended loop can allow entry of the substrate when activated (Thomas *et al.* 2013).

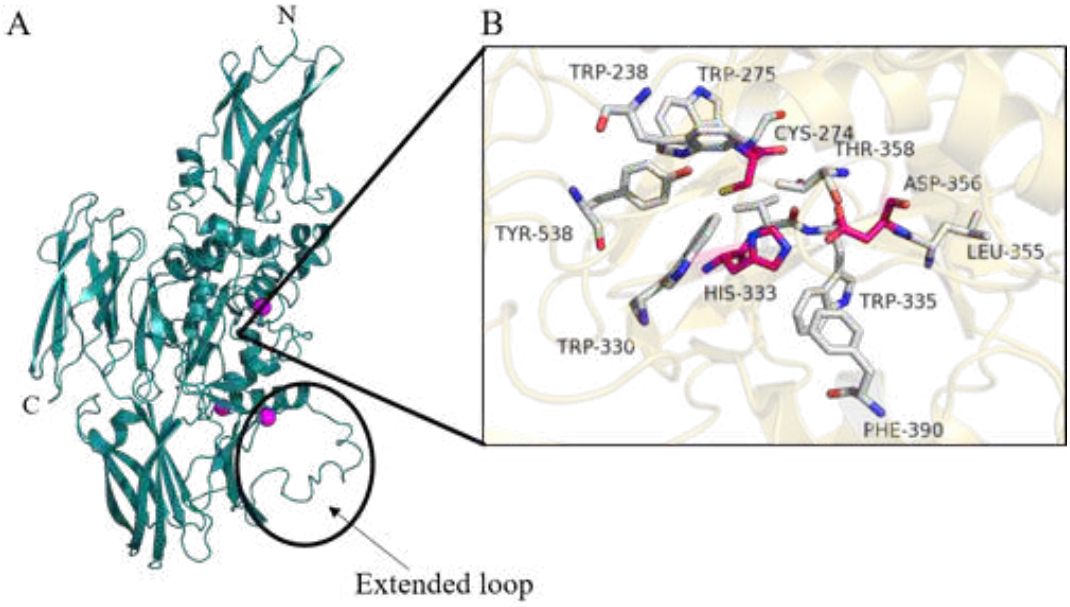


Figure 1.9: TG6 homology model.
A: TG6 aligned with and modelled on active TG3 (3xCa^{2+}, PDB: 1L9N, (Ahvazi *et al.* 2002)) using CLUSTAL Omega and input into SWISS model(Sievers *et al.* 2011; Bienert *et al.* 2017; Waterhouse *et al.* 2018). Structure in cartoon representation and Ca^{2+} shown as spheres (magenta). B: The active site residues (white) and catalytic triad residues (pink) of TG6.

As for TG2 and TG3 the catalytic triad, C274, H333 and D356, is situated in a hydrophobic cavity formed by the core domain and β-barrel 1 (Thomas *et al.* 2013). The sulfhydryl group of cysteine can form a thiolate-imidazolium ion pair with H333. N2 in the histidine ring hydrogen bonds with the side chain oxygen of D356. There are two tryptophan residues W238 and W330 buried near the surface which have a role in the gating of the hydrophobic pocket. Modelling of TG6 using TG3 as a template (PDB: 1L9N) and sequence alignment with other TGs enabled the residues W238, W275, W330, V334, W335, L355, T358, F390 and Y538 that form the active site to be predicted (Fig. 1.9), (Thomas H., 2004). The residues are identical to TG3

and also contain valine (also present in TG3) in place of cysteine seen in TG2 at the same position.

1.5 Ca^{2+} activation of TGs.

TG activation is dependent on the binding of three Ca^{2+} in three distinct pockets (Fig.1.10). Ca^{2+} site 1 is permanently occupied with $Ca^{2+,}$ which binds with high affinity and is important for structure stabilisation (Ahvazi *et al.* 2002; Ahvazi *et al.* 2003). The binding of two additional Ca^{2+} is predicted to cause activation by destabilising the hydrogen bonding in β barrel 1 which in turn destabilises the nucleotide binding site, lowering the affinity of TG for GTP (Im *et al.* 1990). The Ca^{2+} binding interactions have been characterised in TG3 (Ahvazi *et al.* 2002). For TG2 and TG6 sequence alignment approaches and modelling have been used to predict the residues involved (Table 1.3 and Fig.1.10) (Ahvazi *et al.* 2002; Kiraly *et al.* 2009; Thomas *et al.* 2013). Overall, each TG is predicted to contain three Ca^{2+} binding sites identified by alignment (sequence and structural) with TG3 (Fig. 1.2 and 1.10A) (Datta *et al.* 2006; Kiraly *et al.* 2009)

1.5.1 TG2 Ca^{2+} binding.

Basal intracellular levels (within the cytosol) of calcium are between 20 nM to 100 nM, which can increase to 500 nM upon activation of inositol,4,5-triphosphate (IP_3) and ryanodine receptors on the endoplasmic reticulum (ER), (Jeitner *et al.* 2009). Using isothermal titration calorimetry (ITC) the K_D of WT TG2 for Ca^{2+} was 0.1 ± 0.03 μM for the 'strong binding site' (site 1) and 4.6 ± 1.3 μM for 'weak binding sites' (sites 2 and 3) (Kiraly *et al.* 2009). Upon Ca^{2+} binding the β barrels can rotate around 180 °, in the plane of the catalytic and β sandwich domain. This movement exposes the active site, activating enzymatic activity (Pinkas *et al.* 2007). I416 to S419 and L577 to E579 form an antiparallel β-sheet pair that stabilises the GTP binding site. When Ca^{2+} binds, the hydrogen bonds in the antiparallel β strands are disrupted. This results in the release of GDP from the nucleotide binding site. Ca^{2+} could also modify the position of the glutamic acid residues, further disrupting the nucleotide binding pocket (Liu *et al.* 2002).

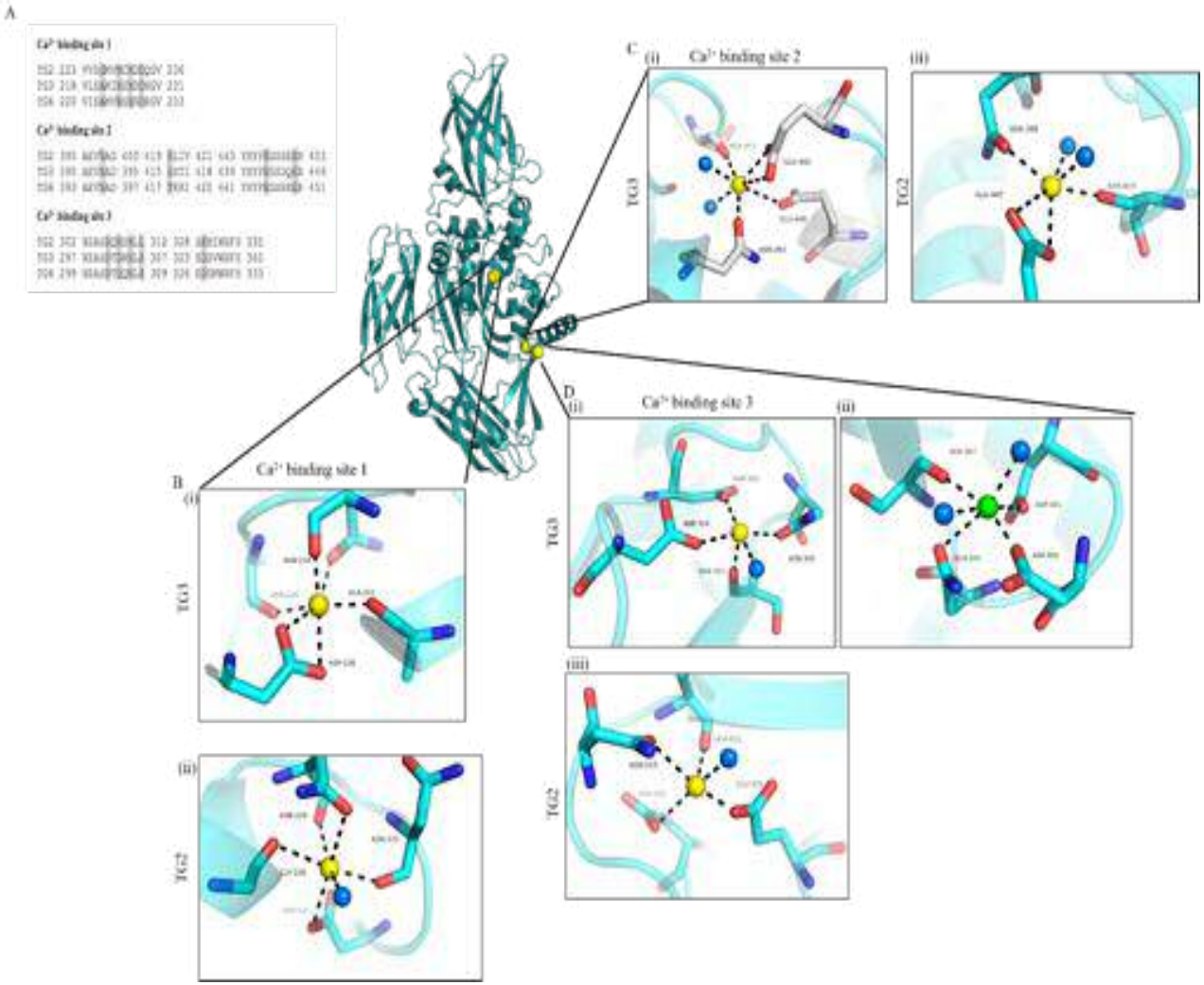


Figure and TG
A: Alig
CLUST
Residue
the stab
sites. Ca
bound
bound T
is in ca
residues
cyan, ni
(yellow)
shown
highligh
binding
A221,
Ca^{2+} is
D233 ar
TG3 Ca
E448 a
stabilise
D: Ca^{2+}
stabilise
D324 ar
by D30
In TG2
L312, E

To identify the residues involved the stabilisation of TG2 Ca^{2+}, sequence alignment was carried out with TG3 (Datta *et al.* 2006). The TG2 Ca^{2+} binding sites were visualised with structural alignment with TG3 (Fig. 1.8B). TG2 Ca^{2+} binding site 1 is made up of the main chain carbonyl groups of G226, N229 and N231 and the sidechain carbonyl groups of D233 and N229. These residues provided five interactions which, stabilise Ca^{2+} and a sixth interaction is provided by H_2O (Fig. 1.10B (ii)). Ca^{2+} is stabilised by the residues N398 (side chain carbonyl), S419 (main chain carbonyl) and E447 (2x side chain carbonyl) and two H_2O mediated interactions in Ca^{2+} binding site 2 (Fig. 1.10C (ii)). E452 is predicted to contribute in place of H_2O but is not seen in this model. In Ca^{2+} binding site 3, Ca^{2+} is stabilised by five interactions (Fig. 1.10D (iii)). This involves the main chain carbonyl of L312, side chain carbonyl groups of D306, N310 and E329 plus one H_2O mediated interaction. An interaction is predicted with N308 by sequence alignment with TG3 but this not observed in this model.

Table 1.3: Ca^{2+} binding site of TG2, 3 and 6.

TG	Ca^{2+} binding site					
	Site 1	**Position**	**Site 2**	**Position**	**Site 3**	**Position**
TG2	G226	O	N398	OD1	D306	OD2
	N229	O, OD1	S419	O	N308	OD1
	N231	O	E447	OE1, OE2	N310	OD1
	D233	OD2	E452	OE1, OE2	L312	O
					E329	OE1
TG3	A221	O	N393	OD1	D301	OD1
	N224	O, OD1	S415	O	D303	OD2
	N226	O	E443	OE1, OE2	N305	OD1, ND2
	D228	OD1, OD2	E448	OE1	S307	O
	N229	O			D324	OD2
TG6	A223	O	N396	OD1	D303	OD2
	N226	O	T417	O	D305	OD2
	N228	O	E445	OE1, OE2	N307	OD1, ND1
	D230	OD1, OD2	E450	OE1	S309	O
					D327	OD2

Residues involved in the stabilisation of Ca^{2+} in TG2 (predicted based on sequence alignment, TG3 (Ahvazi *et al.* 2002) and TG6 (Thomas *et al.* 2013).

1.5.2 TG3 Ca^{2+} binding.

The three Ca^{2+} binding sites in TG3 have been characterised using X-ray crystallography (PDB: 1L9M and 1L9N, (Ahvazi *et al.* 2002). Ca^{2+} is present in Ca^{2+}

binding site 1 of the inactive zymogen and binds with high affinity (K_D of 0.3 μM), indicating it is necessary for the stabilisation of TG3 but not sufficient for its activation (Ahvazi *et al.* 2002; Ahvazi *et al.* 2003). Ca^{2+} is stabilised by the main chain carbonyl oxygen of A221, N224, N226 and N229, the carbonyl side chain oxygen of D224 and an interaction with a H_2O molecule, in an octahedral configuration. Activation requires cleavage of at S469, but the protease involved is still unknown. However, the cleavage can be replicated *in vitro* using dispase (Ahvazi *et al.* 2002). When activated, Ca^{2+} is stabilised in a heptacoordinate manner in site 1. Ca^{2+} interacts with the main chain carbonyl oxygens of A221, N224 and N226 and the carbonyl side chains of N224 and D228 and H_2O (Fig. 1.10B (i)) (Ahvazi *et al.* 2002). In addition, activated TG3 binds two further Ca^{2+}. Ca^{2+} binding sites 2 and 3 have much lower affinity than Ca^{2+} binding site 1, measured with a K_D of *circa* 3-4 μM (Ahvazi *et al.* 2002; Ahvazi *et al.* 2003). Both Ca^{2+} are stabilised with heptacoordinate geometry (Fig. 10C (i) and D (i)). Ca^{2+} binding site 2 is composed of the main chain carbonyl oxygen of S415, the side chain carbonyl oxygens of N393, E443 and E448 and two H_2O molecules (Fig. 1.10C(i)). Ca^{2+} binding site 3 is made up of the main chain carbonyl oxygen of S307, side chain carbonyl oxygens of D301, D303 and D324, the side chain atoms of N305 and a H_2O molecule (Fig. 1.10D(i)), (Ahvazi *et al.* 2002).

When TG3 is nucleotide bound (GTPγS, GDP and GMP), the coordination of Ca^{2+} binding sites 1 and 2 remained the same as activated TG3 (3x Ca^{2+} PDB: 1L9N) (Ahvazi *et al.* 2002; Ahvazi *et al.* 2004a; Ahvazi *et al.* 2004b). Mg^{2+} binds in place of Ca^{2+} at binding site 3, whereby six shorter interactions are involved in ion stabilisation. These interactions involve the main chain carbonyl oxygen of S307, the side chain carbonyl oxygens of D301 and D303, the side chain atoms of N305 (monodentate) and an interaction with H_2O (Fig. 1.10D (ii)) (Ahvazi *et al.* 2003; Ahvazi *et al.* 2004a). In addition, comparison between inactive and active TG3 showed a difference in the opening of the hydrophobic tunnel of TG3 (Ahvazi *et al.* 2002; Ahvazi *et al.* 2003; Ahvazi *et al.* 2004a). In the active conformation, D324 directly interacts with Ca^{2+} (site 3), causing a shift in the loop (K321-S325), which in turn opens the tunnel. Ca^{2+} binding is essential for tunnel opening and loop cleavage alone is not sufficient. In nucleotide bound TG3 structures the channel is closed and Mg^{2+} does not interact with D324, indicating Mg^{2+} binding is also involved in the regulation of channel opening.

Interestingly, GDP bound TG3 showed both open (3x Ca^{2+}) and closed (2x $Ca^{2+}$and 1x Mg^{2+}) hydrophobic tunnels, further supporting the role of Mg^{2+} in channel regulation (Ahvazi *et al.* 2004a).

1.5.3 TG6 Ca^{2+} binding.

TG6 shows similar concentration dependence for Ca^{2+} as TG2 and TG3 (Ahvazi *et al.* 2002; Ahvazi *et al.* 2003; Thomas *et al.* 2013). There are three Ca^{2+} binding sites in TG6 which have been predicted from alignments with TG3. From the model the highest affinity-binding site is composed of the main chain carbonyl oxygen atoms from residues A223, N226 and N228, and the side chain oxygen of D230. Ca^{2+} binding at site 2 involves the main chain oxygen atom of T417, and the side chain oxygens of N396, E445 and E450 are positioned in a way that coordinates Ca^{2+} (Table 1.1). The third Ca^{2+} binding site is adjacent to H333 and is within a loop region, which is shifted upon activation. The moieties predicted to coordinate Ca^{2+} are the main chain carbonyl oxygen of S309 and the side chain oxygen atoms of D303, D305, D327 and the side chain atoms of N307.

1.6 Nucleotide binding of TG2, TG3 and TG6.

Of the TGs, TG2 (Bergamini *et al.* 1987) , TG3 (Ahvazi *et al.* 2004a), TG4 (Spina *et al.* 1999), TG5 (Candi *et al.* 2004) and TG6 (Thomas *et al.* 2013) bind GTP. In contrast, FXIIIA is not regulated by GTP, as it contains no nucleotide binding pocket (Achyuthan and Greenberg 1987). It is still unknown whether TG1 and TG7 are regulated by GTP.

1.6.1 TG2 nucleotide binding.

Nucleotide bound TG2 assumes a closed compact conformation (Liu *et al.* 2002). A domain exclusion approach was used to identify residues important for GTP binding and hydrolysis (Iismaa *et al.* 1997). Three constructs were used, FL TG2, β sandwich and catalytic core (βCore, res. 1-471) and the catalytic core alone (Core, res. 138-471). TG2 FL and βCore were capable of transamidase activity (βCore was 70% of FL), whereas TG2 Core was not. However, all three constructs showed [^{32}Pα]-GTP photolabeling indicating that the TG2 Core alone can bind GTP.

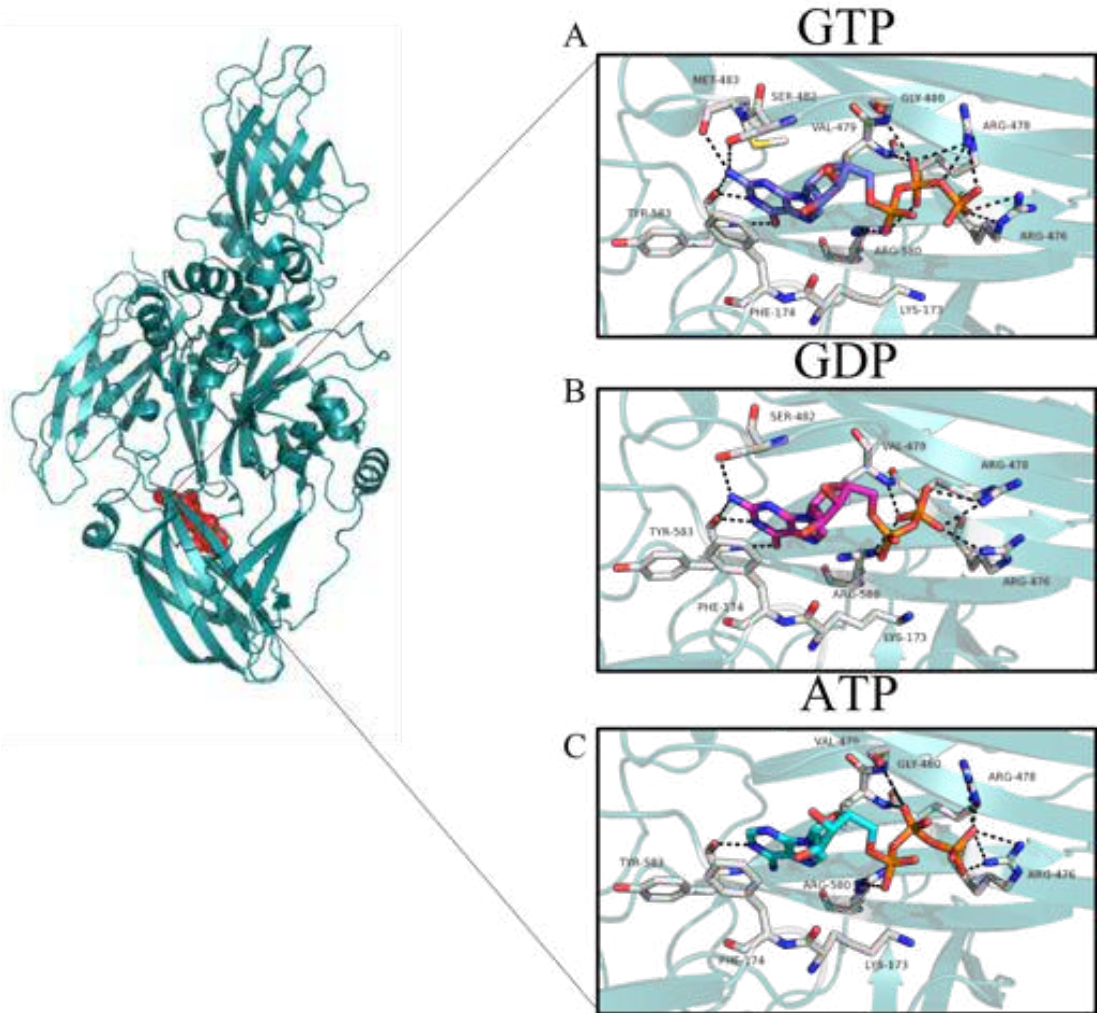


Figure 1.11: Nucleotide binding pocket of TG2.
TG2 has been crystallised in the presence of different nucleotides: GTP (lilac), GDP (magenta) and ATP (cyan) (stick representation). Interacting residues (white sticks) and polar contacts (<3.5 Å, black dotted lines) were highlighted. A: GTP bound TG2 (PDB: 4PYG, (Jang et al. 2014)). B: TG2•GDP (PDB: 1KV3, (Liu et al. 2002)). C: TG2•ATP (PDB: 3LY6, (Han et al. 2010)).

The TG2•GDP X-ray structure showed that GDP stabilises the first and final loops of β barrel 1 (Fig. 1.11B) (Liu *et al.* 2002). The loops linking the third and fourth β strands and the fifth and sixth β strands of β barrel 1 are in a position which occludes the substrate tunnel when TG2 is in the closed conformation (Liu *et al.* 2002). TG2•GTP and TG2•ATP have also been characterised (Fig. 1.11A and C) (Han *et al.* 2010; Jang *et al.* 2014).

The residues, which form the nucleotide pocket, are located within the first and last β strand of β barrel 1 and within the loops between the first and second β strands and

fifth and sixth β strands (R476, R478, V479, S482, M483, R580, L582 and Y583) (Fig. 1.11) (Liu *et al.* 2002). In the TG nucleotide-binding pocket, Y583 forms a hydrogen bond with N1 and N2 of guanine, S482 (via the side chain oxygen) also forms a hydrogen bond with N2 and the backbone nitrogen of Y583 interacts with O6 of guanine. The interactions with S482 and Y583 are observed in the TG2•GTP (Fig. 1.11.A) and TG2•GDP (Fig. 1.11B) but not in the TG2•ATP (Fig. 1.11C) structure which instead has a single hydrogen bond between Y583 and the adenine ring (Han *et al.* 2010; Jang *et al.* 2014). V479 and M483 make polar interactions with guanine in TG2•GTP but no interaction is observed with M483 in the TG2•GDP bound structure (Liu *et al.* 2002; Jang *et al.* 2014). V479 forms a hydrogen bond with the GDP β-phosphate via its backbone nitrogen.

Two residues (K173 and F174) which help compose the pocket are located in the catalytic core domain (Liu *et al.* 2002). F174 stabilises the guanine moiety via hydrophobic stacking interactions and K173 is predicted to interact with the γ phosphate of GTP (Fig 1.12A) but this interaction is not seen in the TG2•GTP structure (Fig. 1.11A) (Liu *et al.* 2002; Jang *et al.* 2014). The binding pocket also contains an abundance of positively charged residues, which can interact with the α and β phosphate. R580 forms two ion pairs with α and β phosphates in TG2•GTP and TG2•GDP, and the backbone amide of R478 also interacts with the GTP β phosphate. (Liu *et al.* 2002; Jang *et al.* 2014). In the TG2•GTP structure, R476 and R478 interact with the γ phosphate (Jang *et al.* 2014). They are conserved in TGs but in some cases lysine/arginine substitutions are made. In addition, as the γ phosphate is in close proximity to K173, R476 and R478, all are predicted to have a role in the GTP hydrolysis mechanism in combination with H_2O. R476 and R478 could stabilise the transition states in the hydrolysis mechanism by acting as an arginine finger (Bourne 1997).

In contrast, ATP is stabilised within the pocket by interactions between G480 and V479 and the β phosphate and R580 does not interact with the α or β phosphates (Fig 1.11C) (Han *et al.* 2010). The reduced number of interactions observed with ATP, compared with GTP and GDP could indicate reasons for lower binding affinity.

1.6.2 TG3 nucleotide binding.

GTPγS, GTP, GDP and GMP have been shown to bind to TG3 with 1:1 stoichiometry (Ahvazi *et al.* 2004a; Ahvazi *et al.* 2004b). Dissociation constants, K_D of GTPγS = 1.6 ± 0.2 μM, K_D of GTP = 1.8 ± 0.3 μM, K_D of GDP = 1.7 ± 0.4 μM and K_D of GMP = 1.6 ± 0.2 μM, were reported. However, this paper was subsequently retracted based on this experiment (Ahvazi *et al.* 2004a; Ahvazi *et al.* 2004b). The inhibition of TG3 activity by nucleotides (GTP, GTPγS, ATP and adenosine 5'-O-(3)-thiotriphosphate (ATPγS)) was also measured (Ahvazi *et al.* 2004a) and showed varying inhibition of TG activity. The inhibition by GTP is reduced at higher Ca^{2+} concentrations (0.60 mM). The potency series in which ATPγS > ATP > GTPγS > GTP was reported (Ahvazi *et al.* 2004a). Further nucleotide binding studies were carried out using boron-dipyrromethane (BODIPY) conjugated GTPγS, whereby BODIPY GTPγS was dissociated using excess nucleotide. Using this approach, K_D values for binding of GTPγS, GTP, GDP, GMP and ATP were reported as 0.4 μM, 0.6 μM, 1.0 μM, 0.4 μM and 28 μM respectively (Ahvazi *et al.* 2004b).

Nucleotide binding was also assessed using X-ray crystallography (Figure 1.12). Based on the various TG3 structures, the GTP binding site contains the residues N168 and R169 from the core domain and K485, K487, V488, L492, R587, I590 from β barrel 1(Ahvazi *et al.* 2004a; Ahvazi *et al.* 2004b). This pocket shows structural rearrangement when compared with TG3 in the absence of nucleotide (PDB: 1L9N, (Ahvazi *et al.* 2002)).

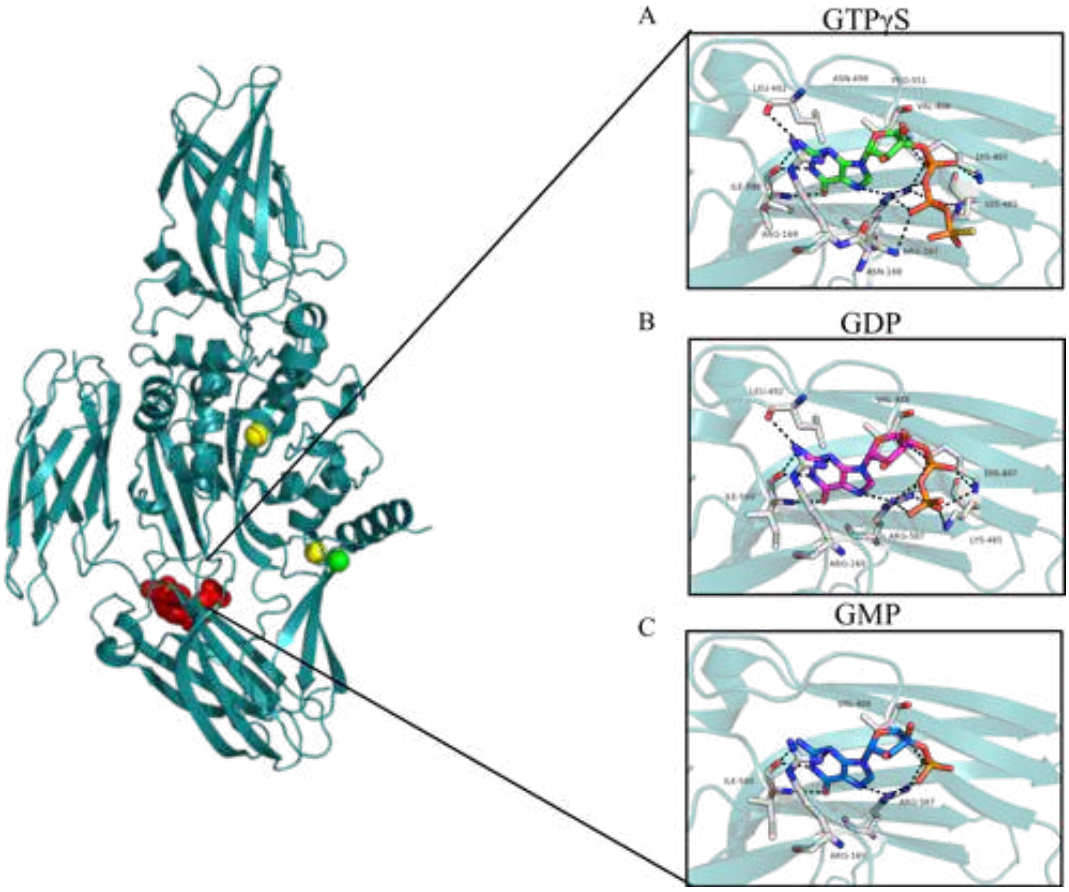


Figure 1.12: Nucleotide binding pocket of TG3.
TG3 has been crystallised in the presence of different nucleotides: GTPγS (green), GDP (magenta) and GMP (blue) (stick representation). Interacting residues (white sticks) and polar contacts <3.5Å (black dotted lines) were highlighted. A: GTPγS bound TG3 (PDB: 1RLE, (Ahvazi *et al.* 2004a)). B: GDP bound TG3 (PDB: 1VJJ, (Ahvazi *et al.* 2004a)). C: GMP bound TG3 (PDB: 1SGX, (Ahvazi *et al.* 2004b)).

In the TG3•GTPγS and TG3•GDP bound structures, the GTP guanine ring is stabilised by I590 through its backbone carbonyl moiety via N1, D588 via its backbone amide to O6, a water mediated interaction to the amide of I590, R587 via N7 and R169 that stacks over the guanine ring (Ahvazi *et al.* 2004a) (Fig 1.12A and B). The opposite side of the guanine ring is stabilised by interactions with V488, L492, I589 and L591. The α and β phosphates interact with K485, K487 and R587. In GTPγS bound TG3 structure, N168, K485 and K587 are in close proximity to the γ phosphate (Ahvazi *et al.* 2004a). K485 and K487 are predicted to act as a lysine finger, which stabilises the transition state of GTP hydrolysis. The lysine finger may have a role in orientating the γ phosphate and stabilising the negative charges. In the hydrolysis mechanism, the side

chain of N168 would be required to act as a general base which interacts with one of the three nearby H_2O molecules, generating the nucleophile required for hydrolysis.

GMP makes different interactions within the TG3 nucleotide binding pocket (Fig. 1.12C) (Ahvazi *et al.* 2004b). The guanine ring is stabilised by I590 (backbone amide and carbonyl interact with O6 and N1and N2, respectively), V488 (via its backbone amide), and L492, I589 and R587 interact with N7. As observed in the GTPγS and GDP-bound structures, R169 stacks over the guanine ring. R587 also interacts with the α phosphate of the GMP and stabilises a water mediated interaction with O6 on the guanine ring.

1.6.3 TG6 nucleotide binding.

Like TG3 and TG2, TG6 activity can also be regulated by GTP. The K_D of TG6 (20 μg/mL) for GTP is 200 μM which is 50x higher than TG2 but similar concentrations are required for inhibition by ATP (1 mM) (Thomas *et al.* 2013). Inhibition of TG6 activity was most potent with GTPγS and the least potent inhibitor was GMP. Inhibitions by the various nucleotides was similar when compared with TG2 but not TG3 (Iismaa *et al.* 2000; Ahvazi *et al.* 2004a).

IMAGE REMOVED DUE TO COPYRIGHT.

Figure 1.13: Nucleotide binding pocket of TG6.
TG6•GDP model interacting residues (grey stick representation, and surface representation blue- white-red) have been predicted based on modelling with TG2•GDP. Image from Thomas *et al* 2013.

The TG6 nucleotide binding site was investigated using the structure of TG2•GDP (Fig. 1.11B, PDB: 1KV3 (Liu *et al.* 2002)). The residues that form this pocket are K170 and H171 from the catalytic core domain, K498 to P505, and K600 to E605 from β barrel 1. The binding site consists of four lysine residues (K170, K498, K500 and K600, Fig 1.13), which are responsible for the stabilisation of the phosphate groups. K498, K500 and K600 are replaced with arginines in TG2 (Fig. 1.11A). V501, P504, P505 and T603 interacts with the guanine moiety via hydrogen bonding with N1, N2 and O6 and the imidazole ring of H171 stacks against the guanine ring. A increased number of hydrogen bonds are seen within the TG2 nucleotide-binding pocket compared with TG6. This difference could explain the lower affinity for guanine nucleotides measured for TG6 (Thomas *et al.* 2013).

1.7 GTPase activity of TG2 and TG3.

TGases with GTP binding properties are predicted to also have GTPase activity. GTPase activity has been identified for TG2 and TG3 but GTPase activity is yet to be confirmed for TG6 (Bergamini *et al.* 1987; Ahvazi *et al.* 2004a; Thomas *et al.* 2013) TG2 acts as a G-protein and GTP hydrolysis is important for receptor signalling via the α1/ β1 adrenergic receptor (Iismaa *et al.* 2000). No β subunit has yet been identified therefore suggesting it acts alone. In contrast, no physiological role for GTPase activity has been established for TG3.

1.7.1 TG2 GTPase activity.

A GTP hydrolysis rate of 0.01 pmol of GTP hydrolysed/pmol protein/minute was measured using either WT TG2 or purified guinea pig liver WT TG2 (Iismaa *et al.* 2000). GTPase and ATPase activity was also measured for each domain of WT TG2 (TG2 FL, βCore and Core, Table 1.4) (Iismaa *et al.* 1997). The inhibition of GTPase an ATPase activity was measured using ATPγS and GTPγS (Table 1.4). Of note both the βcore and the Core have ATPase activity higher than FL TG2, and GTPase activity greater than or equal to FL TG2. This indicates that the β barrel domains are not required for ATP and GTP hydrolysis.

Table 1.4: Hydrolysis of Nucleotides by TG2 domains.

TG2	ATPase Activity (%)	Inhibition with ATPγS (μM)	Inhibition with GTPγS (μM)	GTPase Activity (%)	Inhibition with ATPγS (μM)	Inhibition with GTPγS (μM)
Full length	100	25	50	100	10	8
β core	140	10	26	100	1	4
Core	175	3	9	150	7	11

In addition, GTP and ATP binding and hydrolysis maybe facilitated by Mg^{2+} (Lai et al 1998). Mg-GTP binding was measured with an apparent K_M of 130 μM with a K_{cat} of 0.06 $min^{-1}$and for Mg-ATP a K_M of 38 μM and a K_{cat} of 0.08 min^{-1}. Transamidase activity was inhibited by Mg-GTP, Mg-GTPγS and Mg-GDP, each with a measured IC_{50} of 9 μM and also inhibited by Mg-GMP with an IC_{50} of 500 μM.

1.7.2 TG3 GTPase activity.

TG3 GTPase activity was measured in the presence of 1 mM Mg^{2+} and 0.575 mM Ca^{2+} (Ahvazi *et al.* 2004a). ATP hydrolysis by TG3 was minimal when compared to GTPase activity (10x higher hydrolysis activity) which differs from TG2 and is Mg^{2+} dependent.

1.8 Externalisation of TG2 in P2X7R activation and GTP binding site mutants.

The externalisation of TG2 has been implicated in a number of autoimmune diseases, such as celiac disease (Aeschlimann and Knauper 2017). Our lab has extensively investigated the mechanism of TG2 export in order to understand its role in the inflammatory response (Adamczyk *et al.* 2015). The activation of P2X purinoreceptor 7 (P2X7R) is involved and forms a large membrane pore upon activation with ATP (Adamczyk *et al.* 2015). A number of studies have implicated P2X7R in multiple immune pathways and autoimmune diseases (Aeschlimann and Knauper 2017). Furthermore, the importance of GTP in the mechanism of TG2 export has been shown in chondrocytes (Johnson and Terkeltaub 2005). Upregulated TG2 expression can cause a dysfunction of matrix repair in osteoarthritis patients, for which the export of TG2 is essential. TG2 K173L, a GTP binding site mutant and Y274A (externalisation deficient mutant) were retained inside the cell, whilst WT TG2 and the TG2 active site mutant, C227A, were exported. These observations indicate that GTP may act as a molecular switch for TG2 export but this is not yet fully understood. To investigate the role of GTP in the export of TG2, the GTP binding pocket residues mutants can be targeted. Extensive work has previously been carried out on a selection of TG2 GTP binding site mutants (Iismaa *et al.* 2000; Begg *et al.* 2006a; Begg *et al.* 2006b; Ruan *et al.* 2008; Singh *et al.* 2016). The selected residues were mutated to the equivalent residue in FXIIIA or a residue where the hydrophobicity or polarity was maintained (Table 1.5, (Iismaa *et al.* 2000)).

GTPγS binding assays, GTP photolabelling, ITC, GTP agarose electrophoresis and the inhibition of TGase activity were used to measure the GTP binding of the TG2 GTP binding mutants (Table 1.5, (Iismaa *et al.* 2000; Begg *et al.* 2006a; Begg *et al.* 2006b). ITC measured affinities (K_D) for GTP of 1.6 μM for human WT TG2, 2.9 μM for F174A and 150 μM for R579A (Iismaa *et al.* 2000; Begg *et al.* 2006b). The inhibition

This assay was also carried out at 40% maximal Ca^{2+} stimulated TGase activity for the TG2 mutants (Begg *et al.* 2006b). WT TG2 had an IC_{50} of 27 μM. A significant reduction in IC_{50} was measured for R476A (>350 μM), R478A (>450 μM) and R476A/R478A (>500 μM).

Furthermore, GTP binding was measured with GTP agarose electrophoresis. Interestingly, R579A and R579K had a slower migrating band with and without GTP, compared to GTP bound TG2 (Begg *et al.* 2006a). This observation is consistent with the TG2 R579A and R579K having a slightly expanded closed conformation, which forms independently of GTP binding. In addition, GTP binding was not detectable to TG2 R580A and R580K expressed in HEK293 cells as measured using the incorporation of [α-^{32}P] GTP into TG2 and GTP agarose (Ruan *et al.* 2008; Singh *et al.* 2016).

The GTPase activity of the GTP binding site mutants was also assessed (Iismaa *et al.* 2000). GTPase activity of S171C was equivalent to WT TG2 whilst K173N had impaired activity, K173L had activity that was barely detectable and activity was undetectable for S171E. To address the physiological significance of TG2 GTPase activity the signalling capabilities of the GTP binding mutants (S171C, S171E, K173N and K173L) were assessed upon treatment of the α1/B1-adrenergic receptor with epinephrine (Iismaa *et al.* 2000). IP_3 accumulation for TG2 K173N and S171C was comparable to WT TG2, whereas TG2 S171C caused enhanced accumulation of IP_3 consistent with increased receptor signalling. In contrast, TG2 K173L and S171E showed impaired receptor signalling consistent with no IP_3 accumulation, which indicates a lack of GTP binding and hydrolysis.

F174A (rat)	Abolished[3]	Begg *et al.* 2006a, Begg *et al.* 2006b
F174W (rat)	Equivalent[2]	Begg *et al.* 2006a, Begg *et al.* 2006b
R476A (rat)	Impaired[1]	Begg *et al.* 2006a, Begg *et al.* 2006b
R476A/R478A (rat)	Impaired[1]	Begg *et al.* 2006a, Begg *et al.* 2006b
R476A/R579A (rat)	Impaired[1]	Begg *et al.* 2006a, Begg *et al.* 2006b
R478A/R579A (rat)	Impaired[1]	Begg *et al.* 2006a, Begg *et al.* 2006b
R478A (rat)	Impaired[1]	Begg *et al.* 2006a, Begg *et al.* 2006b
R579A(rat)	Abolished[3]	Begg *et al.* 2006a, Begg *et al.* 2006b
R579K(rat)	Abolished[3]	Begg *et al.* 2006a, Begg *et al.* 2006b
R580A (human)	Abolished[3]	Ruan *et al.* 2008, Singh *et al.* 2016
R580K (human)	Abolished[3]	Ruan *et al.* 2008, Singh *et al.* 2016

List of GTP binding site mutants generated. [1]Impaired, reduced GTP binding compared to WT. [2] Equivalent, activity equal to WT. [3]Abolished, No GTP binding.

1.9 TGs in immune related disease.

There are a number of immune-related disorders facilitated by TGases, these are most commonly mediated by anti- TG antibodies (autoantibodies) (Table 1.6). Celiac disease and gluten ataxia are specifically discussed in Section 1.8.1 and 1.8.2 respectively. Additional, diseases related to TG6 include MS and cerebral palsy are discussed further below (Section 1.8.3-4).

1.9.1 Celiac Disease

Celiac disease is an autoimmune disease and is commonly associated with gastrointestinal problems, such as chronic inflammation and diarrhoea. The most common intestinal consequence is atrophy of the villi of the small bowel leading to the formation of lesions and malabsorption. In addition to these lesions in the intestine, a common indicator is circulating antibodies against deamidated gliadin peptides (AGA) or IgA-class TG2 autoantibodies (Lindfors *et al.* 2009).

Much is still unknown about Celiac disease development, but a principal mechanism has been established. Firstly, gluten is cleaved by digestive enzymes in the bowel to generate a range of peptides. These peptides are internalised into the *lamina propria* where there is a high concentration of TG2. In susceptible individuals, many gluten peptides are toxic and cause an immune response (Beissbarth *et al.* 2005). The most abundant peptides are derived from gliadins (α and γ), which are good substrates for TG2 (Sjostrom *et al.* 1998; Arentz-Hansen *et al.* 2000). Toxic gluten peptides initiate a primary immune response, which propagates via interleukin-15 (Maiuri *et al.* 2000).

Table 1.6: Expression of transglutaminases and associated diseases

Gene	Chromosome	Protein	Expression	Immune disease	Inherited disease	Biological role
TGM1	14q12	TG1	Oesophagus, skin	Celiac disease	Congenital ichthyosis	Cornified cell envelope an barrier formation
TGM2	20q11	TG2	Ubiquitous	Celiac disease,	Early onset diabetes,	Apoptosis, G-protein signalling, matrix stabilisation
TGM3	20p13	TG3	Oesophagus, skin and brain	Dermatitis Herpetiformis	Uncommable hair syndrome	Hair follicle morphogenesis and barrier formation
TGM4	3p21	TG4	Prostate	Autoimmune polyglandularsy type 1, Decreased male fertility	-	Semen coagulation

TGM5	15q15	TG5	Skin, oesophagus	Acral peeling skin syndrome Sporadic schizophrenia	-	Epidermis formation,
TGM6	20p13	TG6	Brain, skin	Gluten ataxia, Gluten neuropathy,	Spinocerebellar ataxia 35, Acute myeloid leukaemia Parkinsons disease	Neuronal differentiation
TGM7	15q15	TG7	Ubiquitous	-	-	-
F13A1	6p25	FXIIIA	Placenta, platelets	Factor XIII subunit A deficiency	Autoimmune haemophilia like disease.	Blood coagulation, wound healing

Expression of TG's in tissues and transglutaminase-associated diseases. Table adapted from Karpati *et al.* 20

In addition, TG2 binds and deamidates the gluten peptide. Hydrolysis is slow and preferentially occurs at pH <7.3, an environment replicated in the intestine (Sollid 2002). The deamidation increases the affinity of the gliadin, by introducing negative charges in the peptide, for presentation by the MHC/HLA complex (Sjostrom *et al.* 1998). Two mechanisms have been suggested for the generation of TG2 autoantibodies. First is the hapten carrier model, where by gluten reactive CD^+ T-cells aid in the recognition of gluten peptide: TG2-gluten complexes (formed via thioester and isopeptide bonds) by TG2-specific B-cells. B-cells bind and internalise these complexes and bound gluten peptide can be released by proteolytic cleavage or thioester/isopeptide bond cleavage. The gluten peptide can then be presented on the cell surface by the HLA complex for recognition by gluten-reactive T-cells (Sollid 2002; Fleckenstein *et al.* 2004; Stamnaes *et al.* 2010). In addition, binding studies with the gliadin mimetic P-DON showed TG2's active site is exposed when bound to gliadin, further supporting the hapten carrier model (Pinkas *et al.* 2007). The second model is molecular mimicry as it has been proposed for the trigger of other autoimmune disease (Lindfors *et al.* 2009). However, there is no evidence to support this model (Sollid and Jabri 2013). Following presentation of the gluten peptide on the cell surface, the peptide is recognised by the $CD4^+$ T-cells, leading to T-cell activation in the *lamina propria* (Lindfors *et al.* 2009). This causes damage to the epithelial cell lining of the intestine wall allowing infiltration of additional gluten peptides to the *lamina propria*. The T-cell activation is followed by the release of pro-inflammatory cytokines, which promotes remodelling of the intestinal tissue (Lindfors *et al.* 2009).

Antibodies are generated against TG2 in the presence of gluten and are cross-reactive with numerous self-proteins e.g. heat shock protein 60 (Hsp60), desmoglein1 and toll-like receptor 4 (TLR4). TG2 antibodies can also cross-react with viral antigens and gliadin-like peptides. This reactivity can increase epithelial cell permeability and cell proliferation in the small intestine, leading to flattening of the villi (Iismaa *et al.* 2009). TG2 autoantibodies are sequestered at the site they are produced and can form deposits in the small intestine and other tissues, e.g. liver, muscle and lymph nodes (Boscolo *et al.* 2010). TG2 autoantibodies are predicted to recognise exposed surfaces of TG2 in the open conformation, when binding gliadin (Pinkas *et al.* 2007).

1.9.2 Gluten ataxia

Using magnetic resonance imaging (MRI), GA has no unique features in the cerebellum to distinguish it from other ataxias, the mean age of onset is 53 and it is not gender specific (Hadjivassiliou *et al.* 2008b). Common manifestations include lower limb ataxia, gaze evoked nystagmus and other ocular deficits. Additional rare movement symptoms include myoclonus, chorea, palatal tremor and opsoclonus myoclonus. Furthermore, it is common for GA patients to have additional autoimmune diseases for example hypothyroidism, type 1 diabetes and pernicious anaemia. Post-mortem analysis showed a consistent loss of Purkinje cells throughout the cerebral cortex (Hadjivassiliou *et al.* 2008b). This observation is supported by evidence suggesting cross-reactivity of antigenic epitopes between Purkinje cell antigens and gluten peptides seen with celiac disease sera using human and rat cerebellum.

A high frequency of patients with neurological symptoms of unknown cause (57%) were found to have anti-gliadin antibodies (AGA) positive sera (Hadjivassiliou *et al.* 1996). However, 12% of healthy controls and 5% with neurological disorders of a known cause also had AGA present in their blood. The original definition of gluten ataxia relies predominantly on the detection of circulating AGA (Hadjivassiliou *et al.* 2008b). However, due to the high prevalence of AGA in the healthy population, AGA positive serology cannot be solely used to identify GA patients (Hadjivassiliou *et al.* 2008b).

Gluten ataxia without enteropathy (Gao) patients shown accumulation of IgA deposits on the cerebellum and brainstem which show co-localisation with TG6 (Hadjivassiliou *et al.* 2008a). There was significant TG6 positivity in the cerebellum, medulla and the perivascular regions, which was not observed in normal cerebellum. The prevalence of TG6 autoantibodies in patients with idiopathic sporadic ataxia and positive for AGA was 73% (Hadjivassiliou *et al.* 2013). 32% of sporadic idiopathic ataxia patients with negative serology for gluten sensitivity were positive for TG6 autoantibodies and only 7% of these had evidence of enteropathy (Hadjivassiliou *et al.* 2013). When comparing the specificity of antibodies for GA, 12% of healthy controls were positive for AGA versus 4% positivity for TG6 antibodies. The circulating TG6 antibodies were only present upon gluten exposure and as a result GA patients symptoms stabilised in

response to a gluten-free diet. The prevalence of TG6 antibodies in a Finnish cohort was only 14% compared with UK cohorts (32% and 73%). This indicates TG6 antibodies can facilitate the diagnosis of GA for some but not all cases.

The hapten carrier model is one of the predicted mechanisms (discussed in Section 1.8.1), which drives the immune response in celiac disease. The ability of TG6 and TG3 to form complexes with gliadin peptides was assessed using biotinylated peptides (Stamnaes *et al.* 2010). TG2 readily forms complexes (via thioester bonds), whereas TG3 is dependent on crosslinking by TG2. Complex formation for TG6 is slower and involves isopeptide and thioester linkage (Stamnaes *et al.* 2010). This could be important in the pathogenesis of GA.

TGs bind gluten peptides with high affinity, which has a role in the pathogenesis of gluten-related disorders in susceptible individuals (Hadjivassiliou *et al.* 2008a; Stamnaes *et al.* 2010). The glutamine-containing substrates overlap between TG2, TG3 and TG6 but they have different levels of specificity for distinct substrate molecules (Stamnaes *et al.* 2010). For example, TG6 can tolerate changes to neighbouring residues in its optimum sequence and therefore has broader specificity. TG6 will readily bind glutamine donor substrates and glutamine acceptors such as lysine but to a lesser degree. The position of proline in the TG substrate peptide in relation to the targeted glutamine is important for recognition by the respective TG. This is due to proline residues restricting the peptide conformation (Stamnaes *et al.* 2010; Fukui *et al.* 2013). When proline is positioned at +1 (e.g. QP) from the reactive glutamine, proline protects glutamine from modification. Prolines situated in positions -1 and -2 (e.g., PQ or PXQ) from the glutamine are also detrimental to substrate recognition. However a proline positioned at +2 (e.g. QXP) has a positive influence on substrate recognition and promotes peptide deamidation (Stamnaes *et al.* 2010). In addition, *N*-terminal extension DQ2-ɤ gliadin has no effect on substrate recognition by TG6, however a *C*-terminal extension of the DQ2-ɤgliadin has a positive influence on substrate recognition. These results suggest that TG6 selectivity is dependent on the primary and secondary structure and the environment of the glutamines in the substrate peptide (Fukui *et al.* 2013).

Screening of TG6 positive peptides showed position +3 from the glutamine was predominantly a hydrophobic residue. Positions -1 and +2 were commonly a serine, proline or glutamate. One peptide analysed (pepY25: DDWDAMDDEQIWF) was weakly cross-reactive with TG3 but bound strongly to TG6. When the reactive glutamine was mutated to asparagine, no reactivity with TG6 was observed. This indicates the importance of the glutamine position for the recognition by TG6 (Fukui *et al.* 2013).

1.9.3 Multiple sclerosis

TG6 has also been associated with MS (Cristofanilli *et al.* 2017). There was a significant increase in the levels of anti TG6 IgG found in the cerebral spinal fluid (CSF) of MS patients when compared to healthy controls, as measured by enzyme-linked immunosorbent assay (ELISA). This change was not observed in serum samples and therefore indicates a CNS-specific response. Primary progressive MS patients showed the highest increase in anti TG6 IgG levels and levels were dependent on the patient being in the active stage of the disease. TG6 reactivity was located to cells with an astrocyte like morphology that expressed GFAP (a reactive astrocyte marker). TG6 expression was also followed through the MS disease process using experimental autoimmune encephalomyelitis (EAE) induced mice. ELISA of spinal cord samples showed a peak in TG6 expression in the brain and spinal cord at day 14. TG6 expression present then plateaued at lower levels by day 21, suggesting TG6 expression is linked to disease progression. Both TG6 and GFAP were upregulated following disease onset suggesting TG6 has a role in astrocyte proliferation in progressive MS

1.9.4 Other TG related immune diseases.

TG6 antibodies have also been detected in patients with alcohol related cerebellar degeneration (Shanmugarajah *et al.* 2016). In the presence and absence of liver disease chronic alcohol consumption provokes an IgA immune response against TG6. This response causes sensitivity to gluten to increase in genetically susceptible individuals. 34% of patients with alcohol related cerebellar degeneration were positive for AGA compared with 12% of the healthy population and 39% of patients had TG6 antibodies compared with 4% of healthy controls.

SCA are a group of hereditary neurodegenerative diseases (Jayadev and Bird 2013). Over thirty different types of SCA have been identified that involve mutations in approximately 24 causative genes. The age of onset can vary depending on the SCA type, and although treatment is available to deal with symptoms there is currently no cure. SCAs are frequently associated with the following symptoms: impaired movement, difficulty speaking, pigmentary retinopathy, dementia and peripheral neuropathy (Wang *et al.* 2010). Atrophy is seen in the brain stem, the cerebellum, spinal tract and peripheral nerves (Liu *et al.* 2013). There are many molecular mechanisms proposed for SCA, e.g. the toxic accumulation of aggregated protein, altered calcium homeostasis, Tau phosphorylation and cytoskeletal abnormalities, but the mechanisms of neuronal cell death are largely unknown.

Mutations causing cerebellar ataxias can be inherited via autosomal dominant, autosomal recessive, X-linked and mitochondrial modes (Jayadev and Bird 2013). However, SCA is predominantly inherited by an autosomal dominant mechanism. The most common mutation is CAG trinucleotide expansions, which can be located anywhere in the causative genes. SCA1, SCA2, SCA3, SCA6, SCA7, SCA12 and SCA17 contain CAG repeats in the coding region of their causative genes. Missense, deletion and insertion mutations are less common, but SCA35 is caused by these type of mutations in *TGM6* (Wang *et al.* 2010).

Mutations in *TGM6* giving rise to SCA35 were identified using various different methods. Mutations D327G, D510H and L517W were identified in a Chinese ataxia cohort via whole exome sequencing (Wang *et al.* 2010; Li *et al.* 2013). Three mutations were also identified in *TGM6* in a Taiwanese (Han-Chinese) cohort

in patients presenting with symptoms of SCA35. These are TG6 N157S (Rao S.H, Columbia University) and TG6 R412C (Kerr M and Khan A, University of Calgary). All identified *TGM6* mutations have been summarise in Table 1.6.

1.10.1 Identification of TG6 mutants in SCA35.

The first identified and the most characterised TG6 mutants in the literature are D327G and L517W. D327G (c.980A>G) is coded for in exon 7 which codes for the central part of the catalytic core domain. It has been identified in affected individuals across two generations of a family with SCA (Wang *et al.* 2010). In the TG6 model (derived from TG3, PDB: 1L9N), D327G is part of Ca^{2+} binding site 3 and thus might interfere with Ca^{2+} binding required for the activation of TG6 (Thomas *et al.* 2013). L517W (c.1550 T>G) was not identified in unaffected members of the SCA family, in addition to 500 healthy controls, but was present in 'clinically unknown' individuals (Wang *et al.* 2010). L517W is located in exon 10 that encodes a β sheet located in the β-barrel 1 domain (Guan *et al.* 2013). MSA of TG6 from various species indicate that the *C* terminal β barrel domains are highly conserved and mutations in this domain are likely to be damaging. However, the molecular consequence of this mutation is uncertain and structural modelling has not predicted any changes (Thomas *et al.* 2013). TG6 L517W was also identified in acute myeloid leukaemia (AML) patients. The significance of this remains to be clarified (Pan et al. 2015).

The mutation D510H (c.1528G>C) was found in two Chinese families and a Han-Chinese population, which had a history of SCA (Li *et al.* 2013; Guo *et al.* 2014). In the Chinese population, onset was from childhood to early 40s and symptoms included cognitive impairment, unsteady gait, tremor, dysarthria and numbness in extremities

(Guo *et al.* 2014). The E574del mutation caused similar symptoms but also included nystagmus and pursuit aberrations.

Table 1.6: Mutation identified in TGM6.

Protein	Codon	Gene position	Mutation	Position	Age of onset	General symptoms	Un sy
-	-	c.7+1G>T	splice donor	N/a	54/59	Dysarthria, gait, tremor, cerebellar ataxia	No
R111C	CGC/TGC	c.331C>T	missense	β sandwich	20-26	Dysarthria, gait, tremor, saccade slowing, myoclonus, cerebellar ataxia	Epi
N137S	AAC/AGC	c.472A>G	missense	β sandwich	30	Cerebellar ataxia	No
Q181H	CAG/CAT	c.543G>T	missense	catalytic core	19	Myoclonus, cerebellar ataxia	No
G279ter	GGA/TGA	c.841G>T	nonsense	catalytic core	40	Dysarthria, gait, cerebellar ataxia	No
D327G	GAC/GGC	c.980A>G	missense	catalytic core	40/41	Dysarthria, saccade slowing, myoclonus, cerebellar ataxia	Oc dys Pse pal
P359L	CCC/CTC	c.1137C>T	missense	catalytic core	17-35	Unpublished (Early onset Parkinsons disease)	N/a

V391M	GTG/ATG	c.1171G>A	missense	catalytic core	36	Dysarthria, gait, tremor, myoclonus, cerebellar ataxia	No
R412C	CGT/GGT	c.1234C>G	missense	catalytic core	deceased age 9	Cerebellar ataxia	Epi sho gro fail
T426N	ACC/AAC	c.1277C>A	missense	catalytic core	Unknown	Cerebellar ataxia	No
Y441C	TAC/TGC	c.1322A>G	missense	catalytic core	54	Dysarthria, cerebellar ataxia	Sw dif spa
R448W	CGG/TGG	c.1342C>T	missense	catalytic core	20	Gait, cerebellar ataxia	No
P493L	CCC/CAC	c.1478C>T	missense	catalytic core	60	Dysarthria, gait, tremor, cerebellar ataxia	Me imp
L502Q	CTA/CAA	c.1505T>A	missense	catalytic core	>50	Cerebellar ataxia	No
L502Q Q652dup	CTA/CAA and CAG/ CAACAG	c.1505T>A and c.1951_1952ins AAC	missense/ insertion	catalytic core	40s	Gait, myoclonus, cerebellar ataxia	No

G508D	GGC/GAC	c.1476G>A	missense	β barrel 1	Unknown	Cerebellar ataxia	No
D510H	GAC/CAC	c.1528G>C)	missense	β barrel 1	12-40s	Dysarthria, gait, tremor, cerebellar ataxia	Nu ext
L517W	TTG/TGG	c.1550 T>G	missense	β barrel 1	40-48	Dysarthria, gait, tremor, cerebellar ataxia	Oc dys Pse pal
V527E	GTG/GAG	c.1476G>A	missense	β barrel 1	Unknown	Cerebellar ataxia	No
E574del	GAAGAC/ GAC	c.1722_1724del	deletion	β barrel 1	56	Dysarthria, gait, tremor saccade slowing, cerebellar ataxia	Ny
Q652dup	CAG/ CAACAG	c.1951_1952ins AAC	insertion	β barrel 2	Unknown	Gait, myoclonus, cerebellar ataxia	No

A large cohort of Caucasians with a family history of an autosomal dominant inheritance and with undefined genetic cause were screened for mutations in many genes including *TGM6* (Tripathy *et al.* 2017). In four centres (Cardiff, Sheffield, Paris and Boston) including our laboratory, five new missense mutations were identified, Q181H, V391M, E406K, Y441C, R448W and L502Q. In addition, a single amino acid duplication was identified (Q652dup). Our patient carried this in isolation and another in combination with the L502Q missense mutation.

Furthermore, a Hispanic patient, presenting with progressive instability and slurred speech and who also had a family history of cerebellar ataxia, was diagnosed with SCA35 (Lin *et al.* 2018). A brain MRI showed significant cerebellar atrophy. Whole exome sequencing identified a deletion of one base pair in *TGM6* (c.841 in exon 6). This results in a frame shift, which causes premature termination of translation, resulting in a protein that is 279 amino acids long. The truncation is in the catalytic core domain, so enzymatic activity is likely to be compromised.

1.10.2 *TGM6* mutants in sporadic and undiagnosed cerebellar ataxia.

Sporadic cerebellar ataxia patients that had pure cerebellar ataxia or a spinocerebellar ataxia phenotype were screened for mutations in *TGM6* using genome sequencing (Sanger method) (Fogel *et al.* 2012). Patients were referred to an ataxia centre in California based on primary adult onset and were negative for the most common genetic ataxias (e.g. SCA2, SCA3, SCA6, SCA7 and Friedreich ataxia). In a cohort of 118 patients presenting with the phenotype of pure cerebellar, spastic ataxia and/or spinocerebellar ataxia, a single patient harboured TG6 G508D and another TG6 V527E, whilst three were found to have TG6 R448W. Furthermore, TG6 G508D was subsequently also reported in a Caucasian sibling pair diagnosed with SCA35 following undisclosed genetic testing (Ruiz-Lopez *et al.* 2017). Symptoms included slurred speech and mild impairment to balance, which deteriorated after 5 years of disease onset.

The splice site mutation c.7 +1G>T and the missense mutation P493L, were identified in undiagnosed cerebellar ataxia patients (Yang *et al.* 2018). A family history of

cerebellar ataxia was apparent in the individual harbouring the splice site mutation (c.7 +1G>T) and they presented with progressive gait unsteadiness, dysarthria and hand tremor. These symptoms were also seen with other relatives and were not present in clinically unaffected family members and 200 ethically matched healthy controls. The missense mutation, P493L, was identified in an individual presenting with progressive gait unsteadiness, dysarthria, hand tremor, memory impairment and also showed mild cerebellar atrophy with MRI (Yang *et al.* 2018). Unaffected family members did not harbour the mutation and it was not identified in 200 healthy controls. This mutation was also identified in one patient presenting with episodic ataxia (Choi *et al.* 2017). An individual harbouring TG6 T426N, was discovered in the whole exome screening of patients characterised with sporadic ataxia (Nibbeling *et al.* 2017). The mutation is located in a highly conserved position in the β barrel 1 domain.

1.11 TG6 mutants in other neurological diseases.

TG6 P359L has been identified in early onset PD. The study investigated the probable autosomal dominant trait in a multigenerational Serbian family (Westenberger *et al.* 2016, unpublished conference abstract). This highlighted 17 possible candidate causative genes expressed in the brain. A missense mutation, P359L, in *TGM6* was identified as the most likely candidate by exome sequencing and by introducing filtering steps which excluded variants in non-conserved regions and non-rare variants. P359L had a high combined annotation dependent depletion (CADD) score (34 out of 39) which is based on the deleteriousness of missense mutations.

1.12. Biochemical analysis of TG6 mutants

Various biochemical and cellular experiments have been carried out in order to characterise the identified TG6 mutants and establish the role they have in neuronal cells death for which two possible mechanisms have been highlighted. Firstly, the TG6 mutants initiate the unfolded protein response (UPR) and subsequent degradation by the proteasome (Tripathy *et al.* 2017). Secondly, the TG6 mutants sensitise the neuronal cells to apoptosis via an unknown mechanism (Guan *et al.* 2013). Further studies have been carried out to investigate these mechanisms.

1.12.1 Investigating subcellular localisation, stability and activity of the TG6 mutants.

In order to establish the mechanism of TG6 mutations in neurological disease, the TG6 mutants were investigated in a cellular context to represent the physiological situation. To establish if the TG6 mutants (D327G, L517W and T426N) impact the localisation of TG6, immunocytochemistry was utilised post transfection into HEK293 cells (Guan *et al.* 2013; Nibbeling *et al.* 2017). TG6 D327G, T426N, L517W were localised to the cytoplasm whereas WT was detected in both the nucleus and the cytoplasm, which suggests WT can be shuttled within the cell. WT TG6 was also detected in the cytoplasm of NIH 3T3 (mouse embryo fibroblast cell line) and Neuro2a cells (Guan *et al.* 2013). The stability of the TG6 mutants (R111C, D327G, T426N, D510H, L517W and E574del) was assessed following cycloheximide (CHX) treatment in HEK293 cells and subsequent Western blotting (Guo *et al.* 2014; Nibbeling *et al.* 2017). All forms of TG6 were expressed to levels comparable to WT TG6 but in all cases the TG6 mutants showed faster degradation and had a markedly reduced half-life. The transamidase activity was measured for TG6 R111C, D327G, D510H, L517W and E574del from HEK293 cell lysates using a biotinylated substrate (Guo *et al.* 2014) 2013). For all TG6 mutants the transamidation activity was compromised. Taken together, these studies indicate that clinically significant TG6 mutations are likely to impact either the structure or function of TG6. These experiments were also carried out using TG6 mutants R111C, Q181H, D327G, V391M, Y441C, R448W, L502Q, L502Q Q652dup, D510H, L517W, E547del and Q652dup and are presented in Section 5.3.5-7.

1.12.2 Initiation of the UPR and targeting to the proteasome.

The levels of different UPR markers in cells expressing TG6 R111C and D327G were investigated (Tripathy *et al.* 2017). Bip, a protein associated with ER stress, was upregulated at both the mRNA and protein levels in cells expressing TG6 R111C (2x and 1.4x) when compared with WT, but no changes were observed for TG6 D327G. An additional two markers of UPR activation, namely spliced X-box binding protein 1 (sXBP1) and CCAAT/enhancer-binding protein (C/EBP) homologous protein (CHOP) were also measured and found to be increased. sXBP1 mRNA levels showed a significant increase (2x) in cells expressing TG6 R111C but also an increase of 1.6x

in cells expressing TG6 D327G, but this was not significant. CHOP mRNA levels increased 2.1x in cells expressing TG6 R111C when compared with WT but no change was seen with TG6 D327G cells.

All SCA patients in this study were heterozygous for the *TGM6* mutations, meaning WT TG6 and mutant TG6 would be present in the same environment (Tripathy *et al.* 2017). Immunoprecipitation showed WT TG6 forms a complex with R111C and D327G and the subcellular localisation of the complex was identified using a proximal ligation assay. WT TG6/TG6 D327G was localised to the nucleus and WT TG6/TG6 R11C interacted in the perinuclear space. The activity of the WT TG6 was investigated following interaction with mutant TG6. WT TG6 alone and WT TG6/ WT TG6 had similar activity, whereas TG6 R111C significantly reduced the activity of WT TG6 by 85%. A reduction in activity of WT TG6 of 58% post TG6 D327G interaction was observed but was not significant. These findings indicate that TG6 R111C has a dominant negative affect on enzymatic function of WT TG6 and indicates a dominant loss of function mechanism may be present in SCA35. These experiments show TG6 mutants localised to the perinuclear regions i.e. TG6 R111C can cause ER stress and be targeted for proteasomal degradation and impact WT TG6 function when in complex via a dominant negative mechanism. TG6 mutants localised to the nucleus i.e. TG6 D327G causes neuronal cell death via a different mechanism.

1.12.3 Assessment of the *in vitro* and *in vivo* viability of TG6 mutants.

The TG6 mutants may have a role in sensitising neuronal cells to apoptosis (Guan *et al.* 2013). The viability of HEK293 and NIH 3T3 cells expressing both WT and TG6 mutants (TG6 L517W and D327G) remained the same but when treated with staurosporine (STS), apoptosis was triggered. Significantly less apoptosis was observed in cells transfected with WT TG6 when compared with TG6 D327G, T426N and L517W (Guan *et al.* 2013; Nibbeling *et al.* 2017). A similar response was observed using a Ca^{2+} ionophore (A23187) to induce apoptosis. To establish if increased caspase activity is responsible for increased apoptosis, capase3/9 levels were assessed (Guan *et al.* 2013). Following STS treatment, there was a decrease in the appearance of caspase-3/9 fragments for WT TG6 compared with controls. Conversely, the TG6 mutants caused an increase in the cleavage of caspase compared with WT. Taken

together, these results indicate that expression of WT TG6 protects the cells from apoptosis, whilst cells expressing the TG6 mutants have increased sensitivity to apoptosis.

In contrast, overexpression of both WT TG6 and the TG6 mutants (TG6 R111C and D327G) in a neuronal cell line suggests increased sensitivity to neuronal cell death (Tripathy *et al.* 2017). The viability of cortical neurons expressing WT TG6, TG6 D327G and R111C showed over expression of WT TG6 reduces viability by 50% and TG6 R111C showed enhanced the neurotoxicity by 76.5% when compared with WT TG6. In contrast, TG6 D327G showed no change when compared to WT TG6. These results indicate that dysregulation of TG6 expression enhances neurotoxicity. When viability was measured in *Drosophila melanogaster*, those flies expressing WT TG6 and TG6 R111C showed increased susceptibility to death in the early stages of life. Death was seen at 17 days (for 25% of flies) and 18 days for flies expressing TG6 R111C and WT TG6 respectively and at 31 days in control flies.

1.13 Expression and purification of TG6.

A baculovirus/*Spodoptera frugiperda (Sf)* 9 cell system has been developed for the generation of recombinant TG6 (DPA lab and Zedira GmbH). This system generates sufficient TG6 for functional assays but would be expensive for crystallisation. As TG6 has identified post-translational modifications, an *E.coli* expression system could be developed for large scale TG6 expression for crystallisation. The current purification method utilises a His_6 tag, allowing purification by immobilised metal affinity chromatography (IMAC). Other chromatographic purification methods include ion exchange chromatography and size exclusion chromatography. Size exclusion chromatography can aid in the purification of TG6. However, TG6 requires substantial concentrations of glycerol to remain stable which is incompatible with this form of chromatography due to the viscosity of glycerol. Ion exchange chromatography is unsuitable for TG6 purification as the pI of TG6 is 6.85, and separation is reliant a pH 4 or 8, for which TG6 is not stable. Consequently, Ni affinity chromatography was selected as the most suitable purification method for TG6.

1.14 Aim of Thesis.

Extensive work has been carried out on TG2 in order to investigate its structure, function and the regulation of its conformation. Previous work has indicated that the conformation of TG2 is important for its export from the cell and this in turn maybe relevant in various pathological conditions. Our lab has investigated in detail the mechanistic aspects underlying the secretion of TG2 from the cell. Of particular interest is the role of GTP binding and hydrolysis in this mechanism. These earlier studies provide an opportunity to develop and optimise functional assays to test the role of GTP in TG2 secretion and to use functional mutants to gauge the sensitivity of the established assays. Furthermore, these assays can be used to carry out biochemical characterisation of TG6 and to elaborate differences between the WT TG6 and the TG6 mutants.

Bioinformatic and biochemical characterisation of WT TG6 and the TG6 mutants at the molecular level would aid in understanding of the underlying mechanism of SCA35. In order to do this, the structure and function of the TG6 mutants requires evaluation. In addition, investigating the role of TG6 in a cellular context would allow biochemical characterisation of the TG6 mutants in a more physiologically relevant model.

Hypotheses

In relation to the TG2 externalisation, I hypothesis the conformation of TG2 is important for export from the cell, rather than a phosphor transfer mechanism via the hydrolysis of GTP. With regards to the TG6 mutants which have been implicated in SCA35 I predict these mutants can be grouped via their impact on the TG6 protein using bioinformatic analysis which in turn will be validated with biochemical analysis. Together this will give us an insight into the potential mechanisms involved in the neuronal cell death observed in SCA35 patients.

Therefore, the following specific aims have been formulated for my thesis:

Aim 1: To investigate the role of nucleotide binding and hydrolysis in the externalisation of TG2. This study included the structural characterisation of WT TG2

and TG2 GTP binding site mutants (K173N, K173L, K173N/F174D and R580A) using circular dichroism (CD). The enzymatic function of TG2 GTP binding site mutants was assayed to assess the impact of mutants on TG2 activity when compared with WT TG2. Furthermore, GTP binding and GTP hydrolysis were used to assess the allosteric regulation of TG2 and the role of TG conformation and/or nucleotide hydrolysis in the mechanism of TG2 externalisation. In addition, insights from these experiments aided the development and optimisation of biochemical assays to investigate the function and regulation of TG6 and the TG6 mutants, using TG2 as a model protein.

Aim 2: To develop an *E.coli* expression system for human TG6 to generate TG6 of suitable quality and quantity for crystallisation. The selected expression plasmids express TG6 fused to a short His_6 peptide and/or other proteins, to aid its purification, expression and stability/folding in solution. The expression level, concentration of purified protein and activity were assessed to evaluate the advantages and disadvantages compared with the eukaryotic baculovirus expression system.

Aim 3: To use bioinformatics to investigate the structure and allosteric regulation of WT TG6 and predict the pathogenicity and structural and functional consequences of the TG6 mutants. Firstly, using the Exome Aggregation consortium (ExAc) and Genome Aggregation database (GnomAD) databases the prevalence of the mutations was extracted and the conservation of the mutants was assessed using MSA. The TG6 mutants predicted to impact function and solubility were evaluated using sequence based methods. TG6 homology models were generated to assess the impact of mutants on both overall and local structure and to identify changes in the allosteric binding sites for Ca^{2+} and GTP. In addition, the homology models were utilised to assess the 3D stability and solubility of WT TG6 and TG6 mutants. This analysis provided a structural framework within which to interpret the biochemical characterisation.

Aim 4: Biochemical characterisation of WT TG6 and the TG6 mutants. These studies were carried out using the methods used and developed in Aim 1. This study involved the expression and purification of the individual mutants. The outcome of these results was compared to predictions made from studies conducted as part of Aim 3.

Furthermore, the TG6 mutants were assessed in a cellular context to represent a more physiological setting.

CHAPTER 2

2. INVESTIGATING THE ROLE OF GTP IN THE SECRETION OF TG2.

2.1 INTRODUCTION

Previous data generated in our lab has indicated that TG2 is secreted by an unconventional mechanism which involves the activation of P2X7R by ATP (Adamczyk *et al.* 2015). It has also been previously demonstrated that the allosteric regulation of TG2 by Ca^{2+} and GTP has a role in its secretion. To understand the mechanism of TG2 secretion, residues in three functional sites (the fibronectin binding site, the active site and the GTP binding site) were mutated (Johnson and Terkeltaub 2005). Mutants were classified as fibronectin binding mutants: A2G and E3A, L7G, E8A, and R9A and L5del, active site mutants: C227G, H335A and D358A, the GTP binding site mutant K173L and the secretion mutant TG2 Y274A. Y274 is present as a *cis* peptide bond and is in close proximity to the active site cysteine residue (C277) (Balklava *et al.* 2002). The *cis-trans* isomerisation in this region is predicted to be important in the exposure of the active site. When mutated to Y274A the peptide bound switches to a *trans* peptide bond (non-proline) which in turn switches off transamidation activity by changing local conformation in the active site. The altered local conformation which results in no secretion

In previous work reported by our lab and others, the externalisation of TG2 and the TG2 mutants has been measured (Johnson and Terkeltaub 2005; Adamczyk *et al.* 2015). One group measured TG2 secretion by its ability to stimulate matrix calcification and induce type X collagen expression (Johnson and Terkeltaub 2005). In this case WT TG2 and the active site mutants were found to induce matrix calcification, whereas this was reduced with K173L and Y274A. In addition, the induction of type X collagen expression was absent for K173L and Y274A, whereas it was present for WT and the active site mutants (Johnson and Terkeltaub 2005). This means the levels of TG2 and the TG2 mutants was assessed in the conditioned media using Western blot analysis. This technique showed comparable externalisation for WT TG2, the active site mutants and the fibronectin mutants. The externalisation of TG2 K173L was dramatically reduced, as was found for TG2 Y274A. This result indicated that GTP was important in the externalisation of TG2 and led us to investigate whether GTP was also necessary for TG2 secretion in our newly established mechanism involving P2X7R (Adamczyk *et al.* 2015).

Two externalisation mechanisms involving GTP have been predicted by our group. 1) The hydrolysis of GTP and the transfer of phosphate to TG2 or an interacting partner provides the energy required for secretion. 2) The conformation of TG2 is important for secretion of TG2 and GTP binding is essential for adoption of the compact conformation. To investigate the mechanism of secretion further, GTP binding mutants were generated and their biochemical properties were analysed (Griffiths R. 2017). Based on previous data investigating GTP binding and hydrolysis, the selected TG2 GTP binding mutants were TG2 K173N, K173L, K173N/F174D and R580A (Iismaa *et al.* 2000; Begg *et al.* 2006b; Ruan *et al.* 2008; Singh *et al.* 2016).

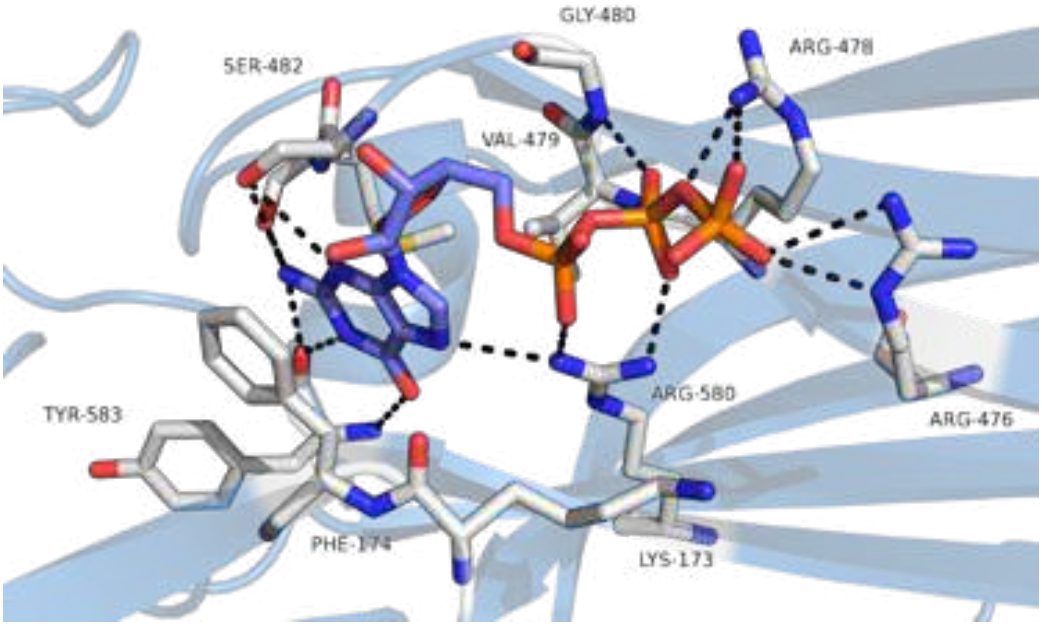


Figure 2.1: GTP binding pocket of TG2.
Figure generated using GTP bound TG2 X-ray structure (PDB: 4PYG, Jang *et al.* 2014). Residues which are important for GTP (lilac) stabilisation are shown (stick representation (carbon: white, nitrogen: blue and oxygen: red). Polar contacts ≤3.5Å are highlighted (black dotted line).

The GTP binding properties of these selected mutants have been investigated previously using various GTP binding assays. For example, ITC, GTP agarose, GTP photolabelling, GTP inhibition assays and BODIPY GTP binding assays (Iismaa *et al.* 2000; Begg *et al.* 2006a; Begg *et al.* 2006b). Using rat TG2, GTP binding was found to be impaired when K173 was mutated to an asparagine or leucine, and was abolished when R580 (human) or R579 (rat) was mutated to an alanine (Iismaa *et al.* 2000; Begg

et al. 2006b). Rat TG2 has previously been assumed to be functionally homologous to human TG2 and therefore data collected using rat models has been extrapolated to human TG2 (Ruan *et al.* 2008). TG2 carrying both K173N and F174D mutations has not been previously investigated, however mutations of these sites have indicated a role in GTP binding (Begg *et al.* 2006a; Datta *et al.* 2007). GTP binding was abolished in the F174A mutant and therefore it was hypothesised that the combination of these two mutations may result in abolished or reduced GTP binding (Begg *et al.* 2006a; Begg *et al.* 2006b). However, the F174W mutant bound GTP to the same degree as WT. As both phenylalanine and tryptophan contain hydrophobic rings, it is likely that this mutation is conservative for the necessary interaction with GTP (Begg *et al.* 2006a).

As the mutants described above differ in their GTP binding abilities, they can be used to help identify the role of GTP in TG2 secretion. Work has been carried out to investigate the GTPase activity of WT TG2 and the GTP binding site mutants (TG2 K173N and K173L) (Iismaa *et al.* 2000). GTPase activity was measured using the charcoal method. This involved the hydrolysis of [γ ^{32}P] GTP in the presence of WT TG2 and the TG2 mutants, and the release of γ ^{32}P was measured by scintillation counting. No published data investigates the GTP hydrolysis by TG2 R580A, but as this does not bind GTP, it is unlikely to have GTPase activity (Begg *et al.* 2006a; Begg *et al.* 2006b; Datta *et al.* 2007). K173N had impaired GTP hydrolysis compared with WT TG2 and the hydrolysis activity of K173L was barely detectable (Iismaa *et al.* 2000).

In addition, the conformational state of TG2 in the presence of GTP has been investigated. TG2 conformation was assessed for WT TG2 and R579A by measuring their sensitivity to trypsin and μ-calpain in the presence of excess GTPγS (Begg *et al.* 2006b). WT TG2 was resistant to trypsin and μ-calpain digestion, whereas R579A was not. This result indicated that TG2 R579A was in a different conformation to WT in the presence of GTPγS, likely due to it being unable to assume the compact conformation. Additional analysis of R579A, using native PAGE, showed a fast migrating species which was somewhat retarded when compared to GTP-bound WT TG2 (Begg *et al.* 2006b). This altered electrophoretic mobility was also observed when R579A was paired as a double mutant with R476A or R478A. This result indicates

that these mutations result in the adoption of a conformation that is not seen for WT TG2, as it is not the same as the slower migrating 'relaxed' open conformation or the faster migrating compact conformation. Sedimentation velocity experiments showed GTP dissociated from R579A faster when compared to WT TG2. GTP binding by R579A is accompanied by a large conformational change which results in a change in the associated sedimentation coefficient from 3.1 S to 2.8 S, which supports a model of rapid exchange between free and bound GTP TG2 structures (Begg *et al.* 2006b). These experiments show R579 is essential for conversion to the compact conformation in rat TG2, therefore R580 is also likely to be essential to formation of the compact conformation in humans. TG2 R580A will therefore be valuable in establishing whether the conformation adopted by TG2 is important in its secretion from the cell.

Published work investigating GTP binding, GTPase activity and conformational regulation of the TG2 GTP binding mutants is useful in informing the best selection of mutants for investigating these functions in TG2 export. Our group has demonstrated that there is differential secretion of the GTP binding mutants and those unable to bind GTP were not detected in conditioned media from HEK293 cells overexpressing both P2X7R and the TG2 mutants, stimulated with ATP (Griffiths R., 2017). Further work investigating the biochemical properties of these mutants can aid in establishing the role of GTP in the secretion of TG2 in response to P2X7R activation. Additionally, the sensitivity of the developed GTP binding/ GTP hydrolysis assays, used in this chapter, can aid in the biochemical characterisation of TG6.

This chapter describes studies towards achieving three aims:

1. Structurally characterise and probe the enzymatic function of TG2 GTP binding site mutants.
2. Assess mechanistic aspects of GTP binding and hydrolysis by TG2 in the context of protein externalisation by cells.
3. Develop and optimise biochemical assays to investigate enzymatic functions of TGs using WT TG2 as a model protein.

2.2 METHODS

2.2.1 Counting cells

Cells were diluted 1:1 with trypan blue (Sigma, T8154) and 10 µL of mixture was added to a haemocytometer. Cells in 4 squares were counted and number of cells per mL calculated using the following equation:

Equation 2:1: Cell counting equation

$$\text{Cells/mL} = \frac{\text{total number of cells}}{\text{number of squares}} * \text{dilution factor} * 10000$$

2.2.2 Externalisation of TG2 mutants

HEK293 cells stably transfected with human P2X7 receptor were seeded at a density of $1.2x10^5$ per well in 0.5 mL Dulbecco's modified eagle's medium (DMEM, Gibco 11960044), containing 10% heat inactivated foetal bovine serum (hiFBS), (Gibco 10082147) and 10 U/mL pen/strep in a 24 well plate (Sarstedt 83.3922) and incubated at 37 °C, 5% CO_2 for 24 hours. Media was changed to fresh DMEM (containing 10% hiFBS, no antibiotics) and the cells were transiently transfected with pcDNA3.1 TG2 (provided by D. Aeschlimann) for the expression of WT TG2 or TG2 GTP binding mutants (K173N, K173N/F174D, K173L or R580A), using Fugene-6 (Promega, E2691). The transfection was carried out as per manufacturer's instructions, at a ratio of 1.5 µL Fugene-6: 0.5 µg DNA. A mixture of Fugene-6, DMEM and TG2 plasmid was incubated at RT for 30 minutes before adding 100 µL of the transfection reaction to each well of the 24 well plate. The cells were then stimulated with 0.1 mM 2' (3')-*O*-(4-benzoylbenzoyl) adenosine triphosphate (BzATP, Sigma B6396, Adamczyk M., 2013) 48 hours post-transfection. The 0.1 mM BzATP was diluted in OptiMEM (Gibco 11058021, prewarmed to 37 °C. 200 µL/well of OptiMEM was used to wash the cells then 250 µL per well OptiMEM + BzATP was incubated with the cells for 10 minutes. Media was collected, and the cells washed with OptiMEM, and fresh OptiMEM without BzATP was added to the cells for 30 minutes and then the media was collected. The collected media was centrifuged at 1500 xg for 10 minutes and the supernatant was kept for analysis. The HEK293 P2X7R cells were lysed with extraction buffer (20 mM HEPES/NaOH pH 7.4, 150 mM NaCl, 1 mM ethylene glycol-bis (β-aminoethylether)-N,N,N',N'-tetraaccetic acid (EGTA), 1% Triton-X-100 and 0.25% deoxycholic acid, 1 mM phenylmethylsulfonyl fluoride (PMSF), 1 mM

N-ethylmaleimide (NEM) and 10% glycerol), and the lysate centrifuged at 15000 xg for 15 minutes at 4 °C. The supernatant was collected for analysis of levels of TG2 expression in the cells.

2.2.3 Sample preparation

Conditioned media samples were lyophilised and reconstituted at 10x the original concentration to allow loading of a sufficient amount of conditioned media (100 µL) for sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE, Section 2.2.4) and Western blot analysis (Section 2.2.5) for TG2 detection. 10 µg of total protein (measured with bicinchoninic (BCA) assay as per manufactur's instructions, Thermo Scientific Pierce, A53227) was loaded for the cell lysate samples.

2.2.4 SDS PAGE analysis.

Precast Novex gels (4-20% Tris-glycine, Invitrogen) were run at 120 V for 2 hours (Invitrogen electrophoresis tank (XCell Sure Lock^{TM} Mini-Cell electrophoresis system and Novex Invitrogen Powerease 500 Electrophoresis power supply) in running buffer (25 mM Tris-HCl pH 8.8, 192 mM glycine, 0.1% SDS). Gels were stained with Coomassie brilliant blue (1g/L Coomassie brilliant blue, 50% methanol and 10% glacial acetic acid) or used for Western blot analysis.

2.2.5 Western blot analysis.

Proteins were transferred on to nitrocellulose membrane (Amersham protran NC, Thermo Fisher), for 2 hours (constant current of 120 mA), using transfer buffer (25 mM Tris, 192 mM glycine, 20% methanol). Ponceau S solution (0.1% Ponceau S, 5% acetic acid in H_2O) was used to visualise transfer efficiency and mark the protein ladder (LMW-SDS marker kit, GE Healthcare 17044601). Membranes were blocked overnight at 4 °C in 5% milk (non-fat milk powder, Fisher) in Tris buffered saline (TBS, 20 mM Tris-HCL pH 7.4, 150 mM NaCl).

The membrane was stained using mouse monoclonal anti-TG2 antibody (20 ng/mL, CUB7402, Invitrogen MA512736) and was incubated for 1.5 hours at RT. The membrane was then washed (3x) with TBS. Anti-TG2 antibody was detected by incubation of membrane with polyclonal rabbit anti-mouse antibody conjugated to

horse radish peroxidase (HRP, 2 µg/mL, Dako P0260) for 1 hour at RT. The membrane was washed in TBS-Tween 20 (TBS-T, TBS containing, 0.05% Tween 2 0) (3x). ECL Prime (GE health care), with a ratio of 1:1 solution A: solution B was added to the membrane for 2 minutes. Bands were captured on X-ray film (Amersham Hyperfilm, GE healthcare).

2.2.6 Bacterial expression of human Transglutaminase-2

BL21 competent *E. coli* transformed with pMAG21.2 TGM2 cDNA were selected on carbenicillin (100 µg/mL, Fisher Bioreagents, BP2648) agar plates (Heil A., 2015). Overnight cultures were prepared from colonies (37 °C, 200 rpm) and used to inoculate a larger culture. When the optical density at 600 nm (OD_{600}) = 0.6 (Beckman DU800 spectrophotometer, 1cm cell path length) protein expression was induced using L-Rhamnose (0.5 g/L, Alfa Aesar A16166.14) and the culture was incubated for ~12 hours (20 °C, 150 rpm). Cells were harvested using a Sorvall RC6 centrifuge (3,000 xg, 4 °C, 20 minutes).

2.2.7 Purification of His6-tagged human TG2

The cell pellet was thawed and resuspended in lysis buffer (50 mM Na_2HPO_4 pH 8.0, 300 mM, NaCl, 4 °C) and homogenized (Stansted fluid power (SFP) C-10-60 homogenizer, 1 bar, 2 °C) or sonicated (60%, 30 seconds on and 30 seconds off for 10 minutes). The homogenate was centrifuged for 30 minutes, at 4 °C, and 11500 xg. The supernatant was loaded onto a 5 mL His trap™ HP column (GE Healthcare, 17524801) using an AKTA Prime system (GE healthcare) at 4 °C. The protein was eluted using a gradient prepared from lysis buffer and elution buffer (50 mM Na_2HPO_4 pH 8.0, 300 mM NaCl, 300 mM imidazole, 4 °C). Samples from each step were analysed using SDS-PAGE Coomassie brilliant blue R staining (as described in Section 2.2.4). Protein fractions were pooled and diluted 1 in 3, using 20 mM Tris-HCl pH 7.2, 1 mM EDTA buffer for ion exchange chromatography. A Resource Q15 column and an AKTA Purifier chromatography system (GE Healthcare) at RT was utilised. A 10 mL Superloop™ (GE Healthcare) was used for sample injection. Fractions were collected following a linear gradient of low salt buffer (20 mM Tris-HCl, pH 7.2, 100 mM NaCl, 1 mM EDTA) and high salt buffer (20 mM Tris-HCl, pH 7.2, 1 M NaCl, 1 mM EDTA). Absorbance was monitored at 280 nm. Specific protein

concentration was measured using a DU800 spectrophotometer (as described in Section 2.2.9). The 280 nm peak fractions were analysed using SDS-PAGE analysis (Section 2.2.8). TG2 GTP binding variants were purified using the same method (as described in Section 2.2.7).

2.2.8 SDS-PAGE analysis of purified TG2

5 µL, 10 µL of purified protein sample or 2 µg of purified TG2 protein in equivalent 2x sample buffer (25 mM Tris-HCl pH 6.8, 3.9 mM EDTA, 4% SDS, 30% glycerol, 0.3% bromophenol blue) 5% β-mercaptoethanol (Sigma, M6250)) was loaded. 5 µL of LMW-SDS marker kit (GE Healthcare 17044601) in an equivalent volume of 2x sample buffer was loaded. Gels were run as described in Section 2.2.4.

2.2.9 Protein concentration determination

Final protein concentration was determined on a Beckman DU800 spectrophotometer using the wavelength scan setting and a 1cm cuvette. The concentration was calculated using the Beer's Law equation

Equation 2:2: Beer's Law equation.

$$\text{Absorbance}_{280\text{nm}} = \text{molar extinction coefficient} \times \text{path length (cm)} \times \text{concentration (mg/mL)}$$

2.2.10 Isopeptidase activity assay

Transglutaminase isopeptidase activity was measured using an established fluorescence assay (Adamczyk *et al*. 2013). The quenched substrate A102 (50 µM, DMSO, optical grade, Zedira) was diluted to 6.25 nM in assay buffer (62.6 mM Tris-HCl, pH 7.5, 125 mM NaCl, 12.5 mM glycine methyl ester) at 37 °C. 20 µg/mL of TG2, 5.1 mM DTT (diluted from a 500 mM stock) and 80 µL A102 assay buffer were loaded into a 96 well plate (Nunc 96-well black optical bottom plates, Thermo Scientific 165305). 2 mM $CaCl_2$ (or H_2O for control) was injected into each well after 5 minutes and the change in fluorescence was measured. Fluorescence was measured using a FLUOstar Omega plate reader (BMG Labtech) with λ_{ex} 320 nm and λ_{em} 440 nm. The readings were baseline (initial 10 cycles) and blank (H_2O injected) corrected and input into Graphpad Prism V6 for analysis. The rate of reaction was calculated using linear regression and percentage activity is presented as a function of WT TG2.

2.2.11 MDC activity assay

Monodansylcadaverine (MDC) incorporation into N, N-dimethyl casein by TG2 gives a measure of its transamidation activity (Lorand *et al.* 1971). To measure real-time amine incorporation, 20 μg/mL TG2 was added to a 96 well plate (Black optical bottom, Nunc) in a ratio of 1:10 of TG2:assay buffer (50 mM Tris-HCl pH 8.0, 10 mM Glutathione (reduced, Sigma G4251), 2.5% glycerol, 2.5% DMSO, 20 μM N,N-dimethyl casein (Sigma C9801) and 25 μM MDC (Sigma 30432)). The 96 well plate and assay buffer were prewarmed to 37 °C before use and the reaction was initiated by injection of 2 mM $CaCl_2$. The fluorescence was measured using the FLUOstar Omega, using bottom optics (λ_{ex} 332 nm and λ_{em} 515 nm). Results were analysed as described in Section 2.2.10.

2.2.12 Circular Dichroism of TG2 GTP binding variants

WT TG2 and TG2 GTP binding variants were purified and diluted to 0.5 mg/mL for structural analysis using circular dichroism spectroscopy (CD). CD spectra (260-190 nm) were collected using a combination of 0.02 cm and 0.1 cm quartz cuvettes at 4 °C, using an Aviv Model 215 spectropolarimeter (Aviv Biomedical Inc., Lakewood, NJ). The TG2 was buffer exchanged into 10 mM Tris-HCl pH 7.5, 50mM NaCl, 0.5 mM EDTA and diluted to a concentration of ~6 μM, as measured with a DU800 UV spectrophotometer (as described in Section 2.2.9). With the 0.1 cm cuvette, a spectrum was measured using 3 second accumulation time per point, 0.1 nm intervals and 1 nm bandwidth. Parameters for the 0.02 cm cuvette were the same except a 6 second accumulation time was used. Buffer baselines were recorded in both cuvettes and smoothed (weighted over 9 neighbours, Prism 4.0a 2003) and subtracted from the data. The data from both cells was combined to produce a spectrum ranging from 190-260 nm and was normalised to the mean residue ellipticities ($[\theta]_{MRW}$). Dichroweb used the CDSSTR algorithm and SP175 data set to calculate the α-helix and β-strand content (Sreerama and Woody 2000; Whitmore and Wallace 2004, 2008).

2.2.13 Ion exchange chromatography

Chromatograms associated with each ion exchange chromatography purification step (Section 2.2.7) of each TG2 protein were combined for comparison. The $Abs_{280\ nm}$ was plotted against elution volume on Prism 7.

2.2.14 Size exclusion of TG2 proteins using gel filtration

The size and shape of the TG2 proteins (0.5 mg/mL) were analysed using size exclusion chromatography, using a Superdex 200 10/300GL column (GE Healthcare). The column was equilibrated in 20 mM Tris-HCl, pH 7.4, 300 mM NaCl with a flow rate of 0.4 mL/minute. Absorbance was measured throughout at λ 280 nm and λ 254 nm. $Abs_{280\ nm}$ was plotted against volume (mL) for each TG2 protein for comparison.

2.2.15 GTP/ATP isopeptidase inhibition assay

WT TG2 and for all the GTP TG2 mutants isopeptidase activity (Section 2.2.10, Adamczyk *et al*. 2013) was measured in the presence of increasing concentrations of GTP (0-2 mM) or ATP (0-8 mM). The inhibition of only WT TG2, K173N and K173N with ATP was measured. Each reaction was baseline and blank corrected, as described in Section 2.2.10. The data was analysed using linear regression (data between 1000-3000 seconds, to ensure a linear portion of the graph was used) and the gradient of each reaction was used to normalise data to between 0 and 1. Data was fitted with a sigmoidal dose response curve to measure the apparent K_D (half maximal activity).

Equation 2:3: Sigmoidal dose response curve equation.

$$Y = Y_{min} + \frac{Y_{Max} - Y_{min}}{1 + 10^{(\log EC_{50} - X)}}$$

2.2.16 Native PAGE analysis of GTP binding site mutants

4 µg of TG2 was mixed with an equivalent volume of 2x sample buffer (100 mM Tris-HCl pH 6.8, 15% glycerol, 0.02% bromophenol blue, 10 mM DTT). Samples containing GTP (25 µM, Sigma G5884) were incubated for 10 minutes prior to loading. 0.25 µM GTP was also included in the running buffer (25 mM Tris-HCl pH 7.8, 192 mM Glycine) for those samples pre-incubated with GTP. All samples (with

and without GTP) were run on 10% Tris-glycine gels (Invitrogen, XP00102BOX) at 100 V, 35 mA for 3 hours on ice.

2.2.17 GTP binding studies for TG2 GTP binding site mutants

CD spectroscopy was used to monitor structural changes associated with GTP binding. Measurements were taken at 218 nm (minimum point) in a 0.1 cm quartz cuvette for 60 seconds following addition of GTP. Both 0.8 mM (WT TG2 and R580A) and 0.4 mM (TG2 K173N, K173L and K173N/F174D) GTP stocks in 10 mM Tris-HCl pH 7.5, 50 mM NaCl, 0.5 mM EDTA were used, and 1 µL of GTP was added sequentially. Measurements were taken for the buffer alone as a control and was subtracted from all following measurements. The concentration (µM) was calculated for TG2 and GTP for each step of GTP addition using Equation 2.4. MRW: Mean residue weight.

Equation 2:4: Conversion to Molarity.

$$\text{Molarity } (\mu\text{M}) = \frac{\text{Concentration (mg/mL)}}{\text{MRW (g/mol)}} \div 10000$$

The dilution of TG2 and GTP upon the addition of GTP was calculated using Equations 2.5 and 2.6. The GTP was measured in buffer without TG2. The contribution of GTP to the TG2 spectra was calculated using Equation 2.7 and was subtracted.

Equation 2:5: Impact of dilution on TG2 concentration.

$$\text{TG2 } (\mu\text{M}) = \frac{\text{TG2 } (\mu\text{L})}{\text{Total volume } (\mu\text{L})} \times \text{MRW (g/mol)}$$

Equation 2:6: Impact of dilution on GTP concentration.

$$\text{GTP } (\mu\text{M}) = \frac{\text{GTP } (\mu\text{L})}{\text{Total volume } (\mu\text{L})} \times \text{MRW (g/mol)}$$

Equation 2:7: GTP contribution.

$$\text{GTP contribution} = \text{Gradient of GTP buffer spectrum} \times \text{GTP } (\mu\text{M})$$

From the CD signal (mdeg) the $[\theta]_{MRW}$ (deg cm^2 mol^{-1}) was calculated using Equation 2.8.

Equation 2:8 Calculation of $[\theta]_{MRW}$.

$$[\theta]\text{MRW deg cm}^2\ \text{mol}^{-1} = \frac{\text{CD signal (mdeg)} \times \text{MRW (g/mol)} \times 100}{\text{Cell path length (cm)} \times \text{concentration (M)}}$$

The $[\theta]_{MRW}$ at 218 nm was plotted against GTP concentration (Graphpad Prism 4.0a 2003). Curves were fitted using Equation 2.9, referring to one binding site and the K_D was calculated.

Equation 2:9: Calculation of the K_D.

$$[\theta] = \frac{B_{Max} \times [GTP]}{K_D + [GTP]}$$

A spectrum (203-260 nm) was measured before and after the addition of the maximal amount of GTP to further identify the change in the secondary structure. A spectrum of buffer containing the equivalent amount of GTP was subtracted from the titrated GTP spectrum.

2.2.18 Binding studies of GxP for WT TG2

To identify whether a change in the number of phosphate groups on the guanine moiety affected the CD signal, GTP was added to ~0.2 mg/mL WT TG2 in two concentrations: 3-6 μM and ~150-370 μM (in excess of TG2) from stock solutions at 1 mM and 10 mM respectively. The two stock concentrations allowed the final concentrations to be achieved with minimal dilution of TG2. GTPγS (Sigma G8634,), GDP (Sigma G7127) or GMP (Sigma G8377) were also added to WT TG2 in a similar manner. Final concentrations of components are detailed in Table 2.1.

Table 2.1: GxP binding CD set up.

Nucleotide	TG2 (µM)	Low Nucleotide (µM)	High Nucleotide (µM)	$CaCl_2$ (mM)
A: GTP	6	6	370	5.5
B: GTPγS	6	3	150	5.5
C: GDP	6	6	345	5.5
D: GMP	6	5	302	5.5

2.2.19 Nucleotide binding of TG2 using SWISS Model

All generated structures were visualised using the PyMOL molecular graphics system (version 2.0, Schrodinger, LLC). Nucleotide binding was compared using various structures solved using X-ray crystallography (PDB: 4PYG, 3LY6 and 1RLE) (Ahvazi *et al.* 2004a, Han *et al.* 2010; Jang *et al.* 2014). First, the impact on GTP binding by the TG2 GTP binding site mutants was predicted. The protein sequence of GTP bound TG2 (PDB: 4PYG, (Jang *et al.* 2014)) was mutated and input into SWISS Model (Guex *et al.* 2009; Benkert *et al.* 2011; Bertoni *et al.* 2017; Bienert *et al.* 2017; Waterhouse *et al.* 2018). The resulting homology model was visualised using PyMOL.

GTP bound TG2 (PDB: 4PYG, (Jang *et al.* 2014) and ATP bound TG2 (PDB: 3LY6 (Han *et al.* 2010)) were visualised with PyMOL. To generate GTPγS bound TG2, GTP bound TG2 and TG3 bound GTPγS (PDB: 1RLE, (Ahvazi *et al.* 2004) were structurally aligned using SWISS Model and superimposed. To understand the GTP hydrolysis mechanism, GTP bound TG2 was also visualised with co-crystallised H_2O.

2.2.20 Molecular docking of nucleotides

Alternatively, as there is no X-ray crystal structure available for TG2• GTPγS, molecular docking studies were also carried out. To generate GTPγS as a docking ligand, GTP was edited using Molecular Operating Environment (MOE, 2013.08), whereby an oxygen on the γ-phosphate was substituted with sulphur. The GTPγS ligand and TG2•GTP (PDB: 4PYG, (Jang *et al.* 2014)) were input into Maestro11 (Schrodinger, Release 2018, LLC, New York, NY). Both molecules were prepared for docking using protein prep wizard. This included minimisation of hydrogens and optimisation of side chains to reduce steric clashes (Sastry *et al.* 2013). A receptor grid (X, Y, Z coordinates) of TG2•GTP, excluding the existing GTP was calculated. The grid established the coordinates of the nucleotide binding pocket. GTPγS was docked

using the ligand docking tool (Glide, Schrodinger, LLC, New York, NY, 2018 (Friesner *et al.* 2004; Halgren *et al.* 2004; Friesner *et al.* 2006)). Extra precision (XP fitting) was required for successful docking. Ten conformational predictions were generated and similar conformations were grouped together. This approach was also to dock ATP, GTP and GTPγS into the GTP binding pocket of TG3 (PDB 1L9N and 1RLE, (Ahvazi *et al.* 2003; Ahvazi *et al.* 2004)

2.2.21 BODIPY GTP binding assay

BODIPY GTPγS (Invitrogen G22183) was used to measure the fluorescence over time as a result of guanine association. Measurements were carried out at 25 °C in a black optical bottom 96 well plate (Thermo Scientific) using a FLUOstar Omega (λ_{ex} 480 nm and λ_{em} 520 nm, gain 2500). 90 µL of assay buffer (20 mM Tris-HCl pH 7.4, 100 mM NaCl and 0.2 mM $MgCl_2$) containing 20 µg/mL TG2 (WT or mutant) was used to measure GTPγS association. 10 µL BODIPY GTPγS (50 nM) was injected and the association was measured over 250 seconds. The dissociation of BODIPY GTPγS involved injection of excess unlabelled GTPγS (100 µM, 10 µL) and fluorescence was measured over 250 seconds. BODIPY GTPγS injection without TG2 was used as the control. BODIPY GTP (Invitrogen G12411) binding was measured using the same method and using unlabelled GTP to dissociate. GTPγS binding was also assessed in this way, with varying concentrations of TG2 protein (40-2.5 µg/mL).

2.2.22 BODIPY turnover assay

Reactions were set up as described in Section 2.2.21. BODIPY GTP dissociation was measured over 16 hours instead of by injection of unlabelled excess GTP (Willard *et al.* 2005).

2.2.23 GTPase Glo™ assay

Intrinsic GTPase activity was measured using the GTPase-Glo™ assay (Promega V7681, Schematic Fig. 2.2) (Mondal *et al.* 2015). First a 2x GTP solution was made using GTPase/ GTPase activating protein (GAP) or guanine nucleotide exchange factor (GEF) buffer (Table 2.2). TG2 (0-1000 ng) was incubated at 37 °C with an equivalent volume of the 2x GTP solution for between 0-16 hours, measured at 2 hour intervals, in a white 96 well plate (Corning, 3912). After incubation the 25 µL GTPase-

Glo™ reagent (reconstituted, Table 2.2) was added and incubated with shaking at RT for 30 minutes. This step converts the remaining GTP into ATP. 50 µL of detection reagent (supplied by kit, contains luciferase) was injected and luminescence signal was captured over 20 minutes using the FLUOstar Omega (lens filter setting). The ratio of 1:1:2 (GTPase reaction: GTPase-Glo™ reagent: Detection reagent) was maintained in these experiments. For this assay the optimal GTP concentration was determined using a standard curve. Briefly, the luminescence was measured for GTP alone at different final concentrations (0-20 µM) to establish optimum GTP concentration. To optimise the amount of TG2 required for the assay, a titration of WT TG2 from 0 ng to 1000 ng, following incubation for 3 hours was carried out to identify the useable amount of TG2 in relation to luminescence signal. In addition, the incubation of TG2 with GTP was carried out with both the GTPase/GAP buffer and GEF buffer to identify the optimal buffer for TG2 and the reaction time was monitored every 2 hours to identify the optimal reaction time (up to 16 hours) where the baseline luminescence is representative of little/no remaining GTP.

Table 2.2: Buffers used in GTPase-Glo™ assay.

Solution	Components
GEF buffer	50 mM Tris-HCl pH 7.5 50 mM NaCl, 10 mM $MgCl_2$ 1 mM EDTA
GTPase/GAP buffer	50 mM Tris-HCl pH 7.5 50 mM NaCl 5 mM $MgCl_2$ 20 mM EDTA
2X GTP solution	10 µM GTP 1mM DTT in GTPase/GAP buffer or GEF buffer
GTPase-Glo™ buffer	Components not supplied
GTPase-Glo™ reagent	2 µL 500X GTPase-Glo™ reagent 0.5 µL 10mM ADP 998 µL GTPase-Glo™ buffer
Detection reagent	Components not supplied

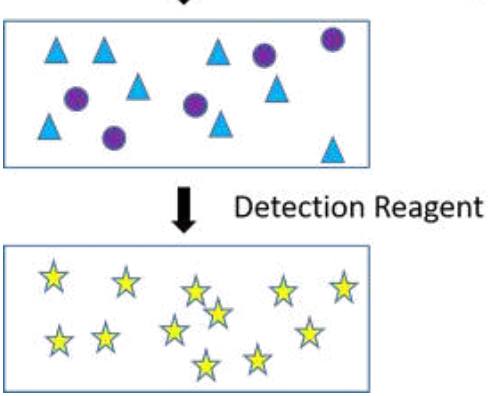


Figure 2.2: Schematic of GTPase-Glo™ assay.
TG2 was incubated with GTP for a set period of time (0-16 hours, at 2 hour intervals). The remaining GTP is converted to ATP using the GTPase-Glo™ reagent. The ATP is used to power a luciferase in the detection reagent which in turn generates a luminescence signal (Image adapted from Promega).

2.3 RESULTS

2.3.1 Assessing the secretion of the GTP binding site mutants

In order to assess the impact of the TG2 GTP binding mutants on the secretion of TG2, the levels of TG2 in conditioned media were measured with Western blot (Fig. 2.3, work carried out by Rhiannon Griffiths).

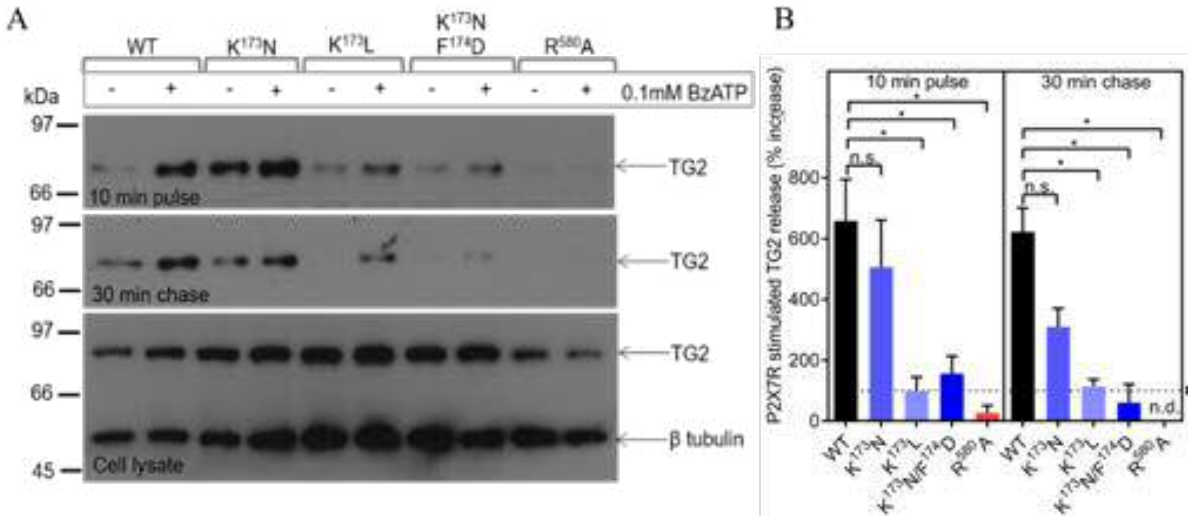


Figure 2.3: Secretion of TG2 GTP binding site mutants by HEK293 P2X7R cells. TG2 was transfected into HEK293 P2X7R cells and treated with 0.1mM BzATP for 10 minutes. The conditioned media was collected and BzATP removed. P2X7R activation was followed for a further 30 minutes and conditioned media was collected. Cell lysates were collected for assessment of TG2 expression level. A: Western blot analysis of conditioned media and cell lysates which assessed the secretion of TG2. B: Quantification of Western blot using densitometry. The percentage increase of stimulated versus unstimulated cells was calculated for each mutant. Data shown are mean ± SEM. Significance was measured using ANOVA, followed by Tukey's post-hoc. n.d.: not detected, n.s: not significant, *: $p<0.05$, **: $P<0.005$ (Work carried out by Rhiannon Griffiths).

Upon stimulation of HEK293 P2X7R cells with BzATP, there in an increase in the amount of WT TG2 detected in the media, in both the 10 minutes pulse and the 30 minute chase (once agonist has been removed) (Fig. 2.3A, compare lanes WT – and + BzATP, for the respective TG2). This response was also seen with the TG2 K173N mutant, but secretion was only slightly reduced and was not statistically significant (Fig. 2.3). In contrast, for TG2 K173L and TG2 K173N/F174D a significant reduction in the secretion of TG2 was observed and little/no secretion was detected with TG2 R580A (Figure 2.3A, R580A lane – and + BzATP). These results indicate that three out of four of the TG2 GTP binding site mutants have reduced secretion compared

with WT TG2, suggesting that GTP binding has a role in the mechanism of TG2 export.

2.3.2 Molecular modelling to predict GTP binding in the TG2 mutants

Molecular modelling of the TG2 GTP binding mutants was used to predict their impact on GTP binding. Modelling utilised TG2•GTP (PDB: 4PYG, (Jang *et al.* 2014)) and images were generated with the respective mutation introduced (Fig. 2.4). WT TG2 has interactions between F174, R476, R478, V479, G480, S482, M483, R580, Y583 and GTP within the GTP binding site (Fig. 2.4A). TG2 K173N, does not interact with the phosphate groups of GTP, and instead the N173 side chain amine interacts with a hydroxyl group on the GTP guanine ring (Fig. 2.4B). As a result, the guanine moiety is further stabilised. This model suggests GTP binding equivalent to or greater than WT TG2. No interaction with the γ-phosphate or the guanine moiety of GTP was observed with the neutral side chain of TG2 K173L (Fig. 2.4C). Consequently, GTP has reduced stabilisation and therefore GTP binding may be reduced. TG2 K173N/F174D, retains the interaction between the N173 amine and hydroxyl group of guanine (Fig. 2.4D). D174 stacks over the guanine ring and interacts with the hydroxyl on guanine via its side chain carboxylate. D174 may not be sufficient to replace hydrophobic interactions of F174 and therefore GTP binding may be reduced. WT TG2 makes three interactions with the R580 side chain amine (Fig. 2.4A), which stabilise the guanine ring, and the α and β phosphates of GTP (Fig. 2.4E). When mutated to A580 all three interactions are lost, predicting reduced GTP binding.

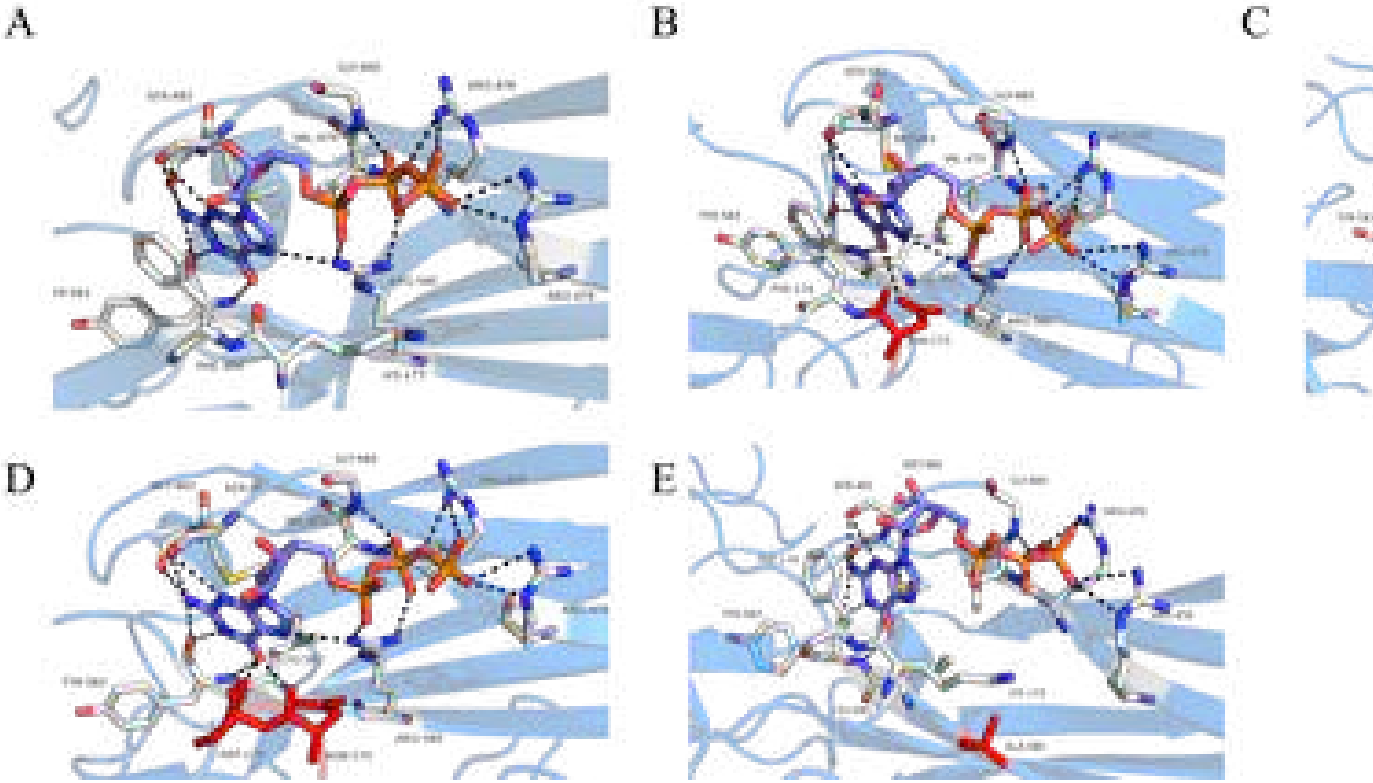


Figure 2.4: Molecular modelling of TG2 GTP binding variants.

The selected residues in the GTP binding pocket were mutated in the GTP bound TG2 structure (PDB: 4PYG, is shown in cartoon representation (blue) and GTP (carbon: purple, oxygen: red, nitrogen: blue and phosphate: white) are show in stick representation. Polar contacts of ≤3.5Å are highlighted (black dotted lines). The muta TG2. B: TG2 K173N. C: TG2 K173L D: TG2 K173N/F174D D: TG2 R580A.

2.3.3 Generation of WT TG2 in a prokaryotic expression system

To investigate the biochemical properties of WT TG2 and the TG2 GTP binding mutants, they were expressed and purified following heterologous expression in recombinant *E. coli* cells. Samples were collected at multiple points throughout the purification (i.e. supernatant, flow-through, 30 mM imidazole wash, 150 mM elution of TG2 and 300 mM imidazole wash) for SDS-PAGE and functional analysis to identify any losses. SDS-PAGE analysis showed an intense band of ~70 kDa indicative of successful TG2 expression in the 150 mM imidazole elution fractions (Fig. 2.5A). The 150 mM fractions containing the eluted TG2 also contain a large number of contaminating protein bands. Following anion exchange chromatography the eluted fractions were again analysed by SDS-PAGE Coomassie staining (Fig. 2.5B). Additional bands were still identified in the eluted fractions. However, these protein bands have been previously analysed and identified as fragments of TG2 using multiple steps of mass spectroscopy (ms/ms) (D. Aeschlimann personal communication).

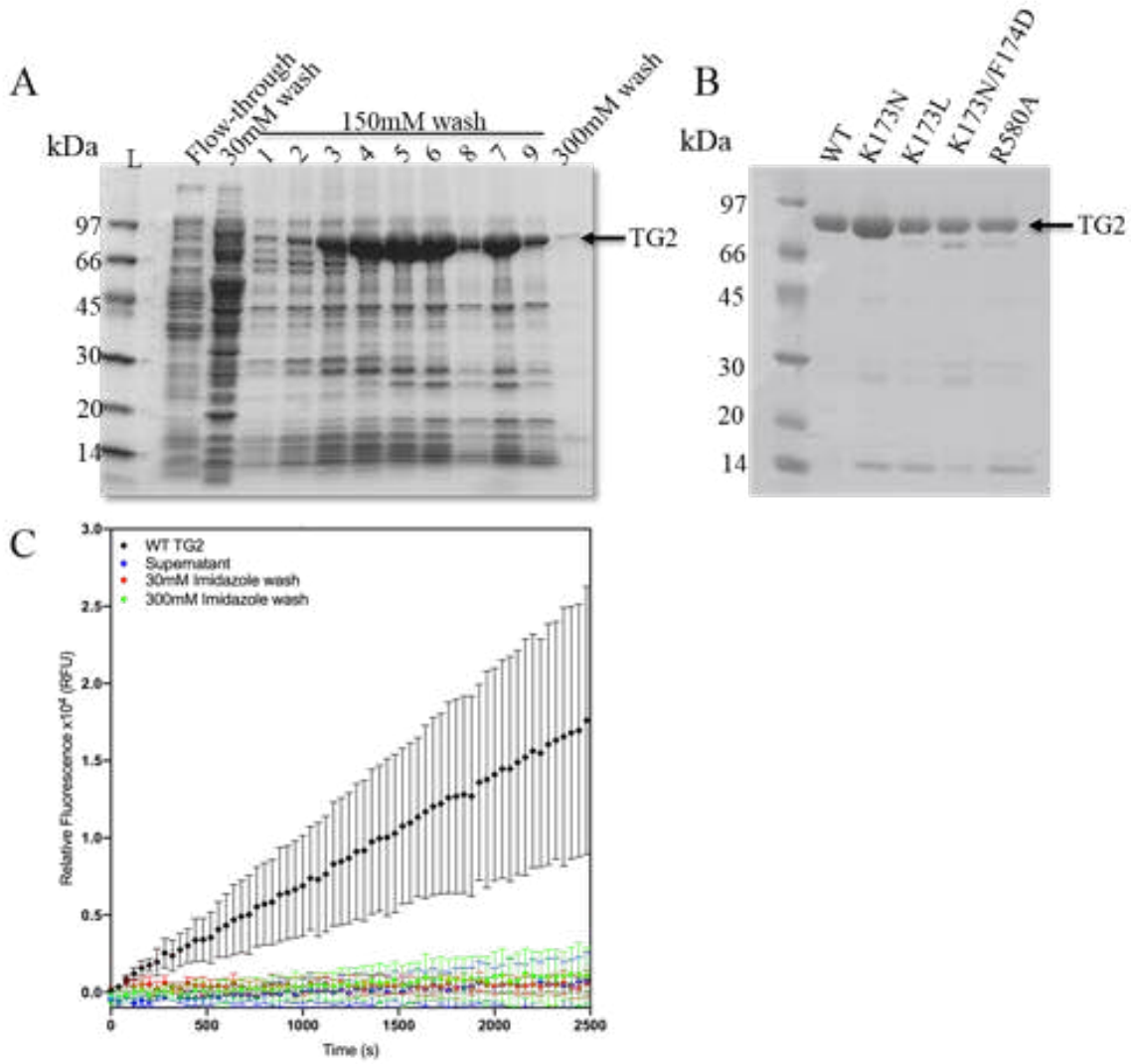


Figure 2.5: Expression and purification of TG2 GTP binding variants.
Samples were collected throughout the purification (supernatants, flow-through, 30 mM wash, 300 mM wash and purified TG2) and analysed using SDS-PAGE analysis, and the isopeptidase activity measured. A: Analysis of 10 µL of collected samples during the His tag affinity purification step, using SDS-PAGE and staining using Coomassie brilliant blue R. B: Analysis of 2 µg of purified WT TG2 and TG2 mutants using SDS-PAGE and stained using Coomassie brilliant blue R.C: Isopeptidase activity of the purified samples (8.3 µL of each sample or 20 µg/mL of purified TG2) was assessed by measuring the fluorescence over 1 hour (λ_{ex} 320 nm, λ_{em} 440 nm) (Adamczyk *et al*. 2013). The reaction was initiated with 2 mM $CaCl_2$. Data represents the mean of 2 wells ± SD (n=2) and are representative of two independent experiments (N=2).

The isopeptidase activity assay was used to assess the function of TG2 and ensure it is correctly folded (Fig. 2.5C) (Adamczyk *et al*. 2013). An increase in the relative fluorescence over time is indicative of isopeptidase activity. The cleavage of the isopeptide bond present in the quenched substrate (Abz-APE (γ-cad-Dnp) QEA) releases the Dnp quencher allowing the Abz group to fluoresce when excited at 320

nm. An increase in the relative fluorescence over time was observed for the purified TG2 indicating it is correctly folded and functional. Other samples collected throughout the purification were found to have little/no activity.

2.3.4 Activity of purified WT TG2 and the GTP binding site variants

In order to clarify whether the respective mutations impacted the activity of TG2, the activity of the purified TG2 GTP binding site mutants was assessed using the MDC amine incorporation assay (measures transamidation activity) and the isopeptidase activity assay (Fig. 2.6A and C, work carried out by Rhiannon Griffiths). For all TG2 GTP binding mutants, transamidation activity was impaired (Fig. 2.6A and B and Table 2.3). TG2 K173N, K173L, K173N/F714D and R580A had significantly reduced isopeptidase activity when compared to WT TG2 (Fig. 2.6 C and D and Table 2.3). Of interest, TG2 K173L show mildly impaired activity with regards to transamidation activity but has significantly reduced isopeptidase activity when compared to WT TG2 (Table 2.3). This result suggests TG2 K173L impacts the binding of selected substrates which promote different enzymatic functions.

Table 2.3 Rate of reaction for TG2 GTP binding mutants.

TG2	MDC activity		Isopeptidase activity	
	Rate (RFU/s)	Activity (%)	Rate (RFU/s)	Activity (%)
WT	29.12 ± 0.66	N/a	34.3 ± 0.92	N/a
K173N	12.27 ± 0.34	42.1	16.2 ± 0.37	47.23
K173L	11.74 ± 0.56	40.3	1.02 ± 0.09	2.92
K173N/F174D	2.11 ± 0.08	7.26	3.77 ± 0.10	11.00
R580A	6.42 ± 0.35	22.0	1.16 ± 0.12	3.50

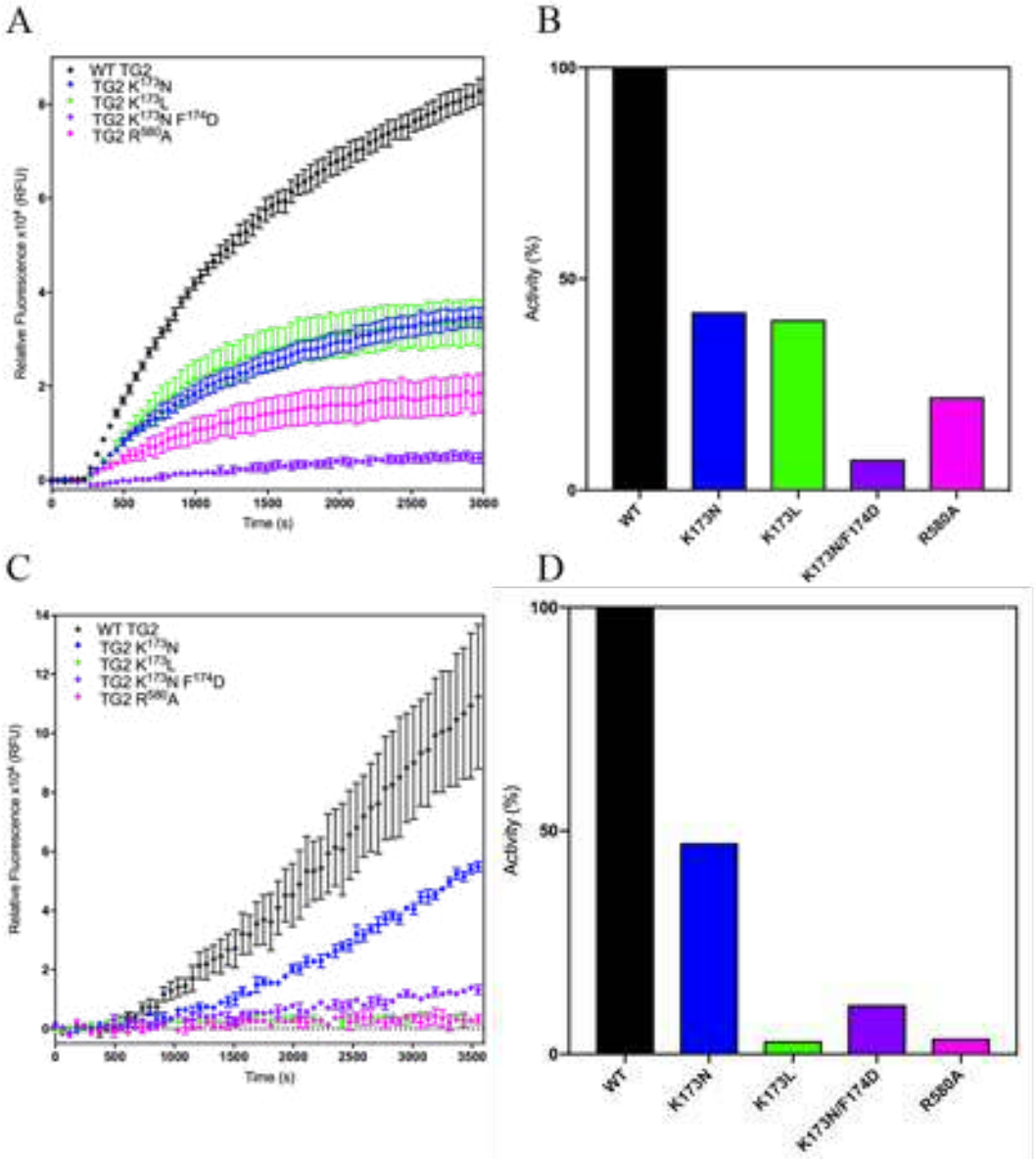


Figure 2.6: Enzymatic activity of the TG2 GTP binding mutants.
The transamidation and isopeptidase activity was measured for the TG2 GTP binding site mutants and compared with WT TG2. Reactions were initiated by addition of 2 mM Ca^{2+}. Data is representative of the mean of 2 wells ± SD (n=2) and representative of two independent experiments (N=2). A: Transamidation activity of TG2 GTP binding site mutants was measured by the incorporation of MDC into N, N, dimethylcasein by TG2 (20 µg/mL). Fluorescence was measured over 1 hour (λ_{ex} 320 nm, λ_{em} 520 nm) B: The rate of transamidation activity was calculated with linear regression and represented as a percentage of WT activity. C: Isopeptidase activity of the GTP binding site mutants (20 µg/mL) showed fluorescence increase following the cleavage of quenched A102 substrate over 1 hour (Adamczyk *et al.* 2013). D: As in B but for isopeptidase activity.

2.3.5 Assessing potential structural changes induced by the introduced mutations using Circular Dichroism spectroscopy

Differences in function observed for the TG2 GTP binding site mutants could be related to secondary changes in structure caused by the mutation rather than the mutation itself. CD (260-190 nm) was utilised to establish whether all the GTP binding mutants were correctly folded. The spectra of all five of the TG2 proteins (WT TG2, K173N, K173L, K173N/F174D, and R580A) indicate similar overall structure (Fig. 2.7). However, the TG2 K173N/F174D spectrum has increased negativity (trouve) at ~190 nm, which could be due to overfilling the 0.2 mm cell, thus increasing the path length of the cell. DichroWeb (CDSSTR, SP175 reference spectra set) was used to calculate the amount of α helices and β strands (Table 2.4) (Johnson 1999; Sreerama and Woody 2000; Whitmore and Wallace 2004, 2008). WT TG2 differs significantly from the TG2 mutants, and variation between TG2 mutants is also observed. The greatest variation is seen in the α-helical content where WT has 9% α helices compared with 15-17% α helices of the TG2 mutants. WT TG2 also had a higher β sheet content (40%) compared to the TG2 GTP binding site mutants (34-35%). When comparing the Dichroweb calculations to TG2 conformations from existing X-ray structures (TG2•GTP, TG2•ATP and open TG2, (Liu *et al.* 2002; Pinkas *et al.* 2007; Han *et al.* 2010; Jang *et al.* 2014)) none exactly match the secondary structure content (i.e. α helical or β strand) predicted by Dichroweb. However WT, K173N and K173L were all reflective of closed conformation i.e. TG2•GTP or TG2•ATP. WT TG2 most closely resembled TG2•ATP with regards to its α helical and β strand content, whilst TG2 K173N and K173L reflected TG2•GTP (Liu *et al.* 2002; Han *et al.* 2010). In terms of α helical content both TG2 K173N/F174D and R580A are similar to the open conformation but β sheet content reflects TG2•GTP (Pinkas *et al.* 2007; Jang *et al.* 2014). This result could represent an intermediate between open and closed conformation in the absence of nucleotide.

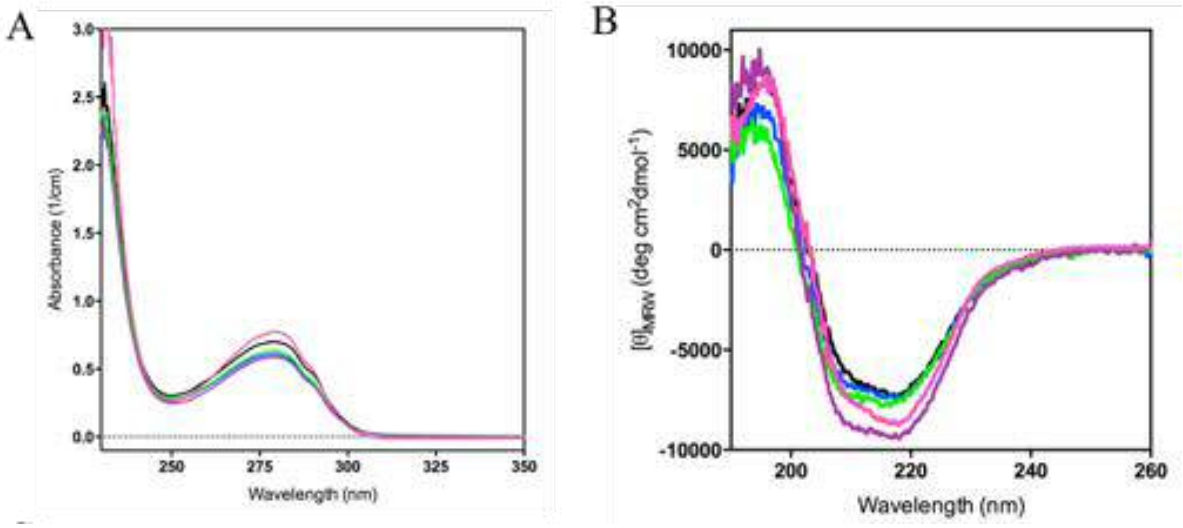


Figure 2.7: Structural characterisation of TG2 GTP binding variants using Circular Dichroism spectroscopy.
Circular dichroism was used assess the secondary structure of the TG2 GTP binding site mutants (K173N (blue), K173L (green), K173N F173D (purple) and R580A (pink)) when compared with WT TG2 (black). A: UV absorbance spectra for WT TG2 and TG2 GTP binding mutants, showing the concentration of TG2 protein used in the CD analysis. B: CD spectra measured for each TG2 variant (0.5 mg/mL) using 0.02 cm and 0.1 cm quartz cells to obtain a spectrum from 190-260 nm (one replicate (n=1)in one independent experiment (N=1).

Table 2.4: Dichroweb secondary structure calculation.

TG2 protein	α helix (%)	β strand (%)	Ref.
WT	9	40	N/a
K173N	15	35	N/a
K173L	15	34	N/a
K173N/F174D	16	35	N/a
R580A	17	35	N/a
WT•GTP	14	36	PDB: 4PYG (Jang *et al.* 2014)
WT•ATP	14	39	PDB: 3LY6 (Han *et al.* 2010)
WT (open)	16	37	PDB: 2Q3Z (Pinkas *et al.* 2007)

2.3.6 Assessing structural changes with chromatography

To further identify changes in the shape and size of the TG2 protein, the chromatograms generated from the ion exchange elution profiles of the TG2 mutants were assessed (Fig. 2.8). When the chromatograms were aligned, a shift (to the right) of the elution profiles was observed for TG2 K173L, K173N/F174D and R580A whilst TG2 K173N showed similar positioning to WT TG2. In addition, TG2 R580A showed a doublet peak rather than a single peak, indicating a high abundance of two individual

species. Size exclusion chromatography was used to further assess any changes in conformation and structure (Fig. 2.9). This analysis showed that all five size exclusion chromatograms overlay suggesting no change in size or shape. Although all samples were centrifuged to removed aggregates, in all cases polymeric structures were observed. WT TG2, TG2 K173N and R580A had a similar shoulder before the main protein peak (monomeric TG2). TG2 K137L and K173N/F174D had a more distinct shoulder. These peaks could represent a higher abundance of the polymeric species compared with WT TG2. Furthermore, TG2 K173N/F174D had an additional broad peak which could indicate a number of different species not present in WT TG2 and the other TG2 GTP binding site mutants.

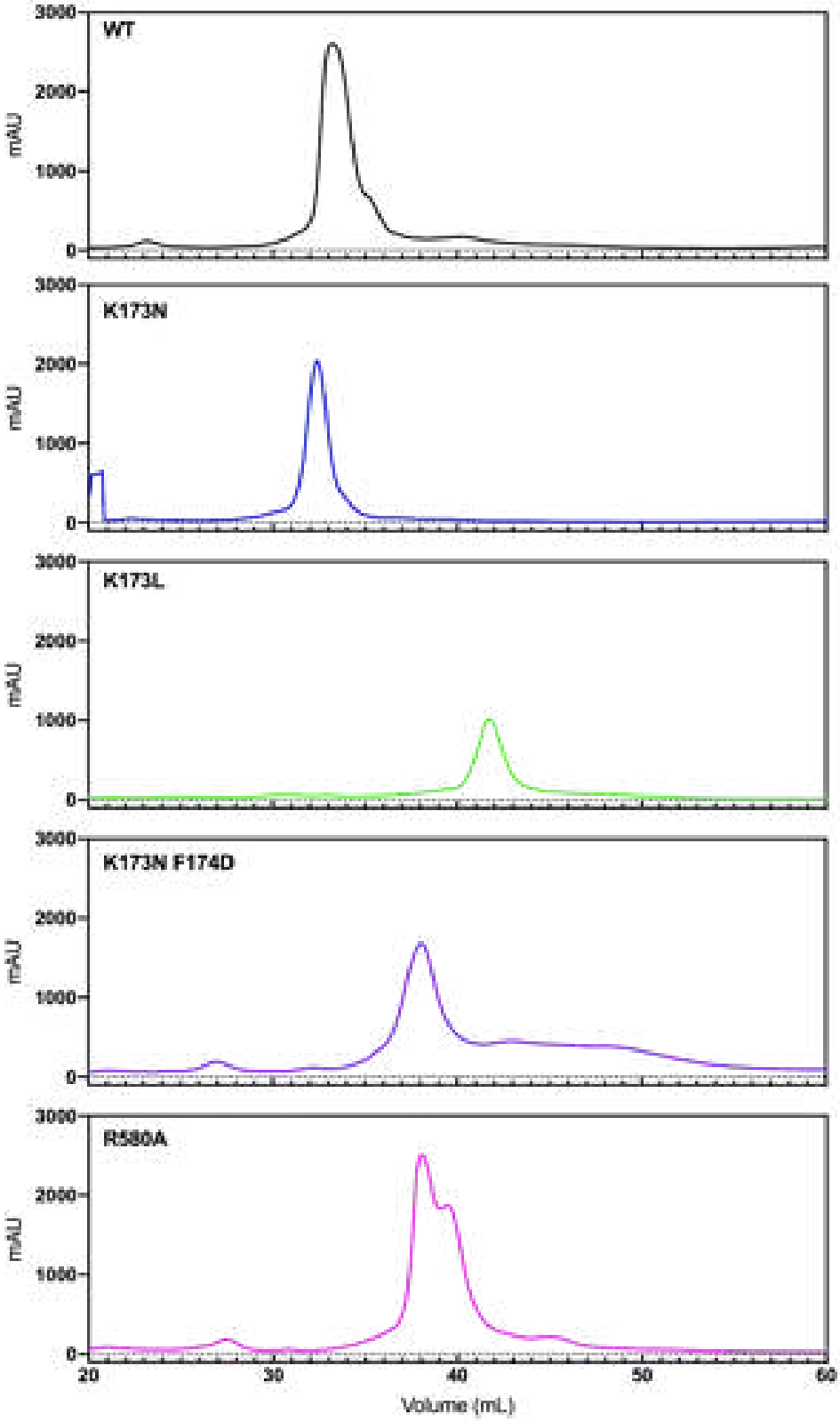


Figure 2.8: Ion exchange chromatogram of TG2 GTP binding mutants.
Ion exchange chromatograms generated during the purification of WT TG2 and the TG2 GTP binding site mutants, using a Resource Q column. Data analysed using Graphpad Prism v.8 and are representative of one replicate (n=1) at least 2 independent experiments (N=2).

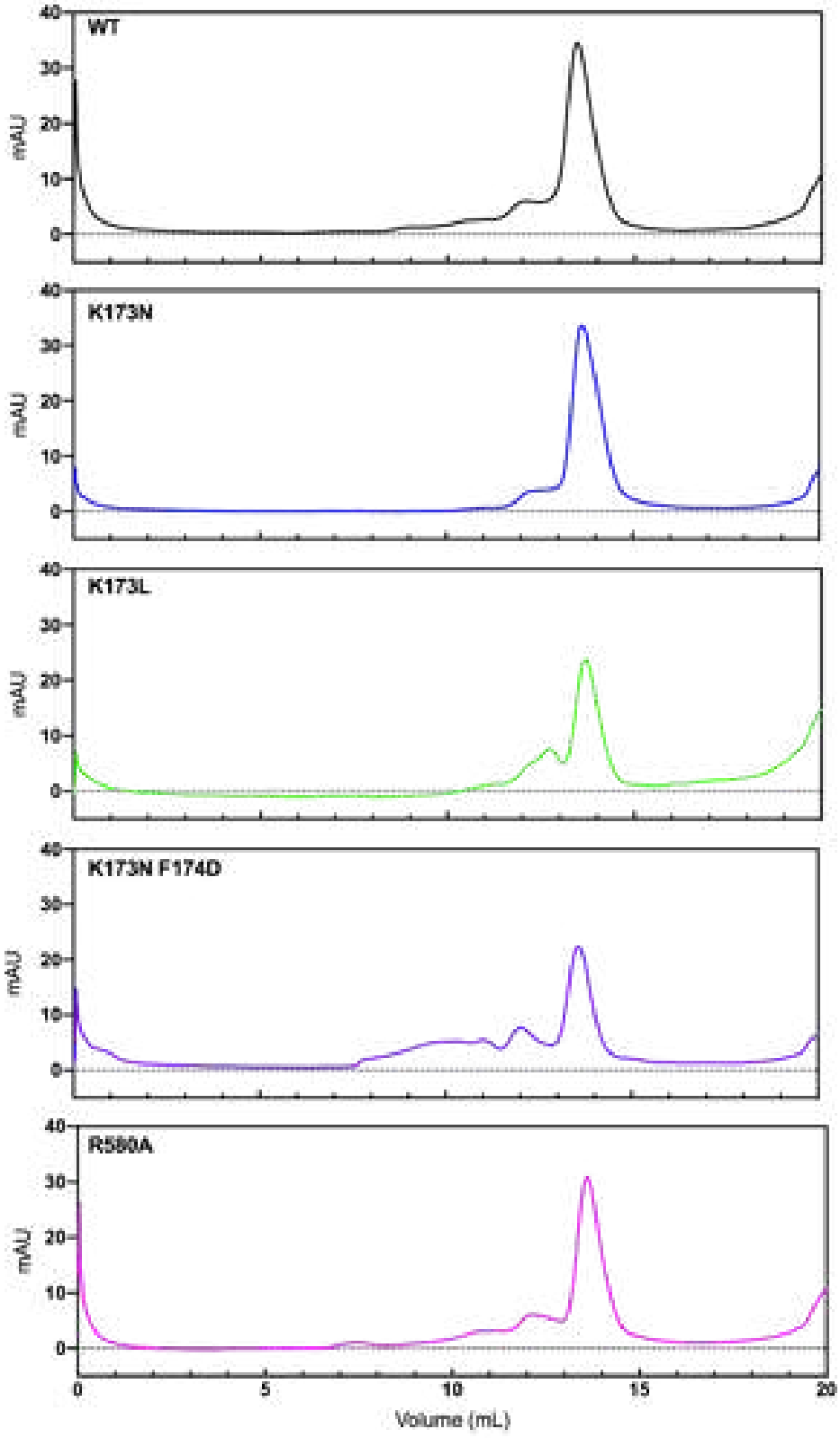


Figure 2.9: Size exclusion chromatogram of TG2 GTP binding mutants.
The size and shape of WT TG2 and the TG2 GTP binding site mutants (0.5 mg/mL) was measured using analytical size exclusion chromatography (Superdex 200 10/300GL column). Data analysed using Graphpad Prism v.8 are representative of one replicate (n=1) in one independent experiment (N=1).

2.3.7 Inhibition of isopeptidase activity of TG2 GTP binding site mutants

Following the structural assessment of the TG2 GTP binding mutants, their GTP binding capabilities were measured. The GTP binding of WT TG2 and the TG2 GTP binding site mutants was first evaluated by the inhibition of TG2 isopeptidase activity. The isopeptidase activity was measured at various concentrations of GTP, 0.25 µM to 600 µM, (Fig. 2.10). This assay showed WT TG2 had an apparent K_D of 22 µM ± 4µM. Both TG2 K173N and K173L had a similar K_D to WT TG2 for GTP (41µM ± 14 µM and 26 µM ± 24 µM, respectively) indicating similar sensitivity to GTP inhibition. This shows these TG2 GTP binding site mutants do not significantly affect GTP binding. In contrast, TG2 K173N/F174D had significantly reduced sensitivity to GTP (K_D of >600 µM), and no reduction in the isopeptidase activity was measured for TG2 R580A at high GTP concentrations.

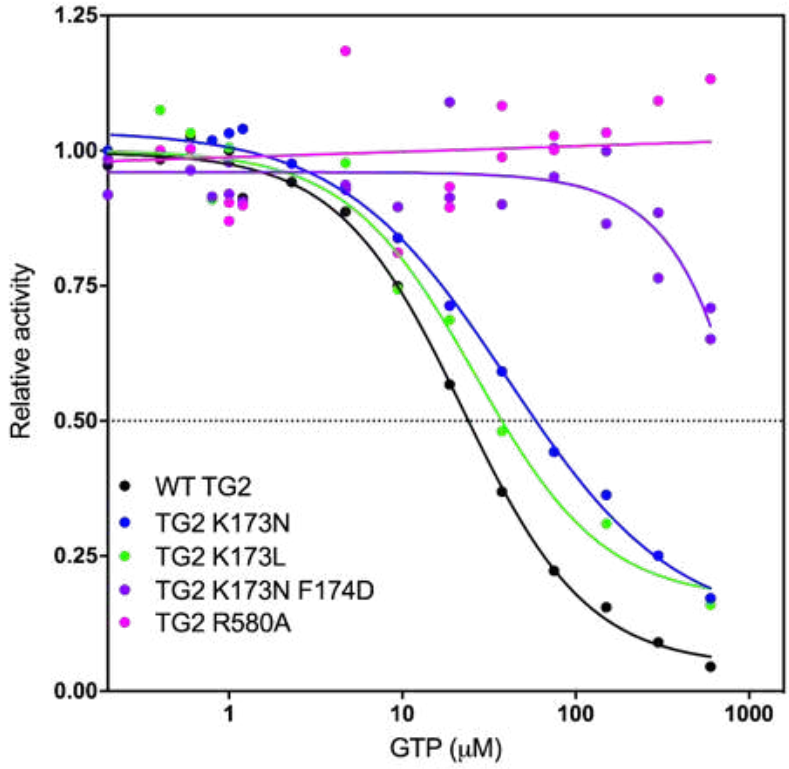


Figure 2.10: Inhibition of isopeptidase activity with GTP.
The isopeptidase activity of the WT TG2 and the TG2 GTP binding site mutants (as in Fig. 2.6) was assessed in the presence of GTP (0 µM to 600 µM). The activity curves were baseline corrected (TG2 with no Ca^{2+} injection). The rate of reaction was measured with linear regression, normalised (between 1 and 0) and fitted to a sigmoidal dose response curve. The apparent K_D was calculated from half maximal

inhibition. The data is representative of the mean of two wells (n=2) and two independent experiments (N=2) (work carried out by Rhiannon Griffiths).

Native PAGE analysis was used to visualise the different conformations of TG2 and the TG2 GTP binding mutants with and without GTP (work carried out by Rhiannon Griffiths, Fig. 2.11). The compact GTP-bound TG2 will migrate faster than the open extended conformation of TG2. In addition, the presence of GTP further increases the negative charge of the protein thus increasing the mobility of the GTP bound moiety. A shift in mobility in the presence of GTP can be observed with WT TG2 (Fig. 2.11, lane WT +/- GTP). A single band is observed without GTP, whereas two species are evident in the presence of GTP (Compare lanes WT (left (-GTP)and right (+GTP)). An equivalent shift in mobility was observed with TG2 K173N and to a lesser degree with TG2 K173L. This suggests TG2 K173N has a similar affinity for GTP to WT TG2 and implies a reduced affinity for TG2 K173L for GTP. In contrast, TG2 K173N/F173D and R580A showed no visible mobility shift with GTP indicating K173N/F174D and R580A are insensitive to GTP consistent with the inhibition of isopeptidase assay. However, TG2 K173N/ F174D showed a minimal shift in mobility when analysed using densitometry (Fig. 2.11B).

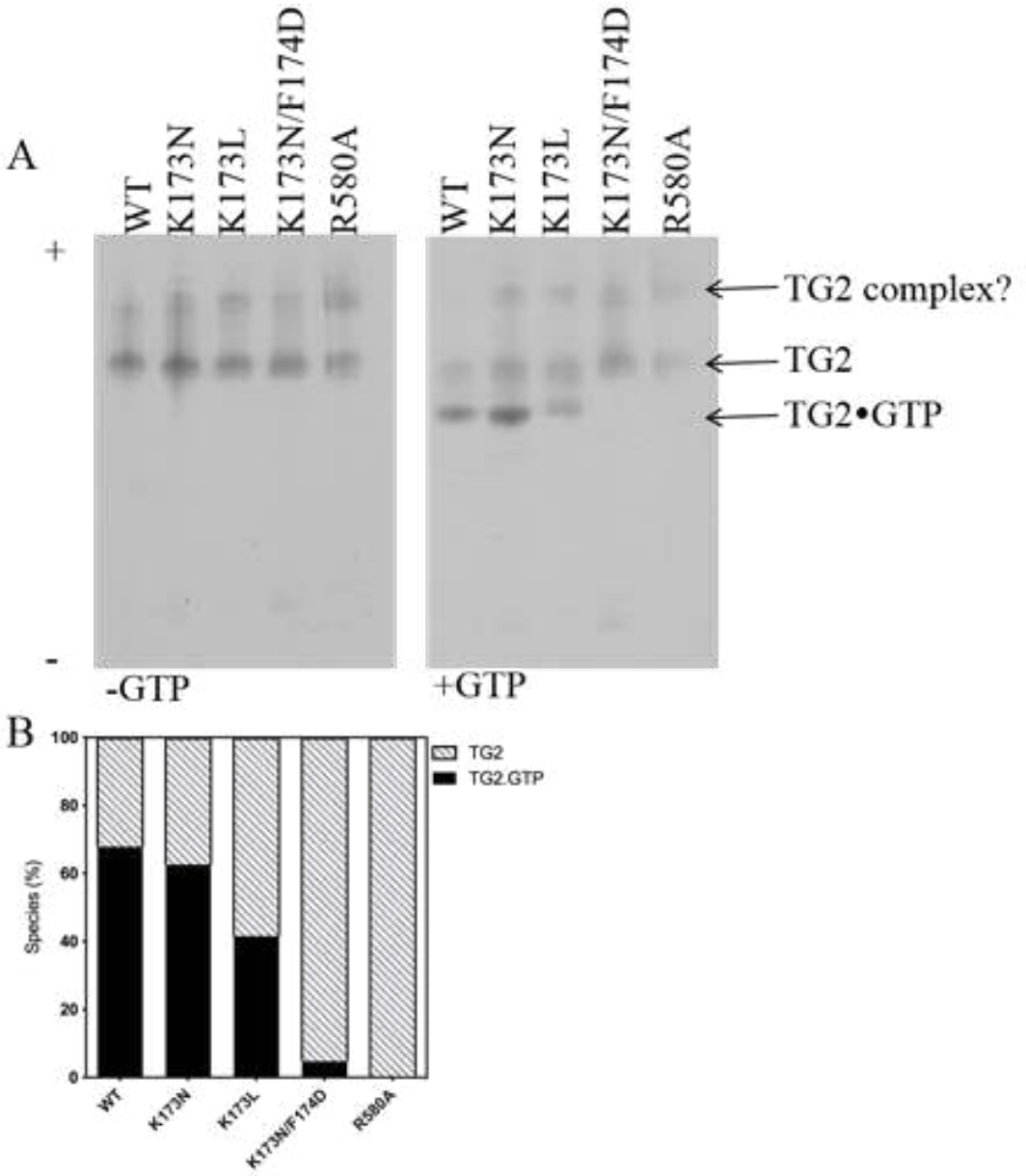


Figure 2.11: Assessing the conformational change of TG2 mutants.
A: Native PAGE analysis of the TG2 GTP binding site mutants. 4 µg of TG2 was run on a 10% Tris-glycine gel under non-reducing conditions. The anode (+) and cathode (-) are indicated. The samples were incubated with 25 µM GTP and the gel was run in in the presence of 25 µM GTP (right) and 0 mM GTP in the sample or chamber (left)(n=1, N=2). Gels were stained using Coomassie R staining. B: Densitometry analysis of the GTP free TG2 and GTP bound TG2. The TG2 complex was not taken into consideration in this analysis (work carried out by Rhiannon Griffiths).

2.3.8. Binding studies of GxP for WT TG2

To identify whether a change in the number of phosphate groups on the guanine moiety impacted the generated CD signal, GTP (Fig. 2.12A), GTPγS (Fig. 2.12B), GDP (Fig. 2.12C) or GMP (Fig. 2.12D) was added to WT TG2 in two concentrations: Low

nucleotide (3-6 μM) and excess nucleotide (~150-370 μM), followed by addition of $CaCl_2$.. In all cases CD spectra indicates the valley becomes more positive between 203-225 nm in relation to increased GxP concentration and marginal differences were observed when comparing the guanine nucleotides. Upon addition of $CaCl_2$, no difference in CD signal is observed for GTP, GDP and GMP. Interestingly, the addition of excess GTPγS then $CaCl_2$ increases the valleys negativity, yielding a minimum comparable to that of low GTPγS. This data could indicate a change in secondary structure under these conditions but would require further investigation.

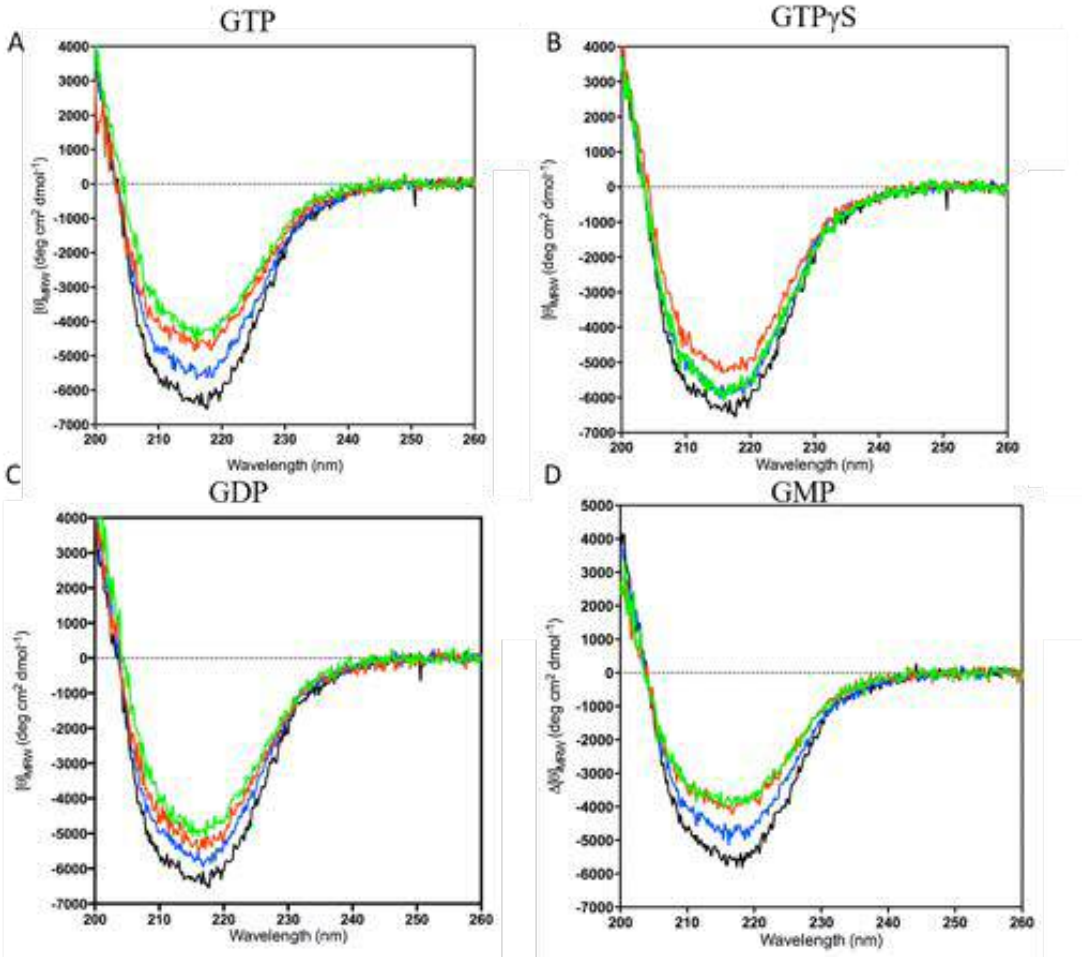


Figure 2.12: Investigating nucleotide binding in TG2 with circular dichroism. Changes in the secondary structure of WT TG2 (0.2 mg/mL) in the presence of guanine nucleotide was investigated. The CD spectra were measured with no guanine nucleotide (black), 1:1 molar ratio of TG2 to nucleotide (blue), excess guanine nucleotide (green) and excess guanine nucleotide plus 5.5 mM Ca^{2+} (red). Data is representative of one replicate (n=1) in one independent experiment (N=1). A: GTP binding to TG2. B: GTPγS binding to TG2. C: GDP binding to TG2. D: GMP binding to TG2.

2.3.9 GTP binding studies for TG2 GTP binding site mutants

In order to assess GTP binding using CD spectroscopy, structural changes were monitored upon titration of GTP (Fig. 2.13). This study shows saturation of GTP binding was not achieved for TG2 K173N/F174D and TG2 R580A over this GTP concentration range.

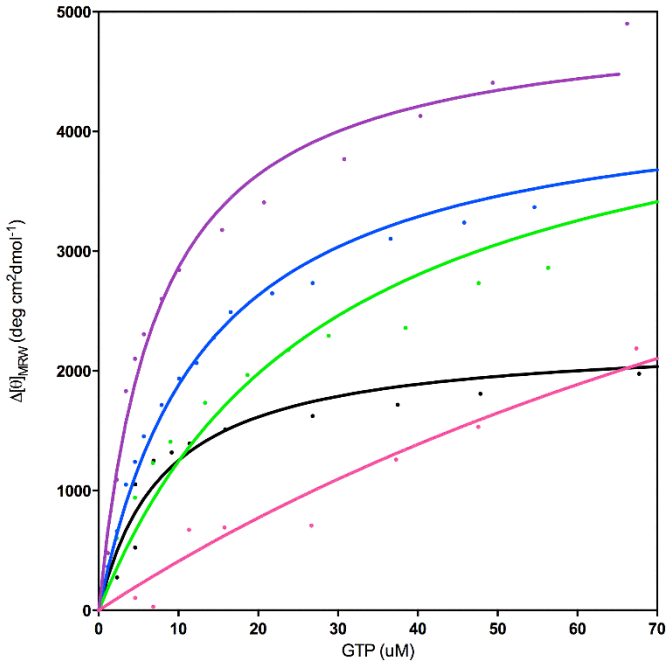


Figure 2.13: GTP binding studies using circular dichroism binding curves. Kinetic analysis of GTP binding to TG2 variants (WT (black), K173N (blue), K173L (green), K173N/F174D (purple) and R580A (pink)) using CD spectroscopy. Measurements were taken at 218 nm for 60 seconds throughout a titration of GTP. Changes in the CD signal were plotted against GTP concentration and binding kinetics were calculated. Binding data and corresponding curve fits derived with non-linear regression with one binding site kinetics. Analysis was carried on Graphpad prism v8 and data is representative of one replicate (n=1) in one independent experiment (N=1).

.

Next, the K_D and B_{max} were calculated (Table 2.5). All the TG2 GTP binding mutants showed measurable binding with the exception of TG2 R580A. TG2 K173N/F174D had the highest calculated affinity whilst TG2 K173N and K173L show a lower apparent affinity for GTP than WT TG2.

Table 2.5: Binding kinetics of GTP using CD spectroscopy.

TG2 protein	K_D (μM)	B_{max} (deg cm^2 dmol^{-1}	r^2
WT	8.1 ± 1.2	2280 ± 83	0.9531
K173N	13 ± 1.2	4380 ± 130	0.9811
K173L	28 ± 4.7	4810 ± 290	0.9621
K173N/F174D	7.4 ± 0.8	4990 ± 171	0.9809
R580A	155 ± 51	6770 ± 1330	0.9677

A spectrum (203-260 nm) was measured before and after the addition of the maximal amount of GTP (Fig. 2.14). This shows the effect of the addition of excess GTP on the CD signal for WT TG2 (Fig. 2.14A) TG2 K173L (Fig. 2.14B), TG2 K173N/F174D (Fig. 2.14C) and TG2 R580A (Fig. 2.14D). TG2 K173N was not included, as a spectrum was not measured after GTP addition. For WT, K173L, K173N/F174D and R580A, a reduction in the negativity of the peak is seen upon addition of GTP. The change upon addition of GTP is extensive and is too big to represent a secondary structure change in TG2. When comparing the α helix and β strand content of the open and closed X-ray crystal structures of TG2, (Table 2.4), no substantial changes were seen, suggesting the change in secondary structure in the CD spectrum is not representative of a change in TG2 (Jang *et al.* 2014).

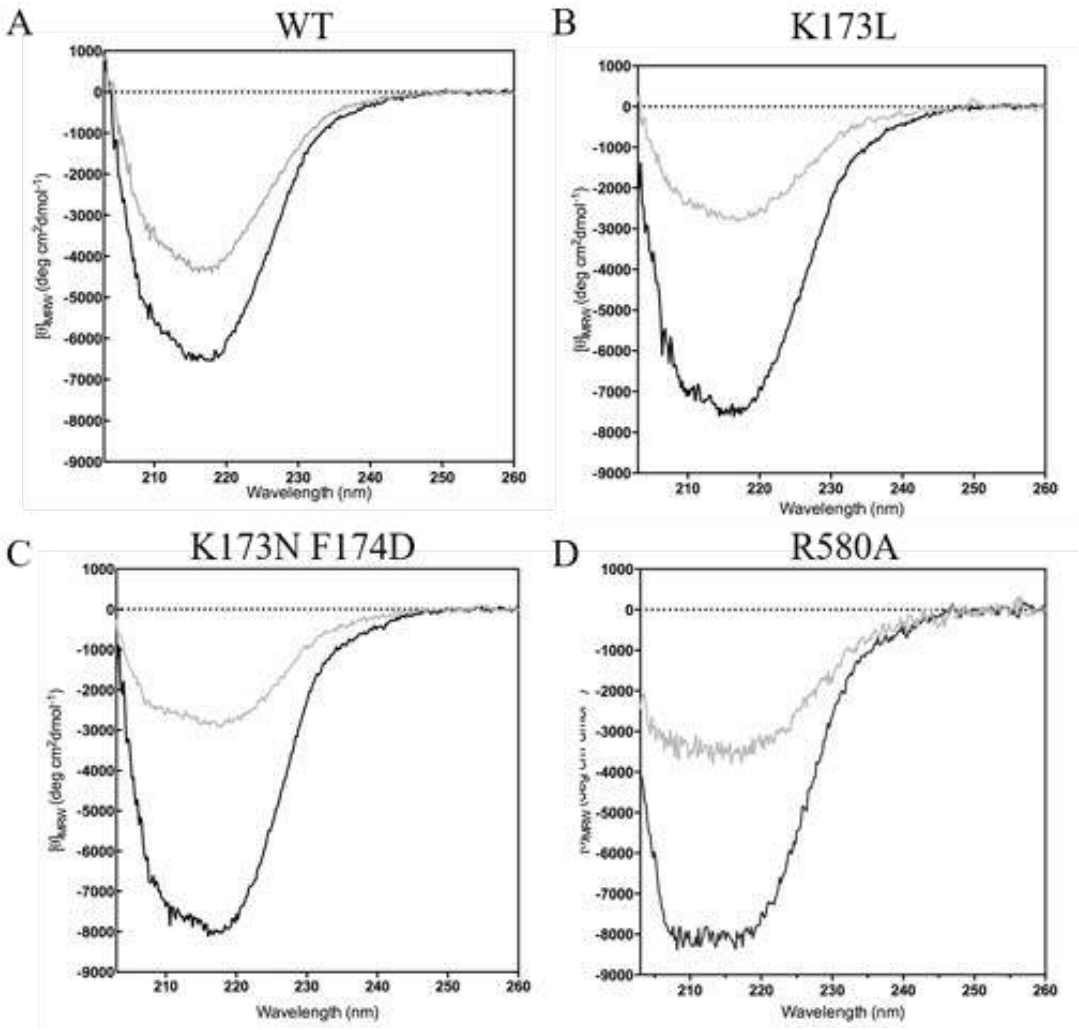


Figure 2.14: GTP binding studies using circular dichroism to assess structural changes.
CD spectra measured before (black) and after the addition of excess GTP (grey) for each TG2 variant (0.5 mg/mL) following GTP titration. Spectra were baseline corrected and calculated contribution of GTP was subtracted. Data is representative of one replicate (n=1) in one independent experiment (N=1). A: WT TG2 with 22x excess GTP. B: TG2 K173L with 44x excess GTP. C: TG2 K173N/F174D with 12x excess GTP. D: TG2 R580A with 60x excess GTP.

Instead the change in signal is likely attributed to an increase in the hydrophobic environment induced by excess GTP. The guanine moiety forms α helical G-quartets when in high concentration which are highly hydrophobic and have their own CD signal (Goncharova *et al.* 2012). A CD spectrum of GTP in buffer alone shows disorder at low concentrations but a signal indicative of structure at higher concentrations (Fig. 2.15). GTP's ability to form its own secondary structure when in sufficient concentration suggests CD is not a suitable method to investigate the interaction of GTP with TG2.

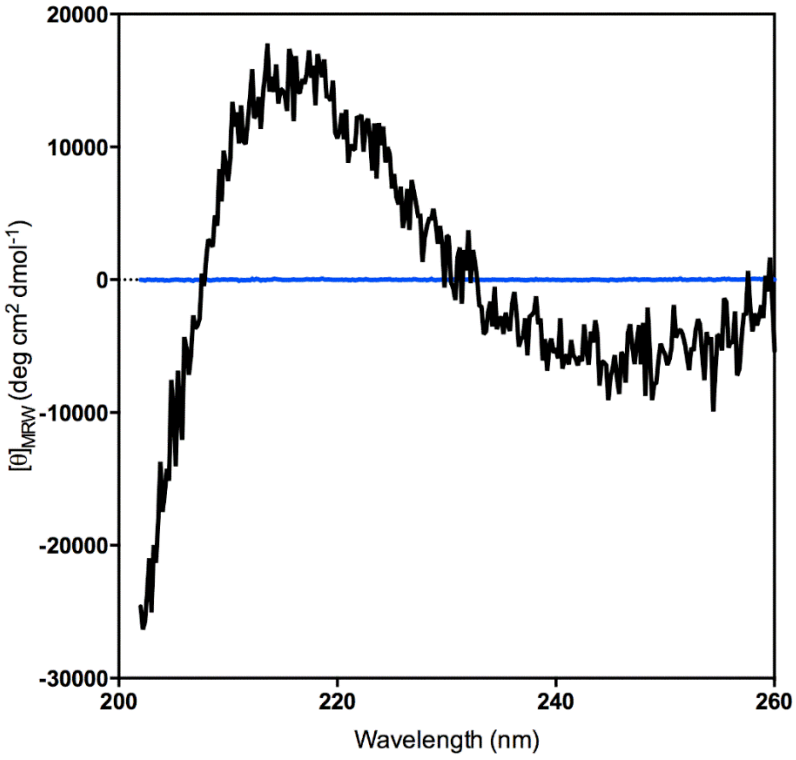
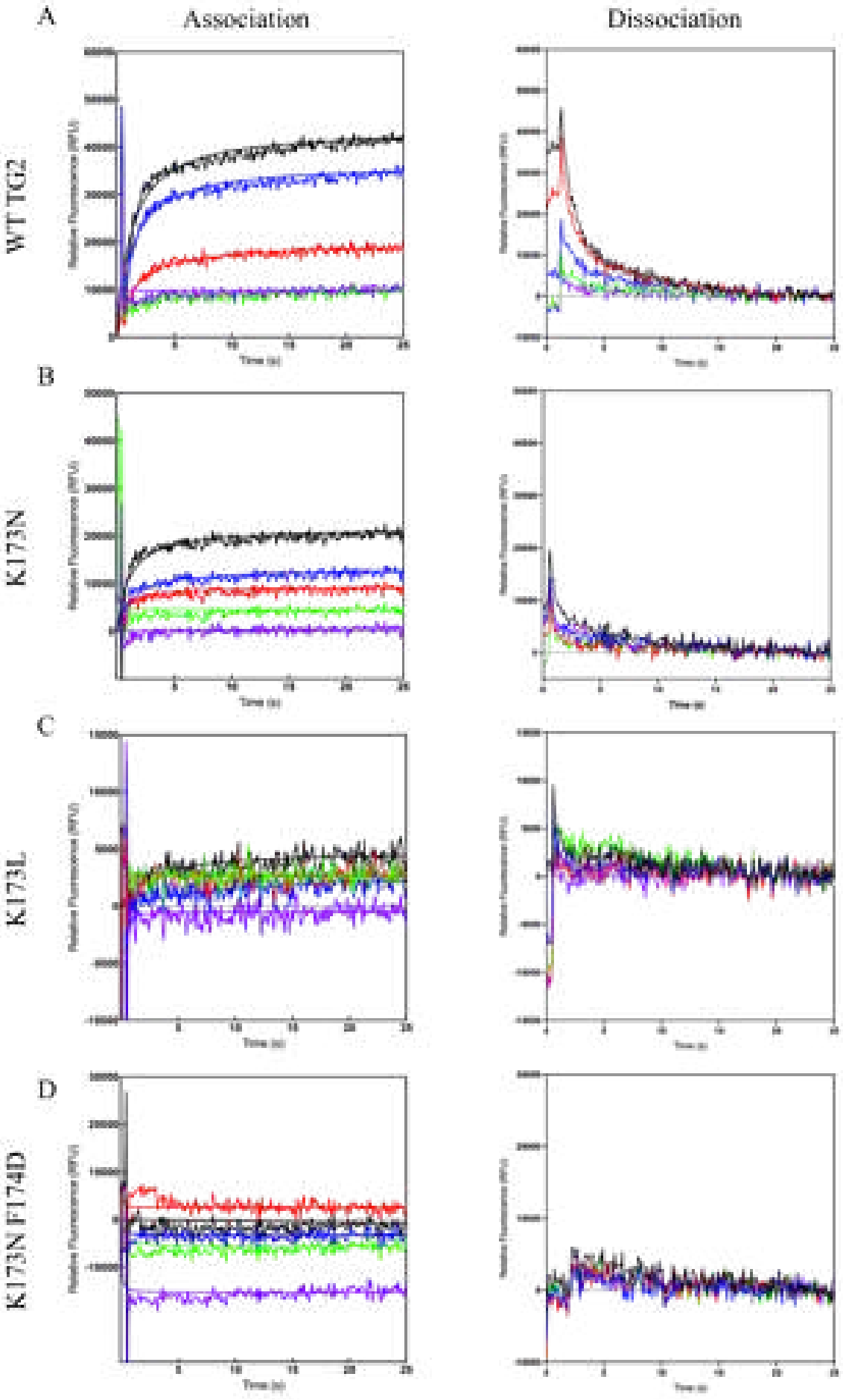


Figure 2.15: GTP secondary structure at high and low nucleotide concentration. A comparison between low GTP concentration (6 µM, blue) and high GTP concentration (370 µM, black). Data is representative of one replicate (n=1) in three independent experiment (N=3).

2.3.10 Measuring the direct binding of GTP binding with BODIPY GTP/GTPγS

Binding of GTP by WT TG2 and the GTP binding mutants was assessed using both BODIPY GTP and GTPγS. Firstly, GTP binding (association and dissociation) was assessed using different concentrations of TG2 (40, 20, 10, 5 and 2.5 µg/mL) (Fig. 2.16). WT TG2 showed a concentration-dependent GTPγS binding (Fig. 2.16A, Association). This is similarly observed for TG2 K173N (Fig. 2.16B, Association). TG2 R580A also showed GTPγS binding, however the greatest binding at 20 µg/mL (Fig. 2.16E, Association) This is unexpected as R580A is not predicted to bind GTP (Begg *et al.* 2006b). No GTPγS association was observed for TG2 K173L and K173N/F174D (Fig. 2.16C and D, Association), suggesting a polar residue (lysine or asparagine) and F174 is important for GTP stabilisation within the nucleotide binding pocket of TG2. With regards to BODIPY GTPγS dissociation, full dissociation was measured with WT TG2 (Fig. 2.16A, Dissociation). Dissociation of GTPγS from K173N (Fig. 2.16B, Dissociation) was comparable to WT but minimal dissociation

was measured for K173L and R580A (Fig. 2.16C and E, Dissociation). As expected, no dissociation was measured for K173N/F174D (Fig. 2.16D, Dissociation).


A
Association
Dissociation
WT TG2
B
K173N
C
K173L
D
K173N F174D

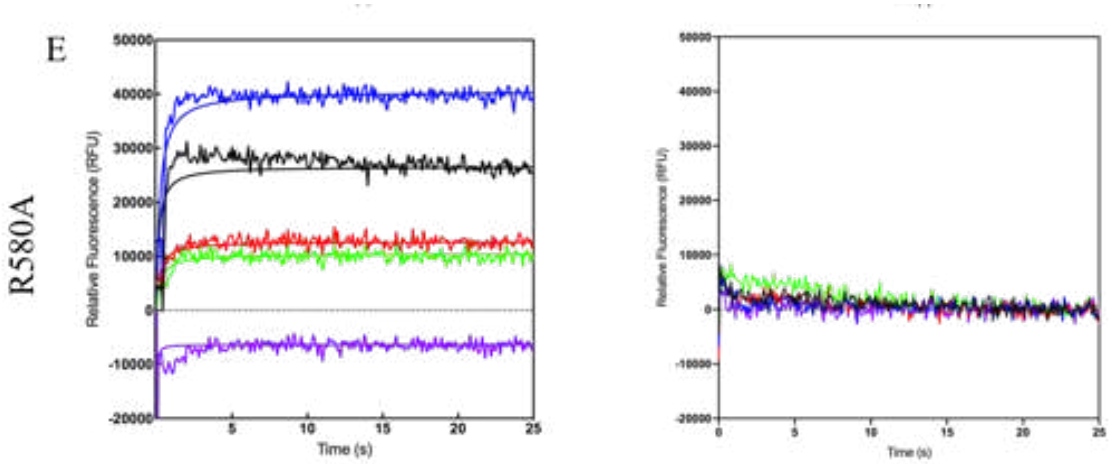


Figure 2.16: Measuring GTP binding with different concentrations of TG2.
Binding of GTP by the TG2 variants was investigated using varying concentrations of TG2 protein (40-2.5 µg/mL). BODIPY-GTPγS (50 nM) association was measured (λ_{ex} 480 nm and λ_{em} 520 nm) with the respective TG2 variant and concentration (40 µg/mL (black), 20 µg/mL (blue), 10 µg/mL (red), 5 µg/mL (green) and 2.5 µg/mL (purple) over 150 seconds. Dissociation was initiated using an excess of unlabelled GTPγS excess (100 µM) to displace BODIPY-GTPγS. All measurements were blank and baseline corrected and are representative of one replicate (n=1) and two independent experiments (N=2). A: Association and dissociation of GTPγS for WT TG2. B: Association and dissociation of GTPγS for K173N. C: Association and dissociation of GTPγS for K173L. D: Association and dissociation of GTPγS for K173N/F174D. E: Association and dissociation of GTPγS for R580A.

showed GTPγS dissociation comparable to FAMBA (Fig 2.17C, purple, pink and red, respectively). This indicates GTPγS no dissociation was measured on this occasion.

In contrast, association with BODIPY GTP was greatest for WT TG2 (Fig. 2.17A, black). Both TG2 K173N and K173L showed reduced binding when compared to WT TG2 (Fig. 2.17A, blue and green). No association was seen for K173N/F174D and R580A (Fig. 2.17A, purple and pink). As with BODIPY GTPγS full dissociation was measured for WT TG2 and K173N. (Fig. 2.17B, black and blue). GTP dissociated faster from TG2 K173L, similar to as seen with BODIPY GTPγS (Fig. 2.17B, green). TG2 K173N F173D and R580A showed no dissociation as no BODIPY GTP binding was measured.

BODIPY GTP and GTPγS shows a different binding affinity with regards to the WT TG2 and the TG2 GTP binding mutants. GTP binding for the WT and the TG2 mutants is similar to that published in the literature whereby GTP binding is impaired for K173L and abolished for R580A (Iismaa *et al.* 2000; Begg *et al.* 2006a; Begg *et al.* 2006b). Alternatively, the TG2 mutants, K173N and K173L, showed higher GTPγS binding when compared with WT TG2. In addition, GTPγS binding was measured for TG2 K173N/F174D and R580A, despite no measurable GTP binding. This suggests GTP and GTPγS have different binding modes within the nucleotide binding pocket promoted by the various TG2 GTP binding mutants.

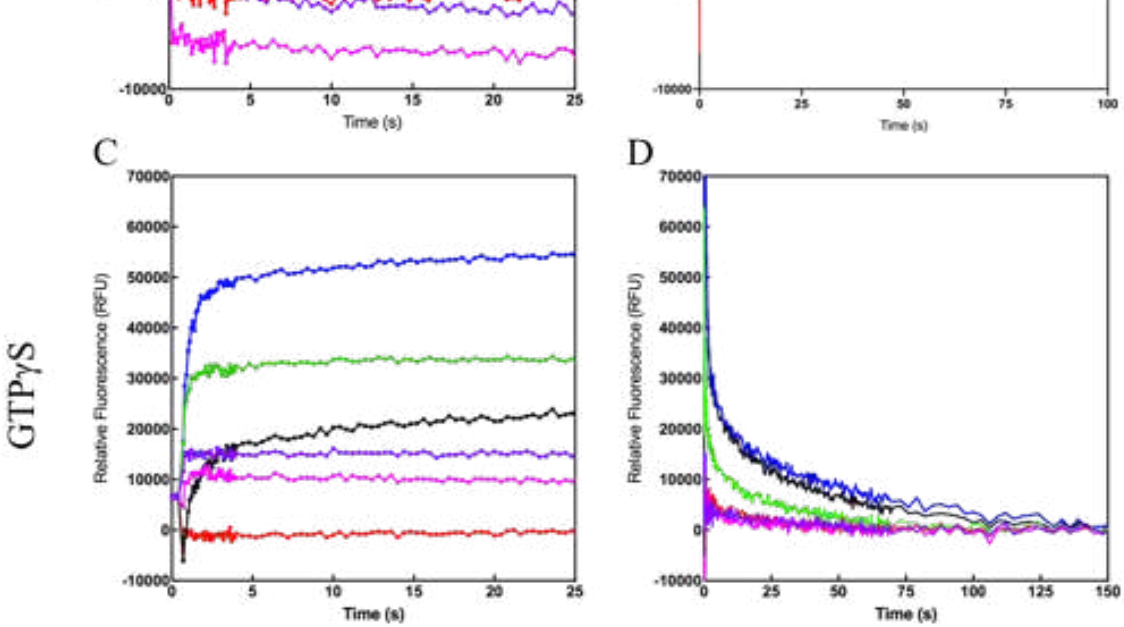


Figure 2.17 BODIPY GTP binding to TG2 GTP binding site mutants
Fluorescently labelled GTP (BODIPY GTP/ GTPγS) was used to compare GTP binding to WT TG2 and the GTP binding site variants. Association was measured with 50 nM BODIPY-GTP/ GTPγS and the respective TG2 (20 µg/mL, WT (black), K173N (blue), K173L (green), K173N/F174D (purple) and R580A (pink)) proteins. FXIIIA (red) was used as a negative control for BODIPY GTP binding. Data is representative of one replicate (n=1) in three independent experiments (N=3). An excess of unlabelled GTP/ GTPγS (100 µM) displaced the BODIPY nucleotide. A and B: Association and dissociation of BODIPY GTP. C and D: Association and dissociation of BODIPY GTPγS.

2.3.11 Docking of nucleotides into TG3

As differences were seen in the binding of GTP and GTPγS between WT TG2 and the TG2 binding site variants, molecular docking studies implementing Glide were used to investigate the different nucleotide binding modes within the nucleotide pocket. WT TG3 where no nucleotide was bound and the pocket differed (PDB: 1L9N (Ahvazi *et*

al. 2003)) was first used as a proof of principle to dock ATP, GTPγS and GTP (Fig. 2.18).

Table 2.6: Conformation prevalence of docked nucleotides.

TG	PDB	Conformation	Percentage of ATP conformations predicted	Percentage of GTP conformations predicted	Percentage of GTPγS conformations predicted
TG3	1L9N	i	20	10	20
		ii	10	N/a	40
		iii	20		30
		iv	30		10
		v	10		N/a
		vi	10		
	1RLE	i	20	60	N/a
		ii	40	10	
		iii	40	20	
		iv	N/a	10	
TG2	4PYG	i	N/a	N/a	10
		ii			10
		ii			80

When ATP was docked into TG3 (PDB: 1L9N (Ahvazi *et al.* 2003)), six different conformations of ATP (cyan) were predicted (Fig. 2.18A and Table 2.6). The docking of GTPγS predicted four different nucleotide conformations (Fig. 2.18B and Table 2.6). In contrast, GTP docking into TG3 only predicted one position for the nucleotide (Fig. 2.18C and Table 2.6). The residues predicted to stabilise nucleotide within the nucleotide binding pocket are N168, R169, K485, K487, V488, L492, R587, I590 and D592. The interactions within the pocket for the respective nucleotides are detailed in Table 2.7.

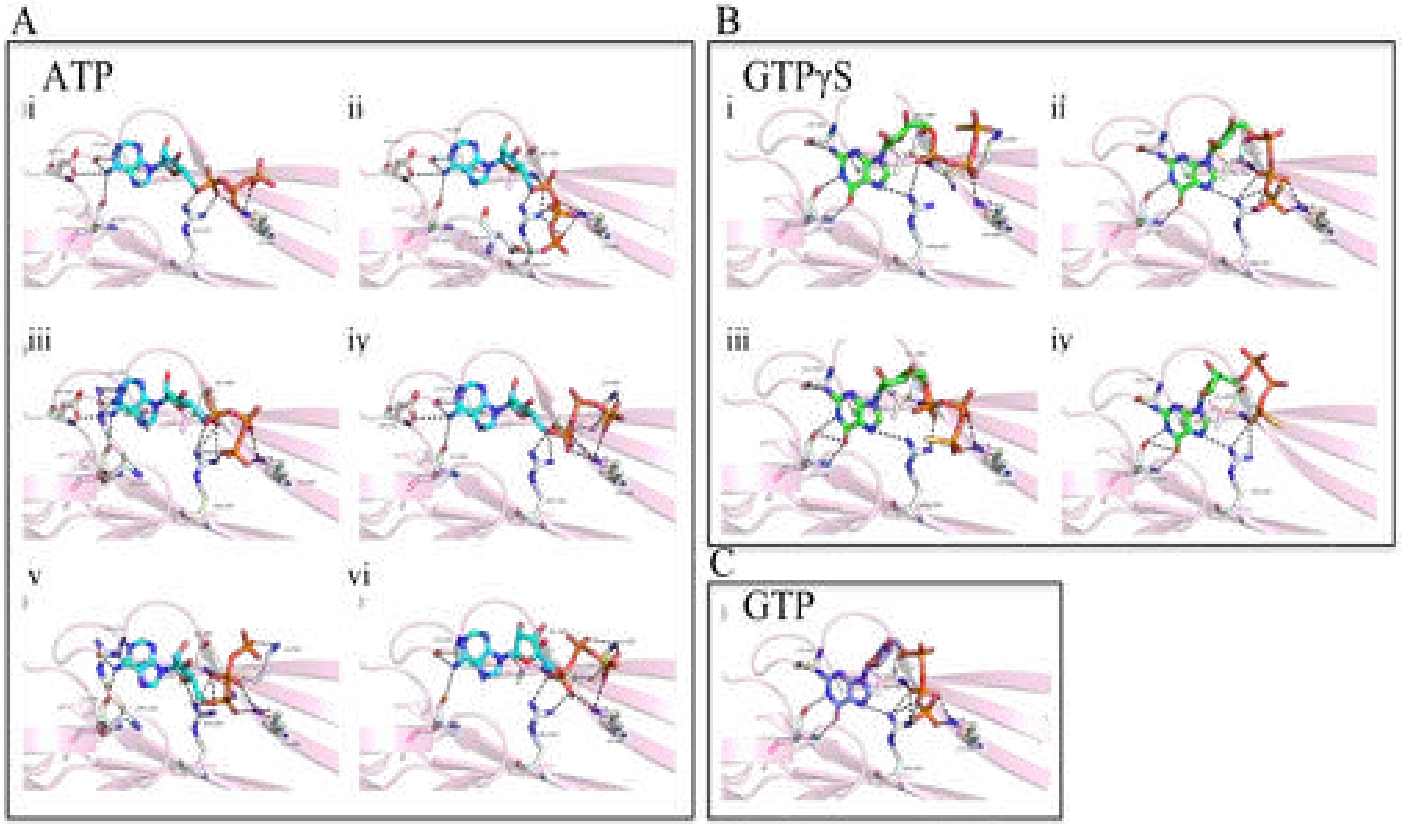


Figure 2.18: Nucleotide docking into Ca^{2+}-bound TG3. The nucleotides ATP, GTPγS and GTP were molecularly docked into a TG3 structure (PDB: 1L9N) which contains no nucleotide. TG3 is shown in cartoon representation (light pink), nucleotide is shown in stick representation and polar contacts with interacting residues (white), nitrogen (blue) and oxygen (red) are highlighted with black dotted lines. A: Molecularly docked ATP (cyan) generated six conformations (i-vi). B: Molecularly docked GTPγS (green) generated 4 conformations. C: GTP (purple) was molecularly docked and generated 1 conformation.

Table 2.7: Interactions within the TG3 nucleotide binding pocket (PDB: 1L9N).

Residue	Position	ATP						GTPγS				GTP
		1	2	3	4	5	6	1	2	3	4	1
R169	Guanine (NH)	x	x	1	x	x	x	x	x	x	x	x
L492	Guanine (NH)	x	x	x	x	1	1	x	x	x	x	x
L492	Guanine (NH3)	1	1	1	1	1	1	1	1	1	1	1
R587	Guanine(NH)	x	x	x	x	x	x	1	1	1	1	1
I590	Guanine (NH3)	1	1	1	1	1	1	x	x	x	x	x
I590	Guanine (NH)	x	x	x	x	x	x	1	1	1	1	1
I590	Guanine (OH)	x	x	x	x	x	x	1	1	2	2	1
D592	Guanine (NH3)	1	1	1	1	x	x	x	x	x	x	x
R587	Ribose	x	x	x	x	1	x	x	x	x	x	x
K485	α-phosphate	1	x	x	2	1	1	x	x	x	x	x
V488	α-phosphate	x	1	1	x	1	x	1	x	1	x	x
R587	α-phosphate	2	2	2	2	1	1	1	x	x	x	x
K485	β-phosphate	1	1	1	x	1	x	1	1	1	x	1
K487	β-phosphate	x	x	x	x	x	1	x	x	x	x	x
V488	β-phosphate	x	x	x	x	x	x	x	1	x	x	x
R587	β-phosphate	x	1	x	x	2	x	x	x	x	x	2
N168	γ-phosphate	x	1	x	x	x	x	x	x	x	x	x
K485	γ-phosphate	1	1	1	1	x	1	x	1	1	x	1
K487	γ-phosphate	x	x	x	1	1	1	1	x	x	x	x
V488	γ-phosphate	x	x	x	x	x	x	x	x	x	1	x
R587	γ-phosphate	x	1	2	x	x	x	x	x	x	2	3

To investigate the accuracy of Glide docking, ATP and GTP were docked into TG3 containing GTPγS as the template (PDB: 1RLE, Ahvazi *et al.* 2004b) (Fig. 2.19). Using this template, the nucleotide, GTPγS could be replaced by another nucleotide and lead to different docking outcomes. The docking of ATP into TG3 showed three different conformations in the GTPγS TG3 structure (Fig. 2.19A and Table 2.6). GTP was also docked into TG3 and gave rise to four different docking outcomes (Fig. 2.19B and Table 2.6). Additional images were captured for GTPγS bound TG3 (PDB: 1RLE, Ahvazi *et al.* 2004b) and GDP bound TG3 (PDB: 1VJJ, Ahvazi *et al.* 2004b) structures for comparison (Fig. 2.19C). The residues involved in the stabilisation of nucleotides within the nucleotide binding pocket were N168, R169, K485, K487, V488, L492, R587 and I590. These residues, with the exception of D592 are the same as those predicted for TG3 with no nucleotide bound (PDB:1L9N (Ahvazi *et al.* 2003)). Interactions involved in the stabilisation of the respective nucleotide are detailed in Table 2.8.

Upon comparison of the docking results, Ca^{2+} bound TG3 (PDB: 1L9N (Ahvazi *et al.* 2003)) was predicted to have the highest affinity for ATP (11 interactions) and GTP (11 interactions) compared with GTPγS (8 interactions) (Table 2.7). GTP and ATP may have similar affinity, however the α-phosphate of GTP is not stabilised, thus impacting affinity. In contrast, TG3•GTPγS (PDB: 1RLE, Ahvazi *et al.* 2004b) is predicted to have lowest affinity for ATP (8 interactions) compared to docked GTP (14 interactions) and the GTPγS (13 interactions) and GDP (13 interactions) bound structures, as it had the least number of interactions (Table 2.8).

The number of nucleotide conformations predicted from docking varied depending on the nucleotide or protein structure used. There is higher variation in the docking results obtained with Ca^{2+} bound TG3 (PDB: 1L9N (Ahvazi *et al.* 2003)) than with TG3•GTPγS (PDB: 1RLE (Ahvazi *et al.* 2004)). This is perhaps due to there being no nucleotide present in the binding site of Ca^{2+} bound TG3 (PDB:1L9N (Ahvazi *et al.* 2003)) on which to base nucleotide binding and/or the binding pocket not being in a nucleotide receptive conformation due to bound Ca^{2+} (Ahvazi *et al.* 2003; Ahvazi *et al.* 2004). Consequently, the docking of nucleotides was carried out using structures co-crystallised with nucleotides to enable more accurate predictions of nucleotide binding.

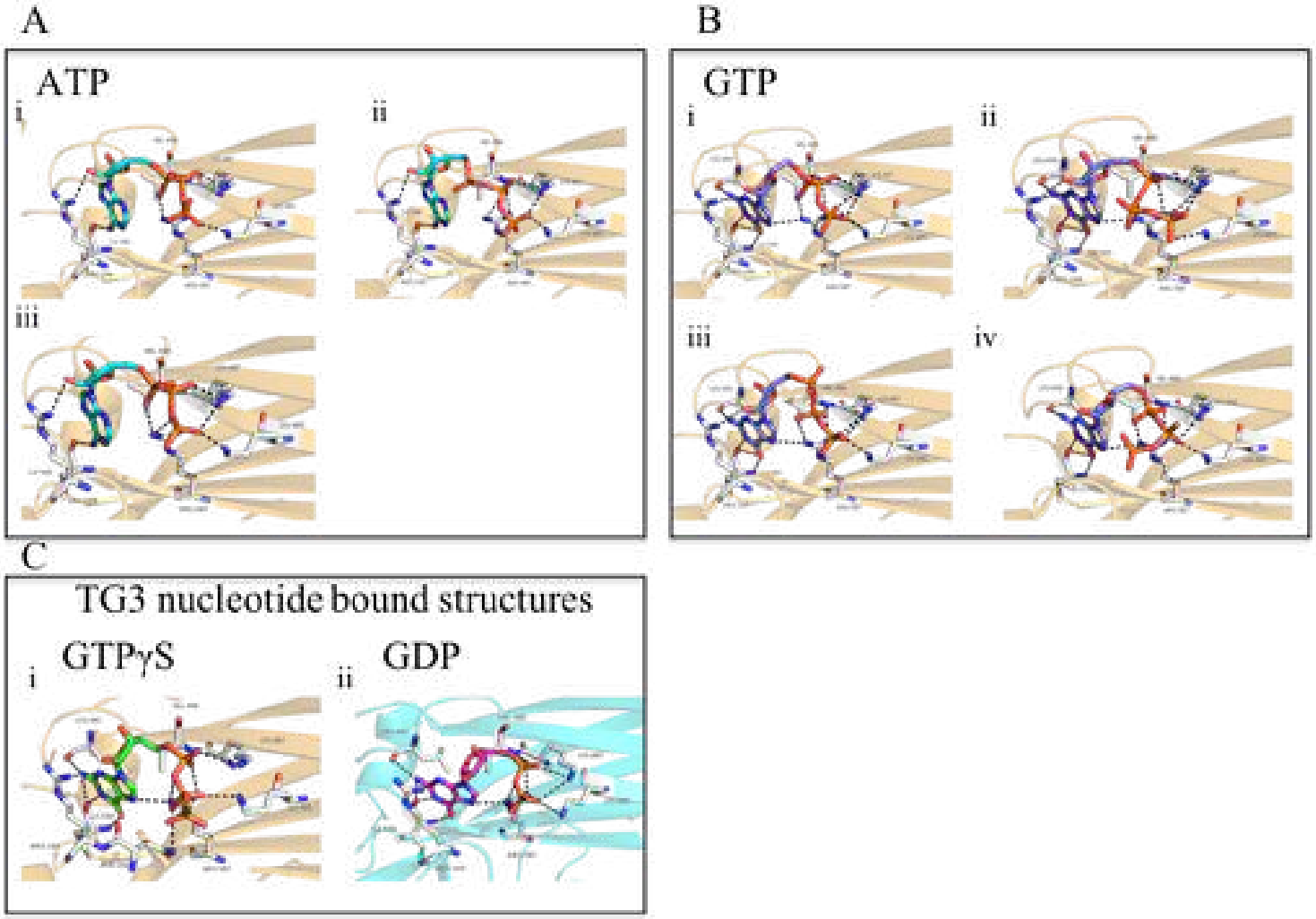


Figure 2.19: Nucleotide docking into GTPγS bound TG3.
The nucleotides ATP and GTP were molecularly docked into a TG3 structure (PDB: 1RLE). TG3 is shown in cartoon representation (dark orange), nucleotide is shown in stick representation and polar contact residues (white), nitrogen (blue) and oxygen (red) are highlighted (black dotted lines) A: Docked ATP (cyan) conformations (i-iii). B: Docked GTP (purple) conformations (i-iv). C: TG3 crystallised with nucleotide i: GTPγS-TG3 (green, PDB: 1RLE) ii: GDP-TG3 (magenta, PDB: 1VJJ).

Table 2.8: Interactions within the TG3 nucleotide binding pocket (PDB: 1RLE).

Residue	Position	ATP			GTP				GTPγS	GDP
		1	2	3	1	2	3	4	1	1
R169	Guanine (NH)	x	x	x	1	1	x	x	1	1
L492	Guanine (NH3)	x	x	x	1	1	1	1	1	1
R587	Guanine (NH)	x	x	x	1	x	1	1	1	1
I590	Guanine (NH)	x	x	x	1	1	1	1	1	1
I590	Guanine (NH3)	1	1	1	x	x	x	x	1	1
I590	Guanine (OH)	x	x	x	2	2	2	2	1	1
R169	Ribose	1	1	1	x	1	1	x	x	x
K485	α-phosphate	x	x	x	x	x	x	1	1	x
V488	α-phosphate	x	x	1	1	1	1	1	1	1
K487	α-phosphate	x	x	x	x	1	x	x	1	1
R587	α-phosphate	x	x	1	1	1	2	2	1	1
K487	β-phosphate	1	1	1	1	x	x	x	x	1
V488	β-phosphate	x	1	x	x	x	x	x	x	x
R587	β-phosphate	x	1	x	1	2	x	x	2	3
N168	γ-phosphate	x	x	x	x	x	x	x	1	x
K485	γ-phosphate	1	1	1	2	1	1	x	x	x
K487	γ-phosphate	1	x	1	2	1	2	1	x	x
R587	γ-phosphate	1	x	1	x	1	x	1	x	x

2.3.12 Docking GTPγS into TG2

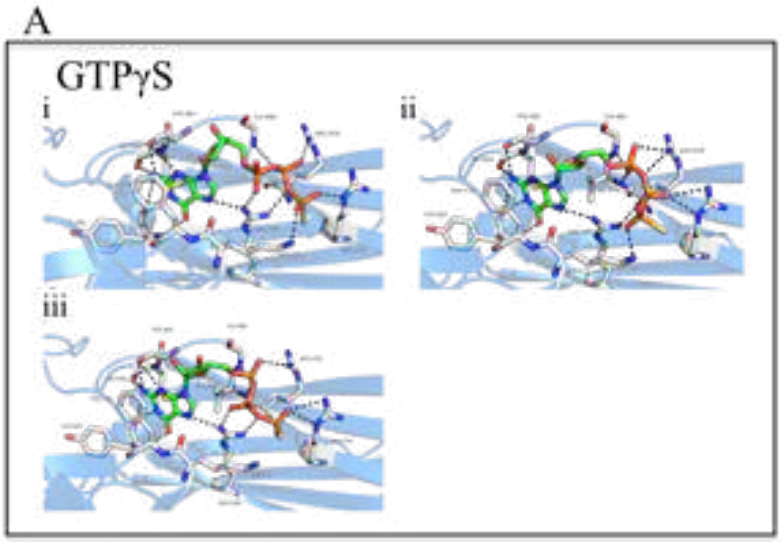
There is currently no TG2 X-ray crystal structure with bound GTPγS so GTPγS was docked into TG2 using GTP-bound TG2 (PDB: 4PYG (Jang *et al.* 2014)) (Fig. 2.20A). From the generated models three conformations were apparent (Fig. 2.20A and Table 2.9). Conformations 1 and 2 both showed interactions between K173, R476 and R580 and the γ-phosphate of GTPγS (Fig. 2.20A (i) and (ii) and Table 2.6 and 2.9). Conformation 3 was observed in 80% of the models and is similar to GTP•TG2 (PDB: 4PYG (Jang *et al.* 2014)) (Fig. 2.20A (iii) and Table 2.6). In this conformation the γ phosphate of GTP makes contact with R476 and R478 but not with K173, and R580 interacts with the α and β-phosphates (Table 2.9). These conformations may represent two different states in the hydrolysis mechanism, the first being the most energetically favourable and the resting state (conformation 3, Fig. 2.20A (iii)). The second conformation (Fig. 2.20A i and ii) only occurs with GTPγS bound. As GTPγS is non-hydrolysable the second conformation may only be captured as GTPγS is 'trapped'.

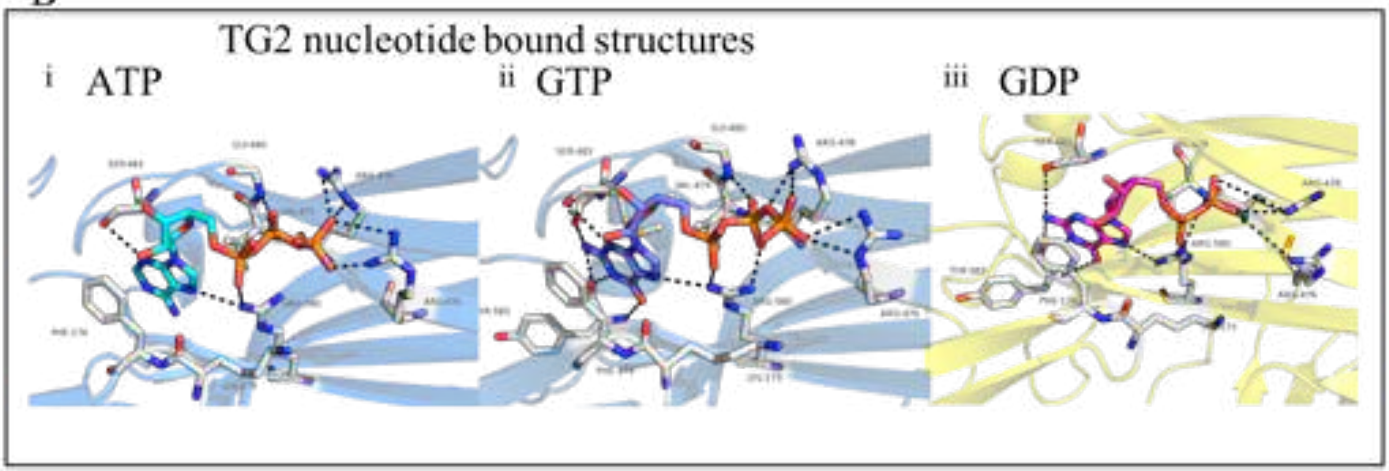
The interactions in the different models and structures are detailed in Table 2.9. This table shows the interactions are the same between GTPγS and GTP (14 interactions). From these results TG2 is predicted to have marginally reduced affinity for GDP (10 interactions, PDB: 1KV3 (Liu *et al.* 2002)), a prediction that has been confirmed experimentally (Achyuthan and Greenberg 1987). GTP-bound and ATP-bound TG2 show different hydrogen bonding patterns (Fig. 2.20B, PDB: 4PYG and 3LY6 (Han *et al.* 2010; Jang *et al.* 2014)). When ATP is present in the nucleotide pocket the hydrogen bonding is maintained between the γ-phosphate and R476 and R478, when compared with GTP. The R580 TG2 mutant interacts with the adenine moiety and the α-phosphate but does not make an interaction with the β-phosphate with ATP (Fig, 2.20B (i)). Numerous polar contacts are lost involving Y583, M483 and L582 when the guanine ring is substituted with the adenine ring of ATP and could result in reduced affinity. As a result, ATP (8 interactions) is predicted to have reduced affinity when compared to GTP and this has been experimentally confirmed (Achyuthan and Greenberg 1987; Iismaa *et al.* 1997).

Table 2.9: Interactions within the TG2 nucleotide binding pocket.

Residue	Position	GTPγS			ATP	GTP	GDP
		1	2	3	1	1	1
S482	Guanine (NH)	1	x	1	x	1	x
S482	Guanine (NH3)	1	1	1	x	1	1
M483	Guanine (NH3)	1	1	1	x	1	x
R580	Guanine (NH)	1	1	1	x	1	1
Y583	Guanine (NH)	1	1	1	x	1	1
Y583	Guanine (NH3)	1	1	1	x	1	1
Y583	Guanine (OH)	1	1	1	x	1	1
S482	Ribose	x	x	x	1	x	x
R478	α-phosphate	x	1	1	x	x	x
V479	α-phosphate	x	1	x	x	x	x
G480	α-phosphate	x	1	1	x	x	x
R580	α-phosphate	1	x	1	1	1	x
R476	β-phosphate	x	2	x	x	x	1
R478	β-phosphate	1	1	x	x	x	2
V479	β-phosphate	x	x	1	x	x	1
G480	β-phosphate	1	x	x	1	1	x
R580	β-phosphate	x	1	1	x	1	1
K173	γ-phosphate	1	1	x	x	x	x
R476	γ-phosphate	1	x	2	2	2	x
R478	γ-phosphate	x	x	x	2	2	x
R580	γ-phosphate	1	x	x	1	x	x

Interactions involved in the stabilisation of GTPγS within the TG2 nucleotide binding pocket compared with ATP (PDB: 3YL6), GTP (PDB: 4PYG) and GDP (PDB: 1KV3).


A
GTPγS
i
ii
iii


B
TG2 nucleotide bound structures
i ATP
ii GTP
iii GDP

2.3.13 Involvement of GTP hydrolysis in the export of TG2

To identify a role for GTP hydrolysis in the export of TG2 from the cell, the potential hydrolysis mechanism was first considered by visualising the TG2 GTP binding site (PDB: 4PYG (Jang *et al.* 2014)) (Fig. 2.21). Two nearby H_2O molecules were identified and were in close proximity to the γ-phosphate phosphoanhydride bond. One H_2O molecule stabilises the α, β and γ phosphates and interacts with R478. Arginines are commonly involved in the stabilisation of the transition state of GTP hydrolysis reactions (Bourne 1997). The other H_2O molecule interacts with K173 and the γ-phosphate of GTP potentially acting as a mediator leading to the interaction between the γ-phosphate and K173. The surrounding H_2O molecules can act to stabilise the phosphate once it has been removed from the nucleotide.

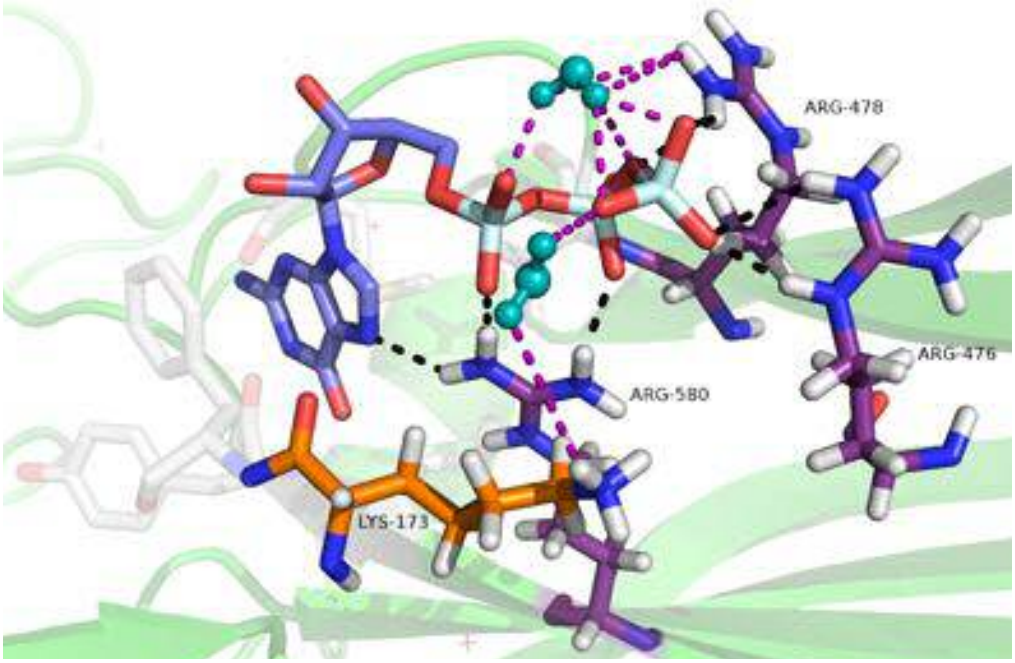


Figure 2.21: TG2 predicted hydrolysis mechanism.
TG2•GTP (PDB: 4PYG) was visualised using PyMOL in the cartoon representation. GTP (carbons in lilac) and TG2 nucleotide binding pocket residues predicted to have a role in GTP hydrolysis (K173, R476, R478 and R580 (carbons in lysine were orange and arginine, purple) are highlighted in stick representation (hydrogens: white, oxygens: red, nitrogens: blue, phosphate: pale blue). Polar contacts are highlighted (black dotted lines). H_2O molecules (teal) are shown in ball and stick representation and polar contacts involving H_2O are also highlighted (purple dotted lines).

2.3.14 Turnover of BODIPY GTP by TG2

To investigate the GTP hydrolysis mechanism, the turnover of BODIPY GTP over time was monitored for WT TG2 and the GTP binding site mutants (Fig. 2.22). There is a decrease in the fluorescence of BODIPY GTP overtime consistent with its dissociation from the binding pocket due to its hydrolysis to BODIPY GDP then BODIPY GMP. Dissociation is more likely due to two hydrolysis steps (fast and slow). First hydrolysis to BODIPY GDP is fast and as GTP and GDP have similar affinity for TG2, a marginal amount of dissociation is measured (Iismaa *et al.* 2000). However, GMP has significantly reduced affinity and hydrolysis to BODIPY GMP and its dissociation from the pocket could represent the measured slow step.

A dramatic reduction in fluorescence is seen overtime with both WT TG2 and TG2 K173N. TG2 K173L shows reduced hydrolysis activity in comparison to WT, whereas fluorescence levels for the TG2 K173N/F174D and TG2 R580A are comparable to FXIIIA (Fig. 2.22). This could be due to reduced GTP binding and therefore a smaller fluorescence spike post-injection which in turn results in reduced hydrolysis. The half-life was measured for the two kinetic events (Table 2.10). WT TG2 and R580A had calculated half-lives for both hydrolysis steps, however the rate of turnover by TG2 R580A was significantly reduced when compared with WT. The first (fast, Table 2.10 (1)) kinetic event was not observed for K173N and K173L. With regards to the second kinetic event (slow, Table 2.10 (2)), turnover was apparent for WT TG2, K173N, K173L and R580A. Interestingly, K173N showed a higher turnover rate when compared with WT and K173L has comparable GTP turnover to WT TG2. In the literature the GTPase activity of TG2 K173N and K173L was impaired (Iismaa *et al.* 2000). No GTP turnover was measured for TG2 K173N/F174D and no data has be published regarding this mutation for comparison.

Table 2.10: Half-life of BODIPY GTP in presence of TG2.

TG2 protein	$t^{1/2}$ (x 10^4 s) 1	$t^{1/2}$ (x 10^4 s) 2
WT	0.7	2.4
K173N	N/a	1.5
K173L	N/a	2.8
K173N/F174D	n.d.	n.d.
R580A	1.8	3.4
FXIIIA	No binding	

n.d.- not detected

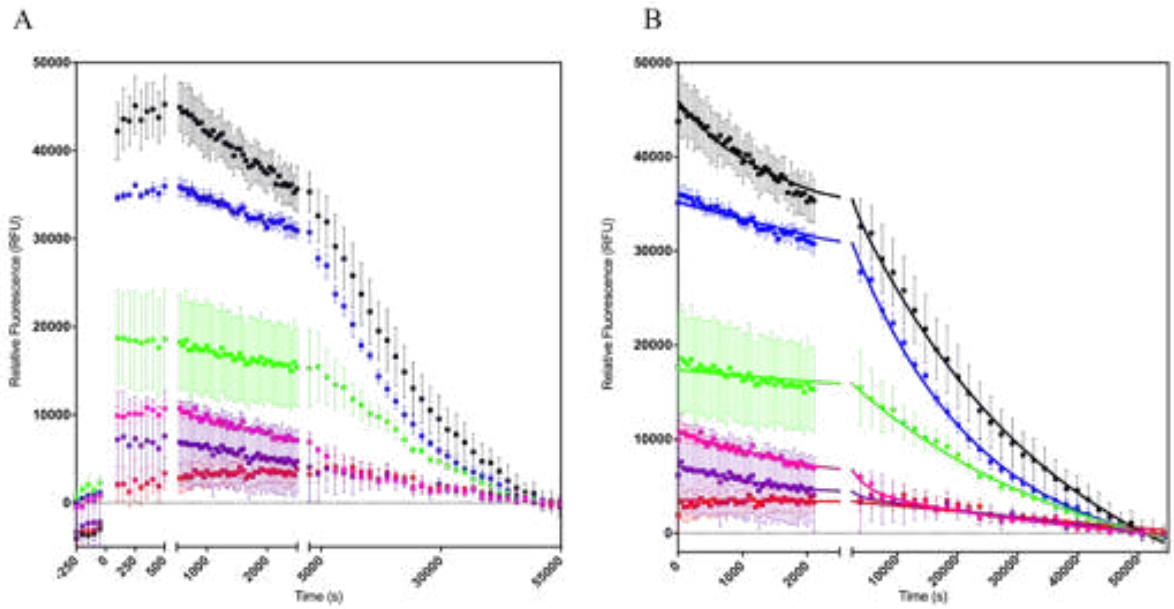


Figure 2.22: GTP turnover for TG2 GTP binding variants.
GTP turnover was measured for WT TG2 (black) and TG2 GTP binding variants, (K173N (blue), K173L (green), K173N/F174D (purple) and R580A (pink)) using BODIPY GTP (50 nM). FXIIIA (red) was used as a negative control for GTP hydrolysis. A: GTP association is followed by a fast turnover and then slow decay. B: Analysis of GTP hydrolysis by TG2 kinetics. Data is representative of two replicates (n=2) in one independent experiment (N=2).

2.3.15 GTPase Glo-assay™ to assess the GTPase activity of TG2

The GTPase activity of the WT TG2 and TG2 GTP binding site mutants was also assessed using the GTPase-Glo™ assay (Promega) (Mondal *et al.* 2015). GTP concentration needs to be in excess in order to not limit the GTPase reaction but minimal as to not saturate the luciferase reaction and therefore present an accurate quantification. The luminescence was measured for GTP alone at different final concentrations (0-20 µM), (Fig. 2.23). This experiment shows that 5 µM GTP is the best compromise for the required criteria and also established the relationship between GTP concentration and luminescence.

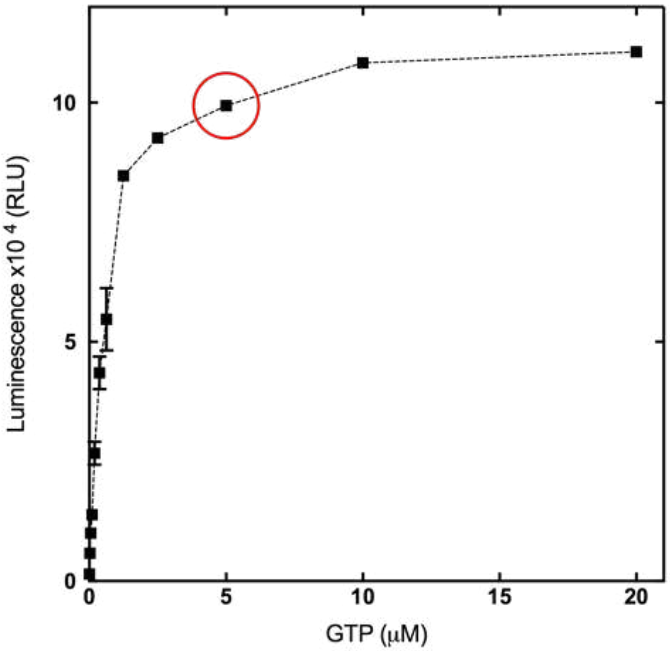


Figure 2.23: Optimising GTP concentration.
Standard curve for GTP conversion to ATP. Red circle represents optimal GTP concentration. Data is presented as means of two wells (n=2) ± SD in one independent experiment (N=1).

The impact of the relevant buffer (GTPase/ GAP buffer and GEF buffer, Table 2.2) on GTPase activity was compared using WT TG2 (Fig. 2.24). This showed GEF buffer supports higher TG2 GTPase activity as indicated by a larger reduction in luminescence signal. Also, the difference between hydrolysis by TG2 at different concentrations is more distinct. Given the known role of Mg^{2+} in the GTPase activity of G-proteins, the increased Mg^{2+} concentration could enhance this activity (Lai *et al.* 1998). This Mg^{2+} dependency also applies to TG2 although no structural explanation for this has been put forth to date. Consequently, GEF buffer was used for future GTPase-Glo™ assays.

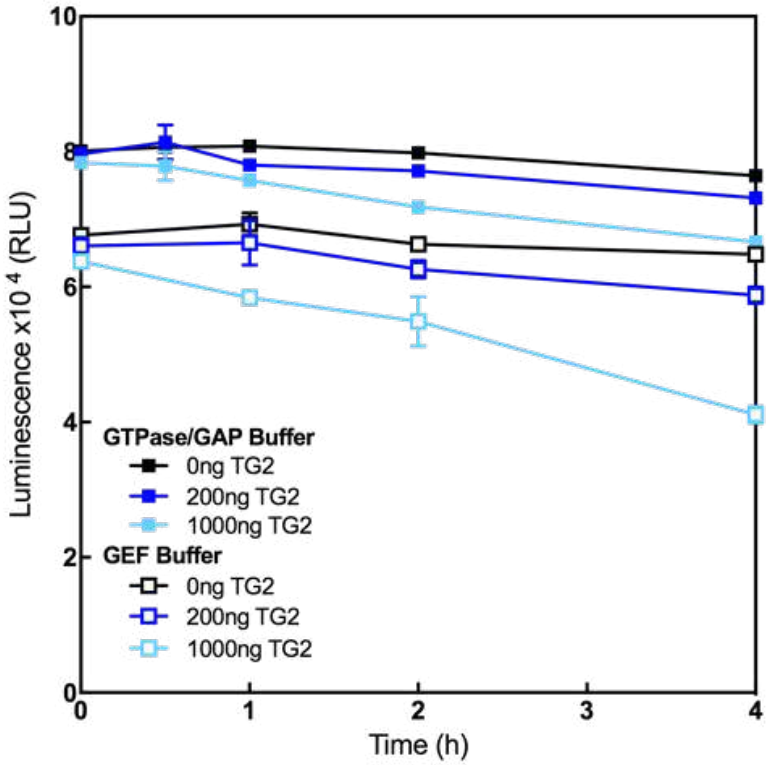


Figure 2.24: Buffer selection for TG2 in the GTP Glo™ assay.
WT TG2 (0 -1000 ng) GTPase activity was investigated with 5 µM GTP in either GTPase/GAP buffer (filled symbols,■) or GEF buffer (open symbols, □) over 4 hours. The reaction was terminated using GTPase-Glo™ buffer containing GTPase Glo reagent and 5 µM ADP. The relative luminescence was measured using FLUOstar Omega plate reader. No enzyme used as a control. Data is presented as means across three wells ± SD (n=3) in one independent experiment (N=1).

The intrinsic GTPase activity for WT TG2 has been reported as low compared to other GTPases (Iismaa *et al.* 1997; Iismaa *et al.* 2000). Therefore, hydrolysis was monitored over 12 hours (Fig. 2.25). This time course shows that 6-8 hours is sufficient to generate signal levels $<5x10^3$ RLU (background level). There is rapid hydrolysis in the first two hours using either 200 ng or 1000 ng TG2. Between 2 to 6 hours the rate of GTP hydrolysis is slower and background luminescence levels were reached at 8 hours. To ensure hydrolysis of GTP was complete, 12 hour incubations were used in all future assays.

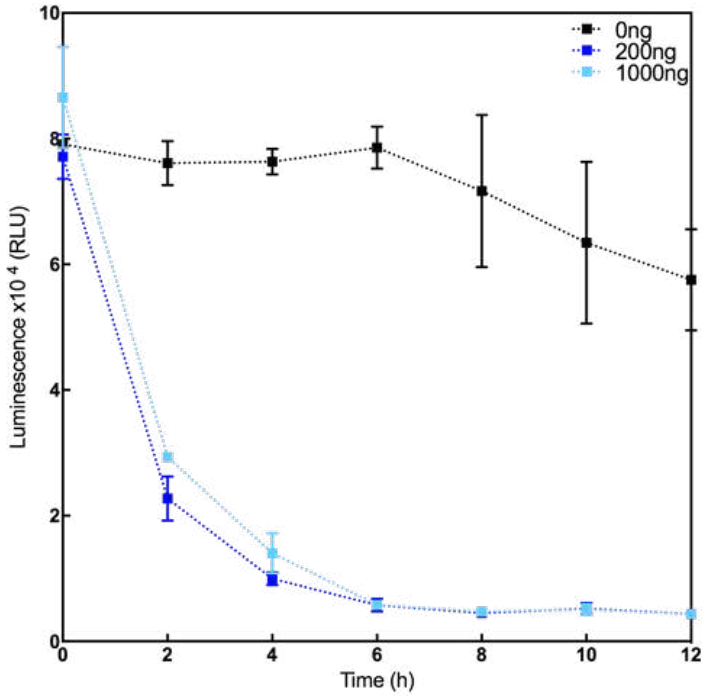


Figure 2.25: Optimisation of the time required for GTP hydrolysis.
GTPase activity was monitored over 12 hours by measuring the luminescence signal generated. Represented data has been averaged over three wells (n=3) ± SD in two independent experiments (N=2).

A titration of WT TG2 (3 hours incubation) was used to identify the useable TG2 concentration range in relation to signal (Fig. 2.26). The lowest amount of enzyme with observable hydrolysis over 3 hours is 37 ng. 1000 ng shows the greatest reduction in signal over time so was used in subsequent studies to assess GTPase activity. Higher concentrations were not possible due to interference with the assay system. In summary, the optimum conditions to assess GTPase activity of transglutaminase were established. The 2x GTP buffer was made up with 10 µM GTP in GEF buffer. The incubation of GTP with TG2 is optimal at 12 hours with 100-1000 ng of TG2.

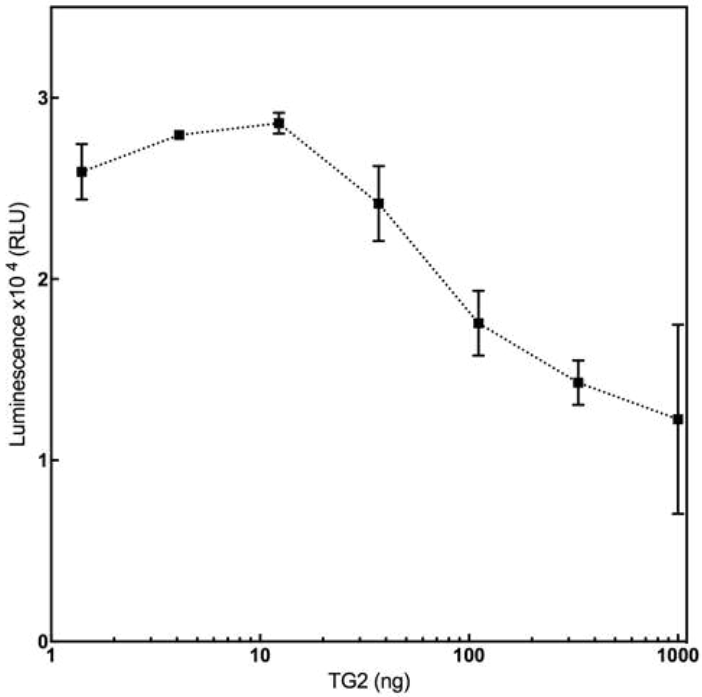


Figure 2.26: Optimisation of amount of TG required.
GTPase activity was measured with increasing WT TG2 (0-1000 ng, decreasing by two thirds each time). Data is averaged over three wells ± SD (n=3) in one independent experiment (N=1).

2.3.16 Investigating the GTPase activity of the GTP binding mutants

With the optimum conditions established, the GTPase activity of the TG2 GTP binding site mutants was assessed over time. As TG2 R580A preparations were very low concentration, 500 ng of each TG2 variant was used (Fig. 2.27). WT TG2 showed a significant reduction in the GTP overtime (Fig. 2.27, black). FXIIIA was used as the negative (no GTP hydrolysis) control (Fig. 2.27, red). As expected, little reduction in luminescence was observed in the absence of TG2 or upon FXIIIA addition over time. The small decrease in luminescence may be spontaneous GTP hydrolysis. TG2 K173N showed reduced GTPase activity compared with WT. Interestingly, the remaining TG2 GTP binding mutants (TG2 K173L (green), K173N/F174D (purple) and R580A (pink)) had increase GTPase activity when compared to WT TG2. The highest GTPase activity was detected for R580A, this is unexpected as R580A has previously been shown to not bind GTP (Fig. 2.17A)(Begg *et al.* 2006a, Begg *et al.* 2006b) .

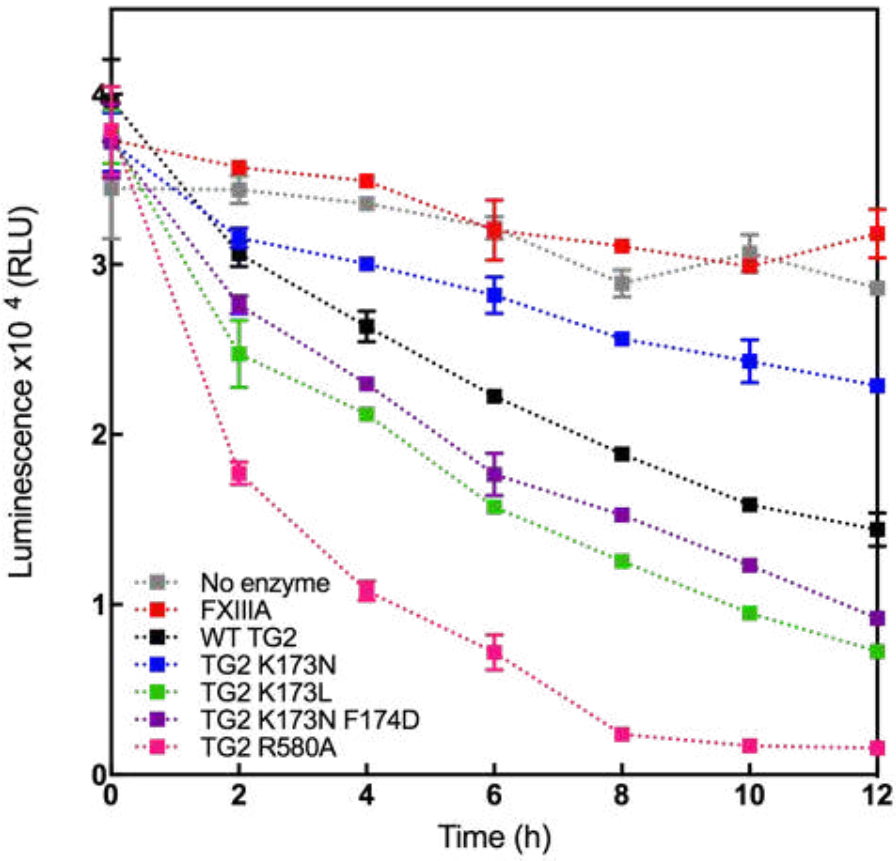


Figure 2.27: GTPase activity of TG2 GTP binding variants with GTPase Glo™ assay.
GTPase activity was measured indirectly using the GTPase Glo™ assay format. 1000 ng of TG2 or FXIIIA protein was incubated with 5µM GTP. The remaining GTP was converted to ATP to power luciferase to generate a luminescence signal. Data is shown as the mean of three wells ± SD (n=3) in two independent experiments (N=2).

To conclude the GTPase assay has shown that TG2 K173L, TG2 K173N/F174D and R580A have a higher rate of hydrolysis relative to WT TG2. TG2 K173N has a lower GTPase activity compared to WT TG2. Interestingly, these results do not match that of the BODIPY GTP turnover assay where limited turnover is observed for TG2 K173N/F174D and TG2 R580A when compared with WT TG2. As the BODIPY GTP hydrolysis is directly linked to its binding it may be possible that the hydrolysis seen in the GTPase-Glo™ assay is not a true representation of GTPase activity.

2.3.17 ATP binding of WT TG2 and TG2 GTP binding mutants

In the GTPase Glo™ assay the remaining GTP is converted to ATP, which is utilised by luciferase to generate a luminescence signal. The TG2 GTP binding mutants could have a higher affinity for ATP than WT TG2. This was investigated by measuring the

inhibition of TG2 isopeptidase activity with varying concentrations of ATP (0-8 mM) (Fig. 2.28).

However, only those mutants with isopeptidase activity could be assayed using this method. These mutants are TG2 K173N and K173N/F174D. There is a small shift (left, increased affinity, Fig. 2.28) in the relative activity of TG2 K173N/F174D when compared to WT TG2. An equivalent shift is seen with TG2 K173N to the right (reduced affinity). The apparent K_D was measured for each of the proteins and were as follows: WT = 1.2 ± 0.14 mM, K173N = 1.8 ± 0.55 mM and K173N/F174D = 0.9 mM (ambiguous). To investigate direct binding of ATP to the TG2 mutants, BODIPY ATP was used (data not shown). However due to the huge difference in affinity of WT TG2 for GTP and ATP (0.2 mM vs 2 mM), optimisation of the assay is required in order to address the signal to noise ratio. Further investigation is required to establish the reasons for the discrepancy between the BODIPY GTP turnover assay and the GTPase-Glo™ assay.

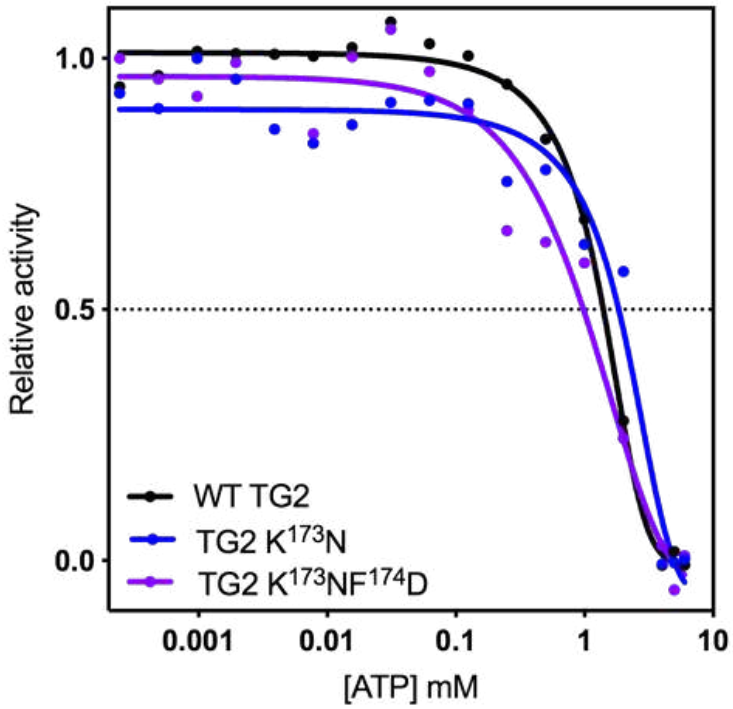


Figure 2.28: Inhibition of isopeptidase activity of TG2 GTP binding variants with ATP.
The isopeptidase activity of the TG2 variants was measured in the presence of ATP (0-8 mM). The apparent K_D was calculated from half maximal inhibition. Mean of two wells (n=2) and representative of two independent experiments (N=2).

2.4 DISCUSSION

From the results presented here it is apparent that GTP binding has a role in the export of TG2 from the cell. GTP is predicted to be important for determining the conformation of TG2 during export and /or GTP hydrolysis to provide the energy required (phosphate transfer mechanism) for the export to take place.

2.4.1 Activity of TG2

The activity of WT TG2 and the TG2 mutants was assessed using the MDC assay (measures binding of second substrate, Fig. 1.4D) and isopeptidase assay (measures binding of first substrate, Fig. 1.4A). Taken together, these results show that TG2 K173L and R580A favoured transamidation over isopeptidase activity and implies differences in substrate selectivity (Fig. 2. 6). This difference may reflect local changes in the glutamine substrate pocket for these TG2 mutants favouring transamidation, which impacts substrate binding. This is a novel observation that has not been noted in the literature. In addition, previous studies showed TG2 R580A (human) has reduced transamidase activity. However, mutation of the equivalent residue in rat TG2 (R579A) had no effect on transamidase activity when compared to WT rat TG2 (Ruan *et al.* 2008). This result implies previous data obtained on rat TG2 R579A may not be representative of human TG2 R580A activity.

2.4.2 Conformation of TG2 GTP binding site mutants

The structure of WT TG2 and the TG2 GTP binding site mutants was assessed using various methods (CD spectroscopy, ion exchange and size exclusion chromatography and Native PAGE). The spectra measured using the CD spectroscopy showed little variation in the α helical and β sheet content when measured using Dichroweb (Table 2.4). This was expected as the predominant change in the conformation between open and closed conformation is in the hinge region between the catalytic core domain and the first β barrel, which involves little change in secondary structure (Pinkas *et al.* 2007). However, when comparing the secondary structure of TG2 K173N/F174D and R580A to existing X-ray crystallography data, they closely resembled the open TG2 structure with reference to their α helical content and may indicate an intermediate conformation (Table 2.4). This is supported by previous data suggesting TG2 R580A assumes a semi compact conformation (Begg *et al.* 2006a). TG2 K173N and K173L

instead are reflective of TG2•GTP (Jang *et al.* 2014). Whilst WT TG2 on this occasion resembles TG2•ATP, despite the lack of nucleotide present (Han *et al.* 2010).

In support of the CD spectroscopy data, the elution profile, from ion exchange chromatography, of the TG2 GTP binding site mutants showed a shift to the right for TG2 K173L, K173N/F174D and R580A, when compared with WT TG2 (Fig. 2.8). This may be indicative of a conformational change, but analytical size exclusion chromatography showed all spectra overlap, implying no change in global size and shape (Fig. 2.9). The lack of change in the observed conformation may be down to the length of the gel filtration column. Increasing the size of the column, whilst maintaining the pore size would improve resolution and allow smaller changes in structure to be captured and therefore could identify an intermediate conformation of TG2 R580A and K173N/F714D (Begg *et al.* 2006a). In addition, inhibitor bound open TG2 could be utilised as a positive control to represent a conformational change. This may allow the observed shoulder peak to be characterised.

In the context of TG2 export, the results obtained from native PAGE analysis indicate the GTP-bound compact conformation is important for export (Fig 2.11). The closed compact conformation can be adopted by both WT TG2, K173N and K173L (to a lesser degree) which is represented by the presence of two species (TG2 and TG2•GTP, Fig. 2.11). The ability to assume the closed compact conformation coincides with the TG2 mutants capable of export from cells (Fig. 2.2). A single species is observed with TG2 K173N/F174D and with TG2 R580A in the presence of GTP. This suggests TG2 K173N/F174D and TG2 R580A cannot form the compact conformation due to lack of GTP binding or GTP cannot sufficiently induce the closed compact conformation and relates to the observed deficiency in export. TG2 K173N/F174D and TG2 R580A may adopt the intermediate conformation suggested by CD spectroscopy (Table 2.3), ion exchange chromatogram data (Fig 2.8) and the literature (Begg *et al.* 2006a). This in turn increases the diameter of TG2 resulting in retention in the cytoplasm. Consequently, the data obtained here supports the importance of TG2 conformation and indicates that it is the main influencer of TG2 externalisation.

2.4.3 GTP binding of the TG2 GTP binding site mutants

Native PAGE analysis indicated a GTP induced conformation is important for the export of the TG2 mutants. To further support the relevance of GTP, GTP binding was assessed in the context of inhibition of isopeptidase activity, and directly using BODIPY GTP. Inhibition by GTP showed TG2 K173N/F174D and R580A were insensitive to GTP (Fig. 2.10). However, the isopeptidase activity for these mutants without GTP was relatively low (Fig. 2.6) which may lead to inaccuracy when measuring the GTP inhibition. Inhibition of WT TG2, K173N and K173L was comparable in this case. Similar results were observed by the inhibition of transamidase activity using GTPγS, where reduced inhibition was measured with TG2 K173N and K173L and no inhibition was measured for rat TG2 R579A (human equivalent: TG2 R580A) (Iismaa *et al.* 2000; Begg *et al.* 2006a). However, given the transamidase and isopeptidase activity of these mutants differ (Fig. 2.6), this could impact GTP inhibition results.

BODIPY GTP binding was measured for WT TG2, K173N and to a lesser degree TG2 Kl73L (Fig. 2.17A). This is as expected as TG2 K173N and K173L have previously been shown to have lower GTP binding affinity (1.5 fold and 6 fold, respectively) than WT TG2 (Iismaa *et al.* 2000). TG2 K173N/F174D and TG2 R580A did not bind GTP and measurements were indistinguishable from FXIIIA. This is consistent with [α-^{32}P] GTP photolabeling experiments where TG2 K173N and K173L had reduced photolabeling, and photolabeling was abolished for TG2 R579A when compared with WT (Iismaa *et al.* 2000; Begg *et al.* 2006b). No data has been published for the TG2 K173N/F174D mutant. However, as F174 is predicted to be important for the stabilisation of the GTP guanine moiety, a reduction in GTP binding has been predicted (Liu *et al.* 2002).

In contrast, the binding of BODIPY GTPγS was measured to all the GTP binding site mutants (Fig. 2.17C). This contradicts the previous data which suggest TG2 K173N/F174D and R580A are unable to bind GTP (Fig. 2.10 and 2.17A). BODIPY GTPγS binding has previously been measured for TG2 K173N but the remaining mutants have not been tested (Datta *et al.* 2007). No binding was measured for conservative mutation of TG2 R580K, suggesting R580A will not bind BODIPY GTPγS (Singh *et al.* 2016). Furthermore, GTPγS filter binding assays showed no

binding to TG2 K173N and K173L but TG2 R580A was not tested (Iismaa *et al.* 2000). Here, GTP and GTPγS binding has been consistently measured for each of the TG2 mutants using the same method which has not been previously done. The data here is inconsistent with regards to whether GTP or GTPγS is used and therefore they should not be used interchangeably to measure guanosine nucleotide binding.

In addition, it is important to take into consideration that the BODIPY moiety is linked to GTPγS via an amide linker on the γ-phosphate where as BODIPY is connected to GTP via an amide linker on the ribose ring (Fig. 2.29). This difference in linkage could influence the binding of BODIPY GTP and could represent altered nucleotide binding pocket influenced by the respective TG2 GTP binding site mutations. As a result, the TG2 K174N /F174D and TG2 R580A mutants may not accommodate, and TG2 K173N and TG2 K173L can accommodate to a lesser degree the bulky BODIPY moiety attached to the ribose ring. It could therefore be hypothesised that the BODIPY GTPγS binding is less affected by the BODIPY moiety (linked to γ-phosphate) as it is not buried within the pocket and has greater exposure to the hydrophilic environment.

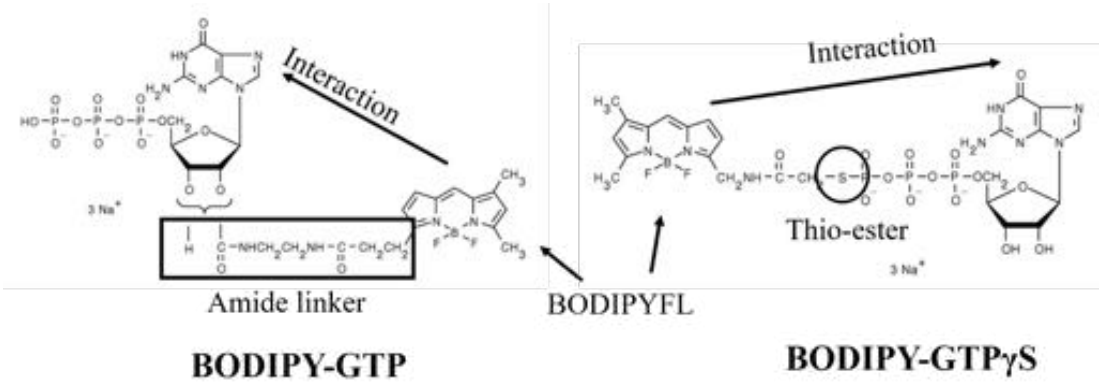


Figure 2.29: BODIPY Guanine nucleotides.
BODIPY has been conjugated to GTP and GTPγS. BODIPY is conjugated to the ribose ring of GTP via an amide linker. BODIPY is conjugated to the γ-phosphate of GTPγS

Molecular docking of GTPγS into TG2 showed two main conformations (Fig. 2.20A). The most energetically favourable conformation is comparable to TG2•GTP (PDB: 4PYG, Fig. 2.20A (iii) and 2.20B (i)). The second conformation shows an interaction between the γ phosphate of GTP and K173 and may represent a GTP hydrolysis intermediate (Fig. 2.20A (i and ii). This is in line with previous literature as K173 is predicted to be important for the binding of the γ phosphate (Iismaa *et al.* 2000; Liu *et*

al. 2002). It indicates GTP is a capable of binding in two conformations but it is possible GTP can only achieve this orientation during hydrolysis. A hydrolysis mechanism of GTP by TG2 is yet to be elucidated, however these molecular docking experiments have given an insight into potential nucleotide binding modes. In addition, the different conformations could be relevant with regards to the different binding of BODIPY GTP and GTPγS.

To conclude, TG2 K173N and K173L are capable of binding GTP and GTPγS as the results presented here suggest. This study has also attempted to define the GTP binding properties of TG2 K173N /F174D and R580A. The majority of results presented here suggest TG2 K173N /F174D and TG2 R580A are unable to bind GTP and this finding is supported by published results. However, binding has been measured for TG2 K173N/F174D and R580A using BODIPY GTPγS (Fig. 2.17 C). There are inconsistencies in the literature with regards to the various assays used to assess GTP and GTPγS binding. Consequently, binding of GTP and GTPγS should be assessed separately.

2.4.4 GTP hydrolysis of TG2 GTP binding site mutants

GTP hydrolysis was measured using the turnover of BODIPY GTP (Fig. 2.22). Turnover was measured for WT TG2, TG2 K173N and TG2 K173L consistent with the BODIPY GTP binding assays. As predicted, minimal binding and turnover is observed for TG2 K173N/F174D and TG2 R580A, as hydrolysis activity is related to BODIPY GTP binding. Instead, GTPase activity was measured with the GTPase-Glo™ assay (Fig. 2.27). These results indicated TG2 K173L, TG2 K173N/F174D and TG2 R580A hydrolyse GTP at a higher rate than WT TG2. These results suggest TG2 K173N/F174D and TG2 R580A mutants are capable of binding GTP contradicting previous GTP binding assays. Reduced GTPase activity of TG2 K173N and severely impaired activity for TG2 K173L has been previously reported (Iismaa *et al.* 2000). Due to weak binding of GTP by TG2 R580A, GTPase activity was predicted to be abolished (Begg *et al.* 2006b). The GTPase activity of TG2 K173N/F174D was expected to be lower than WT TG2 as TG2 K173N had reduced GTPase activity (Iismaa *et al.* 2000). One explanation for the unexpected results is an alternative affect was measured. For example enhanced affinity for ATP during the GTP/ATP

conversion step (Fig. 2.2). However, ATP inhibition experiments show that WT TG2, and the K173N and K173N/F174D mutants have similar affinity for ATP (Fig. 2.28). Interestingly, TG2 K173N/F174D shows significantly reduced affinity for GTP (Fig 2.10) but ATP binding remains unaffected (Fig 2.28).

Another explanation is that published results have indicated that the Core and the βcore are able to bind and hydrolyse GTP (Iismaa *et al.* 1997). This observation indicates that the closed conformation of TG2 is not required to bind and hydrolyse GTP. In addition, an increased rate of GTP hydrolysis was measured for the TG2 Core (50% greater than WT TG2) and the βcore showed GTP hydrolysis rates similar to WT TG2. The Core and the βcore show ATPase activity, 75% and 40% higher, respectively than WT TG2. The increased GTPase activity and ATPase activity of the Core domain alone is relevant to the activities of TG2 K173N/F174D and TG2 R580A as they are suggested to adopt an intermediate conformation. The exposed core domain could allow unregulated GTPase activity and explain the increased GTP hydrolysis activity of TG2 K173N/F174D and TG2 R580A. Moreover, TG2 R580L and TG2 R580K are capable of hydrolysing GTP at higher rates than WT TG2, suggesting the presence of arginine in this position has a role in regulating GTP hydrolysis (Datta *et al.* 2007).

2.4.5 Conclusion

Taken together, the importance of GTP binding in the secretion of TG2 has been established. The data supports the relevance of TG2 conformation in this mechanism however, further work is required specifically to analyse TG2 conformation. In addition, the importance of GTP hydrolysis is yet to be established. With this in mind the following points need to be addressed. Functional and structural characterisation of TG2 R580A is needed to identify whether data obtained on rat TG2 R579A is relevant. Further structural characterisation of TG2 GTP binding site mutants is required to determine if TG2 K173N/F174D and TG2 R580A adopt an intermediate conformation. These structural studies will provide additional evidence to support the importance of conformation in TG2 export. Additionally, more sophisticated methods are required to accurately measure the GTP and GTPγS binding capabilities of the TG2 GTP binding site mutants. These experiments will identify if GTP and GTPγS binding can confidently be interchanged when assessing binding kinetics in TG2 and can detect if the BODIPY moiety is hindering GTP binding of selected TG2 binding

site mutants. The binding modes of GTP and GTPγS observed with molecular docking approaches need to be confirmed in order to formulate a novel mechanism of GTP hydrolysis in TG2. One approach would be to identify changes in the nucleotide binding pocket using GTPγS and various transition state inhibitors ($GDP.AlF_4$/ MgF_3) using X-ray crystallography (Bigay *et al.* 1985, Graham *et al.* 2002). Moreover, additional work is required to establish if the GTPase-Glo™ assay is representative of the GTPase activity, ATPase activity or an unknown mechanism of the TG2 GTP binding site mutants. It is unlikely GTP hydrolysis is involved in TG2 secretion however, alternative assays need to be developed to more confidently establish the involvement of GTP hydrolysis in TG2 export. For example, the use hydrolysis inhibitors such as GTPγS or transition state stabilisers.

.

CHAPTER 3

3. THE DEVELOPMENT OF METHODS FOR EXPRESSION AND PURIFICATION OF WT TG6.

3.1 INTRODUCTION

Limited information is available on the physiological role of TG6 in the CNS and as a consequence further investigation is required. To aid in the understanding of WT TG6 and the TG6 mutants associated with SCA35, its functional and structural characteristics need to be determined. To date the structure of TG6 has not been determined, which limits our understanding of its structure and function. However, a significant amount of research has been directed toward the biochemical and physiological characterisation of TG isoforms TG2 and TG3. This data in turn can be translated and exploited in order to help progress our understanding of TG6.

Our laboratory has successfully developed an *E. coli* system for expression and purification of TG2 but not for the expression of TG6. Instead, a successful baculovirus/*Sf9* expression system for WT TG6 has been developed. This approach yields WT TG6 with ~95% purity and is sufficient for biochemical assays to probe TG6 function and small-scale structural studies such as size exclusion chromatography and CD spectroscopy. However, this approach is expensive and time consuming when compared to the *E. coli* expression systems. In addition, the purified His-tagged TG6 has previously been shown to be unstable at 4 °C (DPA lab, unpublished observation) and is very aggregation prone, so it would be valuable to investigate ways to increase TG6 solubility. In an *E. coli* expression system, solubility tags can be utilised to create TG6 fusion proteins that have increased stability or solubility. As there are no known post-translational modifications of TG6, it would be beneficial to develop a prokaryotic expression system for TG6. This system would aid in the production of larger amounts of TG6 for downstream biochemical analysis and crystallisation studies.

The advantages of developing an *E. coli* expression system are as follows. *E. coli* can be cultured and grown with ease in large scale quantities (Sezonov *et al.* 2007). High cell densities can be obtained, which in turn allows greater protein yields to be achieved. Furthermore, the growth media is inexpensive and readily available. In

addition, *E. coli* can be easily manipulated as optimised expression plasmids are readily available and transformation of plasmid DNA is fast (Rosano and Ceccarelli 2014).

However, the expression of proteins in *E. coli* is limited with respect to post-translational modification and protein folding, as *E. coli* lacks the necessary machinery. The accumulation of misfolded protein results in formation of inclusion bodies (Hartley and Kane 1988; Carrio and Villaverde 2002). In order to reduce the formation of inclusion bodies, a lower temperature of induction and a longer induction time can be used (Schein and Noteborn 2019). These changes promote a lower rate of protein synthesis, allowing the protein to fold properly and decrease aggregation due to reduced hydrophobic interactions at lower temperatures (Baldwin 1986). In contrast, the low temperatures also lead to lower growth rate, lower protein synthesis and therefore low protein yields (Rosano and Ceccarelli 2014). Furthermore, if soluble protein is achieved, the protein may adopt a stable conformation that does not reflect the native conformation and therefore may still be inactive (Martinez-Alonso *et al.* 2008).

TG6 can be fused to other proteins that can aid in its purification, solubility and folding, and expression. As examples, the hexa-histidine (His tag), maltose binding protein (MBP) and glutathione S-transferase (GST) (Terpe 2003) tags can all aid purification. The His tag is unlikely to interfere with the solubility and folding of the protein or the enzymatic activity, however larger tags such as MBP and GST may require cleavage and an additional purification step. The solubility and folding of TG6 can be enhanced using various solubility tags. For example MBP, N-utilisation substance A (NusA) and thioredoxin (Trx). MBP, NusA and Trx induced the greatest increase in the solubility of their fusion partners compared with other tags (Terpe 2003; Nallamsetty and Waugh 2006). The mechanism of increased folding and solubility is unknown for NusA or Trx, but MBP has been shown to have intrinsic chaperone activity (Kapust and Waugh 1999; Costa *et al.* 2013). Again due to their large size they could interfere with structural studies and hence would require cleavage.

Protein tags could also be introduced using the baculovirus/ *Sf9* system, however this process is more time consuming as it involves the manipulation of the pFastbac plasmid, generation of the recombinant bacmid and the transfection of the bacmid to produce the baculovirus. In contrast, *E. coli* expression plasmids require a one-step cloning approach and would therefore be of greater value in assessing the effect of the protein tags on the solubility of TG6.

In this Chapter, studies are described towards achieving three aims:

1. Investigate various protein tags for their ability to aid in the solubilisation, folding and expression of WT TG6.
2. Investigate and optimise the expression and purification of TG6 fusion proteins produced in *E. coli* and examine whether the TG6 fusion proteins are functional and correctly folded.
3. Investigate the biochemical properties of WT TG6 generated using a *Sf9*/baculovirus expression system.

3.2 METHODS

3.2.1. Cloning of human TG6 cDNA into *E. coli* expression plasmids

The primers, TG6 forward (5-'CGAAGTCGACATGGCAGGGATCAGAGTCACCAAGG-3') and reverse (5'-CCTAGCGGCCGCTCACTTGGCAGTGGCCACATGGACG-3') (Sigma), template pIND-TG6 plasmid (made in house) and Phusion® High-fidelity DNA polymerase (ThermoFisher Scientific, F530) were used to generate the TG6 cDNA.

A two-step thermo-cycling protocol was selected due to the high melting points of the primers (Table 3.1, (Chester and Marshak 1993). PCR products were run on a 0.9% agarose gel and the TG6 cDNA was subsequently extracted and purified (QIAquick® Gel Extraction kit, QIAgen, 28115). Purified TG6 cDNA and the selected expression plasmids (Gifted by Michail Nomikos and Tony Lai: pETMM11, pETMM20, pETMM41, pETMM60 and pHis-SUMO-intein (HSIE (Wang *et al.* 2012)) Fig. S1) were digested using the restriction enzymes *Sal*I and *Not*I (NEB, R0138 and R0189) for 3.5 hours at 37 °C. The restriction enzymes were heat inactivated (65 °C) and the DNA was incubated in calf intestine phosphatase (1.5 hours, 37 °C, NEB M0290). QIAquick® PCR Purification kit (QIAgen, 28104,) was used to purify the insert and vectors. Purity was confirmed by gel electrophoresis (0.9% agarose). Ligation mixes were set up containing ~250 ng of the desired expression plasmids and the insert was added as a 3:1 ratio (vector: insert). T4 DNA ligase (NEB, M0202) was used for ligating TG6 cDNA into the vectors and were incubated overnight at 16 °C.

Table 3.1: PCR conditions for WT TG6 cDNA amplification.

PCR	Time (s)	Temperature (°C)	Cycles
Initial denaturing	30	98	1x
Denaturing	10	98	30x
Annealing	33	72	
Final extension	600	72	1x

3.2.2 Initial expression of WT TG6 in *E. coli*

TG6 plasmids were transformed into MultiShot™ TOP10 competent *E. coli* (Invitrogen, C40005) and were grown on antibiotic-containing agar (kanamycin, 50 µg/mL, Sigma 60615) for the modified pETMM vectors, and ampicillin (100 µg/mL, Sigma A1593) for modified pHSIE vectors (Wang *et al.* 2012). Colonies were isolated

and overnight cultures (5 mL, containing 50 μg/mL kanamycin or 100 μg/mL ampicillin, Fisher 10419313) were grown for plasmid purification. Plasmid DNA was extracted from the colonies (QIAprep® Spin Miniprep Kit, 27104) and digested with *Sal*I and *NotI* to confirm integration of the TG6 coding sequence. The purified plasmids were transformed into BL21-CodonPlus competent *E. coli* cells (Agilent, 230240) and incubated on kanamycin (50 μg/mL) or ampicillin (100 μg/mL) containing agar. Single colonies were selected to grow small overnight cultures, which were then used to inoculate 500 mL cultures. Cultures were grown at 37 °C, 200 rpm until the OD_{600} absorbance was ~0.6. Protein expression was induced using isopropyl β-D-1-thiogalactopyranoside (IPTG, 0.1 mM), Fisher 10214633) and incubated at 16 °C, 200 rpm overnight. The cells were harvested by centrifugation at 3000 xg for 25 minutes.

3.2.3 Optimisation of human TG6 expression in *E. coli*

Following preliminary expression trials, the expression of the various TG6 fusion proteins was optimised. One 500 mL culture at required density (OD_{600} = 0.6) was split into 20 mL cultures and different concentrations of IPTG (0-1 mM final concentration) were used to induce protein expression. 1 mL samples were collected after 16 hours, centrifuged at 14000 xg for 60 seconds and the supernatant was removed. Using the method described above, the temperature for the protein synthesis was optimised for selected IPTG concentrations (0.01 mM, 0.1 mM and 0.5 mM) at 16 °C (16 hours), 30 °C (4 hours) and 37 °C (3 hours). After this screen, the induction time at 16 °C and 20 °C was then optimised. Samples were taken from the same culture every 2 hours post-induction over 24 hours.

3.2.4 Initial purification of His-MBP-TG6

Following expression of His-MBP-TG6, the bacterial pellet was resuspended in lysis buffer (detailed in Table 3.2). The resuspended solution was sonicated three times with 10 seconds pulses (amplitude 40%) (Q125 sonicator, Qsonica sonicators) and incubated on ice between pulses. The cell homogenate was centrifuged 48,000 ×g, 35 minutes, 4 °C and the supernatant collected. Amylose resin (1 mL resin volume, NEB E8021) was incubated with the supernatant at 4 °C for 1 hour. Next, the amylose slurry was applied to a column (50 mL empty column) and the resin collected on the column

filter. The amylose resin with bound His-MBP-TG6 was washed using wash buffer (detailed in Table 3.2), until little protein was detected in the flow-through as monitored using a Bradford reagent (diluted 1:3, Bio-Rad protein assay dye reagent concentrate, 5000006). For this assay 100 µL of column flow-through was added to 200 µL of the diluted Bradford reagent in a clear 96-well plate. Bound protein was eluted in 1 mL fractions using elution buffer (detailed in Table 3.2).

3.2.5 Purification of His-TG6, Trx-His-TG6 and NusA-His-TG6

Following expression of His-TG6, Trx-His-TG6 and NusA-His-TG6, the bacterial pellets were resuspended in lysis buffer (detailed in Table 3.2) The following steps were identical to the purification of His-MBP-TG6 (Section 3.2.4) except that Ni Sepharose™ 6 Fast Flow resin (1 mL resin volume, GE Healthcare, 10291210) was used as the affinity matrix. The Ni Sepharose was washed using wash buffer 1 (detailed in Table 3.2) until no protein was detected by Bradford reagent. The resin was washed using wash buffer 2 (detailed in Table 3.2). The bound His-fusion protein was eluted from the resin in 1 mL fractions using elution buffer (detailed in Table 3.2).

3.2.6 Purification of His-SUMO-intein-TG6

Lysis buffer (detailed in Table 3.2) was used to lyse the bacterial pellet (*E. coli* expressing His-SUMO-intein-TG6). The next steps were identical to those described in Section 3.2.4. The resin was collected on the filter and was washed using wash buffer (detailed in Table 3.2). Proteolytic removal of untagged TG6 from the His-SUMO-intein-TG6 fusion protein involved incubation at (4 °C for 3 hours) in 3-4 mL of cleaving buffer (detailed in Table 3.2). Lowering the pH from 8.0 to 6.0 activated intein which cleaved at the intein cleavage site upstream of TG6. Untagged TG6 was released from the Ni resin and collected in 1 mL fractions. Elution buffer (detailed in Table 3.2) was used to recover the cleaved fusion tag from the resin.

Table 3.2: Buffers used in purification of TG6 fusion proteins.

Fusion protein	Buffer	Components
His-MBP-TG6	Lysis	20 mM Tris-HCl pH 7.4, 250 mM NaCl, 1 mM EDTA, 1 mM DTT, Protease inhibitor (1 tablet per 50 mL, cOmplete™ EDTA-free protease inhibitors, Roche Diagnostics) 1 mg/mL lysozyme
	Wash	20 mM Tris-HCl pH 7.4, 250 mM NaCl, 1 mM EDTA
	Elution	20 mM Tris-HCl pH 7.4, 250 mM NaCl, 1 mM EDTA, 10 mM Maltose
His-TG6, Trx-His-TG6, NusA-His-TG6	Lysis	50 mM NaH_2PO_4 pH 8.0, 300 mM NaCl, 20 mM Imidazole, Protease inhibitor (1 tablet per 50 mL, cOmplete™ EDTA-free protease inhibitors, Roche Diagnostics) 1 mg/mL lysozyme
	Wash 1	50 mM NaH_2PO_4 pH 8.0, 300 mM NaCl, 20 mM Imidazole
	Wash 2	50 mM NaH_2PO_4 pH 8.0, 300 mM NaCl, 30 mM Imidazole
	Elution	50 mM NaH_2PO_4 pH 8.0, 300 mM NaCl, 250 mM Imidazole
His-SUMO-intein-TG6	Lysis	50 mM Tris-HCl pH 8.0, 300 mM NaCl, 25 mM Imidazole, Protease inhibitor (1 tablet per 50 mL, cOmplete™ EDTA-free protease inhibitors, Roche Diagnostics) 1 mg/mL lysozyme
	Wash	50 mM Tris-HCl pH 8.0, 300 mM NaCl, 25 mM Imidazole
	Cleavage	40 mM Bis-Tris pH 6.0, 25 mM imidazole PBS (1 tablet: l8 mM NaH_2PO_4, 1.5 mM KH_2PO_4, 137 mM NaCl, 3 mM KCl, per 500 mL H_2O, Oxoid)
	Elution	40 mM Bis-Tris pH 6.0, 300 mM imidazole Phosphate buffered saline (PBS, 1 tablet: l8 mM NaH_2PO_4, 1.5 mM KH_2PO_4, 137 mM NaCl, 3 mM KCl, per 500 mL H_2O, Oxoid)

3.2.7 Optimisation of purification of His-MBP-TG6

To optimise the purification of His-MBP-TG6, the lysis buffer was altered. Post-induction pellets were harvested and resuspended in 50 mL of the selected lysis buffer (Table 3.3). The protease inhibitors (Table 3.3) were added to prevent proteolytic degradation. First the purification of His-MBP-TG6 was carried out using increasing NaCl concentrations 0-500 mM in the lysis buffer (see Table 3.3). Results were analysed using SDS-PAGE analysis and Western blotting (as described in Section 3.2.10-11). Following this, different pH (6.0-8.5) of the lysis buffer were tested (Table 3.3). Finally, additional components were added to aid protein extraction (Octyl- β-glucopyranoside, Triton X100, Tween-20, glycerol or 5 mM $MgCl_2$, Table 3.3). Sonication lysed the cells and the homogenate was centrifuged at 11,500 xg. The supernatant was incubated with 250 µL of amylose resin, per condition, for 2 hours rolling at 4 °C. Following incubation, the resin was collected and washed with wash buffer (Table 3.2, His-MBP-TG6) by resuspension and centrifugation. His-MBP-TG6 was eluted in 1 mL fractions using elution buffer (Table 3.2, His-MBP-TG6).

Table 3.3: Optimisation of His-MBP-TG6 lysis buffer.

Component optimised	Concentration/ pH/Additive	Components
NaCl	0	20 mM Tris-HCl pH 7.4, 250 mM NaCl, 1 mM EDTA, 1 mM DTT, Protease inhibitors (1 µM aprotinin (Sigma, 10820), 0.1 mM AEBSF (Sigma, A8456), 1 µM bestatin (Sigma, B8385) and 10 µM leupeptin hemisulphate (Sigma, L8511))
	50	
	150	
	250	
	300	
	500	
pH	6.0	50 mM NaH_2PO_4, 1 mM EDTA, 1 mM DTT, Protease inhibitors (1 µM aprotinin, 0.1 mM AEBSF, 1 µM bestatin and 10 µM leupeptin hemisulphate)
	6.5	
	7.0	
	7.5	20 mM Tris-HCl 1 mM EDTA, 1 mM DTT, Protease inhibitors (1 µM aprotinin, 0.1 mM AEBSF, 1 µM bestatin and 10 µM leupeptin hemisulphate)
	8.0	
	8.5	
Additives	1 mM Octyl-β-glucopyranoside (Sigma)	20 mM Tris- HCl pH 8.0, 1 mM EDTA, 1 mM DTT, Protease inhibitors (1 µM aprotinin, 0.1 mM AEBSF, 1 µM bestatin and 10 µM leupeptin hemisulphate)
	1% Triton X-100	
	0.5% Tween-20	
	20% glycerol	
	5 mM $MgCl_2$	

3.2.8 Preparation of cell lysates for crude protein analysis

The pre-induction pellet was resuspended in 120 µL 2x sample buffer (200 mM Tris, 4% SDS 12.5 mM EDTA, 30% glycerol, 5 mM Bromophenol blue and 5% β-mercaptoethanol) and post-induction samples were resuspended in 250 µL of 2x sample buffer. Samples were then boiled (95 °C for 10 minutes) and centrifuged (14000 xg, 5 minutes). 20 µL of each sample and 10 µL of molecular weight ladder (Colour Protein standard Broad Range, Neb P7704) or 5 µL of LMW-SDS marker kit (GE Healthcare, 17-0446-01) was analysed using SDS-PAGE and Western blot analysis (as described in Section 3.2.10-11).

3.2.9 Preparation of purified TG6 fusion proteins

Following purification of the five TG6 fusion proteins, they were concentrated (Amicon Ultra 0.5 Centrifugal filters, Merck UFC503008). 10 µL of each sample and 10 µL of molecular weight ladder (Colour Protein standard Broad Range) or 5 µL of LMW-SDS marker kit in 2x sample buffer was analysed using SDS-PAGE and Western blot analysis (as described in Section 3.2.10-11).

3.2.10 SDS-PAGE analysis of bacterial cell lysate

Homemade polyacrylamide gels were made to analyse expression of TG6 fusion proteins from cell lysates and assess the initial purification of the TG6 fusion proteins. Gels consisted of a stacking and separating gel (detailed in Table 3.4). Homemade gels were run at 160V for 50 minutes (BioRad electrophoresis Tank (Protean II) and PowerPac^{TM} Universal power supply) in SDS running buffer (25 mM Tris-HCl pH 8.8 192 mM glycine, 0.1% SDS).

Table 3.4: Recipes for homemade Tris-glycine gels.

Gel (%)	Separating buffer:1.5M Tris pH 8.8, 10% SDS (mL)	Stacking buffer: 1M Tris pH 6.8, 10% SDS (mL)	30% Acrylamide (mL)	10% APS (µL)	TEMED (µL)	Distilled H_2O (mL)
4.5	0.00	1.0	0.4	30	5	1.55
7.5	1.25	0.0	0.9	50	5	2.75
8	1.25	0.0	1.0	50	5	2.70
9	1.25	0.0	1.5	50	5	2.15
10	1.25	0.0	1.6	50	5	2.00

SDS-PAGE was also carried out using the Invitrogen system. This system was used (as described in Section 2.2.4) to analyse the outcome of the optimisation of the lysis buffer optimisation for the purification of His-MBP-TG6.

3.2.11 Western Blotting of bacterial cell lysates

Cell lysate and purified protein samples were transferred for Western blot analysis using either a semi-dry or wet transfer system. SDS-PAGE was carried out on (7.5%

and 9%, Table 3.4) Tris-glycine gels and proteins were transferred to a polyvinylidene fluoride (PVDF) membrane (Immobilon, Millipore IPVH00005), for 1 hour, using the BioRad semi-dry system (transfer buffer: 40 mM glycine, 50 mM Tris, 20% methanol and 0.375 mg/mL SDS). During the optimisation of the purification of His-MBP-TG6, the wet transfer method was used (as described in Section 2.2.5).

Membranes were blocked overnight at 4 °C in 5% milk (non-fat milk powder, Tesco) in TBS (20 mM Tris-HCL pH 7.4, 150 mM NaCl). Two primary antibodies were used to detect different parts of the protein: Mouse anti-His^5 (Table 3.5, Qiagen 34660) or goat-anti human TG6 (Table 3.5, produced in house (Thomas *et al.* 2013)). The primary antibodies (Table 3.5) were diluted accordingly in 5% non-fat milk powder (Tesco) and TBS-T and the membranes were incubated for 1 hour followed by a 1 hour incubation with rabbit anti-mouse HRP (Table 3.5, Santa Cruz Biotechnology, SC-358914) or rabbit anti-goat HRP (Table 3.5, Dako, P0449). Bands were captured on X-ray film (as described in Section 2.2.5)

Table 3.5: Antibodies used in Western blots to detect TG6 fusion proteins.

Antibody	Dilution
Mouse anti-His^5 antibody (monoclonal)	1:5,000
Goat anti-TG6 antibody (polyclonal)	1:300
Rabbit anti mouse HRP antibody (polyclonal)	1: 10,000
Rabbit anti goat HRP antibody (polyclonal)	1: 10,000

3.2.12 Isopeptidase activity of TG6 fusion proteins

Isopeptidase activity of the TG6 fusion proteins (Trx-His-TG6, His-MBP-TG6 and NusA-His-TG6) were measured using an established fluorescence assay (described in Section 2.2.10, Adamczyk *et al*. 2013) instead using the TG6 assay buffer: 62.6 mM Tris-HCl, pH 7.5, 375 mM NaCl, 12.5 mM glycine methyl ester.

3.2.13 Expression of His-tagged TG6 in *Sf9* cells

WT TG6 baculovirus was previously generated by our lab. The *Sf9* cells were seeded at a density of $2x10^6$ cells/mL in suspension for the addition of the baculovirus. The volume of the virus added was calculated using the following equation:

Equation 3.1: Addition of baculovirus.

$$\text{Inoculum required (mL)} = \frac{\text{MOI (PFU/mL)} \times \text{number of cells (cells/mL)}}{\text{titre of viral stock (PFU/mL)}}$$

A multiplicity of infection (MOI) of 0.05 PFU/mL was used and the volume of inoculum depended on the measured titre of viral stock, which varied from batch to batch. The cells were incubated at 27 °C, shaking at 90 rpm for 72 hours. The viability of the cells was calculated pre-purification as this needs to be above 70% to ensure the best yield. Cultures were centrifuged 6000 xg and the pellet retained for protein purification.

3.2.14 Purification of His-tagged human TG6 from *Sf9* cells

The cell pellet was resuspended in lysis buffer (Zedira, confidential) containing protease inhibitors (0.1 mM 4-(2-aminoethyl) benzenesulfonyl fluoride (AEBSF), 10 µM leupeptin hemisulphate, 1 µM aprotinin, 1 µM bestatin). The pellet was sonicated (30 seconds on and 30 seconds off for 10 minutes, amplitude 60%, on ice) and 20% glycerol was added. The cell lysate was centrifuged at 47,800 xg at 4 °C for 30 minutes, and 1 mL of Ni Sepharose resin was added to the supernatant and incubated for 1 hour at 4 °C with agitation using a roller. The resin was recovered by centrifugation (180 xg, 4 °C, 7 minutes) and washed (180 xg, 4 °C, 1 minute) with buffer 1 (Zedira, confidential), buffer 2 (Zedira, confidential) buffer 3 (Zedira, confidential) and buffer 4 (Zedira, confidential). WT TG6 was eluted from the resin using elution buffer (Zedira, confidential). The supernatant was harvested and applied to a PD-10 column containing Sephadex G25 equilibrated in sucrose buffer (Zedira, confidential) to remove the imidazole. Specific protein concentration was measured using a DU800 UV spectrophotometer using absorbance at 280 nm (Section 2.2.9).

3.2.15 SDS-PAGE analysis of WT TG6

Fractions collected throughout the purification were analysed using SDS-PAGE (4-20% Tris-glycine gel). 10 µL of each fraction was analysed (Method as described in Section 2.2.4).

3.2.16 WT TG6 isopeptidase activity

To measure the loss of active WT TG6 throughout the purification, fractions for each of the wash steps were collected and analysed alongside purified WT TG6 for isopeptidase activity. 8.3 µL of each fraction was added in place of 20 µg/mL of TG6 (Method as described in Section 2.2.10) and the TG6 assay buffer was used.

3.2.17 WT TG6 MDC activity.

The MDC activity also measured loss of WT TG6 throughout the purification (Method as described in Section 2.2.11).

3.2.18 Measuring Ca^{2+} binding and GTP binding

Ca^{2+} binding was measured by varying the concentration of $CaCl_2$ (0-32 mM, in a two-fold dilution series), to activate isopeptidase activity using the isopeptidase assay (Methods as described in Section 2.2.10). GTP binding was measured by inhibition of isopeptidase activity using varying concentrations of GTP (0-8 mM, in a two-fold dilution series). As logarithmic graphs cannot display the 0 mM GTP value, the 0 mM GTP was included using the following input equation.

Equation 3.2 : Inclusion of zero value in logarithmic graph

Baseline=IF (HillSlope>0.0, Bottom, Top)
Response=Bottom + (Top-Bottom) / 1+10^ ((LogEC50-X)*Hillslope))
Y=IF (X=ZeroValue, Baseline, Response)

Where the first line defines the baseline of the dose response curve with no added dose, which is equal to the bottom parameter (Bottom) and no added drug is the top parameter is equal to the baseline (Top), which is dependent on a positive or negative Hill Slope. The second line refers to the response for non-zero doses. The third line is the Y value equal to the baseline when X equals the Zero Value. This allows more accurate determination of the EC_{50}.

3.2.19 Assessment of stability of WT TG6 with centrifugation

Lyophilised WT TG6 samples were resuspended, using distilled H_2O, and stored at 4 °C. At 24 hour time points over a 7 day period, the sample was centrifuged to remove large precipitates, and a UV absorbance spectrum was measured.

3.2.20 Assessment of stability of WT TG6 without centrifugation

Stability was measured post purification (day 0). Lyophilised samples were resuspended and incubated at 4 °C over a five day period. Spectra were measured every 24 hours, using a UV spectrophotometer, without centrifugation to measure continuous aggregation of WT TG6.

3.3 RESULTS

3.3.1. Cloning of TG6 with *N*-terminal tags to aid solubility

In order to stably generate large amounts of TG6 in *E. coli* various *N*-terminal tags were investigated (Table 3.6 and Fig. S1). WT TG6 cDNA (2120 bp) was cloned into five different expression vectors each adding a different tag to the TG6 *N*-terminus (Table 3.6).

Table 3.6: Selected plasmids used for the cloning of TG6 cDNA.

Plasmid	Size without TG6 cDNA (bp)	Size with TG6 cDNA (bp)	Fusion Protein	Molecular weight (kDa)
pETMM11-TG6	5371	7491	His-TG6	84
pETMM20-TG6	5693	7813	Trx-His-TG6	95
pETMM41-TG6	6405	8525	His-MBP-TG6	125
pETMM60-TG6	6840	8960	NusA-His-TG6	137
pHSIE-TG6	6497	8617	His-SUMO-intein-TG6 or untagged TG6	125 or 79 (untagged)

The TG6 cDNA was amplified from the pIND TGM6 (plasmid generated in house) using PCR and a band at ~2000 bp (Fig. 3.1A) was gel purified. The individual vectors and TG6 cDNA purity, post digestion with *SalI* and *NotI,* was confirmed by a single band being present in each sample following agarose gel electrophoresis (Fig. 3.1B). Successful ligations were assessed by re-digestion of the recombinant plasmid (Fig. 3.1C). Two bands were identified, one corresponding to TG6 cDNA (~2000 bp) and one corresponding to the vector (~6000 bp).

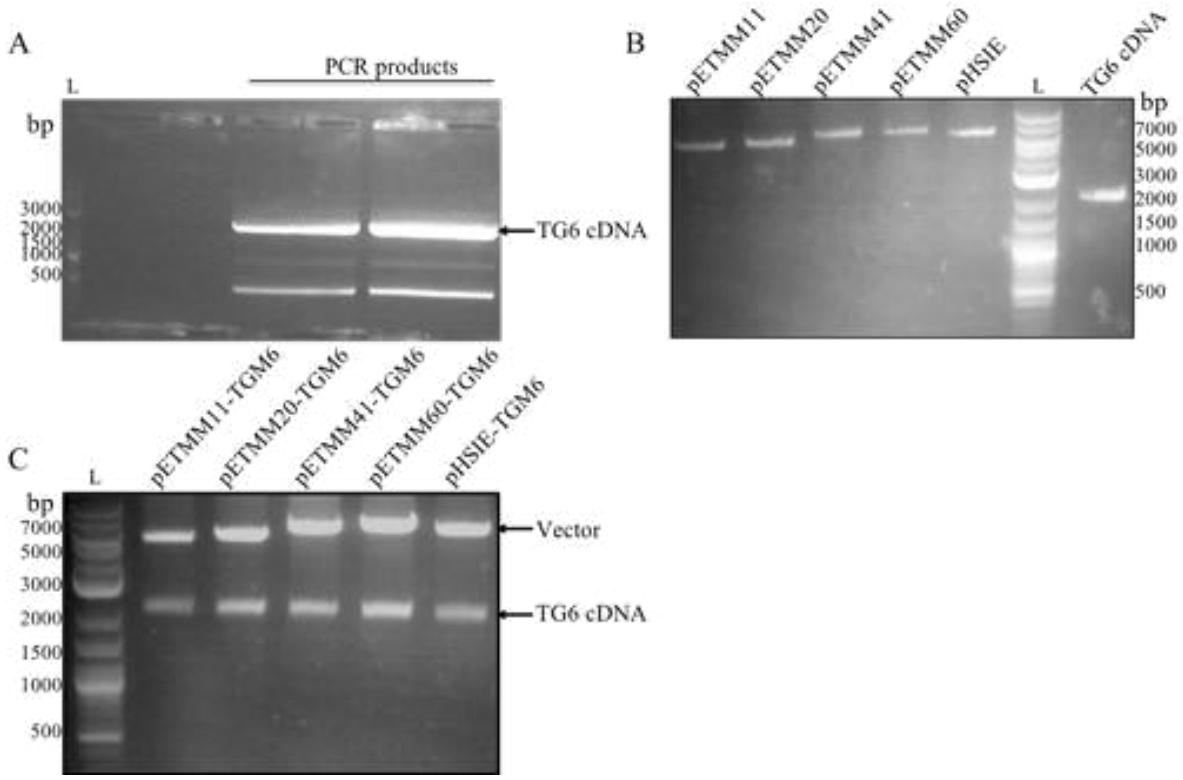


Figure 3.1: TG6 cDNA cloning into selected vectors.
DNA samples were analysed using 1% agarose gel electrophoresis. A: TG6 cDNA was amplified using PCR and the desired band (~2000 bp) was gel purified. B: The amplified TG6 cDNA and the selected expression plasmids were digested using *Sal*I and *Not*I in preparation for ligation. C: Following ligation the generated expression plasmids were digested with *Sal*I and *Not*I to ensure the presence of TG6 cDNA. Data is representative of one replicate (n=1) in one independent experiment (N=1). Work aided by Michail Nomikos and Tony Lai

3.3.2 Expression of TG6 fusion proteins

In order to assess expression of the selected TG6 fusion proteins, samples were taken pre and post induction with IPTG. Samples were analysed using SDS-PAGE followed by Coomassie staining (Fig. 3.2A) or Western blotting (Fig. 3.2B and C). SDS-PAGE followed by Coomassie staining revealed a distinct additional band in the His-MBP-TG6, NusA-His-TG6 and His-SUMO-intein TG6 induced samples, when compared with non-induced controls (Fig. 3.2A). An additional band was not observed in the Trx-His-TG6 and His-TG6 samples (Fig. 3.2A lane His-TG6 and Trx-His-TG6). These two TG6 derivatives have molecular weights of 84 and 95 kDa respectively (Table 3.6) and are masked by other proteins in the cell lysate. Western blot analysis with an anti-His^5 antibody (Fig. 3.2B) confirmed successful expression of all five TG6 fusion proteins. Similar results are seen using the anti-TG6 antibody (Fig. 3.2D), however it did not detect His-TG6 (Fig. 3.2D, lane His-TG6). In all samples, multiple bands were detected with the anti-His^5 antibody. This could be due to degradation of the FL fusion protein rather

than non-specific binding of anti-His[5] as no bands were detected in the non-induced control. A single band is detected in the NusA-His-TG6 and His-SUMO intein-TG6 samples using an anti-TG6 antibody (Fig 3.2D, lanes NusA-His-TG6 and His-SUMO intein-TG6). However, the detected bands, in both cases, do not match their predicted molecular weight 137 and125kDa, respectively (Fig. 3.2B and Table 3.6). This could be due to the different SDS-PAGE gel percentage (4-20%) and molecular weight ladder used. The anti-TG6 antibody detected two bands with Trx-His-TG6, the higher molecular weight band corresponded to the FL TG6 fusion protein (95 kDa, Fig. 3.2D, lane Trx-His-TG6). Multiple bands are detected in His-MBP-TG6 in both anti-His[5] (Fig. 3.2C, lane His-MBP-TG6) and anti-TG6 blots (Fig. 3.2D, lane His-MBP-TG6).

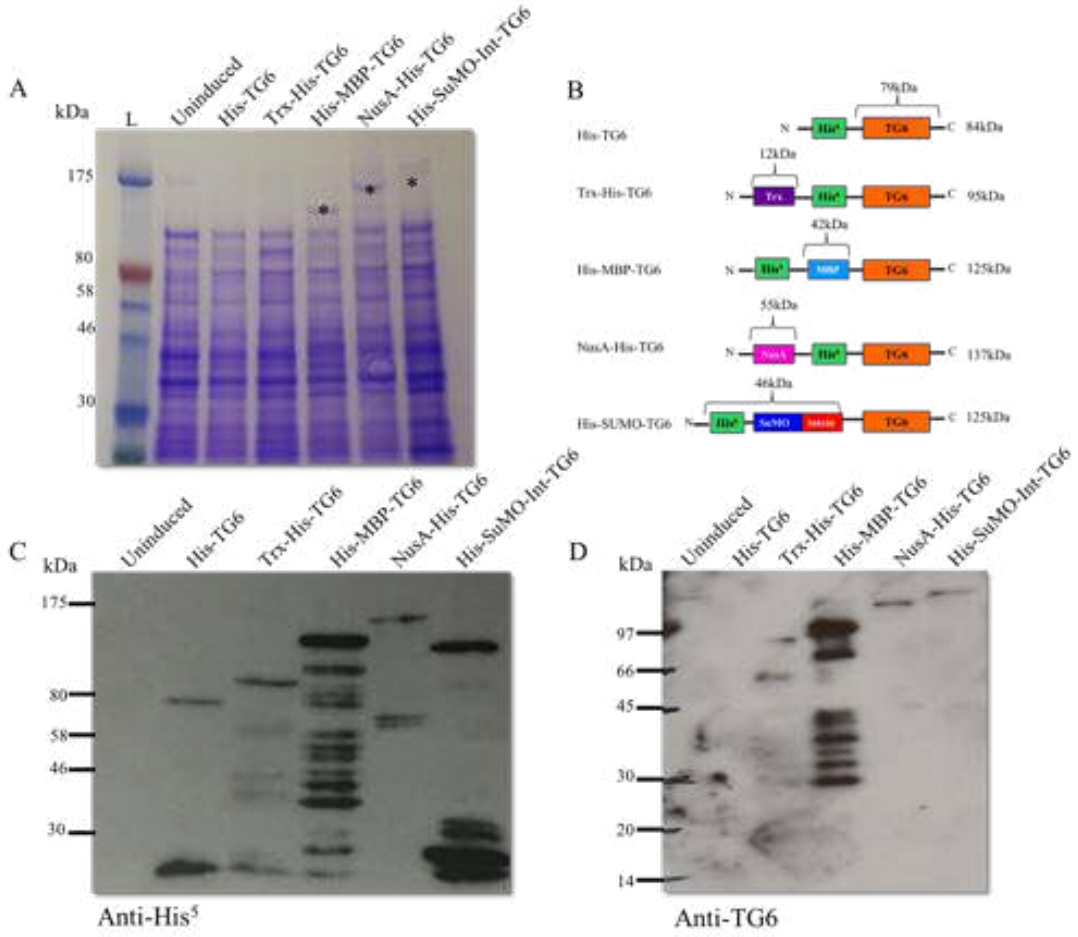


Figure 3.2: Expression of TG6 fusion proteins.
Post-induction (16 hours, 16 °C, 0.1 mM IPTG) for each TG6 fusion protein were collected. A: 20 µL of sample was run on SDS-PAGE (4-20%) under reducing conditions then stained using Coomassie brilliant blue R. Highlighted with * are the extra bands in the post induction samples compared with the uninduced control. B: Schematic of TG6 fusion proteins. C: Western blot (10% SDS-PAGE) of cells lysates. Nitrocellulose was stained with anti His[5] monoclonal antibody. D: Western blot (4-20% SDS-PAGE) of cells lysate samples. Nitrocellulose was stained with anti- TG6 antibody. Data is representative of one replicate (n=1) in one independent experiment (N=1). Work aided by Michail Nomikos and Tony Lai

These results indicate that four out of five (Trx-His-TG6, His-MBP-TG6, NusA-His-TG6 and His-SUMO-intein-TG6) of the TG6 fusion proteins were successfully expressed. However, the expression of His-TG6 is still unclear due to the conflicting results on the anti-His^5 and anti-TG6 blots. The expression of His-TG6 may be below the detection limit of the anti-TG6 antibody due to lower binding affinity when compared with anti-His^5.

3.3.3 Optimising IPTG concentration for TG6 expression

In order to generate higher amounts of TG6 fusion protein, expression was optimised. Altering IPTG concentration will affect the rate of mRNA production and protein biosynthesis, and a lower IPTG concentration can increase the proportion of protein in a stable, 'correctly' folded state (Lebendiker and Danieli 2014). Expression cultures were prepared and induced using final concentrations of IPTG in the 0 to 1 mM range. Cell lysates were collected for SDS-PAGE and Western blot analysis (Fig. 3.3). His-MBP-TG6 showed good expression when using IPTG at 0.1, 0.5 and 1 mM as determined by Western blots for the His-tag and TG6, respectively (band at ~125 kDa) (Fig. 3.3A, C and E). The expression level did not increase further at IPTG concentrations > 0.1 mM which is most evident in the anti-TG6 blot (Fig. 3.3E, lane 0.1mM IPTG). For NusA-His-TG6 (Fig. 3.3B, D and F) similar results were seen. Also, there is a high degree of degradation of both TG6 fusion protein preparations as revealed by the multiple bands detected by the anti-His^5 (Fig. 3.3D) and anti-TG6 (Fig. 3.3F).

3.3.4 Optimising temperature of induction of TG6 expression

To reduce the degradation of the TG6 fusion proteins during expression but maintain the high expression level, different induction temperatures were investigated (16 °C (16 hours), 30 °C (4 hours) and 37 °C (3 hours)). For His-MBP-TG6, fragmentation products were detected by both Coomassie staining (Fig 3.4A) and Western blotting (anti-His^5 and anti-TG6, Fig. 3.4C and E) in samples induced with 0.1 and 0.5 mM IPTG. His-MBP-TG6 showed higher expression when induced overnight at 16 °C but increased degradation was evident (Fig 3.4C and E, lanes 0.1 and 0.5 mM IPTG at 16 °C). Interestingly, at both higher induction temperatures with shorter induction times (30 °C and 37 °C for 3 hours and 4 hours respectively), His-MBP-TG6 was more stable (Fig. 3.4 C and E, lanes 0.1 and 0.5 mM IPTG at 30 °C and 37 °C). NusA-His-TG6 showed the

highest expression following induction at 16 °C for 16 hours, but greater fragmentation was observed by Western blot (Fig 3.4D and F, lanes 0.1 and 0.5 mM IPTG at 16 °C).

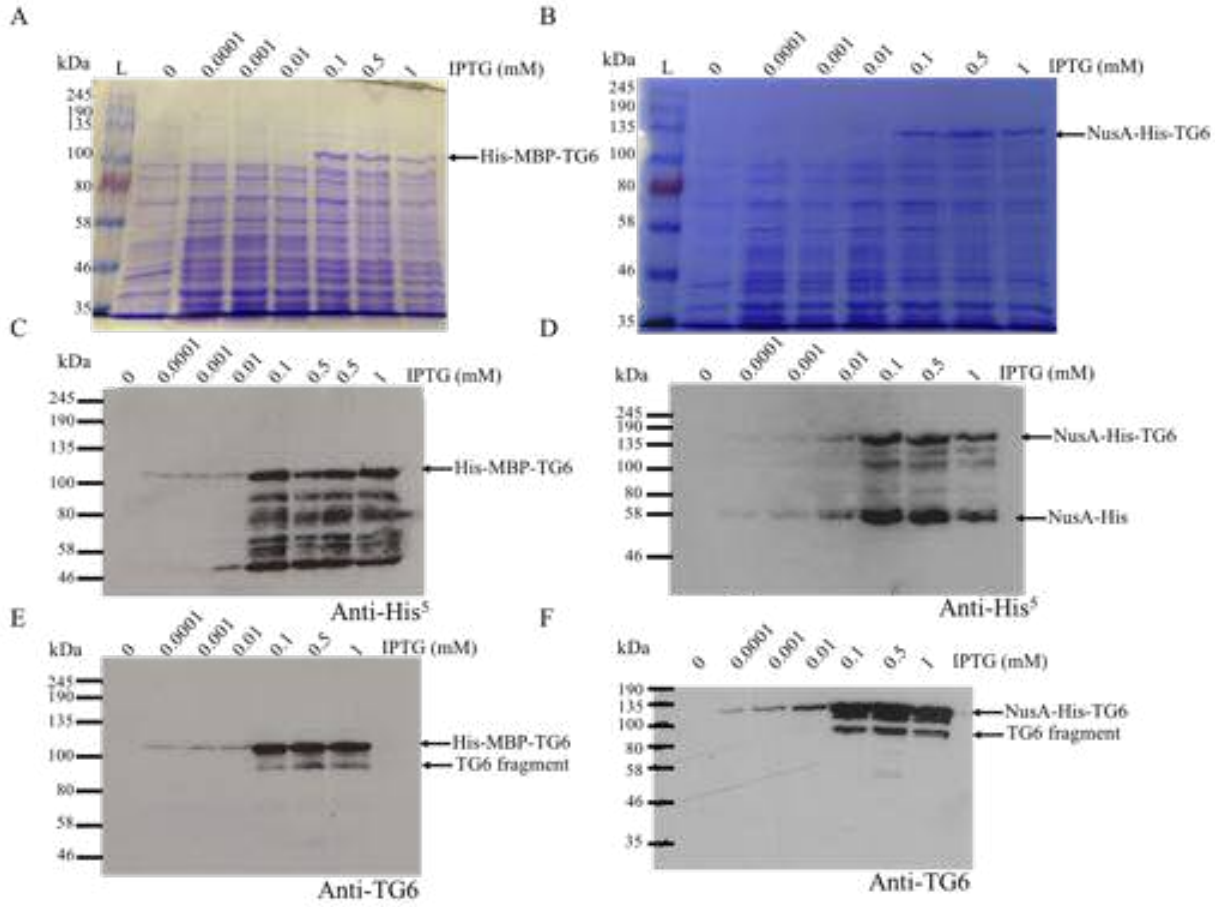


Figure 3.3: Optimisation of IPTG concentration.
TG6 expression was induced with IPTG (0-1 mM) for 16 hours at 16 °C. Cell lysates for each TG6 fusion protein (His-MBP-TG6 and NusA-His-TG6) were collected. Data is representative of one replicate (n=1) in one independent experiment (N=1). A: 20 µL of His-MBP-TG6 cell lysates were analysed by SDS-PAGE (7.5%) under reducing conditions then stained using Coomassie brilliant blue R. B: As in A but NusA-His-TG6 cell lysates were analysed. C: Western blot (7.5% SDS-PAGE) of His-MBP-TG6 cell lysates. Nitrocellulose was stained with anti His^5 monoclonal antibody. D: Western blots as in C but NusA-His-TG6 cell lysates were analysed. E: Western blot (7.5% SDS-PAGE) of His-MBP-TG6 cell lysate but nitrocellulose was stained with anti-human TG6 polyclonal antibody. F: Western blot as in E but NusA-His-TG6 samples were analysed. Work aided by Michail Nomikos and Tony Lai

As observed for the His-MBP-TG6 sample, less degradation was observed at 30 °C and 37 °C when incubated for 3-4 hours (Fig 3.4D and F, lanes 0.1 and 0.5mM IPTG at 30 °C and 37 °C) indicating NusA-His-tagged TG6 was more stable in these conditions. This result could be due to the reduced incubation time as 37 °C also shows the lowest level of FL TG6 fusion protein. A strong band at ~58 kDa is observed in all samples and could be the cleaved His-NusA tag (58 kDa). This result indicates that the FL TG6 fusion protein is highly unstable in any condition (Fig 3.4D).

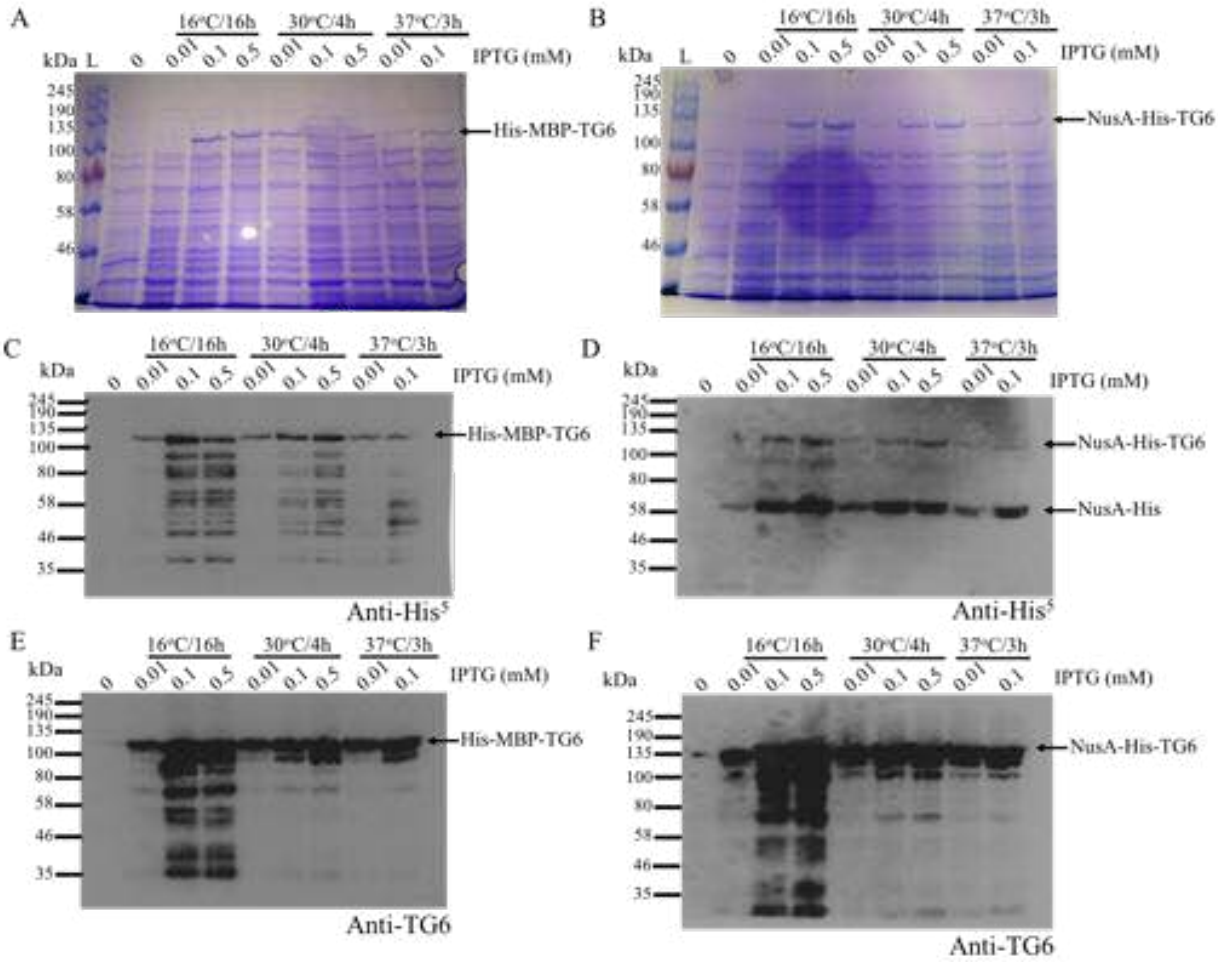


Figure 3.4 Optimisation of induction temperature.
TG6 expression was induced with IPTG (0.1-1 mM) and varying temperatures with respect to time (16 °C, 16 hours, 30 °C, 4 hours and 37 °C, 3 hours). Cell lysates for each TG6 fusion protein (His-MBP-TG6 and NusA-His-TG6) were collected. Data is representative of one replicate (n=1) in one independent experiment (N=1). A: 20 µL of His-MBP-TG6 cell lysate was analysed by SDS-PAGE (7.5%) under reducing conditions then stained using Coomassie brilliant blue R. B: Western blot as in A but NusA-His-TG6 cell lysates were analysed. C: Western blot (7.5% SDS-PAGE) of His-MBP-TG6 cell lysates. Nitrocellulose was stained with anti His^5 monoclonal antibody. D: Western blots as in C but NusA-His-TG6 cell lysates were analysed. E: Western blot (7.5% SDS-PAGE) of His-MBP-TG6 cell lysates but nitrocellulose was stained with anti-human TG6 polyclonal antibody. F: Western blot as in E but NusA-His-TG6 samples were analysed. Work aided by Michail Nomikos and Tony Lai

As a compromise between FL TG6 fusion protein expression and degradation the optimum temperature was determined to be 30 °C for 4 hours. However, higher temperatures have been linked with higher risk of the formation of inclusion bodies. Consequently, protein expression at lower temperatures (16 and 20 °C) but with a reduced induction time to hinder degradation may be beneficial for these fusion proteins.

3.3.5 Optimising induction time of TG6 expression

As previous optimisation steps showed TG6 degradation occurs at 16 °C for 16 hours, samples were taken every 2 hours post-induction for 24 hours. Induction was also carried out at 20 °C as it works effectively for TG2. Samples collected at 0-8 hour time points were analysed and revealed that expression of FL TG6 fusion protein accumulates from 2 to 8 hours (Fig. 3.5). However, both His-MBP-TG6 and NusA-His-TG6 degradation increases with time, but less fragmentation is detected in the anti-TG6 blot (Fig. 3.5C and E). The lower molecular weight band (Fig. 3.5C and E) for His-MBP-TG6 is seen in both blots so it can be assumed that this is a His-MBP-TG6 fragment, and cleavage has occurred at the *C*-terminus of TG6. The optimum induction time was selected as 6 hours being the best compromise between expression and degradation. Similar results were observed for NusA-His-TG6 (Fig. 3.5B, D and E). In this case the anti-His^5 blot showed increased degradation at both 16 °C and 20 °C over time (Fig. 3.5D) but a single band was detected in the anti-TG6 Western blot (Fig. 3.5F). This result indicates that the bands detected by the anti-His^5 antibody in these samples are contaminants.

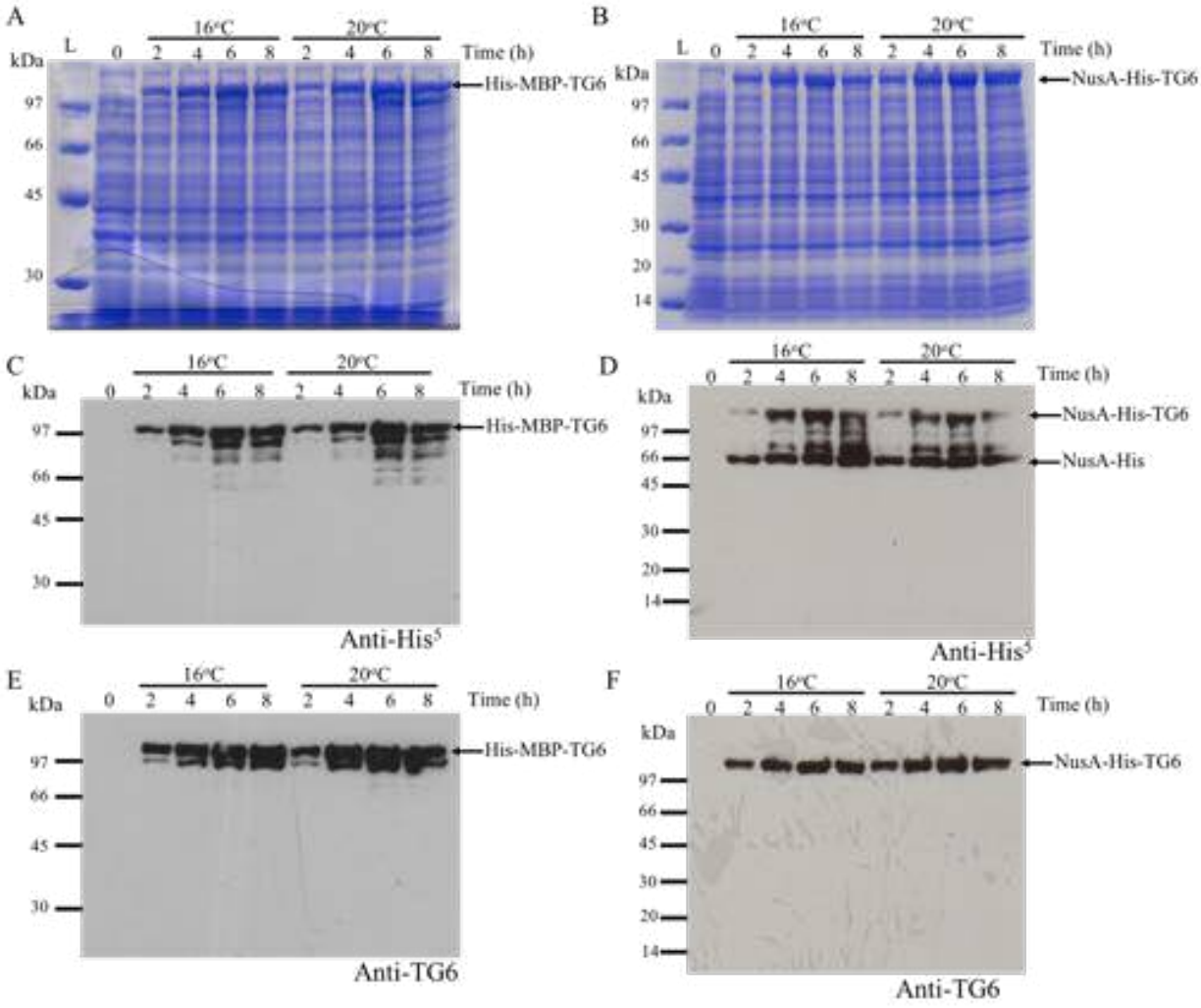


Figure 3.5: Optimisation of induction time.
TG6 expression was induced for varying lengths of time (0-8 hours) at 16 °C and 20 °C. Cell lysates for each TG6 fusion protein (His-MBP-TG6 and NusA-His-TG6) were collected. Data is representative of one replicate (n=1) in one independent experiment (N=1). A: 20 µL of His-MBP-TG6 cell lysates were analysed by SDS-PAGE (4-20%) under reducing conditions then stained using Coomassie brilliant blue R. B: SDS-PAGE as in A but NusA-His-TG6 cell lysates were analysed. C: Western blot (4-20% SDS-PAGE) of His-MBP-TG6 cell lysates. Nitrocellulose was stained using anti His^5 monoclonal antibody. D: Western blots as in C but NusA-His-TG6 cell lysates were analysed. E: Western blot (4-20% SDS-PAGE) of His-MBP-TG6 cell lysates but nitrocellulose was stained with anti-human TG6 polyclonal antibody. F: Western blot as in E but NusA-His-TG6 samples were analysed.

3.3.6 Initial purification using affinity chromatography.

Following the expression of TG6 fusion proteins in *E. coli,* the five TG6 fusion proteins were purified using affinity chromatography. Samples were taken throughout each purification (e.g. supernatant, flow-through, wash steps and elution steps) and analysed using SDS-PAGE followed by Coomassie R staining (Fig. 3.6).

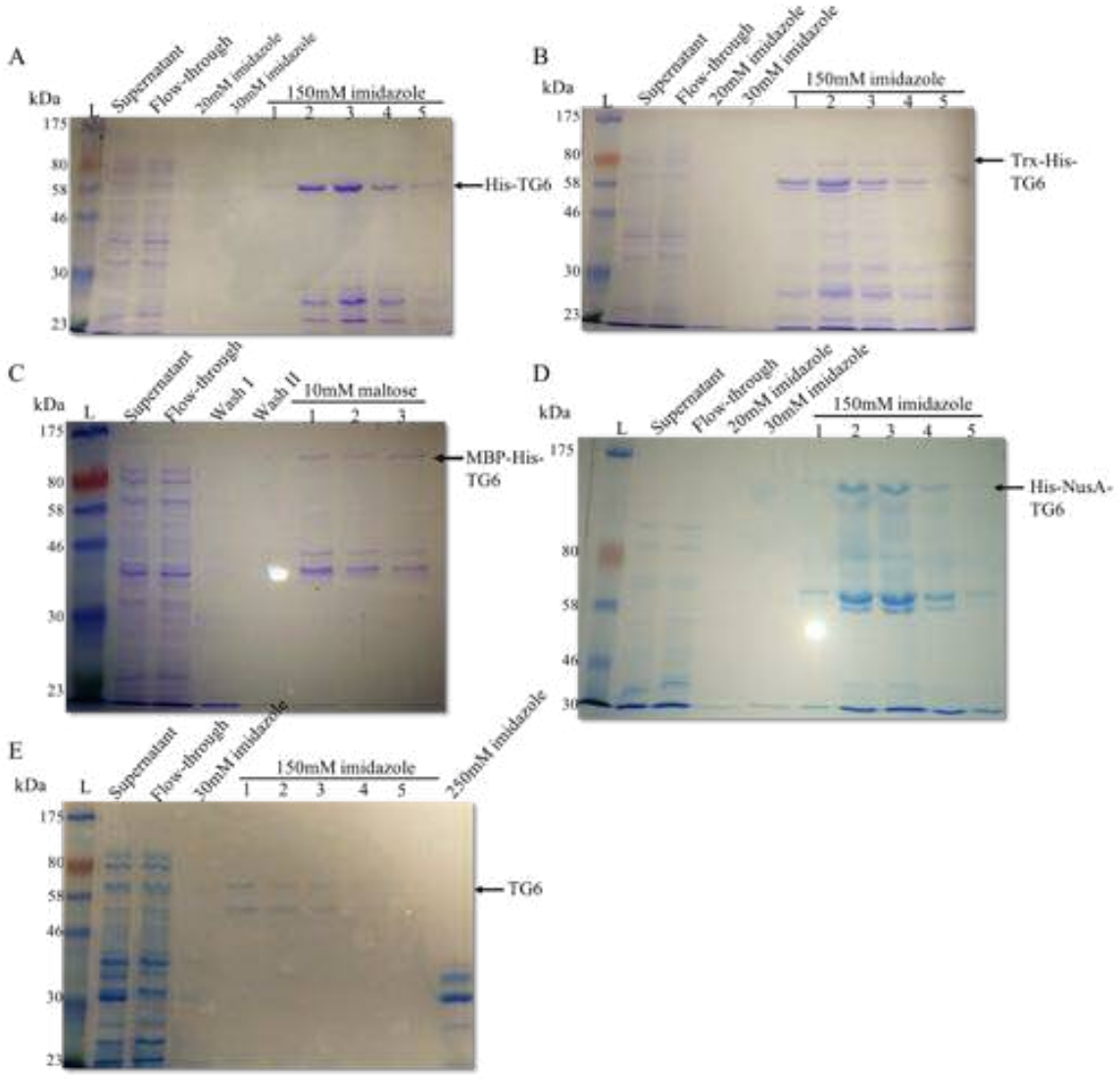


Figure 3.6: SDS-PAGE analysis of the TG6 fusion protein fractions.
20 µL of each purification step were analysed by SDS-PAGE under reducing conditions, then stained using Coomassie brilliant blue R staining. Data is representative of one replicate (n=1) in one independent experiment (N=1). A: 10% SDS-PAGE analysis of His-TG6 purification. B: 10 % SDS-PAGE analysis of Trx-TG6 purification. C: 10% SDS-PAGE analysis of His-MBP-TG6 purification. D: 7.5% SDS-PAGE analysis of NusA-His-TG6 purification. E: 7.5% SDS-PAGE analysis of untagged TG6 purification (cleaved from His-SUMO-intein-TG6). Work aided by Michail Nomikos and Tony Lai

The eluted fractions were then combined and concentrated for comparative analysis using SDS-PAGE and Western blot analysis (Fig. 3.7). Multiple bands were present in all five fusion protein samples. Western blots, using either anti-His[5] or anti-TG6, confirmed the major band corresponding to each fusion protein (Fig. 3.7B and C).

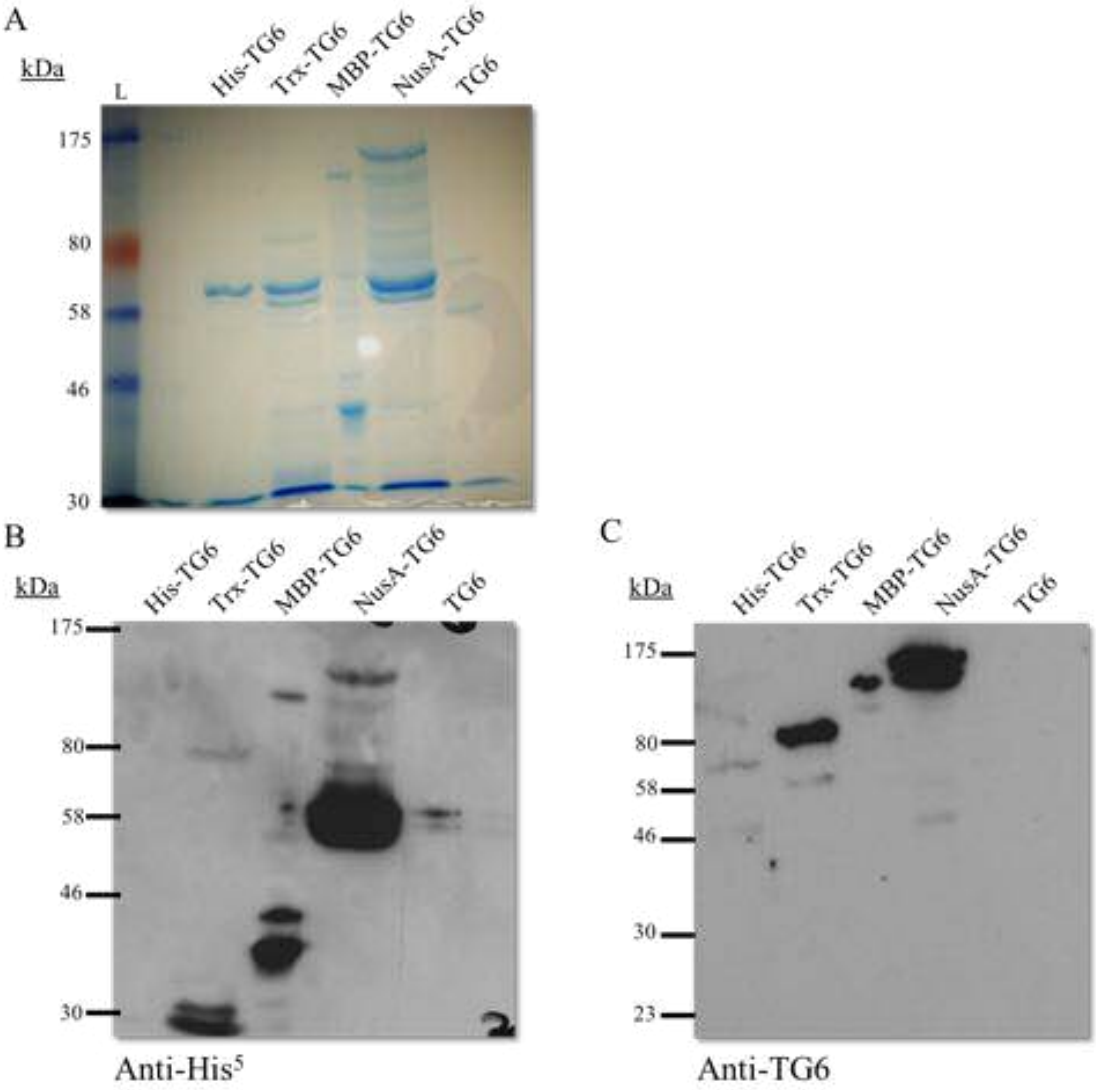


Figure 3.7: SDS-PAGE and Western blot analysis of purified TG6 fusion proteins.
TG6 fusion proteins were purified by affinity chromatography and concentrated.TG6 represents untagged TG6 where by the His-SUMO-Intein tag has been cleaved. Data is representative of one replicate (n=1) in one independent experiment (N=1). A: 20 µL of sample was analysed by SDS-PAGE (9%) under reducing conditions then stained using Coomassie brilliant blue R. B: Western blot (9% SDS-PAGE) of purified TG6 fusion proteins. Nitrocellulose was stained with anti His[5] monoclonal antibody. C: Same as in B but the nitrocellulose was stained with anti-human TG6 polyclonal antibody. Work aided by Michail Nomikos and Tony Lai

The His-TG6 sample showed a band is detected in the Coomassie stained gel at ~73 kDa which is not representative of the predicted molecular weight of 84 kDa (Fig. 3.7A, lane His-TG6). This band is also detected in the anti-TG6 Western blot (Fig. 3.7C, lane His-TG6) but not the anti-His^5 blot (Fig. 3.7B, lane His-TG6). This result suggests that the band is TG6 which no longer harbours the His-tag.

Analysis of the Trx-His-TG6 preparation reveals a weak band at ~92 kDa, indicative of the FL fusion protein (95 kDa) by Coomassie staining (Fig. 3.7A, lane Trx-His-TG6), and anti-His and anti-TG6 blots (Fig. 3.7B and C, lane Trx-His-TG6). Other contaminating bands at ~73 kDa and ~29 kDa were also observed in the Coomassie-stained gel. The band at ~73 kDa was detected by Western blotting using the anti-TG6 antibody (Fig. 3.7C lane, Trx-His-TG6) and the 30 kDa band was detected in the anti-His^5 blot (Fig. 3.7B, lane Trx-His-TG6). Taken together, the results indicate fragmentation of FL Trx-His-TG6.

His-MBP-TG6 samples showed a band at ~122 kDa indicative of His-MBP-TG6 (125 kDa) (Fig. 3.7A, lane His-MBP-TG6) which is detected in both anti-His^5 and anti-TG6 Western blots (Fig. 3.7B and C, lane His-MBP-TG6). Another band in this sample is ~42 kDa (Coomassie and anti-His^5, Fig 3.7A and B, lane His-MBP-TG6) which could represent His-MBP (46 kDa).

In the Coomassie-stained gel, a band was detected at ~137 kDa, representative of NusA-His-TG6 (137 kDa) (Fig. 3.7A, lane NusA-His-TG6). This result was confirmed using the anti-His and anti-TG6 antibodies (Fig. 3.7B and C, lane NusA-His-TG6). An additional band was detected at ~60 kDa, in both the Coomassie and anti-His^5 blot (Fig. 3.7A and B, lane NusA-His-TG6), which could be His-NusA (58 kDa). As the His-SUMO-intein tag had been cleaved during purification, untagged TG6 was generated with a molecular weight of 79 kDa. The Coomassie gel detected a band at ~60 kDa, indicative of untagged TG6 (Fig. 3.7A, lane TG6) but this protein was not detected in the anti-His^5 (as expected) and anti-TG6 blots (Fig. 3.7B and C, lane TG6). Also, a contaminating band is present at 58 kDa in the Coomassie and the anti-His^5 blot (Fig 3.7A and B, lane TG6) but not in the anti-TG6 blot (Fig. 3.7C, lane TG6).

From the five generated fusion proteins two were selected for further investigation: His-MBP-TG6 and NusA-His-TG6 as these showed the best expression from cell lysate preparations (Fig. 3.2 and 3.7). Following purification, Trx-His-TG6 detection by immunoblotting was inconsistent (Fig. 3.7B and C, lane Trx-His-TG6) and the levels of purified His-TG6 and untagged TG6 (from the processed SUMO-tagged fusion protein) were very low as determined by Western blotting (Fig 3.7B and C, lanes His-TG6 and TG6) and therefore were not taken further.

3.3.7 Optimising the lysis buffer: NaCl concentration

Following optimisation of the expression conditions and the completion of purification trials, the purification conditions were optimised for His-MBP-TG6. Buffers with differing NaCl concentration (0 mM to 500 mM) were used to resuspend the bacterial pellets following centrifugation and the purified His-MBP-TG6 was analysed (Fig. 3.8). Irrespective of the NaCl concentration in the extraction buffer a large amount of contaminating or degraded protein still remains, (Fig. 3.8A). The optimal NaCl concentration was 0 mM as shown by Western blot (Fig. 3.8B, lane 0mM). In addition, distilled H_2O was sufficient to give the best yield of His-MBP-TG6. Due to the buffering capacity of Tris and therefore the greater reliability and reproducibly of the conditions, the buffer containing 20 mM Tris-HCl pH 7.4, 0 mM NaCl was selected for further optimisation.

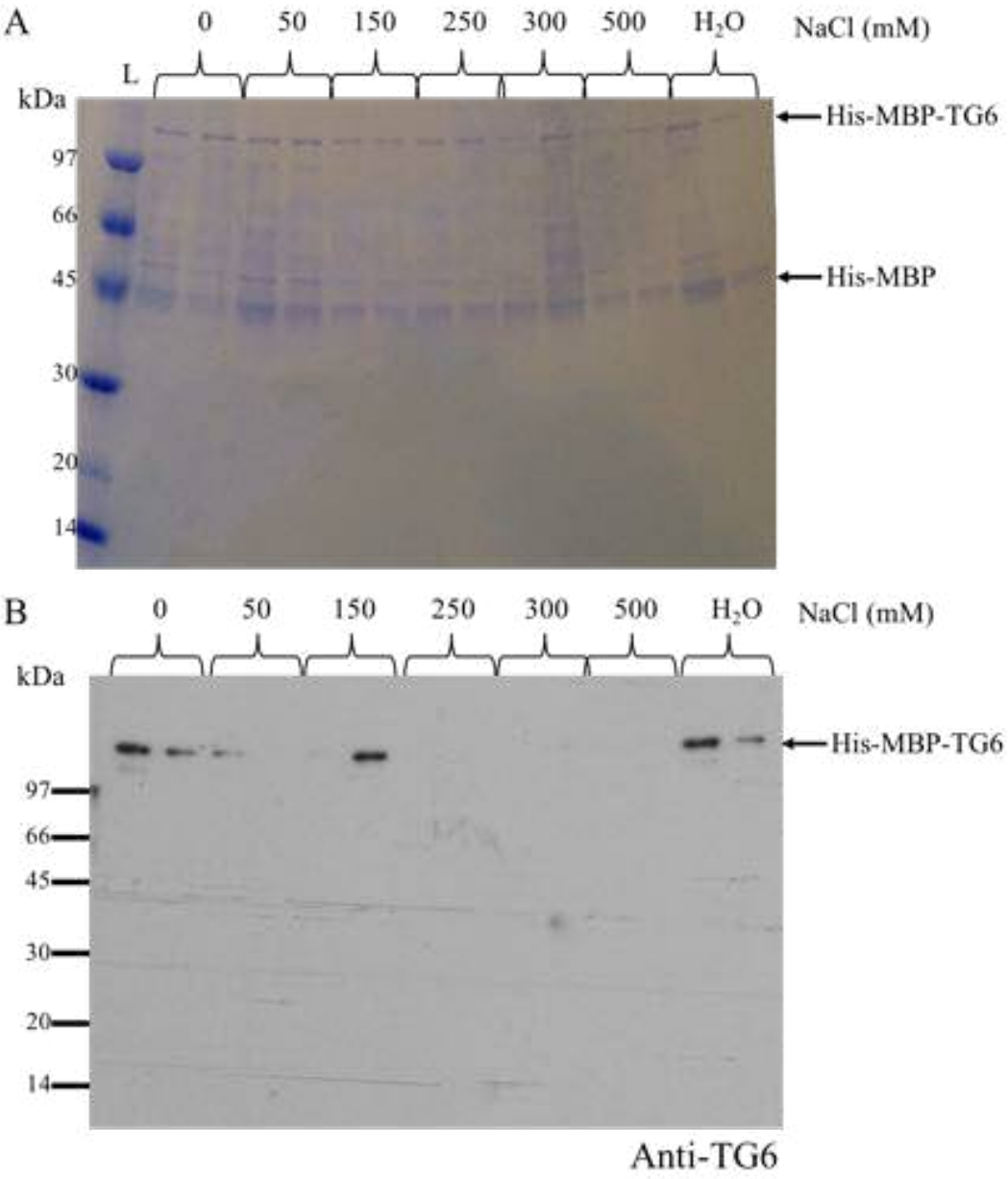


Figure 3.8: Optimisation of NaCl concentration in the lysis buffer.
His-MBP-TG6 was purified using different concentrations of NaCl in the lysis buffer, protein was eluted in two fractions. Data is representative of one replicate (n=1) in one independent experiment (N=1). A: 20 µL of the purified His-MBP-TG6 samples were analysed by SDS-PAGE (4-20%) under reducing conditions then stained using Coomassie brilliant blue R. B: Western blot analysis (4-20%) of purified His-MBP-TG6 samples stained with anti-TG6 antibody.

3.3.8 Optimisation of lysis buffer: pH

Following the optimisation of the NaCl concentration, the pH of the 20 mM Tris, 0 mM NaCl lysis buffer was altered from pH 6.0 to pH 8.0. 50 mM NaH_2PO_4 was selected as the buffering agent for buffers with a pH of 6.0-7.0. SDS-PAGE showed

minimal amounts of His-MBP-TG6 were isolated with buffers at a pH of 6.0, 6.5 and 7.0 (Fig. 3.9A, lanes 6.0, 6.5 and 7.0), which was confirmed by Western blot (Fig. 3.9B, lanes 6.0, 6.5 and 7.0). The optimal pH for protein isolation was pH 8.0 as this had the densest band on both the SDS-PAGE and Western blot. As glycerol has previously facilitated the stability of TG6, it was also included in the buffer at pH 8.0 (Fig 3.9A lanes 8.0 and 8.0 + glycerol). However, a minimal change in the extent of protein extraction was seen with the addition of glycerol. The lysis buffer used for subsequent purifications was 20 mM Tris-HCl pH 8.0, 0 mM NaCl.

3.3.9 Addition of components to the lysis buffer

To further improve the recovery of His-MBP-TG6, the impact of the addition of non-ionic detergents (1 mM Octyl- β-glucopyranoside, 1% TritonX-100 or 0.5% Tween-20), and 2 mM $MgCl_2$ was assessed. These conditions were compared with the addition of 20% glycerol. Non-ionic detergents aid the lysis of the bacterial cell membrane without interfering excessively with protein structure. $MgCl_2$ can aid in solubilisation of proteins and glycerol has previously been shown to aid TG6 purification and solubilisation (unpublished observation, D. Aeschlimann). All purified samples were analysed by SDS-PAGE and Western blot (Fig. 3.10). Addition of any of the non-ionic detergents lead to an increase in the amount of His-MBP-TG6 isolated compared with the no additives control (Fig. 3.10B, lanes Octyl-β, Triton-X100 and Tween-20). Triton-X100 was the most effective detergent and produced a dense band.

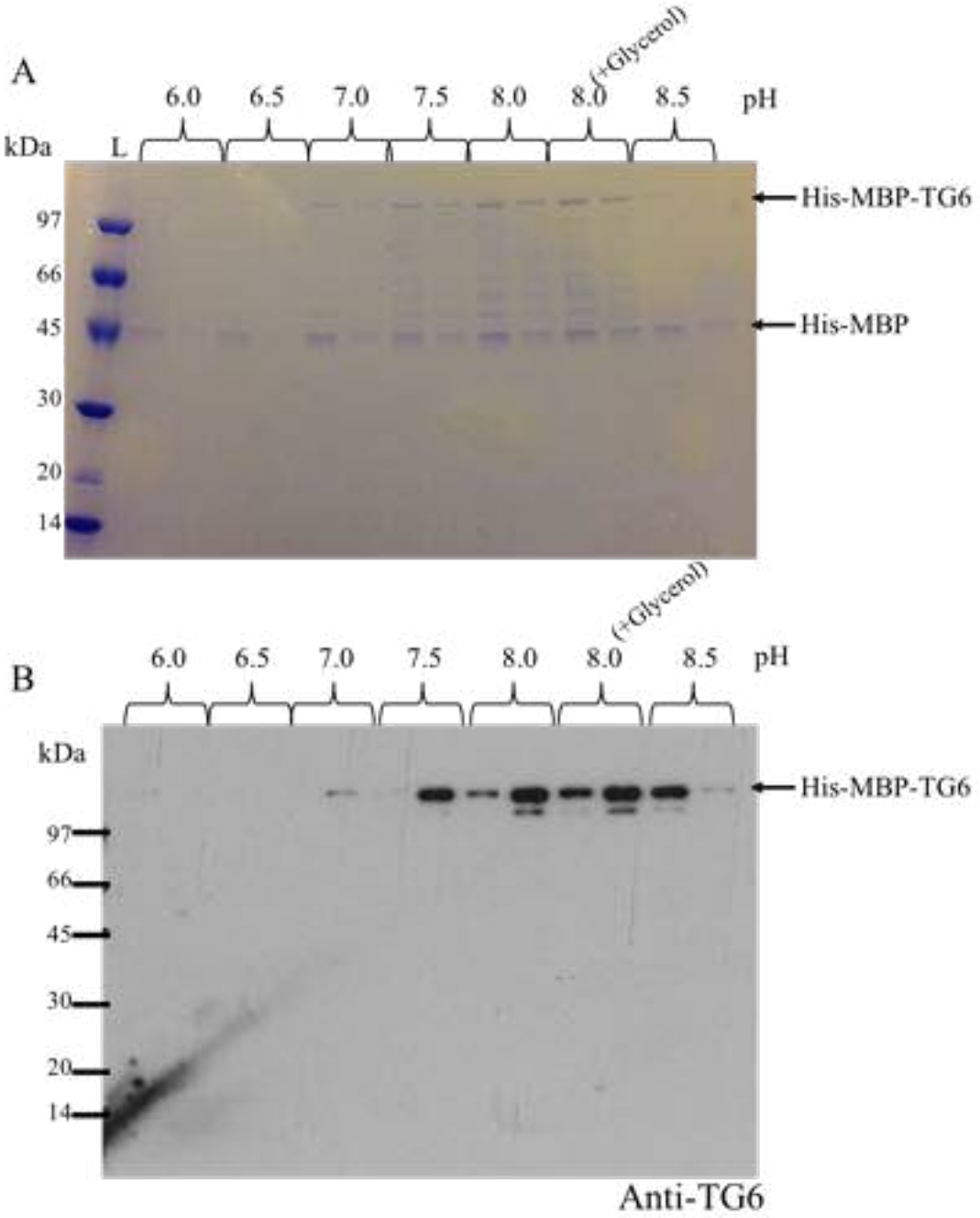


Figure 3.9 Optimisation of the pH of the lysis buffer.
His-MBP-TG6 was extracted using lysis buffer containing 0 mM NaCl and at pH values ranging from pH 6.0 to pH 8.0 in 0.5 pH unit increments. The protein was eluted in two fractions. Data is representative of one replicate (n=1) in one independent experiment (N=1). A: 20 µL of the purified His-MBP-TG6 samples were analysed by SDS-PAGE (4-20%) under reducing conditions then stained using Coomassie brilliant blue R. B: Western blot analysis (4-20%) of purified His-MBP-TG6 samples stained with anti-TG6.

The addition of $MgCl_2$ showed no improvement in protein extraction compared to the non-additive control (Fig 3.10B, lane $MgCl_2$). However, the amount of His-MBP-TG6 isolated in the presence of 20% glycerol (Fig 3.10B lane glycerol) is similar to the non-ionic detergents. Due to possible effects of non-ionic detergents on protein folding, glycerol was the preferred additive.

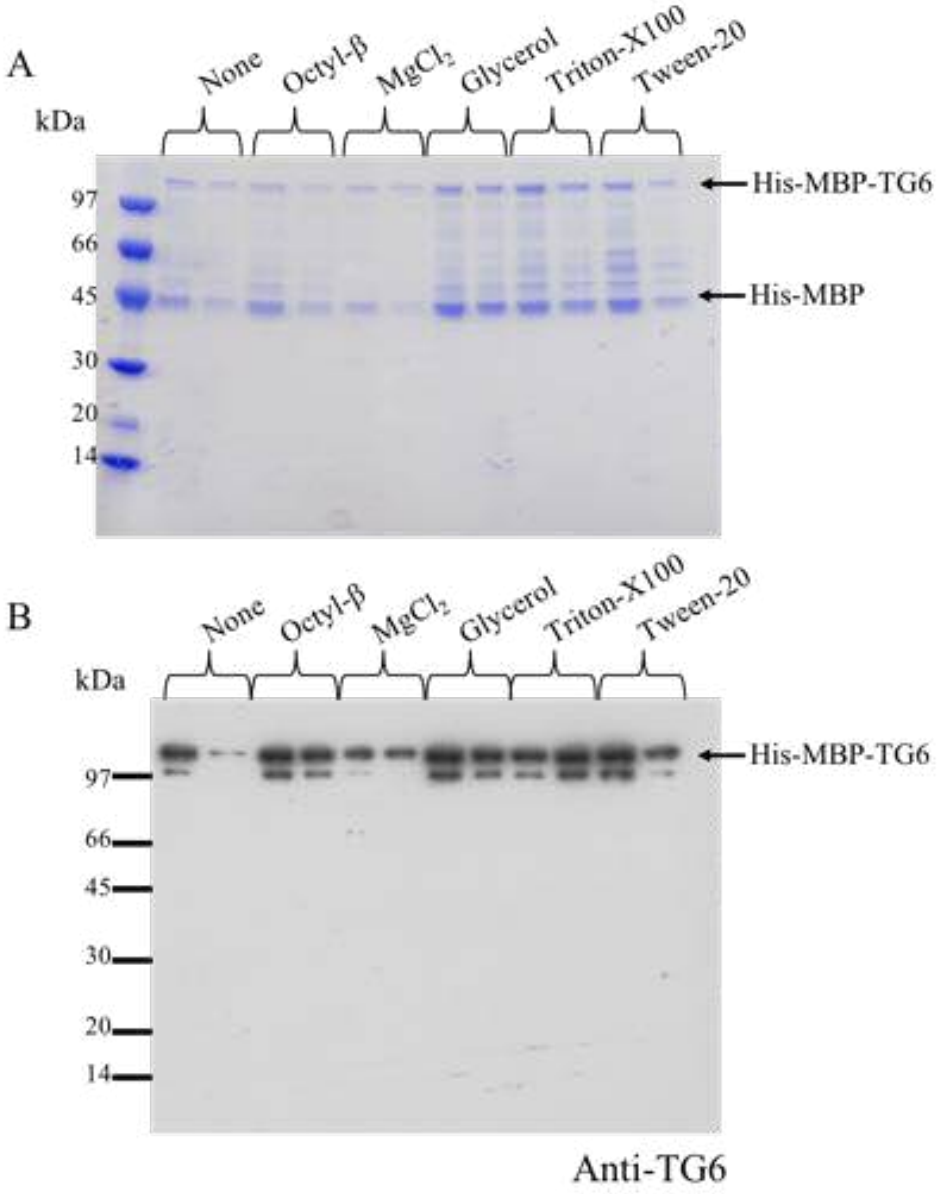


Figure 3.10: Optimisation of the lysis buffer with additives.
His-MBP-TG6 was extracted using lysis buffer (20 mM Tris-HCl pH 8.0, 0mM NaCl) with additional components (1 mM Octyl- β-glucopyranoside, 1% Triton X100, 0.5% Tween-20, 2 mM $MgCl_2$ and 20% glycerol) added to test for improvements in protein extraction. Data is representative of one replicate (n=1) in one independent experiment (N=1). A: 20 µL of the purified His-MBP-TG6 samples were analysed by SDS-PAGE (4-20%) under reducing conditions then stained using Coomassie brilliant blue R. B: Western blot (4-20%) of purified His-MBP-TG6 samples. Nitrocellulose was stained with anti-TG6 polyclonal antibody.

3.3.10 Activity of tagged TG6

In order to assess the impact of the solubility tag on the activity of TG6, the isopeptidase activity was measured for the respective TG6 constructs. The concentration the TG6 fusion protein was measured using a UV Spectrophotometer (Fig. 3.11A). The Trx-His-TG6 curve shows a significant increase in signal intensity at 310 nm indicative of protein aggregation. In contrast, His-MBP-TG6 and NusA-His-TG6 have low absorbance at 310 nm, indicating of little or no aggregation. The increase in aggregation seen with Trx-His-TG6 could indicate concentration-dependent aggregation as the concentration of Trx-His-TG6 is higher than that of His-MBP-TG6 and NusA-His-TG6.

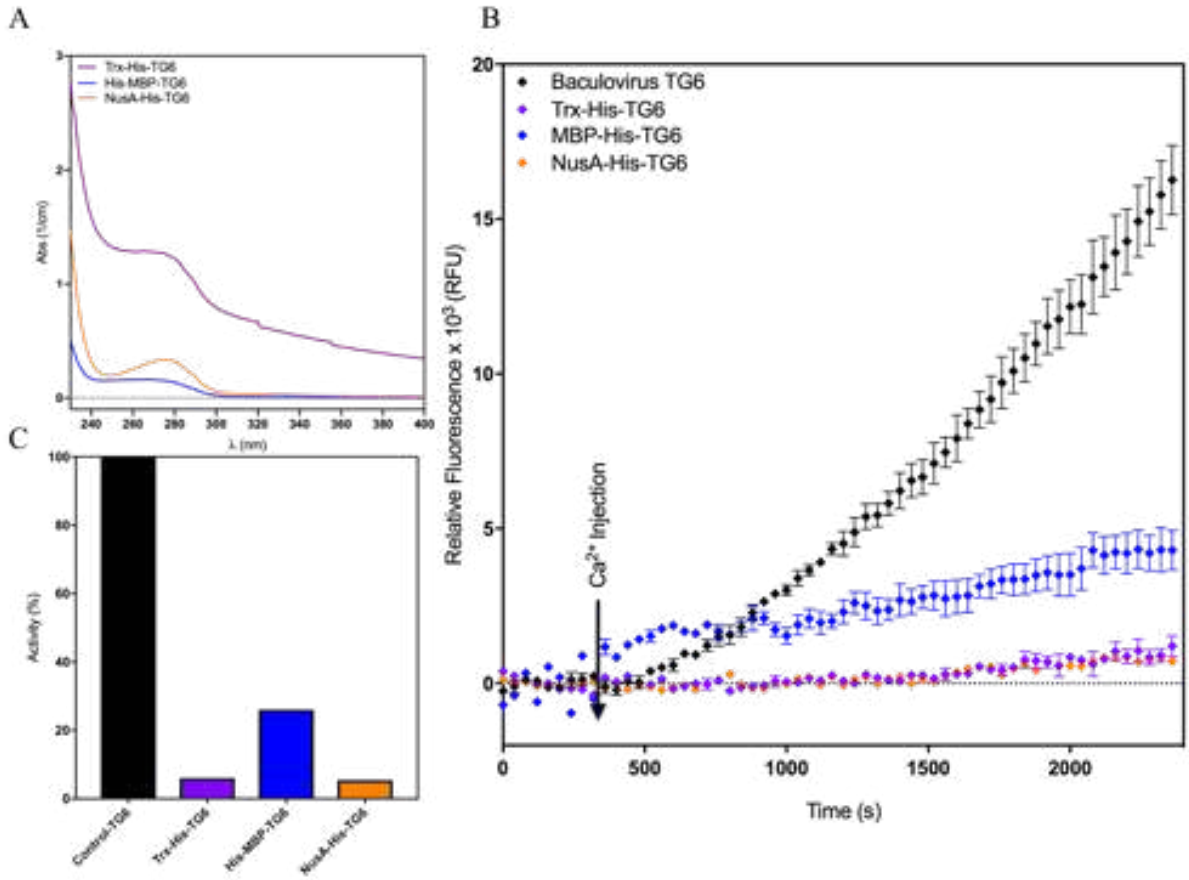


Figure 3.11: Isopeptidase activity of TG6 fusion proteins.
A: Measured protein concentration of concentrated TG6 fusion proteins using UV spectrophotometer (1:100 dilution). B: Isopeptidase activity assay of TG6 fusion proteins and control-His-TG6. Data are the mean of 2 wells ± SD (n=2) and representative of 1 independent experiment (N=1). C: The percentage isopeptidase activity of the TG6 fusion proteins compared to the TG6 control.

Next, the enzymatic activities of the purified fusion proteins (~20 μg/mL) were compared to TG6 generated using the baculovirus/*Sf9* expression system (positive control, Fig. 3.11B). The activity assay showed a much lower relative fluorescence increase for the tagged TG6s when compared with the TG6 control (Table 3.7). The rates of reaction were measured using linear regression and the activities of the TG6 fusion proteins were also expressed as percentage activity, where the TG6 control represented 100% activity (Table 3.7 and Fig. 3.11C). The lower activities of the TG6 fusion proteins are likely due to the incomplete purification of the TG6 fusion proteins: the measured protein concentration is not representative of TG6 alone. In addition, interference from the large solubility tag may also have contributed to the reduced activity.

Table 3.7: Kinetic properties of TG6 fusion proteins.

Protein	Rate of reaction (RFU/s)	Percentage activity (%)
Trx-His-TG6	0.44 ± 0.31	6.03
His-MBP-TG6	1.89 ± 0.74	26.0
NusA-His-TG6	0.40 ± 0.24	5.42
TG6 control	7.26 ± 0.18	100

Although His-MBP-TG6 was successfully extracted, very little FL His-MBP-TG6 was isolated and the preparation was not pure. Overall, the yield of this protein was low. Its relative isopeptidase activity and that of NusA-His-TG6 were both low when compared to baculovirus/*Sf9* cell generated control TG6. Taken together these results indicate that the existing baculovirus/*Sf9* system is a more successful method of generating TG6. The extent to which solubility tags aid expression cannot be confirmed as sufficient protein was not generated. However, the lack of FL protein due to fragmentation suggests the TG6 fusion proteins were unstable and therefore not soluble. As a result, the future generation of WT TG6 was carried out using the baculovirus/*Sf9* cell system

3.3.11 Generation of purified WT TG6

To biochemically characterise WT TG6, it was generated using the baculovirus/*Sf9* cell expression system. Following purification, the purity of TG6 was confirmed using SDS-PAGE analysis (Fig. 3.12A) and the protein concentration was measured using UV spectroscopy (Fig. 3.12B). A single band between 97 and 66 kDa was present corresponding to the predicted molecular weight of His-tagged TG6 of 80 kDa. Multiple samples were also taken during the purification steps to identify where TG6 was lost. A single band corresponding to TG6 was observed in the wash 3 and wash 4 centrifugation steps indicating loss of TG6 in these conditions.

The TG6 isopeptidase activity was measured to confirm that it is correctly folded and active (Fig. 3.12C, Adamczyk *et al*. 2013). Activity was seen for WT TG6 and significantly lower activity was measured in the wash 3 and wash 4 samples, which correspond to the TG6 bands observed by SDS-PAGE. As expected, activity was not detectable in the other purified fractions (supernatant, wash 1:1, wash 2 and wash 1:2). Similar results were seen with the MDC assay (Fig. 3.12D).

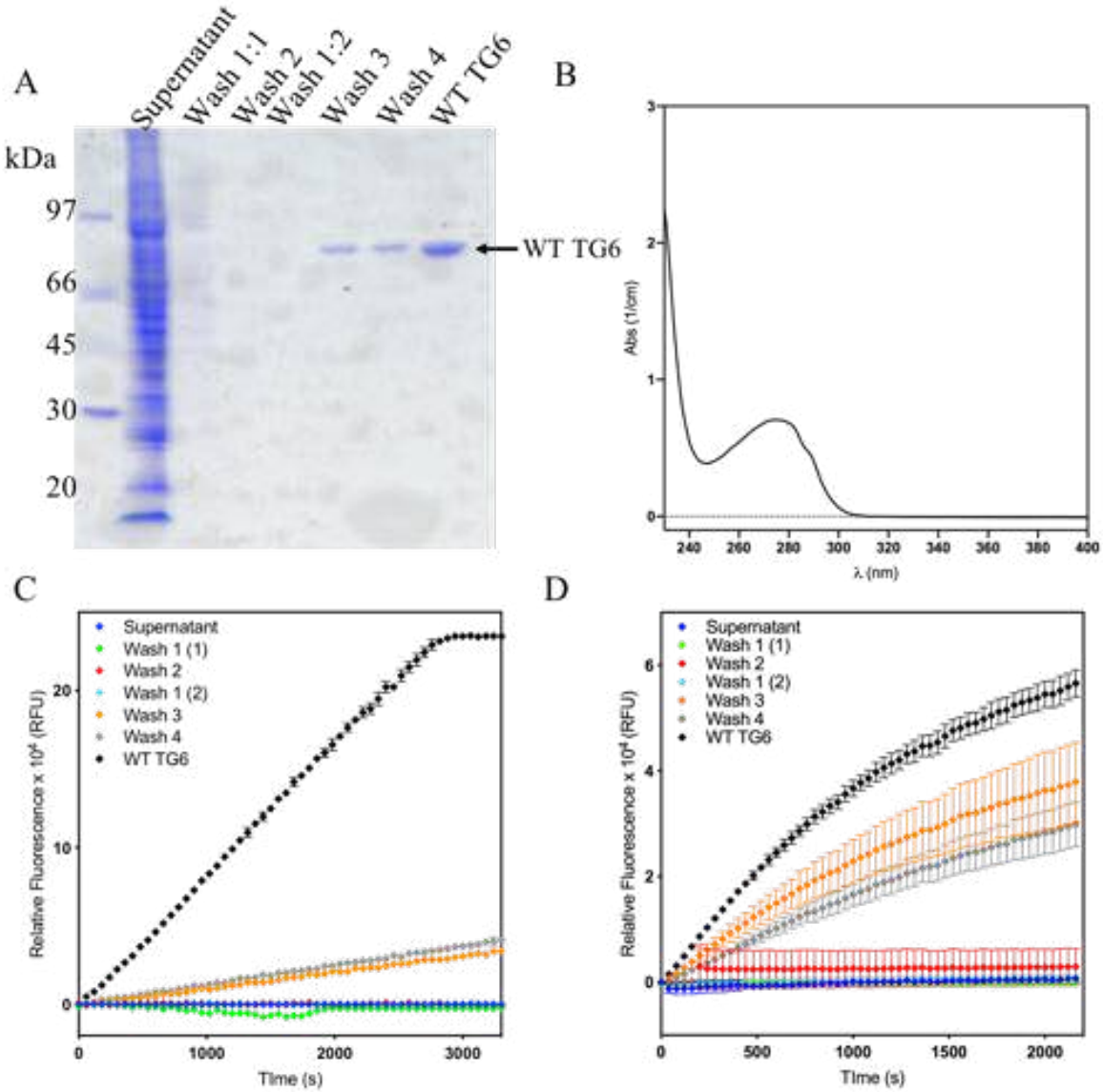


Figure 3.12 Analysis of purified WT TG6.
Samples were collected for analysis at multiple steps of the purification. These samples include supernatant, wash 1 x2, wash 2, wash 3 and wash 4 and purified WT TG6. A: Protein concentration was measured using a UV spectrophotometer and a 1 cm quartz cuvette. B: 10 µL of samples and 2 µg of WT TG6 were analysed on a 4-20% SDS-PAGE gel under reducing conditions, then stained using Coomassie brilliant blue R. C and D: The activity of purified WT TG6 (20 µg/mL) and fractions were assessed using the isopeptidase assay and the MDC assay. Data are the mean of 2 wells ± SD (n=2) and representative of 2 independent experiments (N=2). C: The increase in fluorescence was measured following the cleavage of quenched A102 substrate over 1 hour (λ_{ex} 320 nm and λ_{em} 440 nm). D: Transamidation activity was measured using the MDC assay over 1 hour.

3.3.12 Binding of allosteric regulators to WT TG6

To further characterise WT TG6, the binding of allosteric regulators Ca^{2+} and GTP was investigated in the context of the isopeptidase activity assay (Work aided by R. Griffiths). Ca^{2+} activation generated a concentration-dependent increase in isopeptidase activity (Fig. 3.13A) yielding a Ca^{2+} K_D of 2 mM. In contrast, GTP, GTPγS and GDP binding resulted in a concentration-dependent reduction in isopeptidase activity, with measured K_D s of ~250 µM, ~60 µM and 350 µM, respectively (Fig. 3.13B). TG6 was significantly less sensitive to GMP binding and a K_D could not be determined.

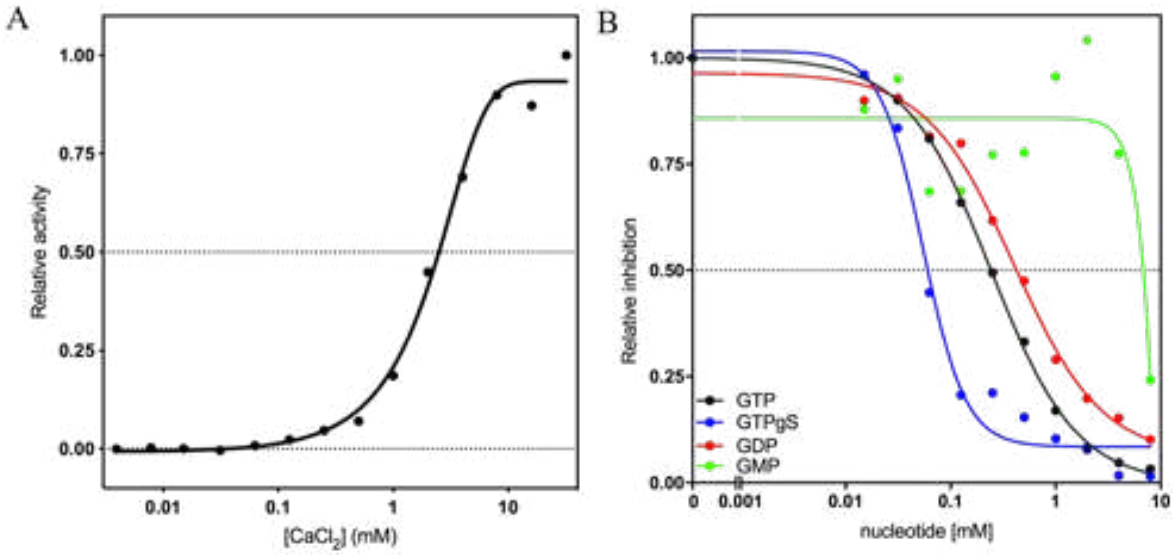


Figure 3.13: Binding of allosteric regulators by WT TG6.
Allosteric regulator binding was assessed using the isopeptidase activity assay. Fluorescence increase was measured following the cleavage of quenched A102 substrate over 1 hour (λ_{ex} 320-10 nm and λ_{em} 440-20 nm). Data are the mean of 2 wells (n=2) ± SD and representative of 1 independent experiments (N=1). A: $CaCl_2$ dependent activation of TG6. $CaCl_2$ concentration was varied from 32 mM to 15.6 µM. B: Nucleotide inhibition of TG6. Nucleotide concentrations (GTP (black), GTPγS (blue), GDP (red) and GMP (green)) were varied from 8 mM to 15.6 µM.

3.3.13 TG6 stability

As WT TG6 is prone to aggregation, the relative oligomerisation of TG6 at 4 °C was followed using UV spectrophotometry over 7 days (Fig. 3.14A and B). The measured peak absorption shows a maximum at ~265 nm which is in agreement with the predicted maximum UV absorption for TG6 expected at 274 nm based on its amino acid composition. The spectra showed that TG6 aggregation increases over 7 days (Fig. 3.14A).

The absorption at wavelengths 310 nm, 280 nm, 265 nm and 240 nm were analysed (Fig. 3.14B). The absorbance at 310 nm measures light scattering indicative of protein aggregation. In ideal circumstances this value would be less than 1 per cent of the maximum absorption at ~275 nm, but as time progresses the TG6 sample absorbance at 310 nm increases. This experiment shows that over the course of 7 days WT TG6 remained stable but showed a small reduction in concentration, most likely due to aggregated protein lost during centrifugation.

The stability over time (at 4 °C, 5 days) was also investigated without centrifugation of the sample. Measurements were taken post-purification (Day 0), and then the stability was followed over 5 days at 4 °C (Fig. 3.14C and D). At day 0 and day 1, the peak maximum remains at 275 nm and the minimum remains at 240 nm comparable to the centrifuged samples (Fig. 3.14A and C). Days 2 to 5 show a change in the shape of the peak where by the minimum is ~250 nm and the maximum is ~270 nm (Fig. 3.14C). Analysis of the absorbance at 310 nm showed an increase in aggregation over time (Fig. 3.14D). A large increase in the absorbance at 310 nm is seen from day 1 to day 2 which plateaus by day 5. These results indicate that TG6 is stable at 4 °C when centrifuged but without centrifugation the TG6 aggregation rapidly increases.

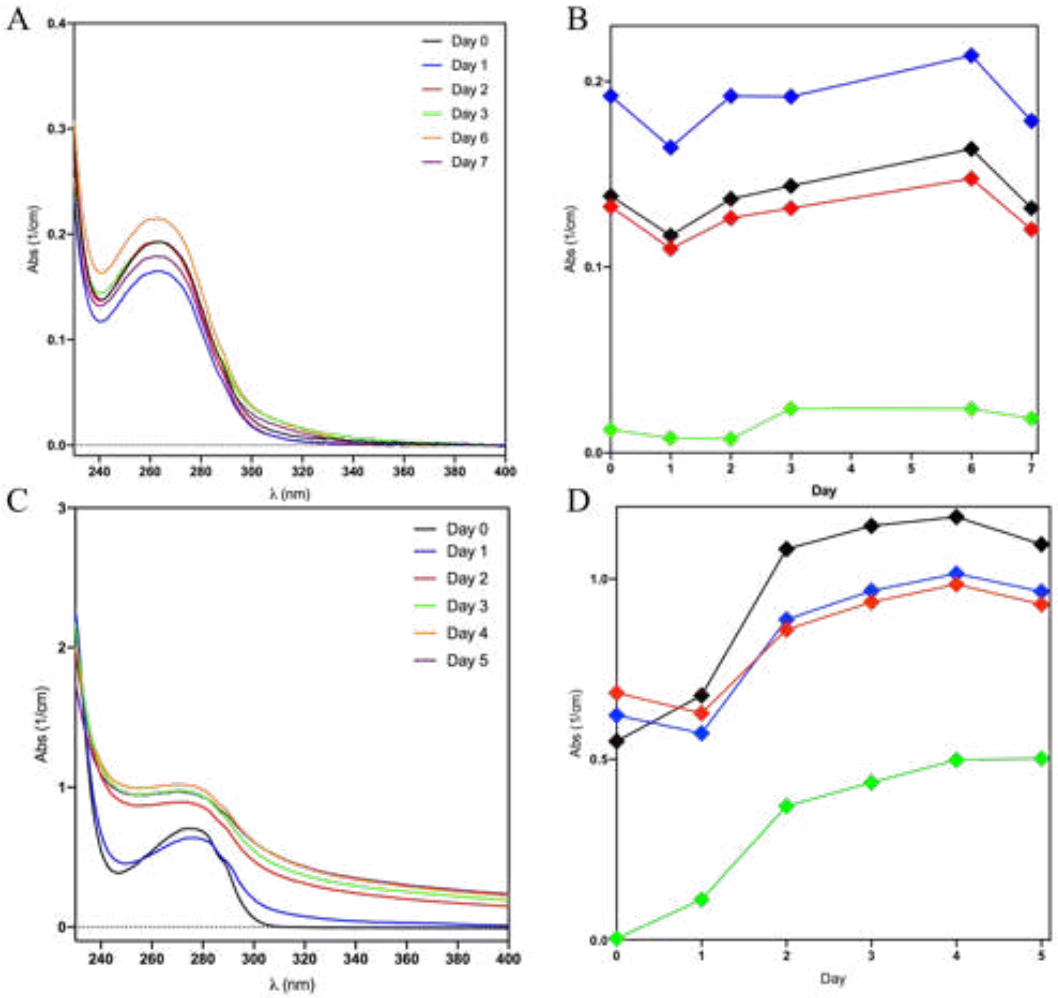


Figure 3.14: UV spectra to measure WT TG6 aggregation over time.
A: WT TG6 samples were stored at 4 °C. Following storage for the indicated times the samples were centrifuged to remove large precipitates, and a near UV absorbance spectrum was measured. B: Time course of the absorbance over a seven day period at 240 nm (black), 265 nm (blue), 280 nm (red) and 310 nm (green) as derived from the spectra shown in A. C: Near-UV spectra recorded every 24 hours following storage of WT TG6 samples at 4 °C. D: Time course of the absorbance over a five day period at 240 nm (black), 265 nm (blue), 280 nm (red) and 310 nm (green) as derived from the spectra shown in C. Data is representative of 1 independent experiment (N=1).

3.4 DISCUSSION

3.4.1 Summary of *E. coli* system

Developing a new successful *E. coli* expression system is useful for producing larger amounts of TG6 for crystallisation. His-TG6 was selected for expression in *E. coli* as its biochemical properties can be directly compared to the His-TG6 generated using the baculovirus/*Sf9* cell expression system. His-SUMO-intein TG6 was of most interest as it could generate untagged TG6. This approach generated recombinant TG6 more representative of TG6 *in vivo* and does not include additional residues that might interfere with structural studies (Wang *et al.* 2012). Both these TG6 proteins were not highly expressed or stable and as a result were not investigated further in the *E. coli* expression system. Taken together these studies indicate that TG6 requires a tag which promotes expression and solubility when expressed in *E. coli* and that it may also require additional chaperones to aid in folding.

Trx, MBP and NusA are endogenously expressed soluble proteins in *E. coli* and can function as general enhancers of solubility when fused to passenger proteins (Nallamsetty and Waugh 2006). In addition they can aid in the overexpression of recombinant proteins and were predicted to help TG6 solubility. Trx-His-TG6 was expressed and resulted in low yields following purification. The UV spectrum shows it is highly aggregated, suggesting that the Trx-His tag does not enhance TG6 solubility and folding. The isopeptidase activity is low when compared to the His-TG6 generated using baculovirus/*Sf9* cells. This low activity could be a consequence of the incomplete purification of Trx-His-TG6 or the presence of aggregated protein.

Both His-MBP-TG6 and NusA-His-TG6 showed good yield compared with the other TG6 fusion proteins and a marginal amount of light scattering at 310 nm in the UV spectrum, indicating the fusion proteins are stable. However, during experiments to optimise their expression both proteins were highly degraded, demonstrating they are unstable and undergo proteolytic cleavage. Additional protease inhibitors could aid in reducing the fragmentation, however inhibitors which target aspartate proteases have previously been shown to impact TG function (personal communication with D. Aeschlimann). Also, the extraction of soluble protein may have been hindered by the formation of inclusion bodies, resulting in lower yields. Protein expression at higher

temperatures has been shown to generate protein at a higher rate but increases the risk of inclusion body formation (Zhou *et al.* 2018). Following optimisation of the expression of His-MBP-TG6 and NusA-His-TG6, optimal conditions were identified as 0.1 mM IPTG, 6 hours at 20 °C. However, a lower temperature may be required to allow sufficient time for TG6 to fold and to reduce the activity of proteases. The expression of TG6 fusion proteins could be improved using auto induction media which allows bacteria to grow to a higher cell density, and by co-expression with molecular chaperones which can aid in the folding of TG6 (de Marco *et al.* 2007; Li *et al.* 2011).

The lysis buffer which proved best at TG6 extraction was 20 mM Tris-HCl pH 8.0, 20% glycerol. The yield of His-MBP-TG6 was improved, but fragmentation was still observed (Fig. 3.10A). Use of a cell homogeniser could enhance the extraction of TG6 with minimal disruption to TG6 folding. A two-step purification using Ni sepharose and amylose resin affinity purification may improve purity. Ni sepharose affinity purification can purify His-tagged proteins <80%, compared with amylose resin that can purify MBP-tagged proteins to <70% purity (Kimple *et al.* 2013). However, an additional purification step will impact yield. Furthermore, His-MBP-TG6 and NusA-His-TG6 had lower isopeptidase activity (0.4-1.9 RFU/s) compared to TG6-control (7.26 RFU/s). The lower activity could be a result of protein fragmentation, which has contributed to the absorbance at 280 nm or the interference of the solubility tag.

The advantages of an *E. coli* expression system are detailed in Table 3.8. To summarise, the cost of materials is relatively small when compared with the baculovirus/*Sf9* cell system and an *E. coli* system is easier to scale up to produce a larger amount of protein. For example *E. coli* can reach a suitable density for protein expression in two days, whereas the *Sf9* cells take 4-5 days to undergo the same number of cell doublings. In addition, in an *E. coli* system protein expression commonly occurs over 16 hours whereas baculoviral protein expression takes a minimum of 72 hours. Furthermore, the initial setup to generate baculovirus is significantly more time consuming than expression plasmid and glycerol stock generation.

However, the *E. coli* expression system is less efficient at producing TG6 (0.1 mg/L of culture) compared with TG6 produced with the baculovirus system (1 mg/L of culture). Taking into consideration the data collected on the *E.coli* expression system further investigation would not generate TG6 comparable to the baculovirus/*Sf9* expression system.

Table 3.8: Comparison of E. coli and baculovirus expression system.

	***E. coli* system**	**Baculovirus system**
Cost	Materials are cheap	Materials are expensive
Set up	Plasmid generation and glycerol stock	Plasmid generation, bacmid generation and baculovirus production
Expansion time	24 hours	96-120 hours
Protein expression time	16 hours	72 hours
Total Time	~3 days	~2 weeks
Protein production	~20 mg/L of culture	~1 mg/L of culture
Expression	Cannot carry out post-translational modifications	Can post-translationally modify proteins

3.4.2 Characterisation of WT TG6

WT TG6 was successfully purified using the baculovirus expression system. Using this method WT TG6 was expressed to 1 mg/L of culture. In addition, the purified WT TG6 was highly active using the MDC and the isopeptidase assays. The binding affinities of Ca^{2+} and guanine nucleotides, was measured providing data for comparison with the TG6 mutants. The binding affinities (K_D) for GTP, GTPγS, GDP and GMP previously measured are 200 μM, 70 μM, 250 μM and 2 mM, respectively and this values are similar to those attained here (Thomas *et al.* 2013). Previous data has not been published regarding the Ca^{2+} binding by TG6, however the data here shows similar Ca^{2+} affinity to TG2 (Iismaa *et al.* 2000).

Assessment of the stability of WT TG6 showed continuous centrifugation reduced protein aggregation (Fig. 3.14B). However, without centrifugation TG6 progressively aggregated until a plateau was reached (Fig 3.14D). The formation of aggregates could be concentration dependent and reach a chemical equilibrium. It is not yet known whether TG6 forms higher ordered structures. Further analysis to investigate oligomerisation could include analytical size exclusion chromatography,

glutaraldehyde crosslinking and CD spectroscopy (Fadouloglou *et al.* 2008; Staub *et al.* 2011; Hong *et al.* 2012; Zhang *et al.* 2018). In addition, CD spectroscopy can be used to assess the secondary structure of WT TG6, prior to structural characterisation by X-ray crystallography.

The work carried out on WT TG6 will be valuable in assessing the biochemical properties of the TG6 mutants. WT TG6 transamidation and isopeptidase activities, the binding of allosteric regulators (Ca^{2+} and GTP) and an assessment of its stability provide a basis from which to compare TG6 mutant structure and function. This analysis could aid the elucidation of the mechanism by which the mutants cause neuronal cell death.

CHAPTER 4

4. USING BIOINFORMATICS TO PREDICT CONSEQUENCES OF DISEASE-ASSOCIATED MUTATIONS ON TG6 STRUCTURE.

4.1 INTRODUCTION

To date at least 22 mutations in *TGM6* have been identified in patients with symptoms matching those of SCA (Wang *et al.* 2010, Li *et al.* 2013, Guo *et al.* 2014, Fogel *et al.* 2012, Nibbeling *et al*. 2017, Ruiz-Lopez *et al*. 2017, Tripathy *et al.* 2017, Lin *et al.* 2018, Yang *et al*. 2018, personal communication with Kuo S.H, Columbia University and Kerr M and Khan A, University of Calgary). Other diseases that have been associated with mutations in *TGM6* include episodic ataxia, early onset PD, and AML (Pan *et al.* 2015; Westenberger *et al.* 2015; Choi *et al.* 2017)

The mutations identified in *TGM6* in the human population can be accessed and categorised using GnomAD and ExAC databases (Lek *et al.* 2016). Both these databases collect and summarise sequencing data from large-scale sequencing projects and make them readily available. ExAC collects exome data, where as GnomAD collects both exome and whole genome sequences. They provide information on the identified variants e.g. number of alleles identified with the variant (allele count), number of alleles tested (allele number) and frequency of allele variant in the population (allele frequency). Information is also provided on the ethnicity of populations tested and the age distribution. Furthermore, the mutation is characterised by type using the annotations missense, splice donor, synonymous and frameshift. This information can aid in predicting if the polymorphism is likely to be benign or pathogenic.

In addition, there are numerous bioinformatic tools that can be used to analyse missense mutations. Assessing the conservation of amino acids mutated in TG6 indicates whether an amino acid is evolutionarily important to the structure and function of the TGs. This can be carried out using MSA, for example comparing TG6 with other TG isoforms or TG6 in different species. This can highlight residues important to TGs in general and also those that are specifically important to TG6. CLUSTAL Omega and the PANTHER database are frequently used in the literature. CLUSTAL Omega uses both seeded guide trees and Hidden Markov Models (HMM) to generate an alignment of the input protein sequences (3 or more, Sievers *et al*. 2011). This highlights fully-conserved and semi conserved residues (where the property of the residues is retained) in the sequences.

PANTHER is a gene ontology tool that utilises HMM to investigate homologous sequences and classify the overall biological function of the protein and association with biological pathways based on genetic relationships (Thomas *et al*. 2003). The phylogenetic trees supplied are from both large protein families and subfamilies which can provide more precise data.

Sequence based methods such as Polymorphism Phenotyping-2 (PolyPhen-2) and Sorting Intolerant from Tolerant (SIFT) tools are frequently used to predict the functional impact of missense mutations. PolyPhen-2 can predict the impact of amino acid by extracting sequence and structural based features from the human protein (Adzhubei *et al.* 2010). Sequence based analysis includes identification of features e.g. metal binding sites and actives sites, from the UniProt/ SwissProt database and screening of these features to identify if the polymorphism is located within these sites. Also, regions within the protein such as transmembrane, intramembrane and propeptide regions are also investigated. In addition, the likelihood of the polymorphism occurring in the specific residue position within the polypeptide sequence is compared to other homologues, which have 30-94% sequence identity. A comparison is then made between WT and mutated protein and a big positive difference are indicative of a mutation rarely observed in the homologous proteins. Structural based analysis includes the identification of structures in the protein structure database, which have a sequence identity of ≥50%, minimal hit length of 100, and maximal number of gaps is 20. The Dictionary of Secondary Structure in Proteins (DSSP) database is also utilised to obtain additional parametres for structural analysis such as secondary structure, solvent accessible surface area and phi-psi dihedral angles. PolyPhen-2 predicts the probability the mutation is damaging and also estimates the false positive rate and the true positive rate.

SIFT is able to predict whether an amino acid substitution in a protein will have a phenotypic impact (Sim *et al.* 2012). It is based on the conserved amino acids residues, which are important for function, and non-conserved amino acids, which are not important with regards to protein function. The query sequence is used to identify similar sequences, which are directly related or may share similar function. These selected sequences are aligned with the query sequence. The MSA is used to calculate the normalised probability of the substitutions.

Multiple bioinformatic methods are available to assess the stability change induced by the introduction of mutations. MUpro can predict the stability based on both the primary sequence and the tertiary structure (Cheng *et al*. 2005). However, tertiary structure is not essential for a stability prediction. This utilises both support vector machines and neural network algorithms to calculate the change in free energy. I-Mutant (1.0 and 2.0) can also be used to predict changes in stability induced by point mutations. It bases the prediction on primary and tertiary structure, however a structure deposited in PDB is required for stability predictions based on tertiary structure. Sequence based stability analysis calculates the free energy change which indicates the mutated residue as destabilising or stabilising (Capriotti *et al*. 2005). FoldX can estimate the free energy change between the folded and unfolded state (ΔG) of a protein using a 3D structure or model (Schymkowitz *et al*. 2005). The change in the change of the free energy ($\Delta\Delta G$) can be used to calculate if the introduced mutation is stabilising ($\Delta\Delta G < 0$) or destabilising ($\Delta\Delta G > 0$).

Bioinformatic tools can also be used to predict the aggregation propensity of proteins using its primary sequence with AGGRESCAN and its tertiary structure using AGGRESCAN3D (A3D) (Chonchillo-Sole *et al*. 2007, Zambrano *et al*. 2015). AGGRESCAN calculates and reports: the aggregation-propensity value (a3v) for each residue in the sequence (experimentally derived values), the peaks in the polypeptide over the calculated aggregation threshold and the hotspots identified in the protein. The average a3v (a4v) is calculated over a sliding window of 5, 7, 9 or 11 residues primed against a database of 56 amyloidogenic proteins, where the hotspots were experimentally known. A region in the polypeptide sequence was deemed a 'hotspot' if there were five or more sequential residues which had a4v which was over the calculated hotspot threshold value (HST) and did not include a proline. The HST is calculated from the average of a3v for the 20 amino acids and by weighting them based on their frequency in the SwissProt database. A3D uses a similar approach but takes into consideration the 3D structure of the molecule (Zambrano *et al*. 2015).

In addition, homology modelling is a popular way of investigating the impact of mutations on the protein. Homology models can be generated using different methods for example SWISS model and Prime homology (Jacobson *et al*. 2002; Jacobson *et al.* 2004). Using existing X-ray structures of other protein family members, i.e. TG2 and TG3, the TG6 structure can be modelled. Using such models, changes in interactions in functionally important binding sites can be assessed for the respective mutant proteins.

For the specific tools employed here, one disadvantage is that the model backbone is static and therefore any overall changes in the protein backbone cannot be assessed using this method.

Bioinformatic analysis has previously been carried out on the TG6 mutants (R111C, D327G, D510H and L517W) which involved the use of SIFT, PolyPhen-2, AGGRESCAN, FoldX and homology modelling (Li *et al*. 2013, Thomas *et al.* 2013, Guo *et al*. 2014). However, the respective tools have not been consistently used for analysis of TG6 mutations known at the time. TG6 R111C and D510H were predicted to be pathogenic by SIFT and PolyPhen-2 but this was not carried out for D327G and L517W (Guo *et al.* 2014). AGGRESCAN analysis was carried out for R111C and D510H whereby a change in aggregation propensity was only apparent in D510H (extension of hotspot by 1 residue, 509-520 instead of 510-520) (Li *et al.* 2013). In addition, the calculation of ΔG, using PopMusic 2.0, predicted an energy change of 0.08 kcal/mol for TG6 D510H compared to WT TG6. This could represent an impact on local folding and surface potential thus destabilising TG6. Molecular modelling techniques were carried out with TG6 D327G, D510H and L517W. TG6 D327 is part of Ca^{2+} binding site 3, so is predicted to impact Ca^{2+} binding at this site (Thomas *et al.* 2013). Modelling of D510H, showed D510 could form one hydrogen bond with S568 whereas H510 is predicted to form two hydrogen bonds with T566 instead (Li *et al*. 2013). The change in size and the charge (from negative to 10% positive and 90% negative) of the residue, from aspartate to histidine was predicted to alter protein folding. No impact of the L517W substitution could be predicted using homology modelling (Thomas *et al*. 2013).

From this limited and inconsistent bioinformatics analysis it is difficult to form hypothesis on the impact of the TG6 mutations on protein function. Consequently a more thorough and consistent method is required to accurately predict the biochemical consequences of the mutations on TG6. Furthermore, recent work in the group had identified a number of additional mutations in *TGM6* in SCA patients (Tripathy *et al*. 2017) and the consequence of these sequence alterations was entirely unclear at the time. Therefore, a combination of bioinformatics tools were used to investigate the folding, stability and solubility, and functional activity of the TG6 mutants. In addition, TG6 homology models were used to complement these basic tools and assess in more detail the impact of the TG6 mutations on their local environment, residue-specific interactions, and on important functional sites including the three Ca^{2+} sites and GTP binding site. In

combination, these results can provide hypotheses and these will then be tested biochemically.

Therefore, the specific aims of the chapter are:

1. To assess the impact of the TG6 mutations on the structure of TG6 using bioinformatic tools which assess the function, stability and solubility of proteins.
2. To use homology modelling approaches to:
    a. Analyse the impact of the mutation on its local environment which can predict possible structural implications.
    b. To assess the impact of the mutation on function with regards to Ca^{2+} binding and GTP/GTPγS binding as this could highlight any impact on the regulation of TG6
3. To formulate hypotheses for subsequent testing using biochemical characterisation methods.

4.2 METHODS

4.2.1 Generation of homology models

TG6 models generated included open TG6 (TG2 structure, PDB: 2Q3Z (Pinkas *et al.* 2007)), Ca^{2+} bound TG6 (TG3 structure, PDB: 1L9N (Ahvazi *et al.* 2003)), Ca^{2+}/Mg^{2+} and TG6•GTPγS (TG3 structure, PDB: 1RLE was published in a retracted paper, however the data regarding this structure is believed to be accurate (Ahvazi *et al.* 2004a)) and TG6•GTP (TG2 structure, PDB: 4PYG (Jang *et al.* 2014)).

4.2.1.1 SWISS Model.

The protein sequence of TG6 was aligned with TG2 (PDB: 4PYG and 2Q3Z (Jang *et al*. 2014, Pinkas *et al.* 2007)) or TG3 (PDB: 1L9N and 1RLE (Ahvazi *et al.* 2002, Ahvazi *et al.* 2004a) sequences using CLUSTAL Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) (Ahvazi *et al.* 2002; Ahvazi *et al.* 2004a; Pinkas *et al.* 2007; Sievers *et al.* 2011; Jang *et al.* 2014). The alignment was input into SWISS model and homology model was generated based on the aligned template structure, using default parametres (Guex *et al.* 2009; Benkert *et al.* 2011; Bertoni *et al.* 2017; Bienert *et al.* 2017; Waterhouse *et al.* 2018). The resulting structures were visualised with PyMOL Molecular Graphics System (Version 2.0, Schrodinger, LLC). The nucleotides were manually docked into the pocket by superimposing the homology model with its respective template and polar contracts were identified.

4.2.1.2 Prime Homology

Maestro11 (Release 2018) Prime Homology (https://www.schrodinger.com/Prime/) software was used to generate TG6 homology models based on TG2 (PDB: 4PYG, Jang *et al*. 2014) and TG3 (PDB: 1L9N, Ahvazi *et al.* 2002), using default parametres (Ahvazi *et al.* 2002; Jacobson *et al.* 2002; Ahvazi *et al.* 2004a; Jacobson *et al.* 2004; Han *et al.* 2010; Jang *et al.* 2014). The structures were prepared and formatted for docking with protein prep wizard tool (https://www.schrodinger.com/Maestro) and manually aligned for homology modelling (Sastry *et al.* 2013). The resulting structures were visualised with PyMOL Molecular Graphics System (Version 2.0, Schrodinger, LLC).

4.2.1.3 Manual docking

TG6•GTPγS (TG2 template, PDB: 4PYG) and TG6•GTP (TG3 template, PDB: 1RLE, Ahvazi *et al.* 2004a) structures were not available so GTP and GTPγS were manually

docked into their respective models (Ahvazi *et al.* 2004a; Jang *et al.* 2014). This included superimposing the homology model with the template harbouring the desired nucleotide. For example, TG6•GTPγS was generated by superimposing the TG6 homology model (using TG2 as the template) and TG3•GTPγS (PDB: 1RLE, Ahvazi *et al.* 2004a). GTP and the TG3 protein backbone were subsequently deleted. The same approach was used to generate TG6•GTP (using TG3 as the template).

4.2.2 Validation of TG6 homology models

Model validation was carried out using the structural assessment tool available at: https://swissmodel.expasy.org/assess. This generated various parametres which indicated the fit of the model to the template.

4.2.3 Molecular Modelling of TG6 mutants

TG6 homology models in combination with other altered amino acid sequence were used to generate models of the TG6 mutants. Each model was visualised using PyMOL Molecular Graphics System (Version 2.0, https://pymol.org/2/). Changes in the local environment of the mutation and changes in the Ca^{2+} binding sites and GTP binding site were analysed by assessing changes in the polar contacts (<3.5 Å) and measuring distance between relevant side chain groups and the interacting moieties of the co-factors.

4.2.4 Databases

Information available on the TG6 mutations was extracted from the ExAc (http://exac.broadinstitute.org/about) and GnomAD databases (http://gnomad-old.broadinstitute.org/) (Lek *et al.* 2016).

4.2.5 Multiple Sequence alignment

MSA was carried out for the human TGases (1-7) using CLUSTAL Omega (Sievers *et al.* 2011). Sequences were obtained from UniProt: TG1: P22735, TG2: P21980, TG3: Q08188, TG4: P49221, TG5: O43548, TG6: O95932 and TG7: Q96PF1. The full conservation of the residue is denoted as: *, strongly conserved properties of the residue of the residue is denoted as conserved (:) and weakly conserved properties of a residue is denoted as semi-conserved (.). These residues were highlighted (grey). In addition, all the TG6 mutants analysed herein, were highlighted (cyan), the Ca^{2+} binding sites (red) and the GTP binding residues (purple). This was repeated for TG6 from different species. The homologous sequences from different species were obtained from a blast search and are

as follow: *M. musculus* (Q8BM11), *R.norvegicus* (F1M4T7), *H.sapiens* (O05932), *P. troglodytes, C. lupis familiaras* (F1PKV0), *A.melanoleuca* (G1LY70), *B. Taurus* (F1N2I1) and *S. scrofa* (A0A2987BLG6).

4.2.6 Functional analysis with PolyPhen-2.

The TG6 protein sequence (O95932) was input as the query and the point mutations were selected consecutively and submitted (http://genetics.bwh.harvard.edu/pph2/) (Adzhubei *et al.* 2010). The mutations were sorted into either 'benign', 'possibly damaging' or 'probably damaging' based on the false positive rate threshold calculated. However, these predictions are made on polypeptide sequence alone and do not take into consideration the 3D structure of TG6. Furthermore, frameshift mutations cannot be analysed by this method.

4.2.7 Functional analysis with SIFT

The TG6 protein sequence was input to the SIFT single protein tool as the query (https://sift.bii.a-star.edu.sg/) (Sim *et al.* 2012). The substitutions were listed in the box and the query was submitted. The substitutions were classified into tolerated (probability of ≥0.05) or deleterious (probability of <0.05). This method relies on polypeptide sequence and does not take into account the 3D structure of TG6 and was also unable to analyse frameshift mutations.

4.2.8 Stability analysis with FoldX

The stability of the TG6 mutants was calculated using the FoldX plugin on YASARA (http://foldxsuite.crg.eu/) (Schymkowitz *et al.* 2005). FoldX can estimate the ΔG of a protein using the equation:

Equation 4.1: FoldX stability equation.

$$\Delta G = a \cdot \Delta G_{vdw} + b \cdot \Delta G_{solvH} + c \cdot \Delta G_{solvP} + d \cdot \Delta G_{wb} + e \cdot \Delta G_{hbond} + f \cdot \Delta G_{el} + g \cdot \Delta G_{Kon} + h \cdot T\Delta S_{mc} + k \cdot T\Delta S_{sc} + l \cdot \Delta G_{clash}$$

a-l: Relative weight of different energy terms in calculation.
$\Delta G_{clash:}$ Measure of steric overlap in a structure.
ΔG_{el}: The electrostatic contribution of side chains. This is calculated from Coulombs law.
ΔG_{hbound}: Free energy difference between the formation of an intra-molecular hydrogen bonds compared with an inter-molecular hydrogen bond.

whilst ΔS_{mc} and ΔS_{sc} are calculated theoretical estimates. ΔG was calculated for each TG6 mutant in the open and closed conformation. The difference between WT TG6 and mutant TG6 was calculated to establish the ΔΔG i.e. $\Delta\Delta G = \Delta G_{mut} - \Delta G_{wt}$.

4.2.9 Solubility analysis using AGGRESCAN and A3D

AGGRESCAN is an online tool, which can be used to identify aggregation hotspot within proteins with respect to their polypeptide sequence (http://bioinf.uab.es/aggrescan/) (Conchillo-Sole *et al.* 2007). The WT TG6 and TG6 mutant protein sequences were input into the query box and submitted. The a4v were calculated over a sliding window of 11 residues from the experimentally derived a3v values. Aggregation prone hotspots were denoted as five or more residues over the HST. The HST was calculated using a4v for the 20 amino acids and by weighting them based on their frequency in the SwissProt database.

AGGREGSCAN3D (A3D) improves upon the AGGREGSCAN predictions by taking into consideration the 3D structure of the protein (Zambrano *et al.* 2015; Gil-Garcia *et al.* 2018). The aggregation propensity of individual residues is calculated from the surrounding spherical region (centre of the sphere being the Cα carbon) with a radius of 10 Å, giving rise to a structurally corrected aggregation value (A3D score). The formula used to calculate the A3D score is as follows.

Equation 4.2: A3D score calculation.

$$A3D\ score = Agg_i \cdot \left(\alpha \cdot e^{\beta RSAi}\right) + \Sigma\left[Agg_e \cdot \left(\alpha \cdot e^{\beta RSAe}\right) \times \left(\gamma \cdot e^{-\delta dist}\right)\right]$$

Agg_i: The intrinsic aggregation propensity of the residue in the centre of the sphere.
RSA_i: The relative surface area exposed to solvent (calculated using the Lee and Richards method (Lee and Richards 1971)

Agg_e: The intrinsic aggregation propensity of additional residues included in the sphere.
RSA_e: The relative surface area exposed to solvent and dist is the distance to the central residue.

The exponential of RSA, α (0.0599) and β (0.512) and are important for correcting the higher contribution of exposed residues to the aggregation over residues that are buried whilst the exponential γ (1.291) and δ (-0.256) correct the distance parametre, which allows the residues closer to the centre of the sphere to impact the local aggregation greater than residues more distant. The output of A3D includes A3D average (mean of A3D scores of residues in protein) and total score (sum of A3D scores of residues in protein) and a PDB file which provides a visual representation of the A3D score represented by a colour gradient.

4.3 RESULTS

4.3.1 Investigating WT TG6 using molecular modelling

Three models were generated using SWISS Model and two were generated using Prime homology available on Maestro 11 (Fig. 4.1). Prime homology modelling allows greater flexibility in the backbone of the protein and maximised the interactions with ligands (i.e. GTP/GTPγS) and ions (Mg^{2+} and Ca^{2+}) that were co-crystallised in the template structure. In contrast, homology models generated by SWISS models are static and flexibility is only observed in the residue side chains. In all TG3 based models the extended loop domain (grey circle) was not resolved, in the SWISS model generated homology models the loop was disordered and solvent exposed and therefore is likely inaccurate. In the prime homology generated models, the Ca^{2+} bound TG6 model was analysed with the extended loop deleted (Figure 4.1D), whilst the loop on GTP•TG6 model is disordered but appears to be buried.

The homology models were validated and their quality assessed to identify whether they are an accurate representation of the template (Table 4.1). There is no way to assess if the models represent TG6, as there is no X-ray structure available. For the validity of the TG6 homology model, the following parametres were calculated: MolProbity, clash score, Ramachandran favoured, Ramachandran outliers, C-β deviations, bad bonds, bad angles, *cis* non-prolines, *cis* prolines, twisted non-prolines and twisted prolines. The MolProbity score represents the crystallographic resolution that the measured quality would be expected. In ideal circumstances this would be as low as possible. The clash score is representative of non-hydrogen bond clashes (0.45 Å), which is 0 for an ideal fit. A good fit is representative of a Ramachandran favoured of > 98%. Ramachandran outliers and rotamer outliers were a measure of outliers for resolutions below 3.0 Å. Ideally, this needs to be < 0.2% and < 1%, respectively. C-β deviations show the best fit at 0 and identify residue position deviations of > 0.25 Å from the template. Both bad bonds and bad angles should be 0 and are representative of > 4 σ from the template. Cis/twisted prolines/non-prolines are defined <30 ° (*cis*) and < 150 ° (twisted) from the template and 0 represents the best fit to the template.

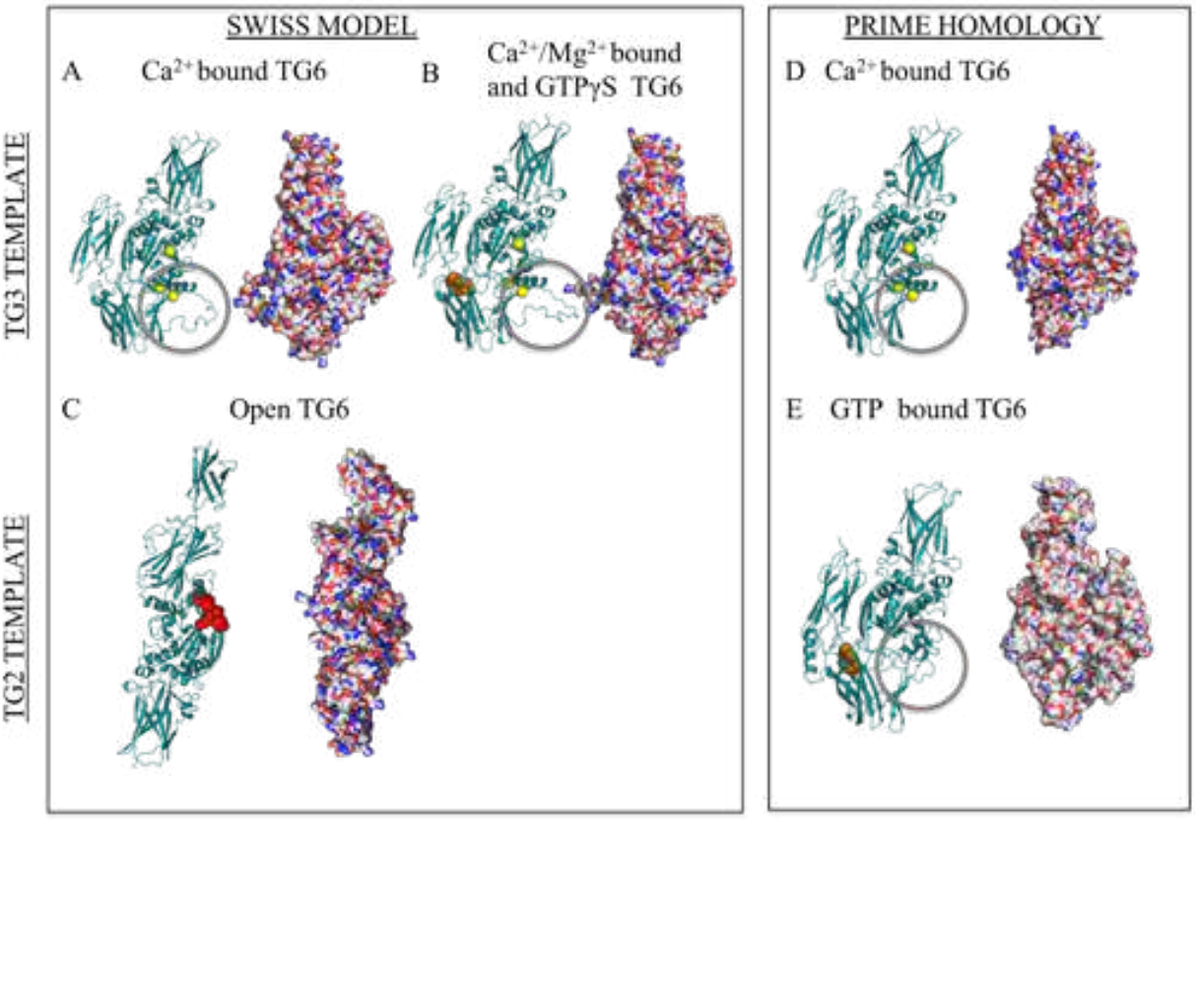


Figure 4.1

TG6 homol
analysis us
Model (A-
Images ar
representat
red: oxyge
Ca^{2+}/Mg^{2+}
or GTPγS
spheres) ar
red sphere
highlights
crystal stru
'closed' Ca
homology
(1x) and G
homology
A but was g
extended
removed du
model. E:
GTP boun
domain (gr

The calculated values are similar when comparing the models generated with Prime homology compared with SWISS Model, with the least accurate being Ca^{2+} bound TG6 generated with Prime homology (Fig. 4.1D), with respect to the MolProbity score and clash score. This could be due to deletion of the loop domain during model generation. For all the TG6 homology models, the Ramachandran favoured and Ramachandran outliers were not within ideal range (> 98% for Ramachandran favoured and < 0.2% Ramachandran outliers). Furthermore, none of the TG6 homology models were within the optimal range for rotamer outlier (< 1%). Multiple differences were observed for all TG6 homology models with respect to C-β deviations and bad bond *cis* non-prolines and *is* prolines, which indicate the template is not optimum for TG6.

The quality of the models was also assessed for the TG6 models, this was done by calculating the QMEAN. QMEAN can be used to assess the global quality of the structure. A value of zero represents the model is similar to the experimental structure. Deviation from zero indicates a lower quality model. The QMEAN calculation utilises the C-β, all atom, solvation and torsion values. The C-β value is the distance dependent evaluation of the frequency of the interactions centres of C-β atoms with reference to secondary structure. In addition, the all atom value also assesses the distance of the interactions. Solvation is a measure of residue burial with reference to the frequency of the interactions centres (C-β). This measures the solvent accessibility of a given residue compared with other residues in the structure. Torsion is the measure the torsion angle of a given residue compared with other residues in the structure. This measures local geometry of three consecutive amino acids and recognises the nativity of the target protein. In all cases a value close to zero is ideal for a high quality homology model. All values for the TG6 homology models show they are of low quality compared to the template.

In combination, this shows these homology models are not ideal for the modelling of TG6, however this could due to the homology for TG2 and TG3 (~50%) and SWISS Model more closely represents the template structure when compared to homology models generated with Prime homology. However, differences in Prime homology TG6 models could be due to increased flexibility in the backbone of the template, whereas the backbone of the SWISS Model generated models remains static. Variation between the template and the TG6 homology models are shown in Fig. 4.2.

Table 4.1: Assessment of validity and quality of TG6 homology models.

		SWISS Model			Prime Homology		Ideal value
		1L9N	1RLE	2Q3Z	1L9N	4PYG	
Validity	MolProbity score	1.35	1.65	1.97	3.50	2.19	Low as possible
	Clash Score	1.22	4.24	6.14	55.12	2.96	0
	Ramachandran favoured (%)	92.02	93.00	92.42	93.72	87.46	>98
	Ramachandran outliers (%)	1.36	1.29	1.43	1.20	3.99	<0.2
	Rotamer outliers (%)	1.07	0.83	1.65	14.04	4.44	<1
	C-β deviations	10	4	5	7	12	0
	Bad bonds	0	0	3	16	1	0
	Bad angles	92	45	44	92	52	0
	Cis non-prolines	5	2	3	2	1	0
	Cis prolines	2	1	1	1	1	0
	Twisted non-prolines	1	0	2	1	1	0
	Twisted proline	1	0	1	0	0	0
Quality	QMEAN	-3.09	-2.86	-3.23	-3.01	-4.58	0
	C-β	-2.08	-2.02	-2.65	-2.21	-2.29	0
	All atom	-0.99	-1.03	-1.55	-1.98	-1.54	0
	Solvation	-0.54	-0.80	-0.64	-1.04	-1.76	0
	Torsion	-2.49	-2.20	-2.48	-2.20	-3.52	0

Ideal values highlighted in yellow.

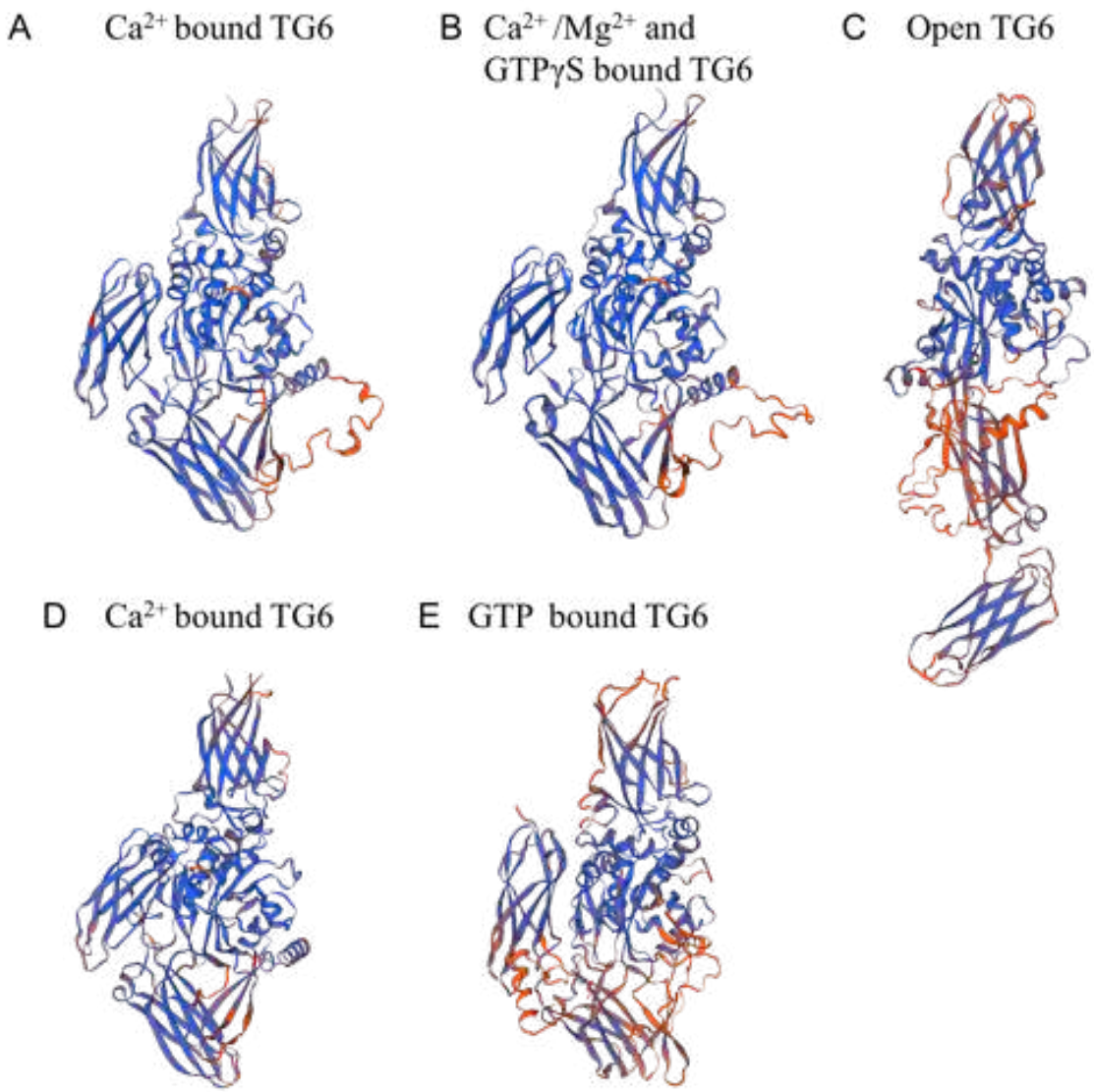


Figure 4.2: Variation of TG template compared with TG6 polypeptide sequence. TG6 homology models generated by SWISS Model (A-C) and Prime homology (D and E) are shown in cartoon representation. Areas coloured in orange show variation between the template and TG6 polypeptide sequence with respect to template structure. A: Ca^{2+} bound TG6 modelled on TG3 (Fig 4.1A). B: Ca^{2+}/ Mg^{2+} and GTPγS TG6 modelled on TG3 (Fig. 4.1B). C: Open TG6 modelled on TG2 (Fig. 4.1C). D: Ca^{2+} bound TG6 modelled on TG3 (Fig. 4.1D). E: GTP bound TG6 modelled on TG2 (Fig. 4.1E).

Analysis of the three Ca^{2+} binding sites highlighted the residues involved in the coordination of Ca^{2+} (Fig.4.3). SWISS Model generated Ca^{2+} bound TG6 (Fig. 4.1A) showed Ca^{2+} binding site 1 was made up of: A223 (backbone carbonyl), N226 (carboxyamide sidechain and backbone carbonyl), N228 (backbone carbonyl) and D230 (2 interactions with carboxyl side chain) which contribute six interactions to stabilise Ca^{2+} (Fig. 4.3A). Ca^{2+} binding site 2 is made up of: N396 (side chain carboxyamide), T417 (backbone carbonyl), E445 (2 interactions via carboxyl side chain) and E450 (carboxyl side chain). This contributes five interactions to the stabilisation of Ca^{2+}, a further backbone interaction or H_2O is predicted to stabilise the Ca^{2+} but was not shown in this model (Fig. 4.3B). The third Ca^{2+} binding site is made up of the residues: D303 (side chain carboxyl), N307 (carboxyamide side chain), S309 (backbone carbonyl) and D327 (side chain carboxyl) which contributes four interactions to the stabilisation of Ca^{2+} (Fig. 4.3C).

A further two interactions are predicted to form between Ca^{2+} and the protein backbone or H_2O. The Prime homology generated Ca^{2+} bound homology model (Fig. 4.1D) led to similar result regarding the side chains involved in Ca^{2+} stabilisation. In Ca^{2+} binding site 1, there is an additional interaction with Ca^{2+} N229 (side chain carboxyamide) and the side chain position has changed (Fig. 4.3D). In Ca^{2+} binding site 2, an additional interaction is seen with E450 (carboxyl side chain) (Fig. 4.3E). In Ca^{2+} binding site 3 all six interactions are present to stabilise Ca^{2+} (Fig. 4.3F). Two interactions are made with D327 (carboxyamide side chain) and an interaction with D305 (carboxyamide side chain), not seen in SWISS model homology model. This is consistent with the Ca^{2+} predicted binding sites of TG6 following alignment with TG2 and TG3 (Fig. 1.10 and Table 1.3), except interaction and with N229, in Ca^{2+} binding site 1, was not predicted.

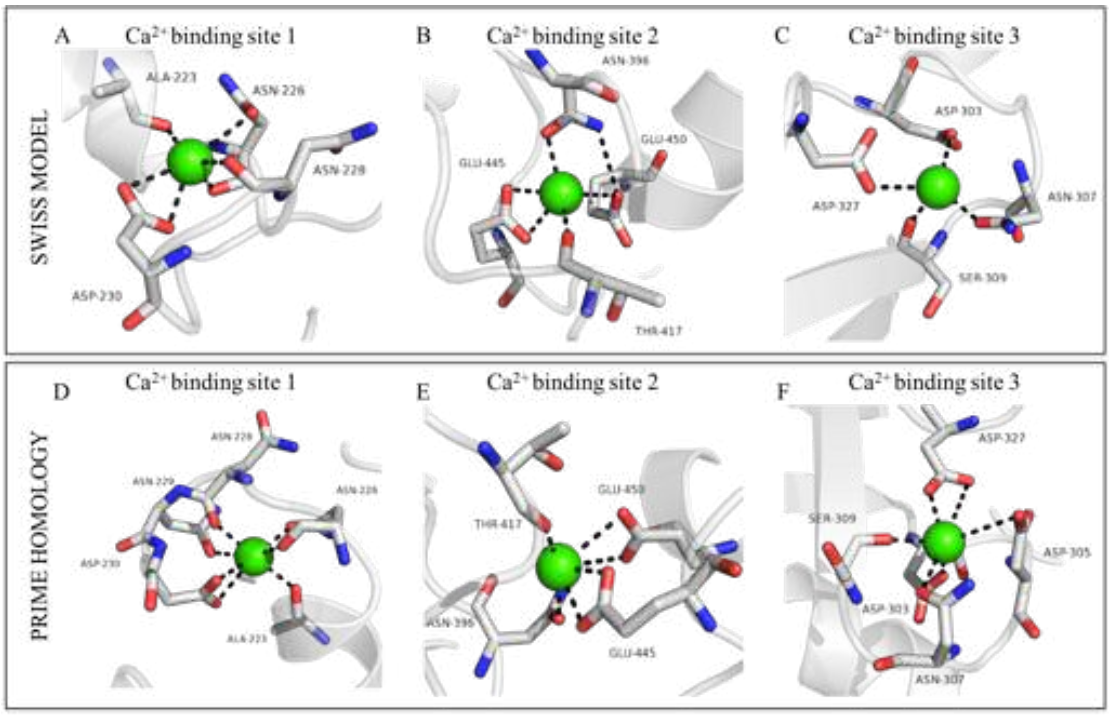


Figure 4.3: Predicted Ca^{2+} binding sites of TG6.
TG6 Ca^{2+} bound homology model generated using SWISS Model (A-C) and Prime Homology (D-F). The protein backbone (cartoon, white) and the residues (sticks representation, carbon: white, oxygen: red and nitrogen: blue) interacting with Ca^{2+} (green sphere) are shown. Polar contacts ≤3.5 Å are highlighted (black dotted line). SWISS model predicted Ca^{2+} binding sites (A-C). A: Ca^{2+} binding site 1 consists of the residues A223 N226, N228, and D230. B: Ca^{2+} binding site 2 consists of the residues, T417, E445 and E450. This shows five interactions with Ca^{2+}. C: Ca^{2+} binding site 3 consists of the residues D303, N307, S309 and D327. These residues contribute to the formation of four interactions with Ca^{2+}. Prime homology predicted Ca^{2+} binding sites (D-F). D: Ca^{2+} binding site 1 consists of the residues A223, N226, N228, N229 and D230 and shows seven possible interactions with Ca^{2+} in the binding site. E: Ca^{2+} binding site 2 consists of the residues N396, T417, E445 and E450. This shows six interactions with Ca^{2+}. F: Ca^{2+} binding site 3 consists of the residues D303, D305, N307, S309 and D327. These residues contribute to the formation of six interactions with Ca^{2+}.

Next the nucleotide binding site of TG6 was investigated. TG6•GTPγS was generated using SWISS Model and the GTPγS was manually docked (Fig. 4.4A). The residues involved in the stabilisation of GTPγS are: K170 (side chain amine), H171 (NE, imidazole side chain), K498 (side chain amine), K500 (side chain amine), V501 (backbone amide) and T603 (amide and carbonyl backbone) (Table 4.1). The GTP could not be docked into the SWISS model homology model, as it contained a loop which occluded the nucleotide pocket, which prevented assessment of GTP binding (Fig. 4.4B). Using Prime homology, GTPγS was docked into TG6 using TG3 and TG2 as a template (Fig 4.3C-D). These models were similar to the SWISS model generated TG6•GTPγS (Fig 4.4A). However, in this case no interaction is seen with K170 and an interactions is seen with P505 (carbonyl backbone) (Fig. 4.4C and Table 4.1), however, this is not true for TG6•GTPγS (TG2 template, Fig. 4.4D). GTP was also docked into TG6 using a TG3 template (Fig. 4.4E) and a TG2 template (Fig. 4.3F). The occluding loop was moved from the GTP binding site (Fig. 4.4B) allowing the interactions involved with GTP binding to be assessed. The residues responsible for GTP binding are: K170 (side chain amine), H171 (imidazole side chain), W472 (indole side chain), R474 (amine side chain), R475 (amine side chain), L502 (backbone carbonyl and amide), K600 (side chain amine) and T603 (backbone amide and carbonyl) (Fig. 4.4F and Table 4.1). This is not true for TG6•GTP (TG3 template, Fig. 4.4F), where the side chains are not in the correct position to stabilise GTP.

Similar residues stabilise GTP/GTPγS using TG2 as a template which are different to the residues stabilising GTP/GTPγS when TG3 is instead used the template. For example R474 and R475 stabilise the γ-phosphate with the TG2 template but K498 and K500 stabilise γ-phosphate with the TG3 template. These difference are most likely due to the repositioning of the occluding loop domain in homology models based on TG2, to allow GTP docking. There is currently no evidence to suggest there is a reposition of this loop, however, this is not required when the homology models are based on a TG3 template. Currently, there is no TG3•GTP or TG2•GTPγS and only TG3•GTPγS and TG2•GTP X-ray structures are available (Ahvazi *et al.* 2004a; Jang *et al.* 2014).

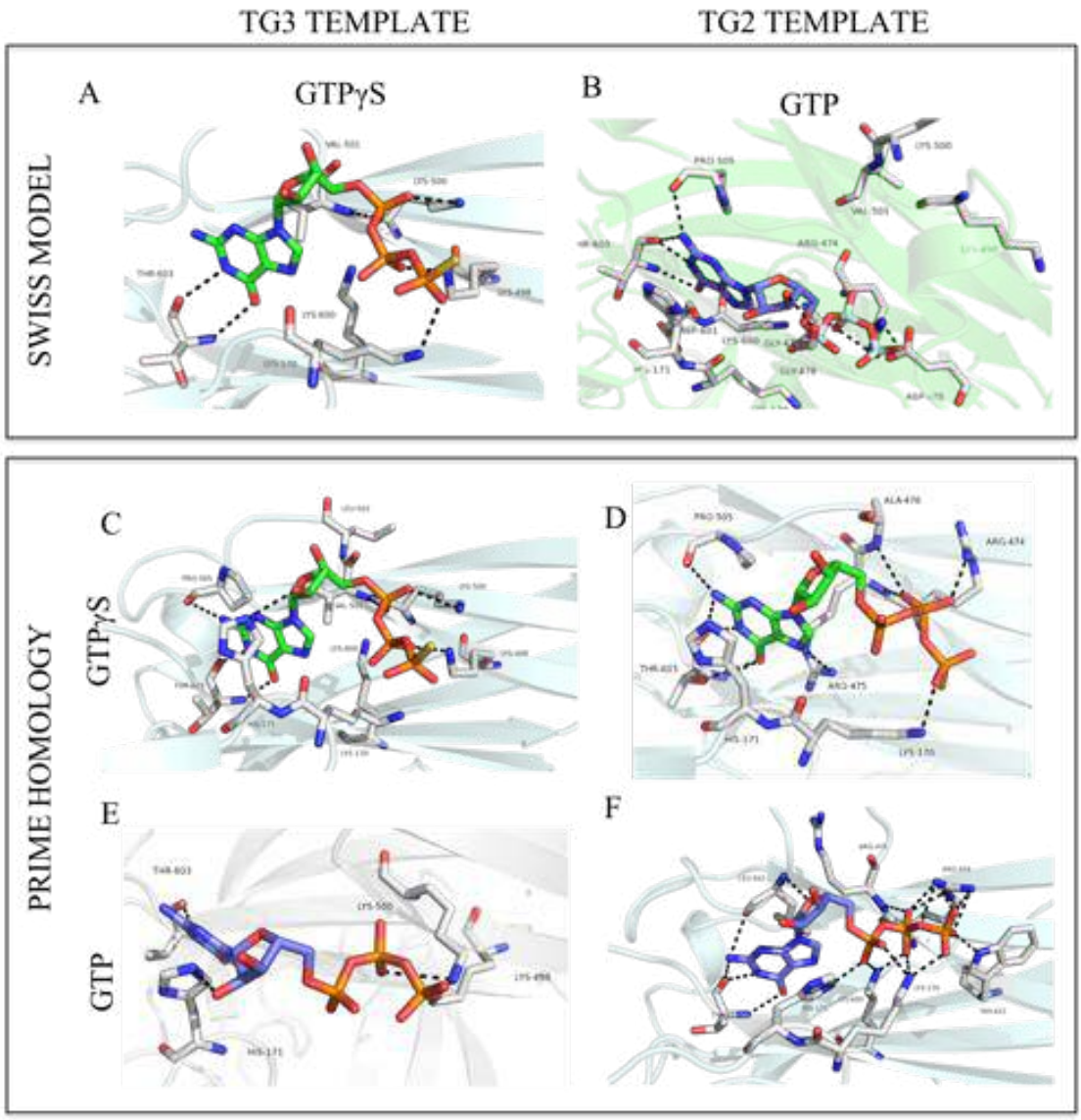


Figure 4.4: Nucleotide binding of TG6.
TG6 protein backbone shown in cartoon representation (pale blue/pale green) and GTPγS/GTP is shown in stick representation (green/lilac: carbon (guanosine), nitrogen: blue, oxygen: red, phosphate: orange and sulphur: yellow). Polar contacts (black dotted lines) between guanine nucleotide and interacting residues (white sticks) are highlighted. SWISS model homology models (A and B) and Prime Homology model (C-F) were generated. A: TG6 was modelled on TG3•GTPγS template (Fig. 4.1B) B: TG6 was modelled on TG2•GTP template (PDB: 4PYG) C: TG6•GTPγS TG6 homology model was generated using TG3•GTPγS as the template (PDB: 1RLE). D: TG6•GTPγS homology model using TG2•GTP as the template (PDB: 4PYG). E: TG6•GTP homology model using TG3•GTPγS as the template (PDB: 1RLE). F: TG6•GTP homology model using TG2•GTP as the template (Fig. 4.1E).

There is no full set of nucleotide bound X-ray structures for the respective TG's to base all TG6 homology models on for direct comparison of GTP/ GTPγS binding by TG6. Consequently, difference in the residues observed stabilising the γ-phosphate are likely template specific rather than nucleotide specific as the backbone of TG6 has not changed, only the nucleotide docked.

Table 4.2: Nucleotide interactions for the TG6 homology models.

		SWISS Model		Prime Homology			
	Template	TG2	TG3	TG2	TG2	TG3	TG3
Residue	Position/ Nucleotide	GTP	GTPγS	GTP	GTPγS	GTP	GTPγS
		Number of interactions					
Lys170	NZ	GTP clashes with loop domain in TG6	1	3	1	x	x
His171	NE		x	1	x	1	1
Trp472	NE		x	1	x	x	x
Arg474	NE		x	1	x	x	x
Arg474	NH1		x	2	1	x	x
Arg474	NH2		x	1	x	x	x
Arg475	N		x	1	1	x	x
Arg475	NH1		x	x	1	x	x
Ala476	N		x	x	1	x	x
Lys498	NZ		1	x	x	1	1
Lys500	NZ		1	x	x	1	1
Val501	N		1	x	x	x	1
Leu502	N		x	1	x	x	x
Leu502	O		x	1	x	x	x
Pro505	O		x	x	1	x	1
Lys600	NZ		x	2	x	x	x
Thr603	N		1	1	1	x	1
Thr603	O		1	2	1	1	1

Interactions (<3.5 Å) with the guanine and ribose moiety (green) and phosphates (orange) are highlighted.

4.3.2 Prevalence of TG6 mutants

There are currently 22 TG6 mutations (Table 1.5), which have been identified and are predicted to cause SCA35 and other neurological diseases. These include nonsense, missense, insertion, deletion and splice donor mutations. Information regarding the prevalence of the TG6 mutants in different populations and ethnicities was extracted from the ExAc and GnomAD databases (Table 4.3 and Table S1.). The allelic frequency of the TG6 mutants differed somewhat depending on the database used (Table S1). TG6 N137S, Q181H and T426N were not found in the GnomAD and ExAc databases and TG6 Y441C, L502Q and E574del were found only in the GnomAD database.

The majority of the mutations are classified as rare (allelic frequency: $\leq 1.0 \times 10^{-3}$) with regards to the total population tested. The TG6 mutants classified as rare: c.7+1G>T, TG6 R111C, G279ter, D327G, P359L, V391M, R412C, Y441C, P493L, L502Q, D510H, L517W and E574del (Table 4.3). An allelic frequency of (1.0×10^{-3} to 1.00×10^{-2}) is classified as uncommon. This included the mutants: TG6 E406K, G508D, V527E and Q652dup (Table 4.3). TG6 R448W was classified as common in the human population as it had an allelic frequency of $\geq 1.0 \times 10^{-2}$ (Table 4.3). In addition, the TG6 mutants are more prevalent in certain populations (based on allele count), which is summarised in Table 4.3.

Table 4.3: Prevalence of TG6 mutations in the population.

TG6 Mutant	GnomAD allelic frequency	ExAc allelic frequency	Classification	Association with specific population
c.7+1G>T	$1.40x10^{-4}$	$1.61x10^{-4}$	Rare	East Asian
R111C	$7.89x10^{-6}$	$8.23x10^{-6}$	Rare	African, East Asian and European (non-Finnish)
N137S	N/a	N/a	N/a	N/a
Q181H	N/a	N/a	N/a	N/a
G279ter	$2.04x10^{-5}$	$8.39x10^{-6}$	Rare	European (non-Finnish) South Asian Latino
D327G	$9.57x10^{-5}$	$9.90x10^{-5}$	Rare	South Asia
P359L	$1.23x10^{-4}$	$7.42x10^{-5}$	Rare	European (non-Finnish) South Asian Latino
E406K	$1.33x10^{-3}$	$1.32x10^{-3}$	Uncommon	All except Ashkenazi Jewish
V391M	$7.81x10^{-4}$	$8.98x10^{-4}$	Rare	East and South Asian European (non-Finnish) Latino
R412C	$2.83x10^{-5}$	$4.12x10^{-5}$	Rare	European (non-Finnish) East Asian
T426N	N/a	N/a	N/a	N/a
Y441C	$1.20x10^{-5}$	N/a	Rare	European (non-Finnish) African
R448W	$1.46\ x10^{-2}$	$2.83x10^{-2}$	Common	All
P493L	$7.89x10^{-5}$	$1.09x10^{-4}$	Rare	East Asian and Latino
L502Q	$7.10x10^{-6}$	N/a	Rare	European (non-Finnish)
G508D	$1.46x10^{-3}$	$1.39x10^{-3}$	Uncommon	European (Finnish and non-Finnish)
D510H	$1.50x10^{-4}$	$1.51x10^{-4}$	Rare	East Asian
L517W	$1.47x10^{-4}$	$1.1x10^{-4}$	Rare	East Asian
V527E	$1.59x10^{-3}$	$1.32\ x10^{-3}$	Uncommon	African European (non-Finnish) Latino Ashkenazi Jewish
E574del	$3.98x10^{-6}$	N/a	Rare	East Asian
Q652dup	$1.28\ x10^{-3}$	$1.34x10^{-3}$	Uncommon	European (Finnish and non-Finnish)

4.3.3 Assessing the impact of mutations using bioinformatic tools and homology modelling

The method of assessing the impact of the TG6 mutations using bioinformatic tools and homology modelling is detailed in Fig. 4.5. To summarise conservation of residues were evaluated with MSA. Bioinformatic tools assessed the function (SIFT and PolyPhen-2), structure (FoldX) and solubility (AGGRESCAN and A3D). This allowed preliminary predictions to be made and the TG6 mutants were grouped according to predicted impact e.g. structural, functional, benign and unknown but possibly damaging. Subsequently, the TG6 homology models generated (Fig. 4.1) were used to assess the local environment of the mutation and investigate the interactions in the functional binding sites (Ca^{2+} and GTP and active sites). The combination of all the bioinformatic results were used to revise the predictions made, which can then be tested biochemically.

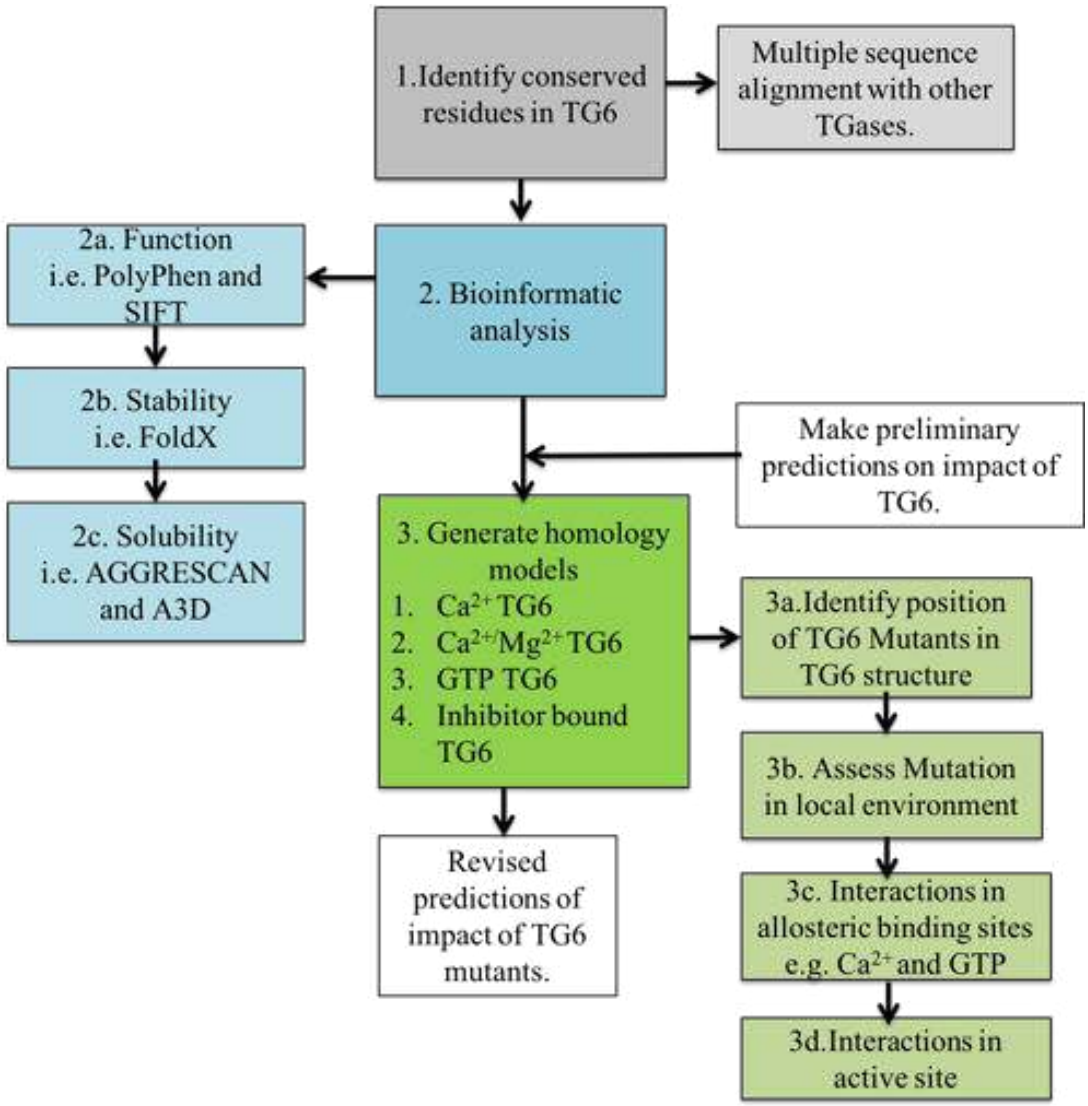
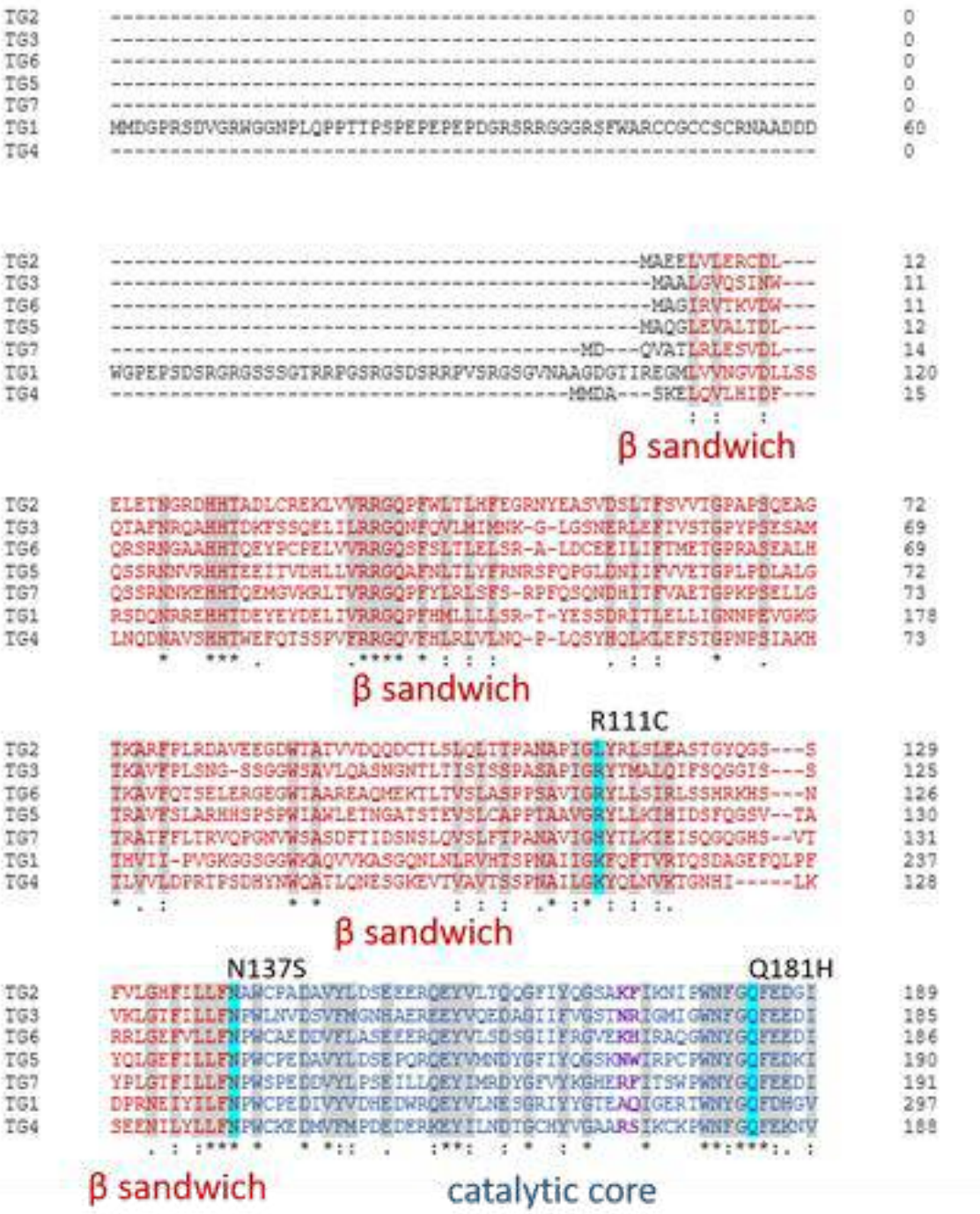
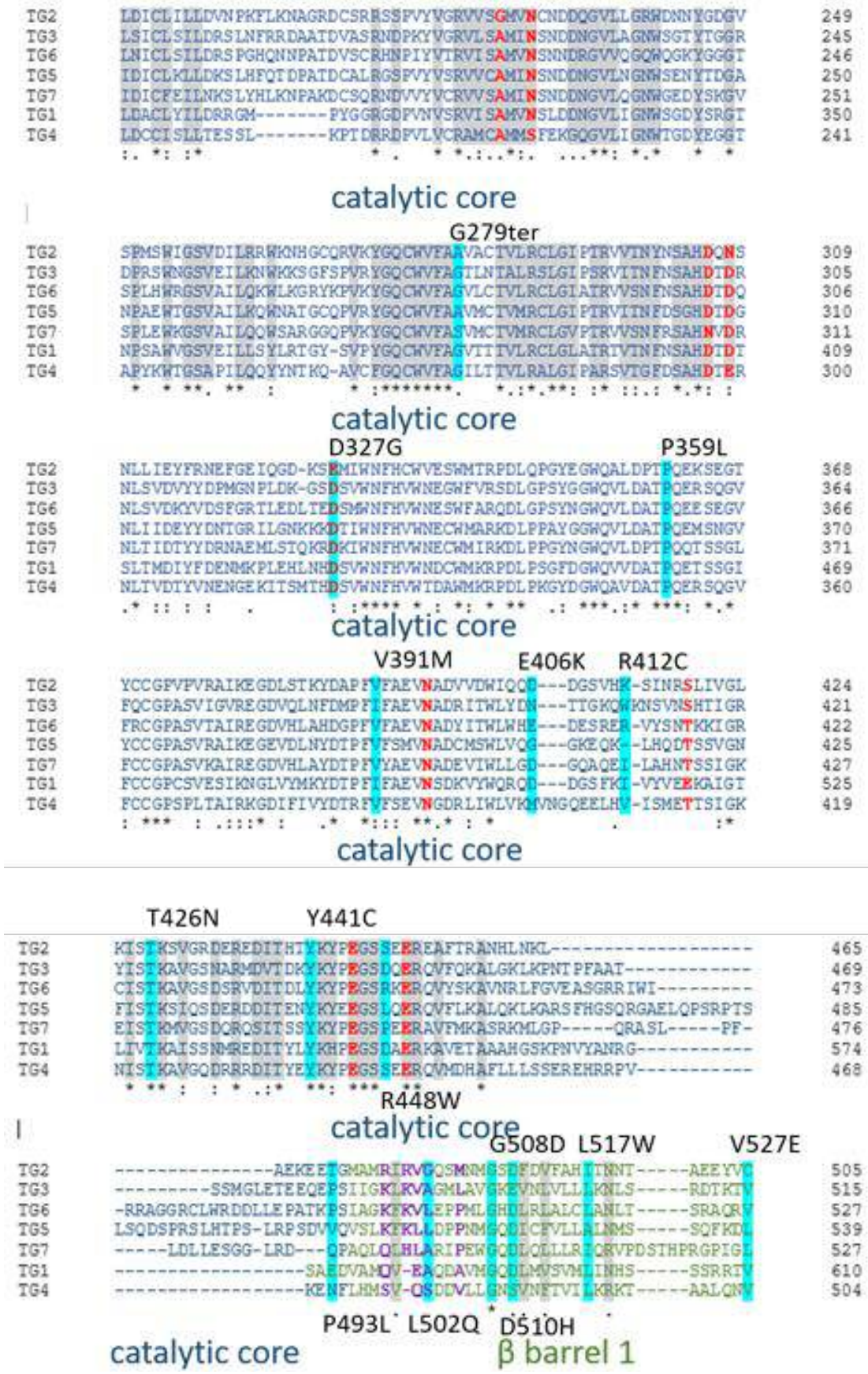


Figure 4.5: Bioinformatic analysis.
Schematic outlining the method of bioinformatic analysis used to predict the impact of TG6 mutations.

4.3.4 Conservation of TG6 mutants in TG family

MSA (CLUSTAL Omega) of human TGases (1-7) was carried out using CLUSTAL Omega, where protein sequences were obtained from UniProt (Fig. 4.6). Of the TG6 residues that are mutated; N137, Q181, P359, T426, Y441 and G508 are completely conserved in other TGases. The TG6 residues D327 and V391 are conserved residues or are substituted with a residue with a similar side chain for example D327 is replaced with a glutamate in TG2, V391 is replaced with isoleucine in TG3 and TG1. TG6 D510 is semi conserved i.e. replaced with a residue with weakly similar properties in other TGases. For example D510 is replaced with other charged/polar residues such as glutamate in TG3 and serine in TG4. The remaining TG6 residues that are mutated: TG6 R111, E406, R412, R448, P493, L502, L517 and V527, are not conserved residues. However, R111 is conserved in TG3 and TG5 and is replaced with other positively charged residues such as lysine and histidine, except TG2 which has a leucine residue in this position. L517 is conserved in TG1, TG3 and TG4 and is replaced with isoleucine in TG2 and TG7 but is alanine in TG5. V527 is conserved in TG1, TG3 and TG4 and is replaced by leucine in TG5 and TG7. There is cysteine at this position in TG2. The residue is conserved at in least two other TGases and is likely to be important as they are located within a β barrel. Five out of 22 of the analysed TG6 residues mutated showed no conservation (E406, R412, R448, P493 and L502) and therefore the impact of these mutants unknown. The polypeptide sequence of TG6 from different species was aligned (Fig. 4.7). This showed all *TGM6* mutations are in conserved positions in TG6. However, very few residues are not conserved in this alignment and therefore functional and structural importance of these residues specific to TG6 cannot be assessed.


β sandwich
β sandwich
R111C
β sandwich
N137S
Q181H
β sandwich
catalytic core


catalytic core
G279ter
catalytic core
D327G
P359L
catalytic core
V391M
E406K
R412C
catalytic core
T426N
Y441C
R448W
catalytic core
G508D
L517W
V527E
P493L
L502Q
D510H
catalytic core
β barrel 1

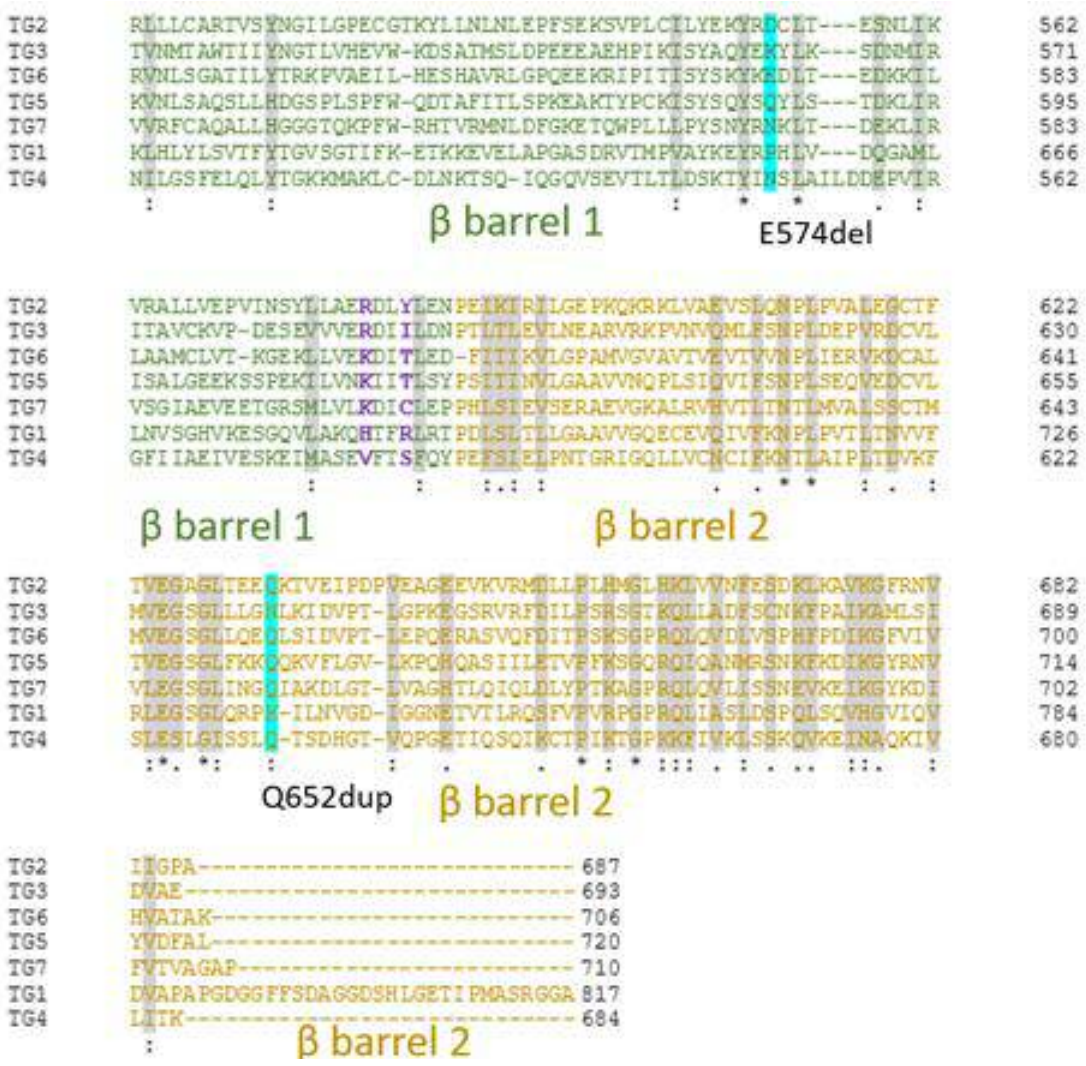
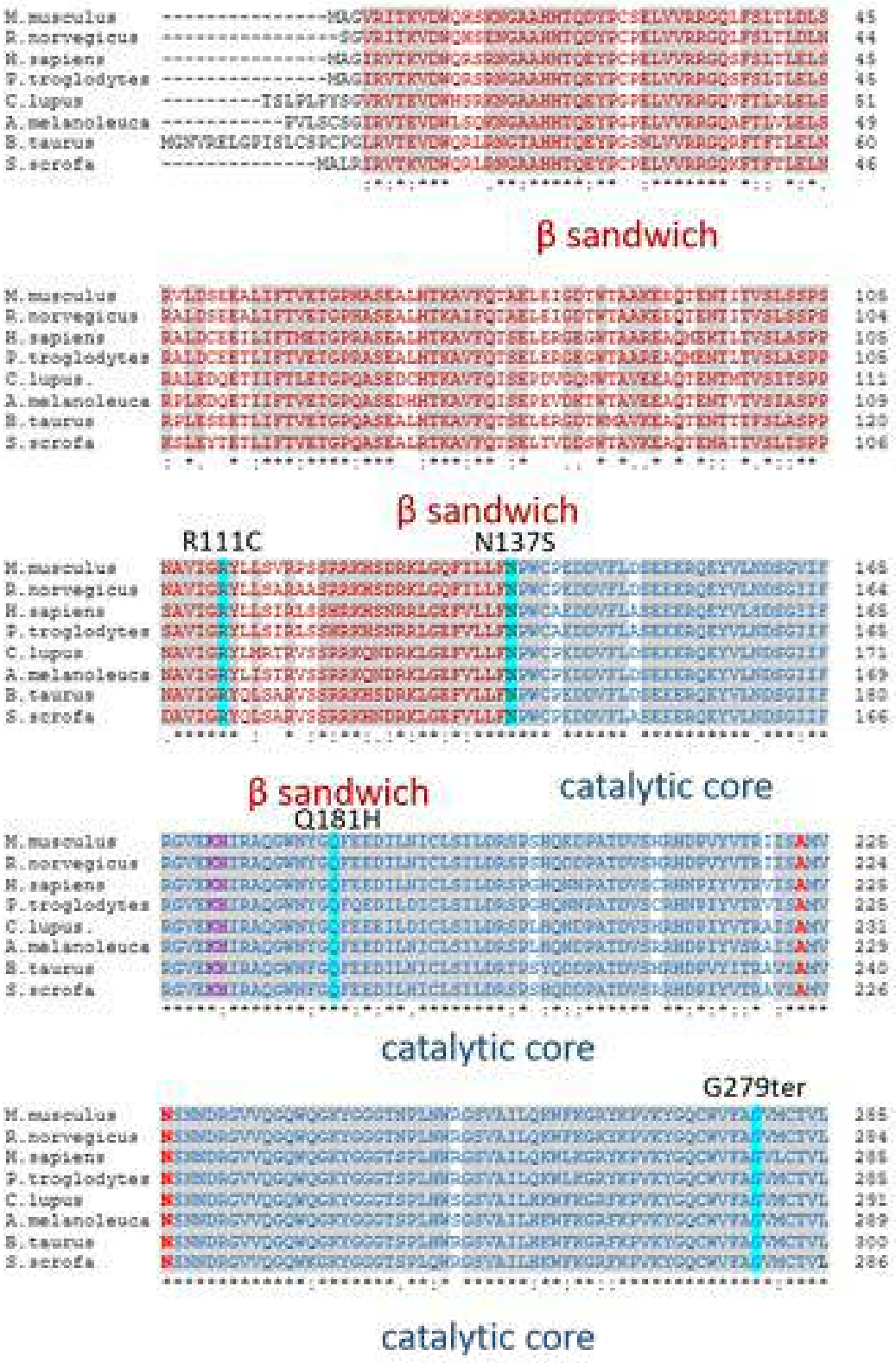
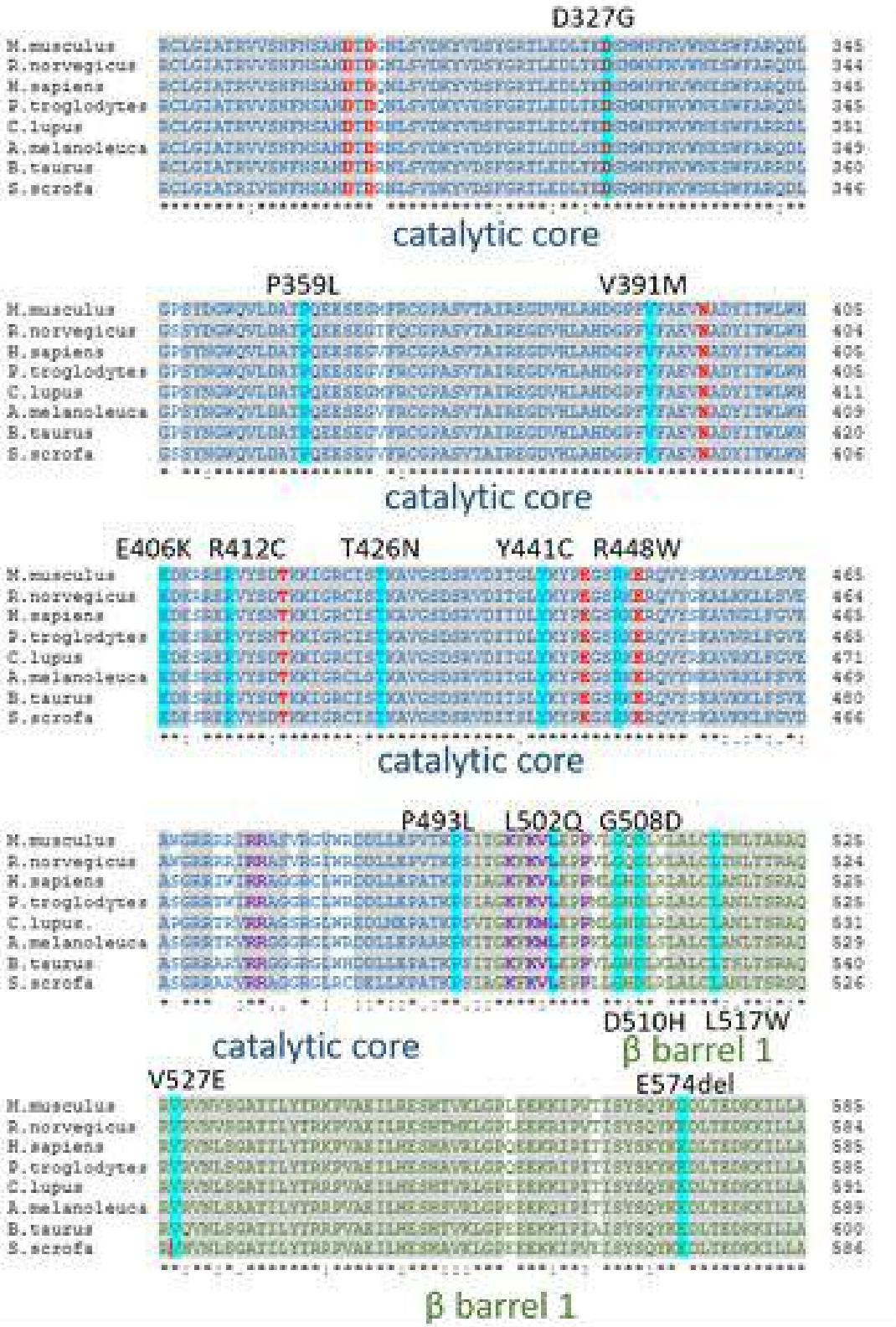


Figure 4.6: Multiple sequence alignment of TGases.
MSA highlights conserved residues between TGases (highlighted in grey), denoted with *, fully conserved residue, :, strongly conserved properties of residues and ., weakly conserved properties of residue. Domains are highlighted β sandwich (dark red, res.3-137), catalytic core (dark blue, res. 138-494), β barrel 1 (green, res. 495-606) and β barrel 2 (dark yellow, res 607-706). Residues important for Ca^{2+} binding are highlighted in red (based on TG2 and TG3) and GTP binding in purple. The location of TG6 mutants are highlighted in cyan.


β sandwich
β sandwich
R111C
N137S
β sandwich
Q181H
catalytic core
catalytic core
G279ter
catalytic core


D327G
catalytic core
P359L
V391M
catalytic core
E406K
R412C
T426N
Y441C
R448W
catalytic core
P493L
L502Q
G508D
D510H
L517W
catalytic core
β barrel 1
V527E
E574del
β barrel 1

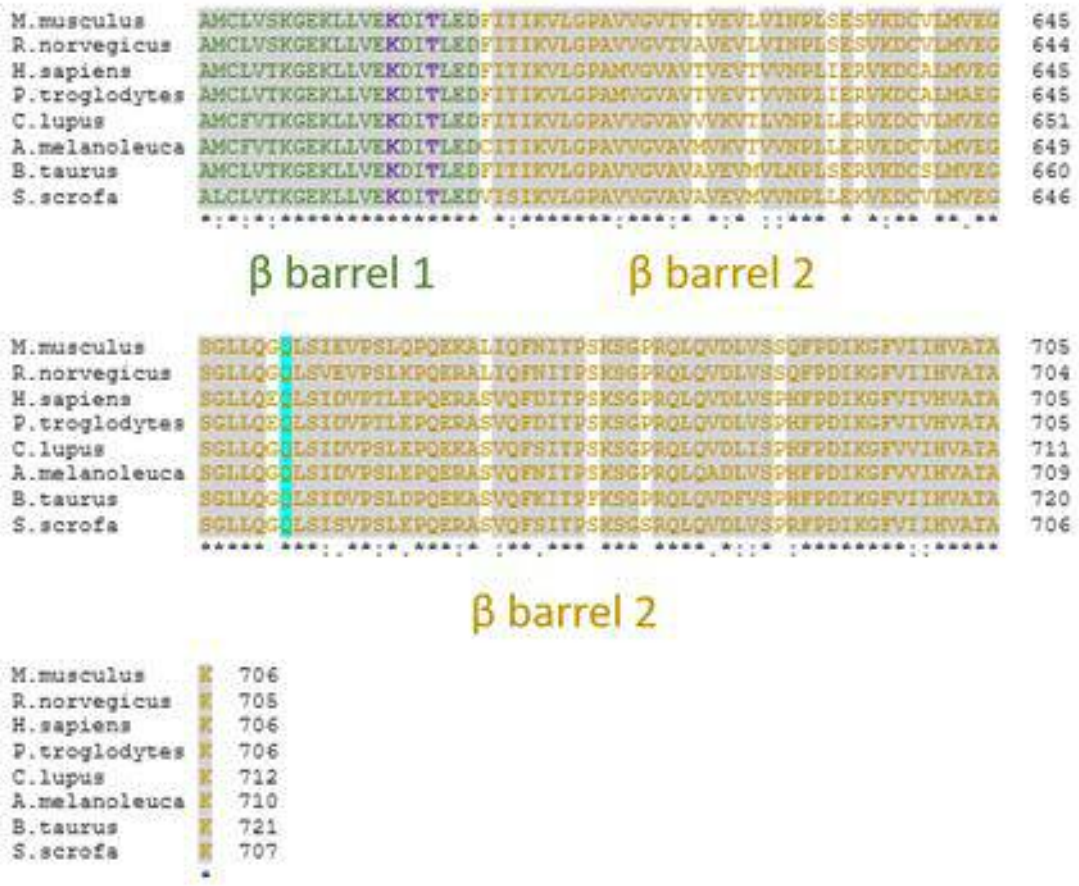


Figure 4.7 Multiple sequence alignment of TG6.
MSA highlights conserved residues in TG6 (highlighted in grey), denoted with *, fully conserved residue, :, strongly conserved properties of residues and ., weakly conserved properties of residue. Domains are highlighted β sandwich (dark red, res.3-137), catalytic core (dark blue, res. 138-494), β barrel 1 (green, res. 495-606) and β barrel 2 (dark yellow, res 607-706). Residues important for Ca^{2+} binding are highlighted in red (based on TG2 and TG3) and GTP binding in purple. The location of TG6 mutants are highlighted in cyan.

Bioinformatic analysis included the use of various tools which generated information of pathogenicity, solubility and stability. PolyPhen-2 and SIFT predicted pathogenicity of the TG6 mutant whilst FoldX was used to assess the stability and AGGRESCAN and A3D was were used to assess solubility. PolyPhen-2, SIFT and AGGRESCAN were deemed as less accurate than FoldX and A3D as they were solely based on the primary sequence whereas FoldX and A3D analysed the 3D structure of TG6. However, the accuracy of the analysis using the 3D structure is dependent on the accuracy of the homology model used.

Table 4.4: TG6 mutants grouped based on impact.

Prediction	Structural	Functional	Benign	Unknown
TG6 Mutant	R111C Q181H P359L Y441C G508D L517W V527E E574del Q652dup	N137S D327G V391M R412C	E406K P493L L502Q	T426N R448W D510H L502Q/ Q652dup

From the outcome of bioinformatic analysis using PolyPhen-2, SIFT, FoldX, AGGRESCAN and A3D, the TG6 mutants were grouped accordingly.

In order to initially categorise the 20 TG6 mutations, previous work on these mutants was taken into consideration. This was primarily linked to previous data regarding D327G and L517W. D327G is present in Ca^{2+} binding site 3 and is deficient in transamidation activity and is therefore predicted to cause impaired function (Thomas *et al.* 2013; Guo *et al.* 2014). TG6 L517W is located in the first β barrel domain in the middle of a β strand. No changes were observed when it was molecularly modelled, however due to its location and the substitution a of flexible aliphatic side chain to a bulky aromatic side chain it is predicted to impact TG6 structure (Thomas *et al.* 2013). TG6 L517W had compromised transamidation activity which is likely due to structural deficits (Guo *et al.* 2014). Specifically those showing decrease solubility and stability were categorised as structural mutations (consistent with TG6 L517W), whilst those which were comparable to WT TG6 with respect to stability and solubility or had increased stability, consistent with TG6 D327G, were categorised as functional mutations. TG6 mutants with inconsistent bioinformatic findings were allocated as 'unknown' and those polymorphisms predicted in the first instance to be tolerated or benign were predicted to be probably benign. Based

Tables 4.5-8.

Analysis using PolyPhen-2 and SIFT indicated the missense mutations (R111C, Q181H, P359L, Y441C, G508D, L517W and V527E) were predicted to affect protein function (scores of 0.00-0.03) by SIFT and were predicted as 'probably damaging' by PolyPhen-2 (scores of 0.986-1.000) (Table 4.5). This analysis could not be carried out for E574del and Q652dup. FoldX calculated the stability for all mutants predicted to affect the structure of TG6 (Table 4.6). The FoldX predictions were categorised based on the known accuracy of FoldX (0.46 kcal/mol, based on standard deviation (Guerois *et al.* 2002; Schymkowitz *et al.* 2005; Studer *et al.* 2014). The following categories are as follows: highly stabilising (< −1.84), stabilising (−1.84 to −0.92), slightly stabilising (−0.92 to −0.46), neutral (−0.46 to +0.46), slightly destabilising (+0.46 to + 0.92), destabilising (+ 0.92 to +1.84), and highly destabilising (> +1.84).

TG6 R111C, was the most stable in both open and closed conformation (ΔΔG = 1.1 and 1.8 kcal/mol, respectively) for the structural TG6 mutant group, but even so TG6 R111C destabilises structure when compared with WT TG6. TG6 Q181H is less stable in closed conformation (ΔΔG = 3.2 kcal/mol, denoted as highly destabilising) and has greater stability in open conformation (ΔΔG = 1.5 kcal/mol) when compared with WT TG6, TG6 P359L was predicted to be highly destabilising in closed conformation (ΔΔG = 3.4 kcal/mol) and destabilising in open conformation (ΔΔG = 1.3 kcal/mol). In both open and closed conformation TG6 Y441C is predicted to be highly destabilising (ΔΔG = 5.0 and 6.4 kcal/mol, respectively) when compared with WT TG6.

TG6 Q652dup was found to be more stable in open conformation (ΔΔG = 20.8 kcal/mol) compared with closed conformation (ΔΔG = 110.7 kcal/mol) and both conformations

for E574del and Q652dup. For example the change in the number of residues between WT and these TG6 mutants can lead to errors in the calculation. Both predictions for closed conformation are significantly different (109-117% change) compared to the other TG6 mutants (2-7% change) but calculations in open conformation are within the range of the other TG6 mutants. This could suggest there are significant changes in the interactions within the 3D structure but is more likely the FoldX calculation is unsuitable for indel mutations.

The solubility of the TG6 mutants was investigated using AGGRESCAN and A3D (Table 4.7-8, Fig. 4.8 and 4.9). AGGRESCAN showed small changes (-0.4 to 0.3) in the solubility of TG6 mutants compared with WT TG6 when looking at the normalised sum of the average aggregation propensity values across the whole polypeptide sequence (NavSS) (Table. 4.7). Marginal changes in NavSS were calculated for the structural mutants, suggesting no impact on the solubility of TG6. In addition, AGGRESCAN can highlight changes in the number or length of aggregation prone areas (hotspots) within the polypeptide. TG6 Q181H causes the extension of a neighbouring hotspot (res. 186-194 was extended to res.184-194) thus potentially decreasing its solubility. However, this is not reflected in the NavSS value which is identical to WT TG6.

AGGRESCAN can also be utilised to analyse the implications of insertion and deletion mutations on the solubility of proteins (Table 4.8). The total hotspot area (THSA) and the area of the aggregation profile above the threshold per residue (AATr) was used as an indicator of changes in the solubility. These values were used to assess solubility changes in TG6 E574del and TG6 Q652dup. No change was seen in the AATr for both TG6 E574del and Q652dup when compared to WT TG6. However, a change in the THSA was calculated for TG6 E574del (ΔTHSA = 0.32) indicating a decrease in the solubility

Table 4.5: PolyPhen-2 and SIFT analysis of TG6 structural mutants.

		R111C	Q181H	P359L	Y441C	G508D	L517W	V527E
PolyPhen-2	Result	Probably damaging	Probably damaging	Probably damaging	Probably damaging	Probably damaging	Probably damaging	Probably damaging
	Score	1.000	1.000	1.000	1.000	1.000	1.000	0.986
SIFT	Result	Affects protein function	Affects protein function	Affects protein function	Affects protein function	Affects protein function	Affects protein function	Affects protein function
	Score	0.03	0.00	0.00	0.00	0.00	0.02	0.03

Table 4.6: FoldX analysis of structural TG6 mutants.

Conformation	Stability	WT	R111C	Q181H	P359L	Y441C	G508D	L517W	V527E	E574del	Q652
Closed TG6	$\Delta G_{folding}$ (kcal/mol)	94	95.8	97.1	97.4	100.4	100.5	98.9	96.8	197.2	204.6
	ΔΔG (kcal/mol)	x	1.8	3.2	3.4	6.4	6.5	4.9	2.8	103.2	110.7
Open TG6	$\Delta G_{folding}$ (kcal/mol)	175	176.0	173.5	176.3	180.0	199.2	185.9	178.3	168.0	195.7
	ΔΔG (kcal/mol)	x	1.1	-1.5	1.3	5.0	24.2	11.0	3.3	-7.0	20.8

Stability ($\Delta G_{folding}$) was calculated for TG6 in 'closed' (TG6-Ca^{2+} bound (Fig. 4.1A)) and 'open' conformatio
for WT and the TG6 mutants respectively. The TG6 mutants were grouped according to their ΔΔG: high
stabilising (−1.84 to −0.92, light green), slightly stabilising (−0.92 to −0.46, yellow), neutral (−0.46 to +0.46,
to + 0.92, light orange), destabilising (+ 0.92 to +1.84, dark orange), and highly destabilising (> +1.84, red).

Table 4.7: Analysis of structural missense TG6 mutants using AGGRESCAN and A3D.

	AGGRESCAN					TG6 (closed) A3D					TG6 (op
TG6	**Number of Hot Spots**	**NavSS**	**ΔNavSS**	**Located within a Hotspot?**	**Result**	**Average score**	**Δ Average score**	**Total score**	**Δ Total score**	**% change**	**Average score**
WT	23	-3.8	N/a	N/a	N/a	-0.817	N/a	-1149.11	N/a	N/a	-0.932
R111C	23	-3.5	0.3	Yes	No change	-0.809	0.008	-1137.46	11.65	1.01	-0.933
Q181H	23	-3.8	0.0	No	Extends hotspot	-0.815	0.002	-1145.98	3.13	0.27	-0.934
P359L	23	-3.5	0.3	No	No change	-0.813	0.004	-1142.78	6.33	0.55	-0.93
Y441C	23	-3.9	-0.1	No	No change	-0.806	0.011	-1132.86	16.25	1.41	-0.935
G508D	23	-4.0	-0.2	No	No change	-0.815	0.002	-1145.72	3.39	0.30	-0.933
L517W	23	-3.8	0.0	Yes	No change	-0.812	0.005	-1140.94	8.17	0.71	-0.935
V527E	23	-4.2	-0.4	No	No change	-0.814	0.003	-1145.1	4.01	0.35	-0.936

Solubility of the structural TG6 mutants predicted by AGGRESCAN and A3D. An increased solubility (bl
highlighted. For A3D, a change in solubility is indicated by a change in total score of 0.5% compared with W

Table 4.8: Analysis of insertion and deletion TG6 mutants with AGGRESCAN.

TG6 Protein	Number of Hot Spots (nHS)	Total Hot Spot Area (THSA)	Change in THSA	AAT per residue (AATr)	Located within a Hotspot?	Consequence
WT	23	77.88	N/a	0.12	N/a	N/a
E574del	23	78.20	0.32	0.12	No	Extends hotspot
Q652dup	23	77.88	0.00	0.12	No	No change

Solubility of indel mutations predicted by AGGRESCAN compared with WT TG6.

The average and total aggregation propensity scores were generated for the TG6 missense mutants and are based on the length of the protein (Table 4.7). As a measure of solubility the Δ in average score (0.002 to 0.011, Table 4.7) and Δ total aggregation (global aggregation) propensity scores (3.13 to 16.25) was calculated. The significance of the change is dependent upon the protein, so in this case a minimum change of 0.5% in global solubility was selected, as default, to represent a change in the aggregation propensity of TG6. This will then need to be adjusted following biochemical testing (Carija *et al.* 2016).

TG6 R111C, P359L, Y441C and L517W showed a significant change in solubility in closed conformation (Δ in total score =11.65, 6.33, 16.25 and 8.17, Table 4.7). In open conformation all missense mutants showed small changes in total score (Δ in total score =-2.29 to 1.48) but the change was not sufficient to affect the solubility of TG6.

A3D also generates a PDB file, containing the mutation, which highlights the aggregation prone regions (replaces the temperature column with A3D score in the file). The generated WT TG6 and TG6 mutant structures were compared for changes in global aggregation hotspots (Fig. 4.8). The structures were visualised in surface representation and were coloured according to b factor i.e. temperature and in this case A3D score. Areas of high solubility (blue) and areas of low solubility and therefore aggregation prone (red) were highlighted (based on a blue, white, red scale). Localised changes in solubility were identified in three out of seven of the TG6 missense mutants (Fig. 4.9). TG6 R111C shows a decrease in solubility at the β sandwich catalytic core junction in open and closed conformation. An increase in solubility was observed in the case of TG6 G508D in the mutation's local environment (β barrel 1) in closed conformation; this was not seen in open conformation. TG6 P359L showed an increase in the solubility in two areas and no changes were evident in aggregation propensity of TG6 Q181H, Y441C and V527E in the local environment of these mutations.

When assessing the stability of WT TG6 in open and closed conformation, it is predicted open conformation ($\Delta G_{folding}$ = 175 kcal/mol) is significantly less stable than closed conformation ($\Delta G_{folding}$ = 94 kcal/mol) as more energy is required for folding. This consistent with closed conformation containing extensive networks of interactions of buried residues, whilst open conformation has a greater surface area exposed to the environment and thus is less stable.

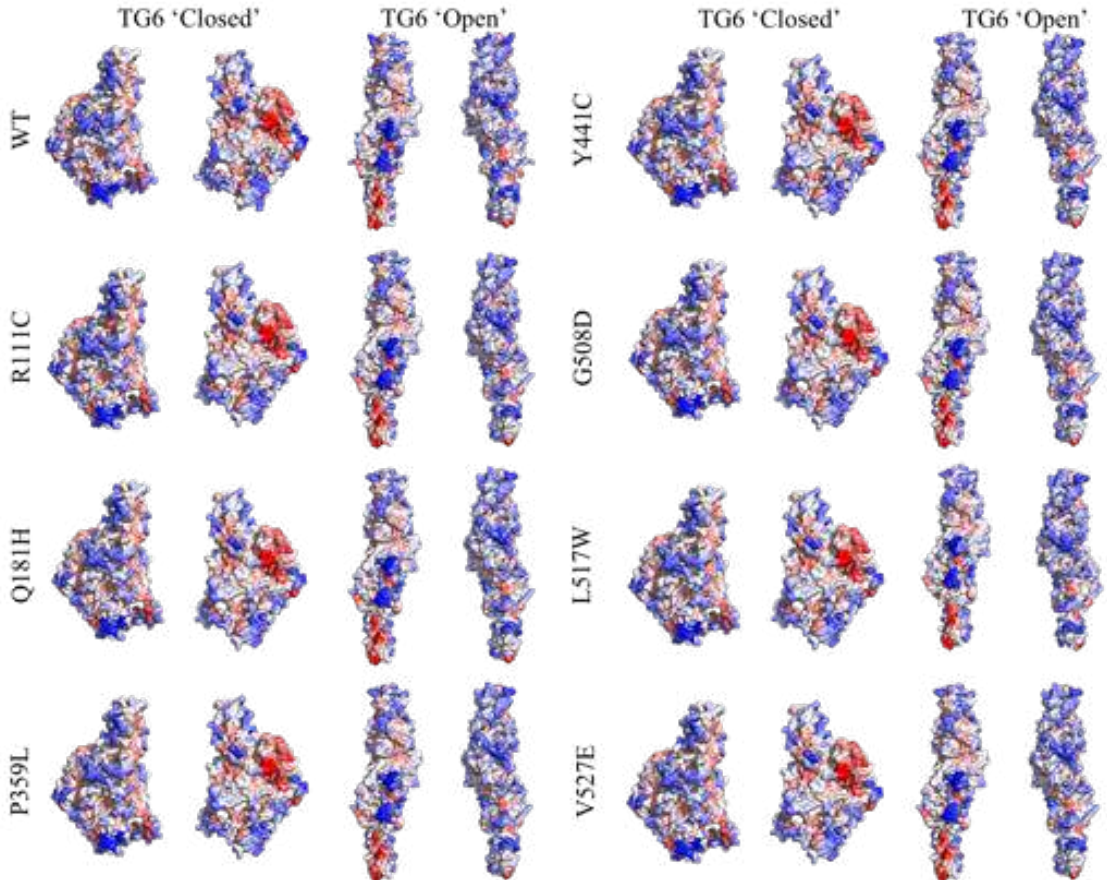


Figure 4.8: Visualisation of aggregation prone regions in structural TG6 mutants. A3D analysis showing the areas of high solubility (blue) to low solubility (red) in both open (Fig. 4.1C) and closed TG6 (Fig. 4.1B) homology models. Generated structures are shown in surface representation and left and right image pair reflects a 180° rotation.

The TG6 mutants were predicted to have structural consequences based predominantly on the FoldX and A3D results, in the context of the open and closed conformations of TG6. Closed conformation is adopted during synthesis and whilst it is intracellular and FoldX indicated all 9 TG6 mutants were destabilising or highly destabilising in this conformation. This suggests structural consequences will be implicated during synthesis and the TG6 mutant will not adopt the conventional closed conformation. The open conformation is predominantly adopted extracellularly and is consistent with enzymatic activation. Using FoldX, 7 out of 9 of these TG6 mutants also were predicted to be destabilising and highly destabilising. As the structure is destabilised in open conformation, the activity of these TG6 mutants may be impaired or abolished. The remaining two mutants, TG6 Q181H and E574del were predicted to have greater stability in open conformation compared to WT TG6. This suggests the structural impact differs for these TG6 mutants and it is likely they retain enzymatic function.

Figure 4.9: Close up of areas affected in TG6 structural mutants.
Changes in solubility were highlighted in the respective structures and the respective TG6 mutants were compared around the site of the mutation with WT. Differences are highlighted with yellow arrows.

A3D predicted decreased solubility in closed conformation for four of the missense mutations (R111C, P359L, Y441C and L517W), suggesting an increased capacity to aggregate compared with WT TG6. Changes in solubility are observed for TG6 in closed conformation but not open conformation. This is unexpected as changes in stability predicted by FoldX indicate a structural impact which in turn may expose buried residues and decrease solubility. The lack of difference may be due to the fixed backbone of the model allowing movement of only the sidechains and therefore no change in hydrophobicity of the environment. Alternatively, the change in the solubility from closed to open conformation could mask the small solubility change observed in closed conformation.

4.3.6 Assessing TG6 mutants predicted to influence TG6 function

The TG6 mutants predicted to impact the enzymatic function are: N137S, D327G, V391M and R412C. PolyPhen-2 and SIFT indicated that all the TG6 mutants predicted to affect function were ‘probably damaging’ by PolyPhen-2 (Table. 4.9, score 0.993-1.00).

Table 4.9: PolyPhen-2 and SIFT analysis of TG6 mutants predicted to have functional deficits.

		N137S	D327G	V391M	R412C
PolyPhen-2	Result	Probably damaging	Probably damaging	Probably damaging	Probably damaging
	Score	0.993	1.000	1.000	1.000
SIFT	Result	Affects protein function	Affects protein function	Affects protein function	Affects protein function
	Score	0.00	0.00	0.03	0.04

Using FoldX, TG6 D327G, V391M and R412C are predicted to have increased stability in closed conformation (ΔΔG = -3.6, -1.1 and -9.4 kcal/mol) and no significant difference in stability in open conformation (ΔΔG = 0.2, -0.2 and 0.3 kcal/mol, Table 4.10). The stability of TG6 N137S (slightly destabilising) was similar to WT TG6 in closed conformation (ΔΔG = 0.7 kcal/mol). Conversely, it is predicted to be highly stable in open conformation (ΔΔG = -57.9 kcal/mol).

Table 4.10: FoldX analysis of TG6 mutants predicted to have deficits in function.

Conformation	Stability	WT	N137S	D327G	V391M	R412C
Closed TG6	$\Delta G_{folding}$ (kcal/mol)	94.0	94.7	90.4	92.9	84.6
	ΔΔG (kcal/mol)	x	0.7	-3.6	-1.1	-9.4
Open TG6	$\Delta G_{folding}$ (kcal/mol)	175.0	117.1	175.1	174.8	175.2
	ΔΔG (kcal/mol)	x	-57.9	0.1	-0.2	0.2

Stability (ΔG) was calculated for TG6 in 'closed' (TG6-Ca^{2+} bound (Fig. 4.1A)) and 'open' conformation (inhibitor bound- TG6 (Fig, 4.1C)) for WT and the TG6 mutants respectively. For details of analysis see Table 4.6.

AGGRESCAN analysis predicted a small change in solubility for TG6 N137S, D327G, V391M and R412C (Δ NavSS = 0.2, 0.2, -0.1 and 0.3, Table 4.11) In addition, the TG6 N137S mutation created a new aggregation prone hotspot (res. 139-144) indicating that this mutant may have a greater propensity to aggregate. The TG6 D327G mutation caused the extension of an aggregation hotspot (hotspot res.329-333, is extended to res.337) thus potentially decreasing its solubility. TG6 V391M shows a shortening of the aggregation prone hotspot V391 is located within (res.392-402 instead of 390-402). However, the changes in predicted hotspots are not reflected in the Δ NavSS and therefore may not contribute to a global change in solubility.

Most of the TG6 mutants predicted to have functional deficits showed marginal changes in aggregation propensity (Table 4.11). However, the TG6 R412C mutations was predicted to affect the aggregation propensity of TG6 to a significant degree (>0.5% change) with a Δ in average score of 0.015 and Δ total score of 22.11, in closed conformation and a Δ in average score of 0.012 and Δ total score of 8.50 in open conformation, respectively.

Table 4.11: Analysis of functional missense TG6 mutants using AGGRESCAN and A3D.

	AGGRESCAN					TG6 (closed, Fig. 4.1A) A3D					TG6 (open, Fig. 4.1C) A3D				
TG6	**Number of Hot Spots**	**NavSS**	**ΔNavSS**	**Located within a Hotspot?**	**Result**	**Average score**	**Δ Average score**	**Total score**	**Δ Total score**	**% change**	**Average score**	**Δ Average score**	**Total score**	**Δ Total score**	**% change**
WT	23	-3.8	N/a	N/a	N/a	-0.817	N/a	-1149.11	N/a	N/a	-0.932	N/a	-653.57	N/a	N/a
N137S	24	-3.6	0.2	Yes	Causes addition hotspot	-0.815	0.002	-1146.07	3.04	0.26	-0.934	-0.002	-654.55	-0.98	0.15
D327G	23	-3.6	0.2	No	Extends hotspot	-0.813	0.004	-1143.71	5.40	0.47	-0.934	-0.002	-654.54	-0.97	0.15
V391M	23	-3.9	-0.1	Yes	Shortens hotspot	-0.814	0.003	-1144.10	5.01	0.44	-0.934	-0.002	-654.89	-1.32	0.20
R412C	23	-3.5	0.3	No	No change	-0.802	0.015	-1127.00	22.11	1.92	-0.920	0.012	-645.07	8.50	1.30

Solubility analysis of the TG6 mutants using AGGRESCAN and A3D. An increase in solubility (blue) and decrease in solubility (red) considered to be significant is highlighted. For details of analysis see Table 4.7

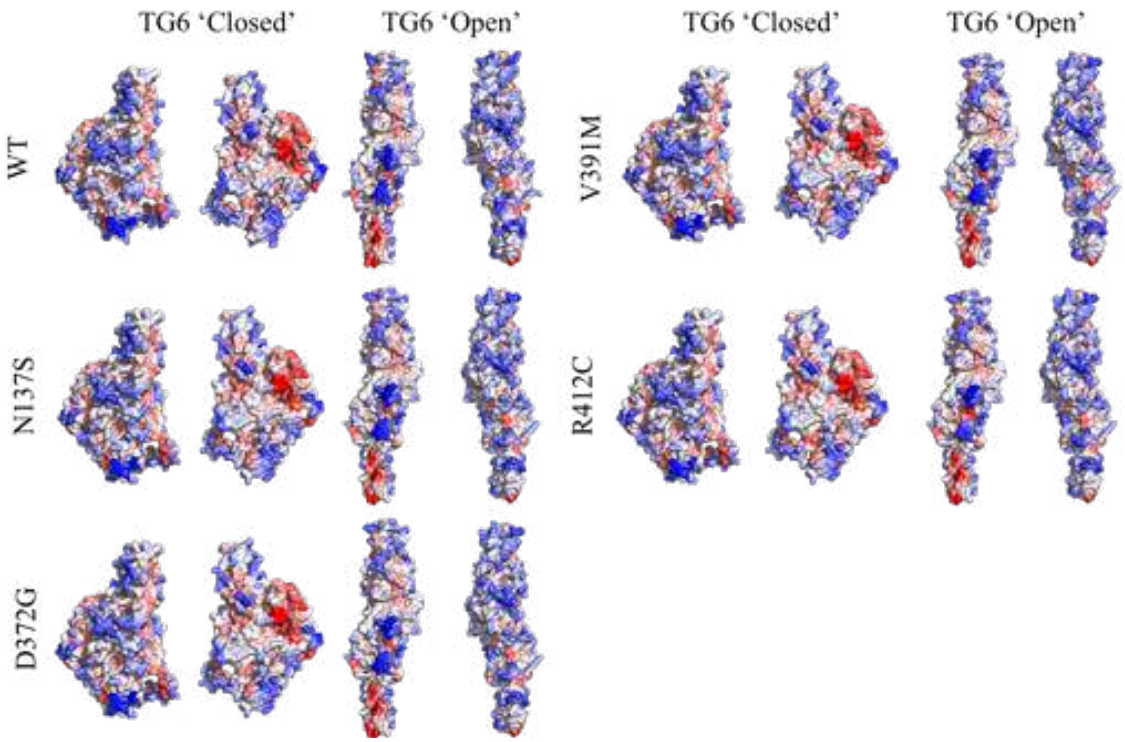


Figure 4.10: Visualisation of aggregation prone region in TG6 mutants.
A3D analysis showing the areas of high solubility (blue) to low solubility (red) in both open and closed TG6 homology models (Fig. 4.1A and C). The structures generated by AGGREGSCAN3D are shown in surface representation. Left and right image pairs represent a 180° rotation.

The output A3D output PDB file were used to visualise changes in aggregation propensity (Fig. 4.10). Small changes in local aggregation propensity were identified in all four of the TG6 missense mutants. TG6 N137S shows an increase in the solubility (red to white) of the local environment of N137 located at the β sandwich and catalytic core junction in both conformations causing changes in neighbouring areas (Fig. 4.11). TG6 D327G, shows no visible change in the solubility in closed conformation but a small change in open conformation (area of red increases in size). TG6 R412C shows two areas of change (Fig. 4.11, R412C 1 and R412C 2) in closed conformation. In the environment of R412 (R412C 1, Fig. 4.11) there is a small decrease in the local solubility (increase in red intensity) when compared to WT TG6. The second area of change is located in β barrel 1 (R412C 2, Fig. 4.11). This area shows a decrease in the local solubility (blue to a paler blue). In open conformation an area of high solubility shows a small change that apparently impacts the surrounding area. The A3D structures showed no change in the solubility in open and closed conformation for TG6 V391M. In conclusion the missense mutants are largely stable and display an aggregation propensity similar to WT, with the exception of R412C where a substantial increase in the latter is seen.

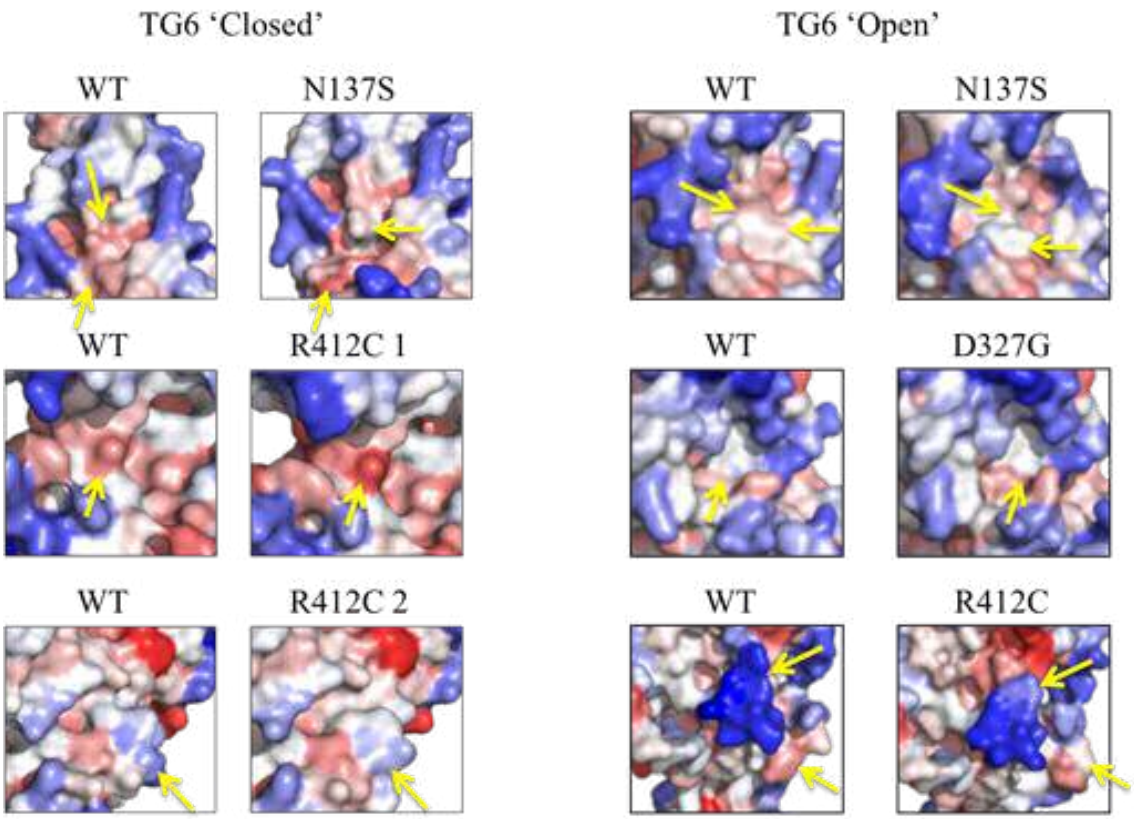


Figure 4.11: Close up of areas affected in TG6 functional mutants.
Changes in solubility were highlighted in the A3D generated structures when the respective TG6 mutants were compared to WT. Changes are highlighted with yellow arrows. All structures were analysed in surface representation with blue indicating areas of high solubility and red indicating aggregation prone hydrophobic areas.

4.3.7 Assessing TG6 mutants with amino acid substitutions that are predicted to be probably benign

Three of the TG6 mutations (E406K, P493L and L502Q) were predicted to be benign. The TG6 mutants were analysed as before, firstly with PolyPhen-2 and SIFT (Table 4.12). All three TG6 mutations, were predicted to be tolerated with SIFT. However, P493L and L502Q were predicted to be damaging by PolyPhen-2 with a score of 0.981 and 0.995, respectively (Table 4.12).

Table 4.12: PolyPhen-2 and SIFT analysis of benignTG6 mutations.

		E406K	P493L	L502Q
PolyPhen-2	Result	Benign	Probably damaging	Probably damaging
	Score	0.002	0.981	0.995
SIFT	Result	Tolerated	Tolerated	Tolerated
	Score	0.79	1.00	0.35

FoldX analysis showed TG6 E406K mutation was predicted to stabilise TG6 structure (ΔΔG = -2.1 kcal/mol) in closed conformation and no significant difference was found in open conformation (Table 4.13). TG6 L502Q is predicted to be destabilising in closed conformation (ΔΔG = 1.2 kcal/mol) and comparable to WT in open conformation (ΔΔG = 0.3 kcal/mol). No significant change in the stability of TG6 P493L was predicted in open or closed conformation when compared with WT TG6.

Table 4.13: FoldX analysis of TG6 mutations predicted to be probably benign.

Conformation	Stability	WT	E406K	P493L	L502Q
Closed TG6	$\Delta G_{folding}$ (kcal/mol)	94.0	91.9	94.7	95.1
	ΔΔG (kcal/mol)	x	-2.1	0.7	1.2
Open TG6	$\Delta G_{folding}$ (kcal/mol)	175.0	174.4	175.8	175.3
	ΔΔG (kcal/mol)	x	-0.6	0.8	0.3

Stability (ΔG) was calculated for TG6 in 'closed' (TG6-Ca^{2+} bound (Fig. 4.1A)) and 'open' conformation (inhibitor bound- TG6 (Fig, 4.1C)) for WT and the TG6 mutants respectively. For details of analysis see Table 4.6.

AGGRESCAN and A3D was used to analyse the aggregation propensity of the TG6 mutants (Table 4.14 and Fig. 4.12). With AGGRESCAN, small changes in the NavSS score were observed for TG6 E406K (0.1), P439L (0.3) and L502Q (0.4) but was insufficient to impact the aggregation propensity. The TG6 L502Q mutation causes the loss of a hotspot (res. 496-503), which supports an increase in solubility for the respective mutant. Whilst P493L caused the extension of a hotspot (res. 496-503 to 494 to 503).

Table 4.14: Analysis of TG6 mutants predicted to be using AGGRESCAN and A3D.

	AGGRESCAN					TG6 (closed, Fig. 4.1A) AD3					TG6 (open, Fig. 4.1C) AD3				
TG6	**Number of Hot Spots**	**NavSS**	**ΔNavSS**	**Located within a Hotspot?**	**Result**	**Average score**	**Δ Average score**	**Total score**	**Δ Total score**	**% change**	**Average score**	**Δ Average score**	**Total score**	**Δ Total score**	**% change**
WT	23	-3.8	N/a	N/a	N/a	-0.817	N/a	-1149.11	N/a	N/a	-0.932	N/a	-653.57	N/a	N/a
E406K	23	-3.7	0.1	No	No change	-0.814	0.003	-1144.21	4.90	0.43	-0.930	0.002	-651.66	1.91	0.29
P493L	23	-3.5	0.3	No	Extends hotspot	-0.812	0.005	-1142.29	6.82	0.59	-0.928	0.004	-650.59	2.98	0.46
L502Q	22	-4.2	-0.4	Yes	Loss of hotspot	-0.814	0.003	-1144.51	4.60	0.40	-0.946	-0.014	-663.28	-9.71	1.49

For analysis details see Table 4.7.

With A3D a significant change was calculated for TG6 P493L in closed conformation, with a change in total score of 6.82, which indicates the aggregation propensity is increased (Table 4.14). Interestingly, TG6 L502Q showed increased solubility in open conformation (Δ in total score: -9.71). Analysing the A3D generated structures, no change was observed for open and closed conformation for TG6 E406K. For TG6 P493L, a decrease in solubility was apparent in open conformation only (area of blue is reduced and red area has increased) (Fig. 4.12). TG6 L502Q showed similar changes in solubility when in open and closed conformation, i.e. an increase in the solubility of β barrel 1 where L502 is located. Although TG6 E406K and P493L were both characterised as benign, TG6 P493L showed significant changes in A3D (Δ total score = 6.82 (closed conformation) which indicated TG6 solubility is likely to be affected.

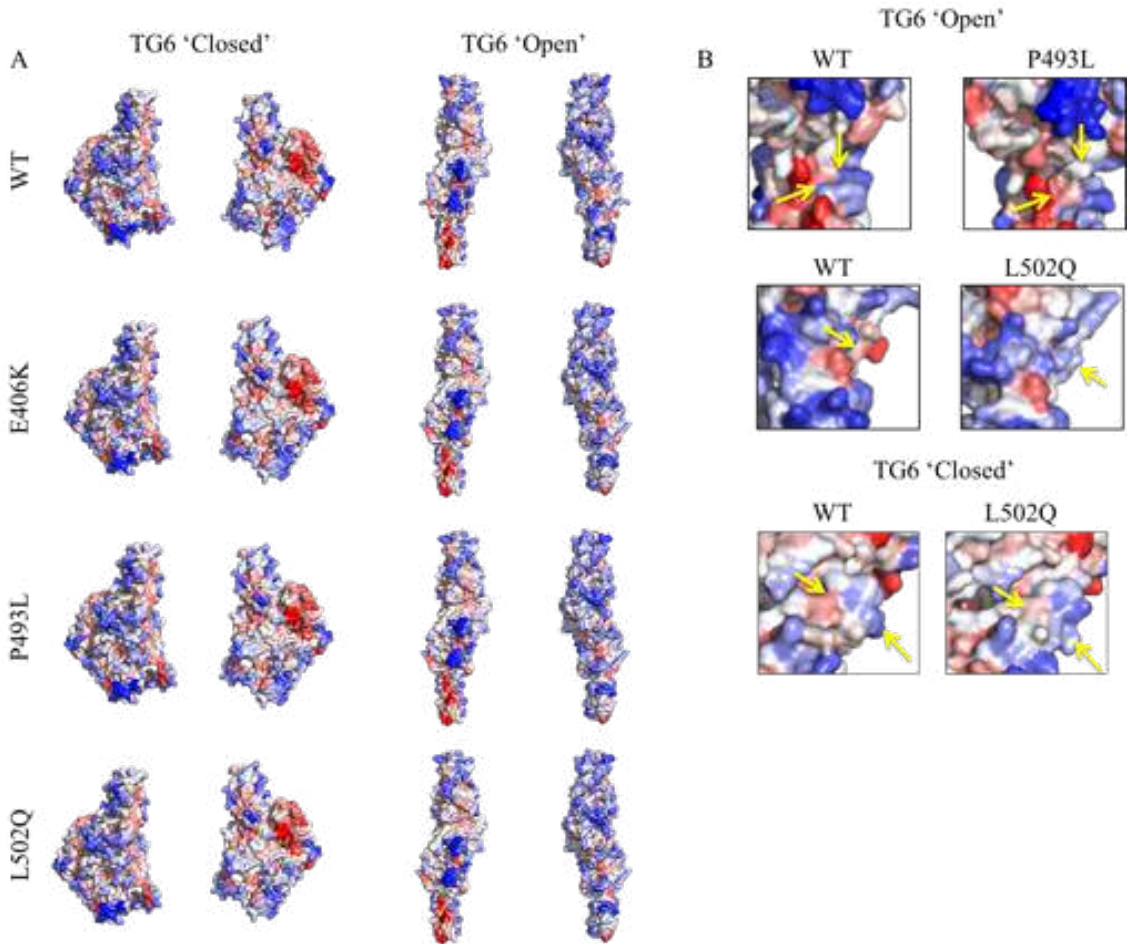


Figure 4.12: Solubility of TG6 probably benign mutants.
A3D analysis showing the areas of high solubility (blue) to low solubility (red) in both open and closed TG6 homology models (Fig. 4.1A and C). A: The structures generated by A3D are shown in surface representation. Left and right image pairs represent a 180° rotation. B: Changes in solubility were highlighted when the respective TG6 mutants were compared to WT (yellow arrows).

4.3.8 Assessing TG6 mutants with unknown but possibly damaging consequence.

Bioinformatic tools were unable to predict the consequence of four TG6 mutations (T426N, R448W, D510H and L502Q Q652dup). The TG6 mutants were analysed as before, firstly with PolyPhen-2 (0.979-1.000) and SIFT (0-0.03) (Table 4.15) which predicted the three missense TG6 mutations to be damaging and affect protein function.

Table 4.15: PolyPhen-2 and SIFT analysis of TG6 mutations of unidentified impact.

	T426N	R448W	D510H
Result	Probably damaging	Probably damaging	Probably damaging
Score	0.979	0.993	1.000
Result	Affects protein function	Affects protein function	Affects protein function
Score	0.00	0.02	0.01

FoldX analysis showed three mutants (T426N, R448W and D510H, Table 4.16) had stability similar to WT (ΔΔG = 0.6, 0.2 and 0.2 kcal/mol, respectively) in closed conformation. In open conformation, TG6 T426N (ΔΔG = 2.3 kcal/mol) appears to be the only unstable of the four TG6 mutants and the two remaining TG6 missense mutants (R448W and D510H) showed stability similar to WT (ΔΔG = -0.7 and 0.3, respectively). In contrast, TG6 L502Q Q652dup is predicted to highly destabilise closed conformation (ΔΔG = 89.0 kcal/mol) but has increased stability in open conformation (ΔΔG = -3.3 kcal/mol) when compared with WT TG6.

Table 4.16: FoldX analysis of benign TG6 mutations of unidentified impact.

Conformation	Stability	WT	T426N	R448W	L502Q Q652dup	D510H
Closed TG6	$\Delta G_{folding}$ (kcal/mol)	94.0	94.6	94.2	183.0	94.2
	ΔΔG (kcal/mol)	x	0.6	0.2	89.0	0.2
Open TG6	$\Delta G_{folding}$ (kcal/mol)	175.0	177.3	174.3	171.6	175.2
	ΔΔG (kcal/mol)	x	2.3	-0.7	-3.4	0.3

Stability (ΔG) was calculated for TG6 in 'closed' (TG6-Ca^{2+} bound (Fig. 4.1A)) and 'open' conformation (inhibitor bound- TG6 (Fig, 4.1C)) for WT and the TG6 mutants respectively. For details of analysis see Table 4.6.

When assessing solubility with AGGRESCAN, small changes in the NavSS score were observed for the four mutants, with TG6 R448W showing the greatest difference (0.3) when compared with WT (Table 4.17). In addition, for TG6 D510H is predicted to extend the hotspot (res. 509-520 instead of 510-520) to include H510. In the A3D analysis, the largest change for closed conformation was for TG6 D510H with a change in total score of 8.00 (Table 4.17), indicating a significant decrease in solubility. The two mutants (T426N and R448W) showed decreased solubility but the change in the total score was <0.5%. For the open conformation, TG6 R448W had the greatest and only significant change in total score of 14.54 when compared to WT TG6.

Changes in the solubility were also assessed using the A3D generated PDB files (Fig. 4.13). The A3D structures showed no change in the solubility in open and closed conformation for TG6 T426N. TG6 R448W showed a region of high solubility (blue) change to a region of low solubility (red) in the environment of R448 in open and closed conformation. No visible change was seen with TG6 D510H in closed conformation however, in open conformation there is a change in an area of high solubility (blue) to lower solubility (pale blue).

T426N, R488W L502Q Q652dup and D510H were grouped as TG6 mutations which have unknown but possibly damaging consequence impact. TG6 T426N is predicted to have decreased stability when compared with WT, which is consistent with the TG6 structural mutants. TG6 R448W is predicted to have increased aggregation propensity in open conformation when compared with WT TG6. This magnitude of change is seen in both structural and functional mutants. TG6 L502Q Q652dup showed a significant decrease in stability. This is more comparable to stability changes observed with the TG6 mutations predicted to affect structure. In addition, it indicates that the Q652dup mutation (predicted to affect structure) may have a greater influence on TG6 than the L502Q mutation (predicted to be benign) in the double mutant and as a result the TG6 L502Q Q652dup should be regrouped into the mutants predicted to affect structure. TG6 D510H had decreased solubility in closed conformation but its stability was unaffected. Due to the conflicting nature of the findings in the bioinformatic analysis it was difficult to place these TG6 mutants into a clear category at this point.

Table 4.17: Analysis of TG6 mutations of unidentified impact, using AGGRESCAN and A3D.

	AGGRESCAN					TG6 (closed, Fig. 4.1A) A3D					TG6 (o
TG6	**Number of Hot Spots**	**NavSS**	**ΔNavSS**	**Located within a Hotspot?**	**Result**	**Average score**	**Δ Average score**	**Total score**	**Δ Total score**	**% change**	**Averag score**
WT	23	-3.8	N/a	N/a	N/a	-0.817	N/a	-1149.11	N/a	N/a	-0.932
T426N	23	-3.9	-0.1	No	No change	-0.814	0.003	-1144.24	4.87	0.42	-0.935
R448W	23	-3.5	0.3	No	No change	-0.816	0.001	-1146.78	2.33	0.20	-0.912
D510H	23	-3.7	0.1	Yes	Extends hotspot	-0.812	0.005	-1141.11	8.00	0.70	-0.932

For analysis details see Table 4.7.

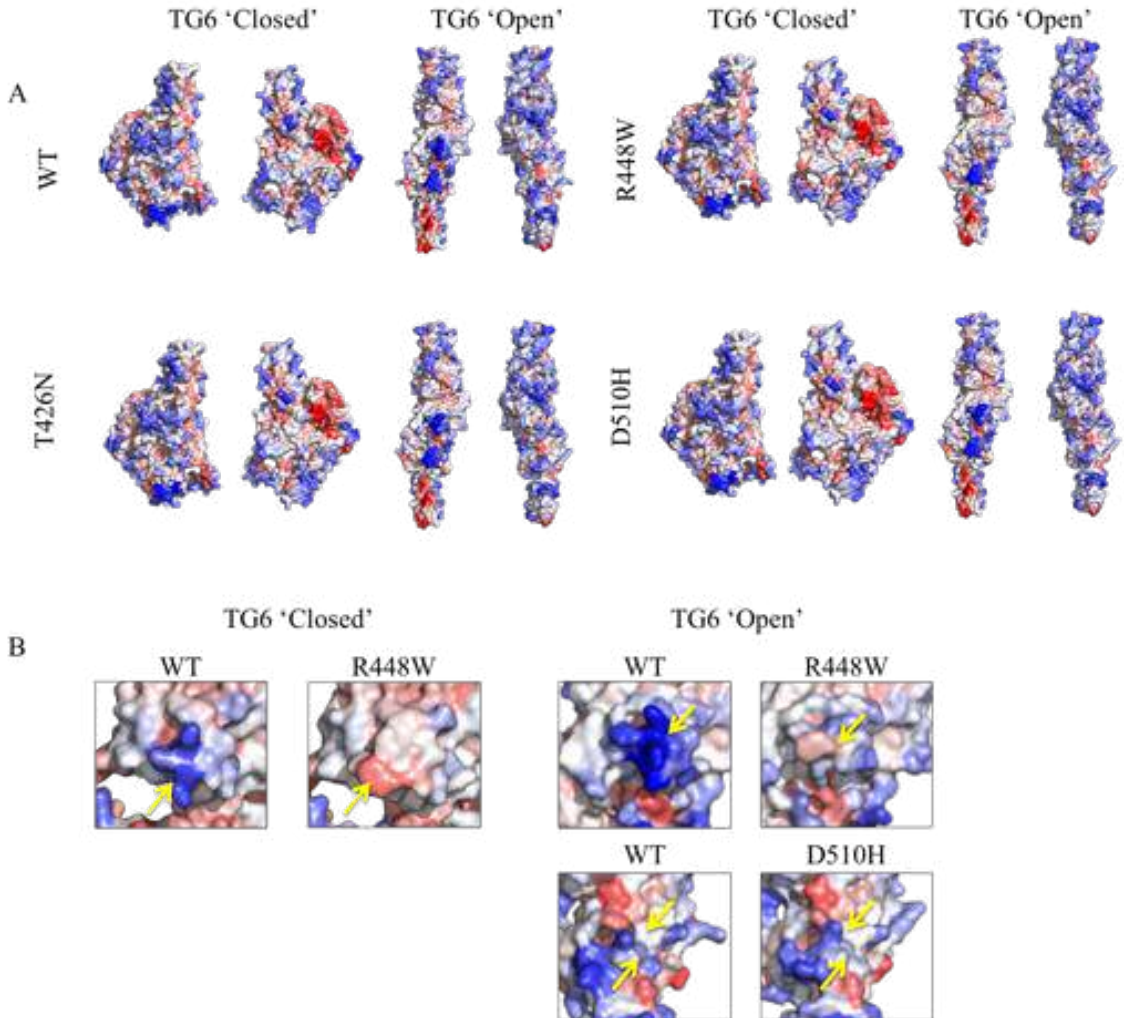


Figure 4.13: Solubility of TG6 'unknown but possibly damaging consequence' mutants.

A3D analysis showing the areas of high solubility (blue) to low solubility (red) in both open and closed TG6 homology models (Fig. 4.1A and C). A: The structures generated by AGGREGSCAN3D are shown in surface representation. Left and right image pairs represent a 180° rotation. B: Changes in solubility were highlighted when the respective TG6 mutants were compared to WT (yellow arrows).

4.3.9 Domain localisation of TG6 mutations

Using the previously generated homology models, the domain localisation of the TG6 mutations were investigated in open (Fig. 4.1C) and closed conformation (Fig. 4.1A). TG6 mutants were identified in all four of the TG6 domains, however they are predominantly located to the catalytic core domain and β barrel domain 1 (Fig. 4.14 and Table 4.18). This suggests mutations located in these domains are commonly pathogenic, likely because they affect sites involved in the enzymatic function and/or regulation of TG6. Mutations in the remaining two domains (β sandwich and β barrel 2) are either tolerated/ benign or fatal and therefore have not been identified in neurodegenerative disease. It is unlikely the mutations are fatal as TG6 knock-out mice are viable (Wellcome Trust Sanger Institute).

Table 4.18: Domain location of TG6 mutants.

Mutation	Location
R111C	β sandwich
N137S	β sandwich
Q181H	catalytic core
Gly279ter	catalytic core
D327G	catalytic core
P359L	catalytic core
V391M	catalytic core
E406K	catalytic core
R412C	catalytic core
T426N	catalytic core
Y441C	catalytic core
R448W	catalytic core
P493L	catalytic core
L502Q	catalytic core
G508D	β barrel 1
D510H	β barrel 1
L517W	β barrel 1
V527E	β barrel 1
E574del	β barrel 1
Q652dup	β barrel 2

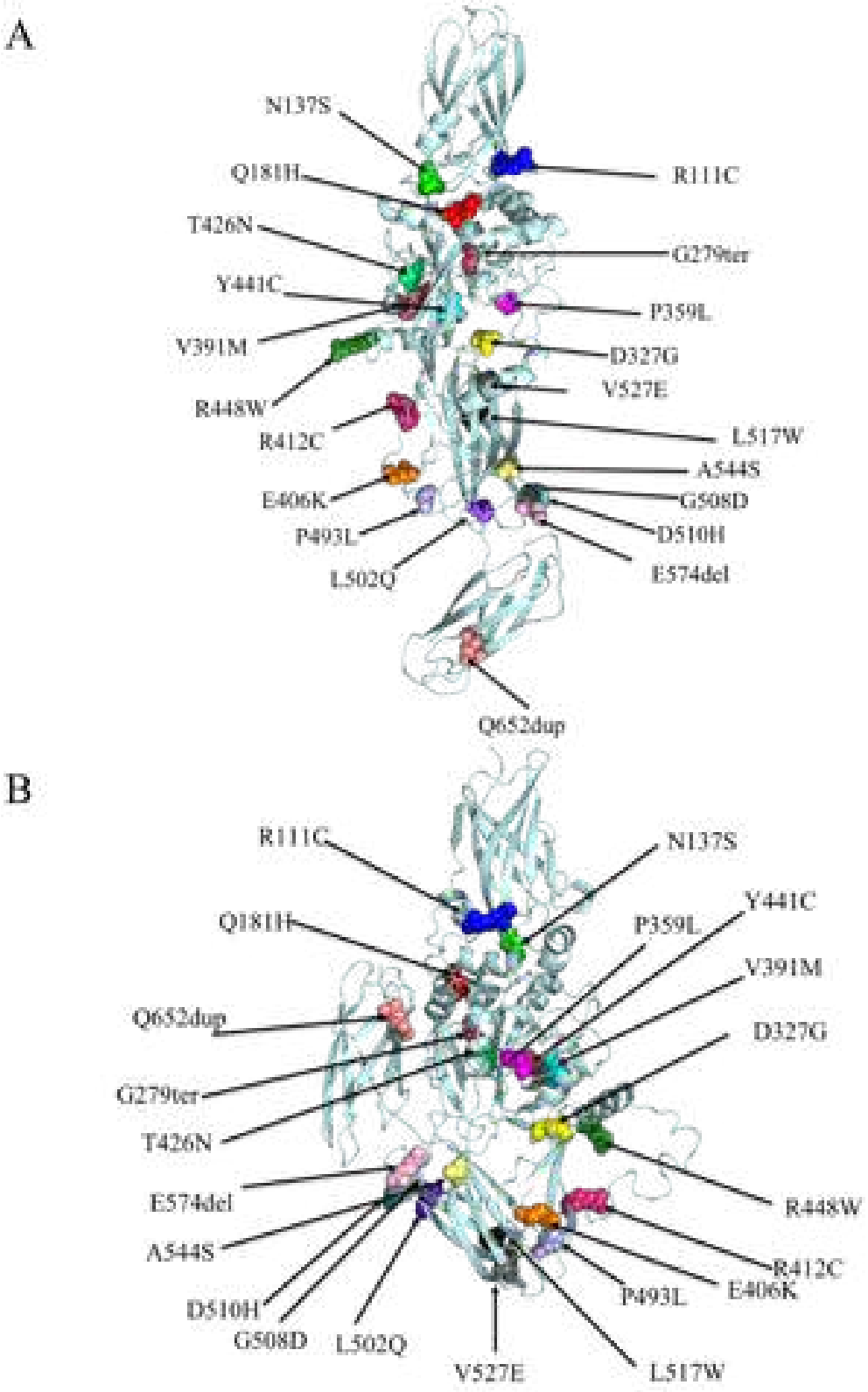


Figure 4.14: Location of mutations in TG6 model.
Mutated residues in TG6 are highlighted (atoms in coloured sphere representation) in both open (A) and closed (B) conformation. The protein backbone is shown in cartoon representation (pale blue). A: Open TG6 model (Fig. 4.1.C). B: Closed TG6 model (Fig. 4.1A)

4.3.10 Investigating the impact of the TG6 mutation on local structural environment

To assess the impact on local conformation, the local environment of the mutation in TG6 was analysed and compared to WT TG6, specifically investigating whether interactions between residues are lost or additional interactions established. The impact of the respective mutation was investigated using the open (Fig. 4.1C) and closed (Fig. 4.1A) TG6 homology models.

4.3.10.1 Structural TG6 mutants.

The TG6 mutants R111C, Q181H, P359L, Y441C, G508D, L517W, V527E, E574del, and Q652dup were assessed both, in Ca^{2+}-bound 'closed' conformation and open conformation (Fig. 4.15).

R111C: In closed conformation, WT TG6 has five contacts possible with the R111 side chain. Contacts are made via the carboxylate side chain on E29 (3 contacts) and on E131 to the amine on R111. Interactions are also made by the side chain carbonyl of N188, to the two amines on the R111 side chain. In addition, the backbone carbonyl of R111 interacts with the hydroxyl group of T60. When R111C is introduced all interactions made with the R111 side chain are lost but the backbone interaction made with T60 is maintained. These contacts take part in the joining of the core domain to the β sandwich. This data suggests that the interaction between the core domain and β sandwich may become unstable which leads to exposure of hydrophobic areas. In open conformation, R111 makes three contacts: two interactions between the carboxylate sidechain of E131 and the R111 side chain. Also, the backbone carboxylate of R111 interacts with the hydroxyl side chain of T60. When mutated to C111 it loses all three contacts with E131 but maintains backbone interaction with T60. The loss of three bonds in the same position could affect the stability in this area.

Q181H: Q181 is located in the core domain and is not positioned near any functional binding sites. In closed conformation, it can form interactions with the carbonyl backbone of V157 or with the hydroxyl side chain of S159 via the Q181 side chain carboxyamide. When mutated to H181 it loses all the interactions observed in WT and instead the imidazole side chain interacts with the backbone carbonyl group of L158. When in open conformation, TG6 Q181 sidechain carboxyl amide group can in addition to the already

described interactions interact with its own backbone amide group. In addition, Y25 interacts with the carbonyl backbone of Q181 via its side chain hydroxyl group. When Q181H is introduced V157 interacts with the imidazole side chain of H181 via its backbone carbonyl and the hydroxyl group of Y25 maintains its interaction with the backbone carbonyl of H181. The other potential side chain interactions are lost.

P539L: TG6 P359 does not interact with any of its surrounding residues. When P359L is introduced no apparent changes are observed. In open conformation, P359 interacts via its backbone carbonyl with the backbone amide of E361; when mutated to L359 this interaction is maintained and an additional possible interaction is made with N226 (backbone carbonyl of N226 to the backbone amide of L359). Both these interactions could act to stabilise open conformation as opposed to closed conformation.

Y441C: Y441 is located close to the active site and interacts with W335, which forms part of the residues relevant to catalysis (Fig. 1.9) (Thomas H., 2004). In closed conformation, Y441 interacts with the indole side chain of W335 via its side chain hydroxyl group. Y441 can also interact via its side chain hydroxyl with the side chain carboxylate of E337. The backbone amide of Y441 interacts with the backbone carbonyl group of T438 or I437. When mutated to C441 it loses the contacts with E337 and W335. The Y441C substitution could therefore impact the enzymatic activity of TG6 due to its association with the substrate binding site. The Y441 side chain hydroxyl group, in the open conformation, interacts with E337 and W335 as in the closed conformation. The backbone amide of Y441 also interacts with the carbonyl group of T438. When mutated to C441 the interactions with the side chain (E337 and W335) are lost. Therefore, the changes in the interactions are similar for both the open and closed conformation.

L517W: In both, open and closed TG conformation, TG6 L517 interacts with K561 via its backbone amide and carbonyl to the backbone amide and carbonyl of K561. When L517W was introduced, no change in these interactions was observed. However, L517 is located in the first β barrel domain and may impact local structure and TG regulation.

G508D: In WT, the backbone amide of G508 interacts with the hydroxyl side chain of Y572 in open and closed conformation. TG6 G508D showed no changes in the interacting residues in closed conformation when compared to WT TG6. In open conformation, this

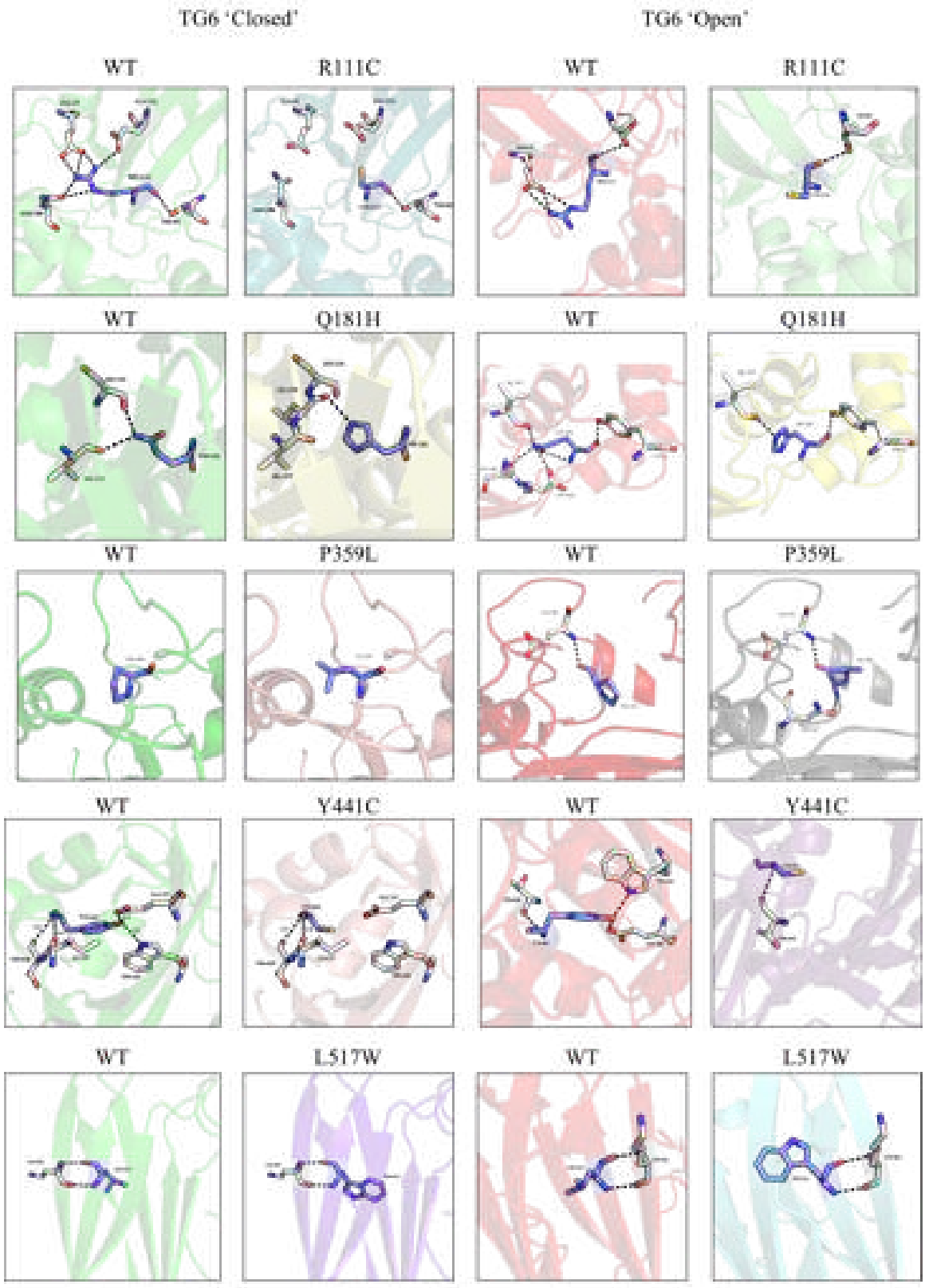
interaction is lost when D508 is introduced and therefore may impact the conformation of the flexible loop it is located within.

V527E: V527 interacts with V553 via its backbone carbonyl and amide to the backbone carbonyl and amide of V553 in open and closed conformation. When V527E is introduced the interactions are maintained in both closed and open conformation.

E574del: In closed conformation E574 does not interact with any of its neighbouring residues and is solvent exposed. When E574 is deleted no change in interactions are observed. However, in open conformation, E574 may make a polar contact with the amine side chain of K573, via its side chain carboxylate. When E574 is deleted, the majority of the local polar contacts are maintained. The polar contact between E574 and K573 is lost, as expected, but is not replaced in the mutant. As this mutation causes a deletion in β barrel 2, it may impact the interface between β barrel 1 and 2.

Q652dup: Q652 is located in a loop that forms part of β barrel 2. It makes a contact with the side chain carboxylate of E644, via its side chain amide when in closed conformation. When duplicated to Q652/Q653, no contacts with other residues are predicted. In open conformation, Q652 makes no contacts with neighbouring residues. When duplicated, the introduced Q653 contacts the backbone amide (via Q653 backbone carbonyl) and backbone carboxylate (backbone amide of Q653) of V643. Q653 also contacts the backbone carbonyl of L654 via its carboxylamide side chain. This could act to stabilise local TG6 conformation due to the gain of additional interactions upon mutation. This residue is in close proximity to the core domain so may be important in the regulation of TG6 activity but no direct interactions between the core domain and these residues within the β barrel were identified.

In conclusion, potentially relevant structural alteration could be identified for R111C, P359L, Y441C and G508D but not for the other mutants despite detailed investigation.


TG6 'Closed'
TG6 'Open'
WT
R111C
WT
R111C
WT
Q181H
WT
Q181H
WT
P359L
WT
P359L
WT
Y441C
WT
Y441C
WT
L517W
WT
L517W

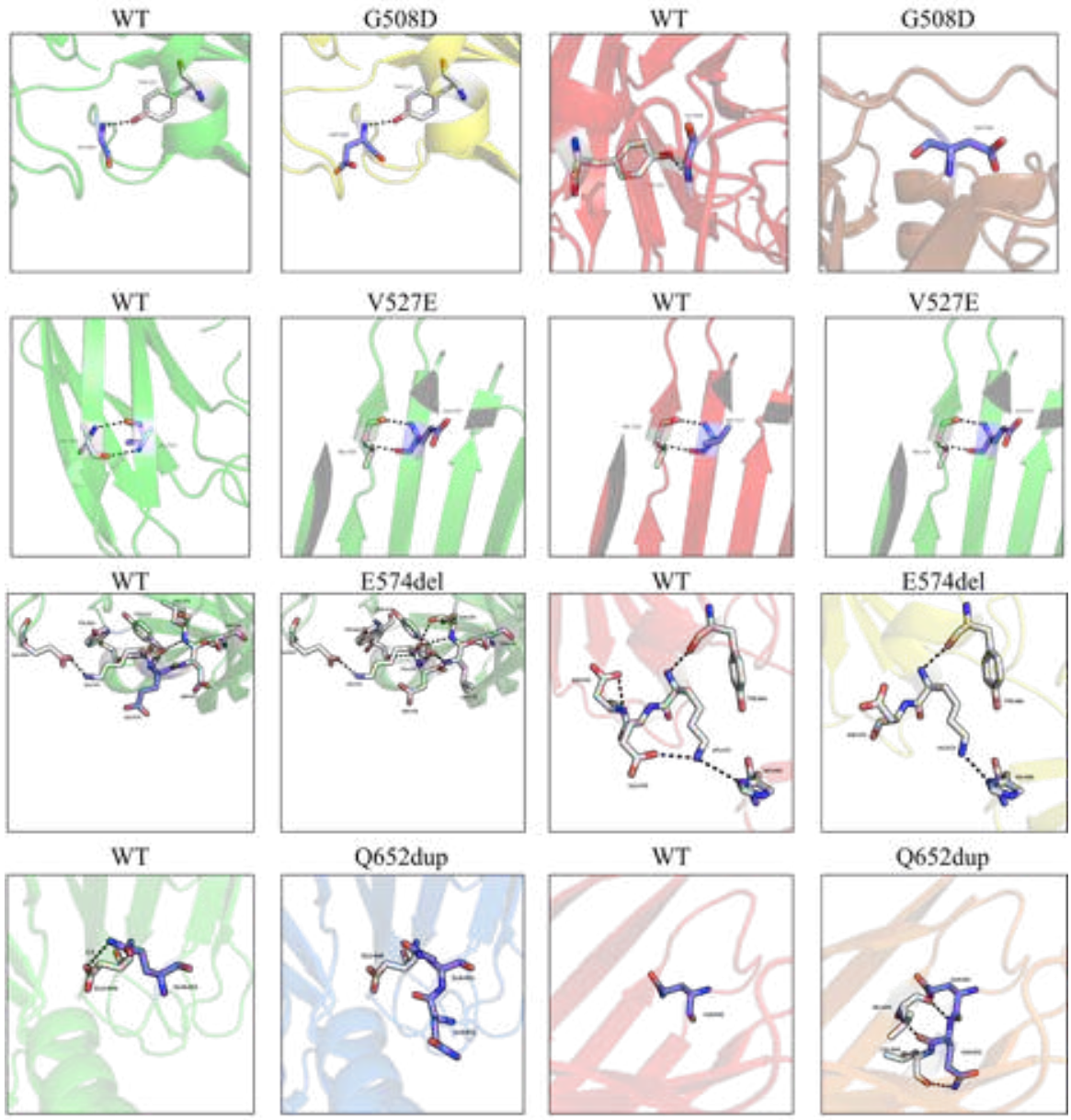


Figure 4.15: Impact of TG6 mutations on local structural environment.
Changes in structure in the vicinity of the mutation were investigated using open (right two columns) and closed (left two columns) TG6 homology models (see Fig. 4.1). Protein backbone is shown in cartoon representation. The mutated residue (carbon: lilac, nitrogen: blue and oxygen: red) and the interacting residues (carbon: white) are highlighted in stick representation. Polar contacts that may occur (≤3.5 Å) are highlighted (black dotted lines).

4.3.10.2 TG6 mutations impacting function

The local environment changes of the TG6 mutants N137S, D327G, V391M and R412C at the site of the mutations were assessed (Fig. 4.16).

N137S: N137 makes five potential interactions. This consists of two backbone interactions i.e. C140 backbone amide interacts with N137 protein backbone carbonyl

group and D143 side chain carbonyl group interacts with the backbone amide of N137. The remaining three interactions involve the N137 side chain: R34 (backbone carbonyl interacts with N137 carboxylamide), Q36 (side chain carboxylamide interacts with N137 carboxyamide) and W139 (backbone amide interacts with N137 carboxylamide group). When S137 is introduced all three side chain interactions are lost. These interactions are part of the interface between the β-sandwich and catalytic core domains, this mutation may destabilise the structure as similar changes were observed for both TG6 conformations.

D327G: In WT TG6 in Ca^{2+}-bound but closed conformation, D327 interacts via its carboxyl side chain with the amide backbone and hydroxyl side chain of T304. D327 also interacts with the hydroxyl side chain of T325 (via its amide backbone). Importantly, D327 forms part of Ca^{2+} binding site 3 and interacts with Ca^{2+} via its side chain carboxylate. In D327G, there is a loss of multiple contacts, i.e. there is no longer any contact made with Ca^{2+} in Ca^{2+} binding site 3 consistent with our previous results (Thomas *et al.* 2013). In addition, G327 loses the two contacts with T304 (side chain and backbone) but maintains its backbone interaction with T325. In open conformation D327 interacts via its backbone amide to the sidechain carboxyamide of Q360. When D327G is introduced, the interactions with Q360 is maintained. However, the Ca^{2+} binding site is lost in open conformation.

V391M: In closed conformation, V391 interacts with V396 backbone amide, via its backbone carbonyl group and interacts with the backbone carbonyl of D387 via its backbone amide, likely contributing to the respective core domain α helix. These interactions are maintained in V391M. When in open conformation, V391 interacts via its backbone carbonyl with the backbone amide of E394 and when mutated to M391 this contact is maintained. Additionally, an interaction via the backbone carbonyl group of M391 and the backbone amide of V359 is possible. This indicates that the mutation causes a minor change in side chain positioning, which in turn has altered the backbone, allowing the original interaction to be maintained for closed to open structure transition.

R412C: In closed conformation, R412 (via side chain amine) forms an electrostatic interaction between the core domain β sheet and the flexible extended loop (D484, carboxylate side chain). It also interacts via its backbone amide, with the carboxylate side chain of E411. When mutated to C412, the electrostatic interaction with the side chain is

no longer observed. Instead the backbone amine and carbonyl groups interact with L403 (via backbone carbonyl and amide) in the neighbouring β strand. In open conformation, R412 does not interact with any neighbouring residues and this is not altered when R412C is introduced. The caveat for these observations being that the extended loop linking the core domain and β barrels, does not have a defined structure.

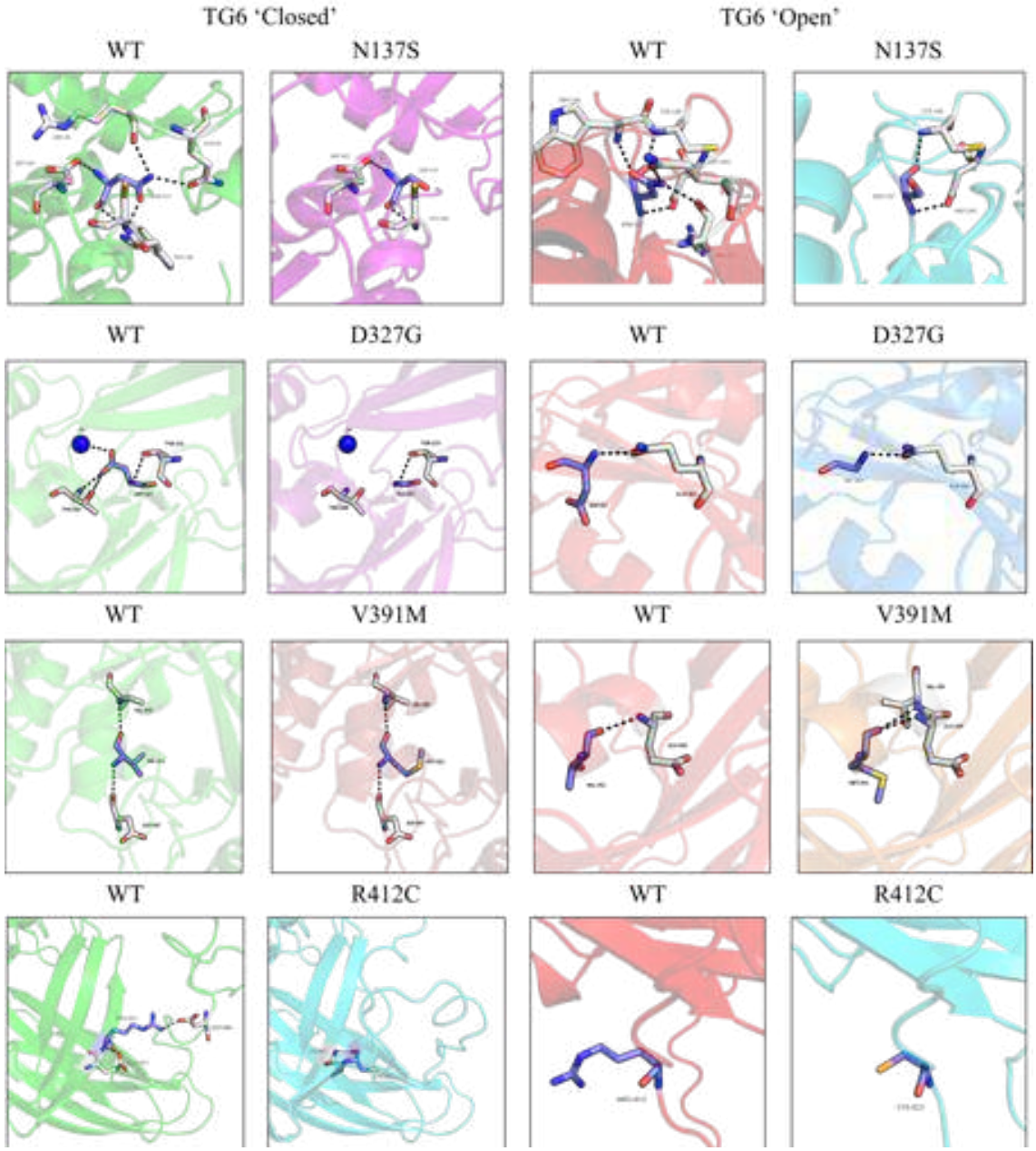


Figure 4.16 Impact of TG6 functional mutants on local structural environment. Changes in the environment of the TG6 mutant were investigated using open and closed TG6 homology models. TG6 proteins backbone is shown in cartoon representation (various colours). The mutated residue (lilac: carbon, nitrogen: blue and oxygen: red) and the interacting residues (carbon: white) were shown in stick representation. Polar contacts ≤3.5 Å are highlighted (black dotted lines).

4.3.10.3 'Benign' TG6 mutation

The local environment of the TG6 mutations predicted to be benign (TG6 E406K, P493L and L502Q) were assessed to investigate the impact on local structure (Fig. 4.17).

E406K: In closed conformation WT TG6, E406 side chain carboxylate interacts with its own backbone amide group. In E406K, the side chain amine group interacts with the D323 carboxylate group. This could enhance the stability of the closed conformations which is consistent with the higher stability predicted by FoldX. In open conformation E406 and K406 do not interact with other residues. Together, this suggests E406K could impair the transition from closed to open conformation due to enhanced stability in closed conformation.

P493L: P493 makes one interaction between the backbone carbonyl group of P493 and S494 hydroxyl group in closed conformation. When L493 is introduced, this interaction is lost and no other interactions are made. P493 makes no interactions with nearby residues, in open conformation, but when L493 is introduced, its backbone carbonyl group interacts with the backbone amide of E605. The latter interaction could extend the surface between β barrel 1 and β barrel 2 and may alter the positioning of these domains relative to others.

L502Q: L502 makes no polar contacts with any other residues. When mutated to Q502, two new contacts are made (to R512 (carbonyl backbone) and E503 (side chain carboxylate)) via its backbone amide and carboxylamide side chain, respectively. It is located in the loop domain between two β strands, which make up part of β barrel 1. The effect of this on the overall structure of the β barrel is not clear. However, this may impact on GTP binding as this part of the protein backbone forms the nucleotide binding pocket. In open conformation, TG6 L502 does not interact with other residues but when mutated to Q502 it forms an electrostatic interaction between its side chain carboxylamide and the side chain carboxylate of E503.

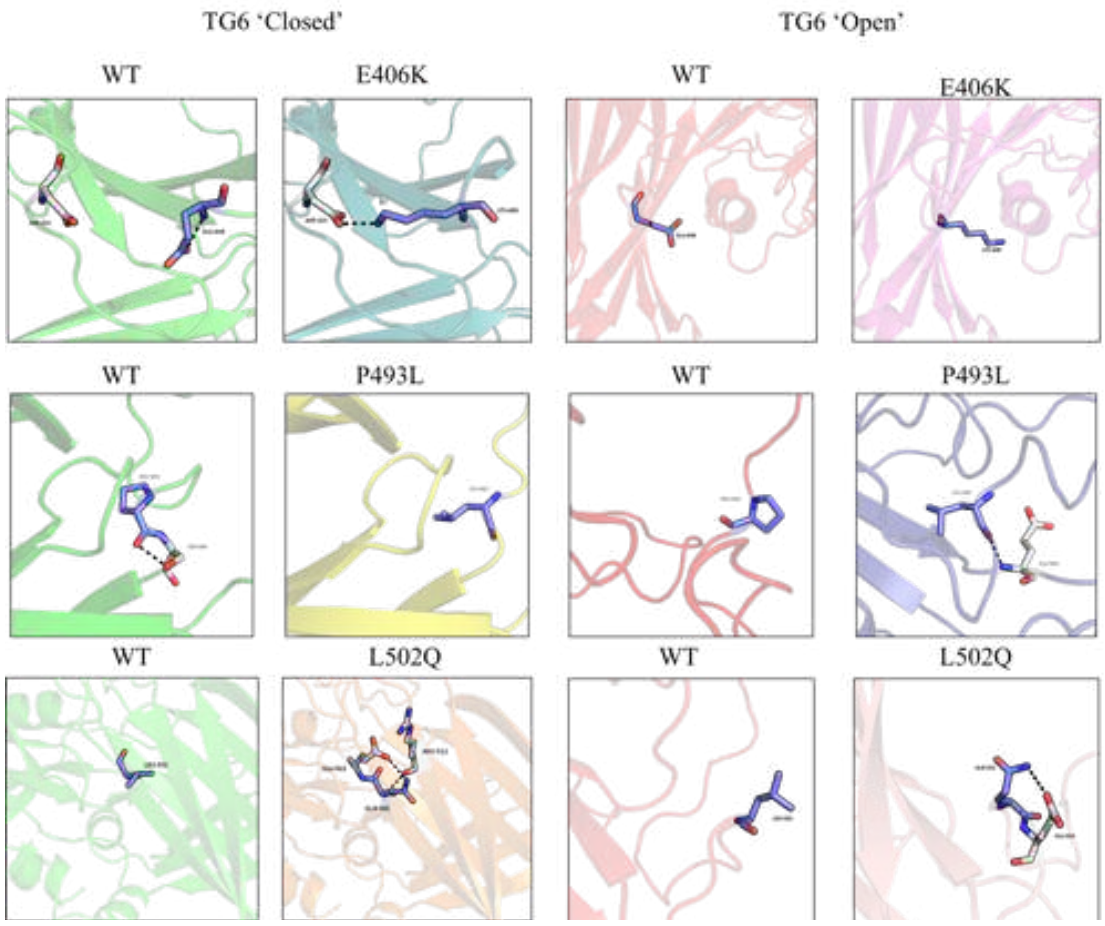


Figure 4.17: Impact of 'benign' TG6 mutations on local structural environment. Changes in the environment of the TG6 mutant were investigated at the site of the mutation using open and closed TG6 homology models. TG6 protein backbone is shown in cartoon representation (various colours). The mutated residue (carbon: lilac, nitrogen: blue and oxygen: red) and the interacting residues (carbon: white) are highlighted in stick representation. Polar contacts ≤3.5 Å are highlighted (black dotted lines).

4.3.10.4 Environment of TG6 mutants with unknown predicted impact

The environmental of mutations with unknown predicted impact (TG6 T426N, R448W and D510H) was assessed (Fig. 4.16).

T426N: T426 interacts via its side chain hydroxyl with the backbone carbonyl of V294 and the backbone amide of K427. The backbone amide of T426 interacts with the backbone carbonyl group of V435 and the backbone carbonyl group of T426 interacts with V435's backbone amide. When mutated to N426 the backbone interaction with V435 is maintained but the side chain interaction with V294 and K427 are lost. Instead the carboxylamide of N426 interacts with the side chain amine of A293. In open conformation, T426 interacts with V294 and K427, via its side chain, as in closed conformation. The same backbone interactions with V435 are also made, similar to closed conformation. When mutated to N426, the interaction facilitated by the side chain

electrostatic interaction with the histidine ring. This could interfere with local structure as a stronger interaction is introduced. This is not consistent with previous data whereby D510 interacts with S568 but when mutated to H510 it forms two possible interactions with T566 (Li *et al.* 2013). In open conformation, D510 makes no interactions with other nearby residues. H510 interacts with the backbone carbonyl of G508, via its backbone amide. G508, is another position for an identified TG6 mutation which could suggest the pathogenic method of D510H and G508D may be linked.

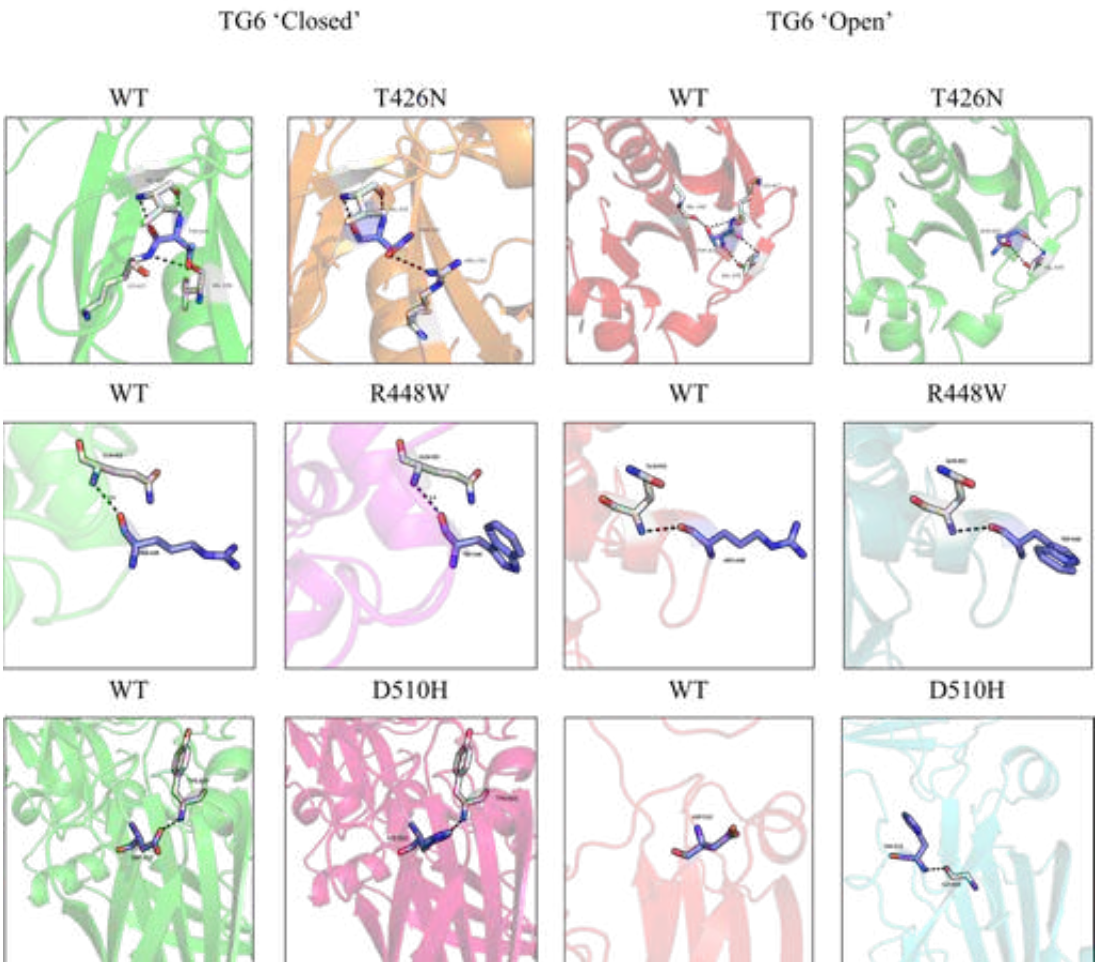


Figure 4.18: Impact of TG6 mutations of 'unknown' significance on local structural environment.
Changes in the environment of the TG6 mutant at the site of the mutation were investigated using open and Ca^{2+}-bound closed TG6 homology models. TG6 proteins backbone is shown in cartoon representation (various colours). The mutated residue (carbon: lilac, nitrogen: blue and oxygen: red) and the interacting residues (carbon: white) are highlighted in stick representation. Polar contacts that may be formed (≤3.5 Å) are indicated (black dotted lines).

4.3.11 TG6 mutant Ca^{2+} binding site analysis.

The impact of the TG6 mutations on Ca^{2+} binding was assessed using the Ca^{2+} bound TG6 models. In this case the SWISS model generated homology models (Fig. 4.1A and B) were used for Ca^{2+} binding site analysis, as the Prime homology model of TG6 was not available at the time. Not all six interactions expected to stabilise Ca^{2+} are seen in these SWISS models homology models but additional interactions of the respective side chains of site 2 and 3 are suggested by the Prime homology model or are H_2O mediated. All interactions in the Ca^{2+} binding sites for WT TG6 have already been discussed earlier in this chapter (Fig. 4.3).

G508D and V527E, indicative of more wide spread conformational adjustment within the core domain. TG6 mutations R111C, Q181H, L571W, E574del and Q652dup were not predicted to impact Ca^{2+} binding and thus TG6 activation. Of the TG6 mutations predicted to impact function, N137S, D327G and R412C could impact Ca^{2+} binding and enzyme activation. TG6 N137S and D327G lose an interaction with Ca^{2+} via E450 at site 2 whereas TG6 R412C loses an interaction with Ca^{2+} via T417. TG6 D327G also loses an additional interaction between D327 and Ca^{2+} at site 3. The majority of the other changes in these mutants result in shorter interactions with Ca^{2+}, which would enhance affinity for Ca^{2+}. V391M showed changes similar to E574del and therefore was not predicted to impact function.

Regarding mutations of unknown effect, TG6 T426N showed a loss of the interaction with E450 in Ca^{2+} binding site 2. Further changes predicted shorter interactions when compared with WT TG6 with this TG6 mutant. However, disrupted binding at site 2 could impact TG6 function. The remaining TG6 mutants (TG6 R448W and D510H) showed no substantial changes. L502Q Q652dup showed minimal changes but was comparable to Q652dup, suggesting a structural consequence of the double mutation.

Of the TG6 mutations predicted to be probably benign, TG6 E406K and L502Q showed similar interactions to WT TG6. However, TG6 P493L unexpectedly showed significant deviance compared with WT TG6, with changes comparable to TG6 G508D and V527E (structural) and to TG6 N137S (functional) and T426N (unknown).

Table 4.19: Ca^{2+} binding site interactions in the TG6 mutants.

				Structural									Functional				Unknown				Benign		
			WT	R111C	Q181H	P359L	Y441C	G508D	L517W	V527E	E574del	Q652dup	N137S	D327G	V391M	R412C	T426N	R448W	D510H	L502Q/ Q652dup	E406K	P493L	L502Q
	Residues	Position	Interaction distance (Å)																				
Ca^{2+} binding site 1	A223	O	2.7	2.7	2.8	2.3	2.7	2.7	2.7	2.7	2.7	2.7	2.7	2.7	2.7	2.7	2.7	2.7	2.7	2.7	2.7	2.7	2.7
	N226	O	2.7	2.7	2.6	2.6	2.6	2.5	2.6	2.5	2.6	2.7	2.5	2.6	2.6	2.5	2.5	2.6	2.6	2.7	2.6	2.5	2.7
	N226	OD1	3.1	3.1	3	2.3	3	2.8	3.1	2.8	3	3.1	2.8	3	3	2.8	2.8	3	3	3.1	3	2.8	3.1
	N228	O	2.8	2.9	2.6	2.3	2.6	2.4	2.7	2.4	2.5	2.8	2.4	2.6	2.6	2.4	2.4	2.6	2.6	2.9	2.6	2.4	2.8
	D230	OD1	2.7	2.6	2.6	2.4	2.6	2.5	2.6	2.5	2.6	2.6	2.5	2.6	2.6	2.5	2.5	2.6	2.6	2.7	2.6	2.5	2.7
	D230	OD2	2.8	2.8	2.8	2.7	2.8	2.8	2.8	2.9	2.8	2.8	2.8	2.8	2.8	2.9	2.8	2.8	2.8	2.8	2.8	2.9	2.8
Ca^{2+} binding site 2	N396	OD1	2.6	2.6	2.6	2.2	2.6	2.5	2.6	2.5	2.6	2.6	2.5	2.6	2.6	2.5	2.5	2.6	2.6	2.6	2.6	2.5	2.6
	T417	O	2.3	2.3	2.3	2.3	2.3	2.2	2.3	2.3	2.3	2.3	2.2	2.2	2.3	x	2.2	2.3	2.3	2.3	2.3	2.3	2.3
	E445	OE1	3	3.2	2.9	2.5	2.9	2.8	3	2.8	2.8	3.1	2.8	2.9	2.9	2.7	2.8	2.9	2.9	3.2	2.9	2.8	3
	E445	OE2	2.9	3	2.9	2.7	2.9	2.8	2.9	2.7	2.8	2.9	2.8	2.9	2.9	2.7	2.8	2.9	2.9	3	2.8	2.8	2.9
	E450	OE1	3.5	3.5	3.5	3	3.5	x	3.5	x	3.5	3.5	x	x	3.5	3.6	x	3.5	3.5	3.4	3.5	x	3.5
Ca^{2+} binding site 3	D303	OD2	2.9	2.9	3	2.4	3	2.9	3	2.9	2.9	2.9	2.9	2.7	2.9	3	2.9	2.8	2.9	2.9	3	2.9	2.9
	D305	OD1	x	x	x	3.5	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
	D307	OD1	2.6	2.6	2.5	2.4	2.5	2.8	2.5	2.7	2.5	2.6	2.8	2.8	2.5	2.6	2.8	3.1	2.5	2.6	2.5	2.6	2.6
	S309	O	2.4	2.4	2.4	2.2	2.4	2.1	2.4	2.2	2.4	2.4	2.1	2.2	2.4	2.3	2.1	2.3	2.4	2.6	2.3	2.2	2.4
	D327	OD1	x	3.4	x	x	x	x	x	x	x	3.5	x	x	x	x	x	x	x	3.4	x	x	3.5
	D327	OD2	3.2	3.5	3	2.6	3	2.8	3.2	2.9	2.9	3.4	2.8	x	3	2.8	2.8	3	3	3.6	3	2.9	3.4

Using the Ca^{2+} bound activated TG6 model (Fig. 4.1A) created with SWISS model, the three Ca^{2+} binding sites were analysed in the TG6 mutants. Changes of 0.1 Å (yellow), changes of ≥ 0.2 Å (orange), loss of interaction (red) and gain of interaction (green) compared with WT TG6 are highlighted.

The impact on Ca^{2+} binding was also assessed using Mg^{2+}/Ca^{2+} and GTPγS bound (inactive) TG6 homology model, whereby Ca^{2+} binding site 1 and 2 are occupied by Ca^{2+} and Ca^{2+} binding site 3 is occupied by Mg^{2+} (Fig 4.1B, discussed in Section 1.5.2). With regards to the structural TG6 mutants, the greatest change in Ca^{2+} interactions was observed in P359L when compared to WT TG6 (Table 4.20). TG6 P359L showed a gain of two interactions with Ca^{2+} via the carboxylate group of E445 at site 2. Moderate changes were observed with TG6 G508D, V527E, E574del and Q652dup. No change was observed with the TG6 mutants R111C and Y441C whereas a contact is lost at site 3 in both TG6 Q181H and L517W (S309 and N307, respectively) was observed when compared with WT TG6. In TG3 the change in the ionic radii from Ca^{2+} to $Mg^{2,}$ results in interchange from heptacoordinate stabilisation of Ca^{2+} changes to six shorter interactions when Mg^{2+} is present (Fig. 1.10D(i-ii)) (Ahvazi *et al.* 2004a). Specifically, a change in interaction with D324 (equivalent to D327 in TG6) causes a shift in a loop domain (K321-S325) which regulates substrate tunnel opening (discussed in more detail in Section 1.5.2).

TG6 N137S and V391M showed the greatest variation among the TG6 mutations predicted to impact function. Both mutants showed the gain of interaction with Ca^{2+} via the carboxylate group on E445 (1 interaction for N137S and 2 interactions for V391M). R412C showed the gain of two interactions in Ca^{2+} binding site 2 via the E445 carboxylate and loss of an interaction in Ca^{2+} binding site 3, via N307. D327G showed little change in Ca^{2+} binding site 3. WT shows no interaction of D327 with the Mg^{2+} present in this site and therefore replacement with G327 causes less disruption of interactions. This suggest N137S, V391M and R412C could impact the regulation of TG6, more specifically the substrate tunnel access.

L502Q Q652dup and T426N showed the greatest variation of the TG6 mutants with unknown impact. L502Q Q652dup and T426N showed a gain of two interactions (E445 in site 2 and D305 in site 3) and multiple changes in interaction length (> 0.2 Å). No change was observed for D510H and R448W. The likely benign TG6 mutant, E406K shows little variability in the Ca^{2+} binding sites when compared with WT TG6. However, TG6 P493L showed greater variability i.e. gain of interaction with carboxylate of E445 at site 2 and multiple changes of >0.2 Å. These changes were again reminiscent of TG6 N137S, T426N, G508D and V527E. In contrast, L502Q had identical interactions to WT.

Table 4.20: Ca^{2+} binding site interactions in the TG6 mutants.

				Structural								Functional			
			WT	R111C	Q181H	P359L	Y441C	G508D	L517W	V527E	E574del	Q652dup	N137S	D327G	V391M	R412C
	Residues	Position	Interaction distance (Å)													
Ca^{2+} binding site 1	A223	O	2.6	2.6	2.6	2.6	2.6	2.5	2.6	2.5	2.6	2.6	2.5	2.6	2.6	2.6
	N226	OD1	3.1	3.1	3.2	2.9	3.1	3.2	3.2	3.2	3.2	3.2	3.2	3.2	3.1	3.1
	N226	O	2.4	2.4	2.4	2.7	2.4	2.6	2.4	2.6	2.6	2.6	2.6	2.4	2.4	2.4
	N228	O	2.5	2.5	2.5	3.0	2.5	3.1	2.7	3.1	3.1	3.1	3.1	2.7	2.5	2.5
	D230	OD1	2.5	2.5	2.5	2.7	2.5	2.6	2.6	2.6	2.6	2.6	2.6	2.5	2.5	2.5
	D230	OD2	2.9	2.9	2.9	2.7	2.9	2.8	2.9	2.8	2.8	2.8	2.8	2.9	2.9	2.9
Ca^{2+} binding site 2	N396	OD1	2.6	2.6	2.6	2.4	2.6	2.4	2.6	2.4	2.5	2.5	2.4	2.6	2.6	2.8
	E445	OE1	x	x	x	3.5	x	x	x	x	x	x	3.6	x	2.7	2.7
	E445	OE2	x	x	x	3.6	x	x	x	x	x	x	x	x	2.7	2.7
	E450	OE2	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.1	3.0	3.0	3.1	3.0	3.2
Ca^{2+} binding site 3	D303	OD2	2.9	2.9	2.9	2.9	2.9	3.0	2.9	3.0	3.0	3.0	2.9	2.9	2.9	2.9
	D305	OD2	2.3	2.3	2.4	x	2.3	x	2.5	x	x	x	x	2.5	2.3	2.3
	D305	OD1	x	x	x	3.0	x	3.0	x	3.0	3.0	3.0	3.0	x	x	x
	N307	OD1	2.9	2.9	2.9	2.5	2.9	2.5	x	2.7	2.5	2.5	2.5	3.1	2.9	x
	S309	O	2.5	2.5	x	2.6	2.5	2.6	2.6	2.6	2.6	2.7	2.6	2.4	2.5	2.5

Using the Mg^{2+}/Ca^{2+} bound, GTPγS bound (inactive, Fig. 4.1B) TG6 model created with SWISS model, the th
in the TG6 mutants. Changes of 0.1 Å (yellow), changes of ≥ 0.2 Å (orange), loss of interaction (red) and g
with WT TG6 are highlighted.

e.g. multiple changes of > 0.2 Å, a gain of an addition interaction with K170 and a loss of interaction with K498 (via β-phosphate) and K600 (via α-phosphate), respectively. In addition, TG6 P359L had multiple changes of >0.2 Å and a loss of interaction with K498 (via β-phosphate). TG6 G508D and V527E lost another interaction between K600 and the α-phosphate of GTPγS. Minimal variation was observed for TG6 R111C, Q181H, Y441C and L517W, although the interaction of the guanine ring with P505 was absent. Hence, the pattern of interactions with nucleotide split mutants into two groups: i. similar to WT and ii. Similar to G508D.

The other TG6 mutants followed the same pattern with N137S, R412C, T426N, P493L, L502Q Q652dup being similar to G508D and D327G, L502Q, V391M, R448W, D510H and E406K being similar to WT and therefore showing no consistency with categorisation regarding the type mutations.

Table 4.21: GTPγS binding site interactions in the TG6 mutants.

					Structural									Functional				Unknown				Benign		
				WT	R111C	Q181H	P359L	Y441C	G508D	L517W	V527E	E574del	Q652dup	N137S	D327G	V391M	R412C	T426N	R448W	D510H	L502Q/ Q652dup	E406K	P493L	L502Q
Nucleotide Position		**Residue Position**		**Interaction distance (Å)**																				
Guanine	N2	P505	O	3.4	x	x	3.4	x	3.3	x	3.4	3.4	3.4	3.4	x	x	3.4	3.4	x	x	3.4	x	3.4	x
	N2	T603	O	3.3	3.3	3.3	3.1	3.3	3.1	3.3	3.1	3.4	3.1	3.1	3.3	3.3	3.1	3.1	3.3	3.3	3.1	3.3	3.1	3.3
	N1	T603	O	3.1	3.1	3.1	2.8	3.1	2.8	3.1	2.8	3.1	2.8	2.8	3.1	3.1	2.8	2.8	3.1	3.1	2.8	3.1	2.8	3.1
	O6	T603	N	3.4	3.5	3.4	3.2	3.4	3.4	3.5	3.2	2.7	3.1	3.2	3.4	3.4	3.2	3.2	3.4	3.4	3.1	3.4	3.1	3.4
α-phosphate	O1A	K500	NZ	2.7	2.7	2.7	3.3	2.7	3.3	2.8	3.3	3.2	3.3	3.3	2.8	2.7	3.3	3.3	2.7	2.7	3.3	2.7	3.3	2.7
	O2A	V501	N	3.3	3.3	3.3	3.1	3.3	3.1	3.3	3.2	3.1	3.1	3.1	3.3	3.3	3.1	3.1	3.3	3.3	3.1	3.3	3.1	3.3
	O2A	K600	NZ	3.5	3.5	3.5	3.5	3.5	x	x	x	3.1	x	x	3.5	3.5	x	x	3.5	3.5	x	3.5	x	3.5
β-phosphate	OB1	K498	NZ	2.8	2.8	2.8	x	2.8	x	2.9	x	x	3.4	x	2.9	2.8	x	x	2.8	2.9	3.4	2.8	x	2.8
γ-phosphate	OG3	K170	NZ (H1)	3.1	3.1	3.2	3.2	3.1	3.2	3.3	3.2	3.4	3.1	3.2	3.2	3.1	3.2	3.2	3.1	3.1	3.1	3.1	3.2	3.1
	OG2	K170	NZ (H2)	x	x	x	x	x	3.5	x	3.5	3.3	3.3	3.5	x	x	3.5	3.5	x	x	3.3	x	3.5	x

Using the Mg^{2+}/Ca^{2+} bound, GTPγS TG6 model (Fig. 4.1B) created with, the GTP binding site was analysed in the TG6 mutants. Changes of 0.1Å (yellow), changes of ≥ 0.2 Å (orange), loss of interaction (red) and gain of interaction (green) compared with WT TG6 are highlighted

The interaction with GTP was assessed using the GTP bound TG6 homology model (Fig 4.1E). The TG6 structural mutants: TG6 R111C, Q181H, Y441C, L517W showed little variation i.e. changes of 0.1 Å (Table 4.22). The greatest change in interactions was seen with Q652dup and E547del when compared with WT TG6. This consisted of a gain of three interactions with the β-phosphate via K170, R474 and K600 and a loss of one interaction with the ribose moiety (via Q502) and one interaction each with the α-phosphate (via H171) and the γ-phosphate (via K170) and numerous changes of >0.2 Å. TG6 P359L, G508D and V527E all showed similar changes which, included a loss of interaction with: the guanine moiety via H171, α-phosphate via K170 and the γ-phosphate via R474 (2 interactions).

For the TG6 mutants predicted to impact TG function, multiple changes in interaction were observed for TG6 D327G. This included the loss of two interactions with the guanine moiety via H171 side chain and T603 amide backbone, gained of five interactions: one with the α-phosphate (via K600), three with β-phosphate (via K170, R474 and K600) and one with γ-phosphate (via R474) and multiple changes of >0.2 Å. These changes may enhance GTP binding of this mutant. Conversely, TG6 N137S, V391M and R412C had little variation when compared with WT TG6.

Furthermore, T426N, R448W and D510H showed little variation when compared with WT TG6 i.e. changes of 0.1 Å. The probably benign TG6 mutant E406K also showed little variation whereas TG6 P493L showed changes in interactions comparable to TG6 G508D and V527E. For both TG6 L502Q (benign) and L502Q Q652dup (unknown), significant changes were observed, whereby GTP was unable to be docked when the mutations were introduced. Instead the GTP had to be manually docked for the measurements to be taken as L502 is located within the GTP binding site. The change in interactions were as follows: loss of three interactions with the guanine moiety via H171, and T603 (2 interactions), loss of interaction with α-phosphate via K170 and H171 and a loss of interaction with γ-phosphate via K170. In addition, a gain of interaction with α-phosphate via Q502 side chain.

From the results presented here it is apparent there are differences between the GTP and GTPγS binding, the use of a TG2 or TG3 template or the use of SWISS model or Prime homology to generate the model. This has been mentioned before in Section 4.3.1 (Fig. 4.4).

Due to the difference in the template and the method of homology model generation it is difficult to interpret the accuracy of the information regarding the nucleotide binding pocket and how the TG6 mutants may impact this. However, the Prime homology generated models have consistently generated more informative models with respect to the Ca^{2+} and GTP binding pockets as the interactions with residues have been optimised with regards to backbone positioning. Taken together the GTPγS docked TG6 homology model should be generated, using Prime homology, and analysed to provide more comparable results and identify differences in GTP and GTPγS binding. Furthermore, this should also be considered in the context of using both TG2 and TG3 as the template, to remove difference associated with template selection.

Table 4.22: GTP binding site interactions in the TG6 mutants.

					Structural									Functional				
				WT	R111C	Q181H	P359L	Y441C	G508D	L517W	V527E	E574del	Q652dup	N137S	D327G	V391M	R412C	T426N
Nucleotide Position		**Residue Position**		**Interaction distance (Å)**														
Guanine	N7	H171	NE2	2.6	2.7	2.7	x	2.7	x	2.7	x	2.7	2.7	2.7	x	2.6	2.7	2.7
	N3	Q502	NE2	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
	N2	Q502	O	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
	N2	L502	O	3.0	3.0	3.0	3.0	3.0	2.9	3.0	3.0	2.6	2.6	3.0	3.0	3.0	3.0	3.0
	N2	T603	O	2.6	2.6	2.6	2.5	2.6	2.5	2.6	2.5	2.6	2.6	2.6	2.7	2.6	2.6	2.6
	N1	T603	O	2.6	2.6	2.6	2.5	2.6	2.5	2.6	2.5	3.1	3.1	2.6	2.6	2.6	2.6	2.6
	O6	T603	N	3.1	3.1	3.1	3.4	3.1	3.0	3.1	3.0	2.7	2.7	3.1	3.2	3.1	3.1	3.1
Ribose	O4	Q502	N	x	x	x	x	x	x	x		x	x	x	x	x	x	x
	O4	L502	N	2.4	2.4	2.5	2.3	2.4	2.3	2.4	2.3	x	x	2.4	2.7	2.4	2.5	2.4
α-phosphate	O1A	K170	NZ	3.5	3.5	3.4	x	3.4	x	3.4	x	3.4	3.4	3.4	2.9	3.4	3.4	3.5
	O2A	H171	NE2	2.9	2.9	2.8	2.8	2.9	3.0	2.9	3.0	x	x	2.9	2.8	2.9	2.9	2.9
	O5	Q502	NE2	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
	O2A	K600	NZ	x	x	x	x	x	x	x	x	x	x	x	3.2	x	x	x
β-phosphate	O1B	K170	NZ	x	x	x	x	x	x	x	x	2.4	2.4	x	3.3	x	x	x
	O2B	R474	NH2	x	x	x	x	x	x	x	x	3.3	3.3	x	3.3	x	x	x
	O3B	R474	NH2	x	x	x	x	x	x	x	x	x	x	x	3.3	x	x	x
	O2B	R474	NE	3.1	3.1	3.0	3.2	3.0	3.2	3.0	3.2	2.9	2.9	3.1	2.8	3.1	3.0	3.1

	O2B	R475	N	3.3	3.3	3.3	3.3	3.3	3.3	3.3	3.3	3.0	3.0	3.3	3.2	3.3	3.3	3.3
	O1B	K600	NZ	x	x	x	x	x	x	x	x	2.8	2.8	x	3.1	x	x	x
γ-phosphate	O2G	K170	NZ	2.8	2.8	2.8	3.0	2.8	3.0	2.8	3.0	x	x	2.8	2.6	2.8	2.8	2.8
	O3G	W472	NE	2.7	2.7	2.7	2.7	2.7	2.7	2.7	2.7	3.0	3.0	2.7	2.6	2.7	2.7	2.7
	O3G	R474	NH1	3.3	3.3	3.2	x	3.2	x	3.3	x	3.2	3.2	3.2	2.9	3.2	3.2	3.3
	O1G	R474	NH1	3.0	3.1	3.0	x	3.0	x	3.0	x	3.0	3.0	3.0	2.7	3.0	3.0	3.1
	O1G	R474	NH2	3.1	3.1	3.0	3.1	3.0	3.1	3.0	3.1	3.0	3.0	3.1	2.8	3.0	3.0	3.1

Using the GTP bound TG6 mode (Fig. 4.1E) l created with Prime homology, the GTP^{+} binding site was analyse
(yellow), changes of ≥ 0.2 Å (orange), loss of interaction (red) and gain of interaction (green) compared with W

4.4 DISCUSSION

4.4.1 Bioinformatic tool selection

A novel bioinformatic analysis method was developed to access the impact of TG6 mutants on TG6 structure and function. SIFT and PolyPhen-2 predicted the functional impact of the TG6 mutants based on their polypeptide sequence (Adzhubei *et al.* 2010; Sim *et al.* 2012). PolyPhen-2 identifies structural characteristic such as metal binding sites and information on the likelihood the mutant is found in other homologs. SIFT uses multiple sequence alignment to identify if the position of the TG6 mutant is conserved in homologs. The majority of the TG6 mutants were classified as 'damaging' and predicted to 'affect protein structure'. These tools are a crude method of predicting if the mutants are pathogenic and could not provide any indication on how the TG6 mutation was pathogenic if predicted to be deleterious. Alone these tools are not sufficient to identify specific impact on protein structure and function, as it does not take into consideration the 3D structure. For example TG6 P493L was predicted to be 'benign' by PolyPhen-2 but the use of additional bioinformatic tools indicated that it may have structural and functional implications. Furthermore, TG6 L502Q was predicted to be 'tolerated' by SIFT but was found to be positioned in the nucleotide binding site and thus is likely to have functional consequences based on our homology modelling. Nonetheless these tools were valuable in the first instance to identify potentially deleterious changes but ultimately were not helpful for classification of TG6 mutations according to mechanistic impact.

FoldX proved to be a valuable tool in identifying the stability of the TG6 mutants by calculating the ΔΔG based on the homology models. The FoldX classification system could clearly identify those mutants that affected and to what degree they affected the structure. In addition, it identified those mutants that had stability similar or increased when compared with WT, which made folding problems less likely as a reason for pathogenicity.

AGGRESCAN and A3D predicted how the mutation would affect the solubility of TG6. AGGRESCAN information was based on aggregation prone hotspots identified in the polypeptide sequence but was deemed of limited reliability due to the lack of integration of information on the structure. A3D incorporates FoldX into its calculation of solubility. A change in solubility (Δ total score of >0.5%), was used to indicate whether the TG6

mutant was less stable. In some cases no change was observed in FoldX but a change was observed in the solubility and *vice versa*. However, A3D is still undergoing optimisation and further testing at present, whilst FoldX is fully optimised. As a result, changes in solubility alone could not be an indicator of structural changes but can aid in the classification of the TG6 mutants.

4.4.2 Validation of TG6 homology models

When assessing the validity and the quality of the TG6 models is was apparent they were not the best fit. However, given that the sequence identity with TG2 and TG3 for TG6 is 50-55%, differences in the structure are expected. To obtain the ideal parameters for a perfect fit, a higher homology between the template and target is required and ideally the TG6 structure would be utilised in this bioinformatic approach. As there is no crystal structure for TG6 predictions, were made using the various structures available.

Prime homology consistently showed greater variation (various parametres) in the validity and quality of the homology models when compared to the SWISS Model generated models. This is likely due the greater flexibility in the backbone which alters to better accommodate the allosteric regulators, resulting in deviation from the template. Despite the variation associated with Prime homology, these models inform on the residues involved in optimal binding of allosteric regulators. In contrast, the backbone is fixed in the SWISS Model generated models and only represent changes in the position of the residue side chains. As discussed in Section 4.3.1 and 4.3.12, the template and program used to generate the models need to be used for all structures (Section 4.4.4). As this cannot be completed using SWISS Model alone, further homology models should be generated using Prime homology. The generated homology models should be compared on alternative homology modelling programs to assess their reliability.

4.4.3 Homology modelling as a tool to assess TG6 mutants

A collection of new homology models were used to analyse the TG6 mutants. Homology modelling by far has proved the most informative method for analysing the impact of mutations. TG6 open and closed homology models were used to identify the structural implications in the environment of the mutation. This was valuable in identifying changes between TG6 mutants and WT TG6 and changes between the open and closed conformations. To predict the functional consequences of the TG6 mutants, the interactions in Ca^{2+} binding sites and nucleotide binding site were investigated. This used

SWISS model generated homology models based on the closed TG structure (Ca^{2+} bound and Mg^{2+}/Ca^{2+} and GTPγS bound). In addition, Prime homology generated closed TG6 (GTP-bound, PDB: 4PYG) was used. In this context, homology modelling was valuable in identifying TG6 mutants which have functional consequence, there was a clear difference between those mutations predicted to exclusively have structural consequences from those with functional impact.

4.4.4 Nucleotide binding and hydrolysis

The different interactions involved in the stabilisation of GTP and GTPγS was investigated. Distinct changes in the residues responsible for the stabilisation of the phosphates were observed (Fig 4.4). This gave rise to the hypothesis of that there are two distinct modes of binding and that the repositioning of residues with functional important e.g. hydrolysis of GTP, similar to TG2. However, these differences could have been observed due to the differing template used to generate the homology model (e.g. TG2 or TG3). To investigate this further Prime homology models can generate models with GTPγS in TG6 using a TG2 (PDB: 4PYG) and GTP in TG6 using a TG3 template (PDB: 1RLE), instead of being manually docked (as in Fig. 4.4). This would increase the accuracy of the predictions made with regards to the nucleotide binding of WT TG6 and TG6 mutants and allow further investigation into the two observed binding modes. Additionally, the nucleotide binding of TG6 depending on a TG2 or TG3 template needs to be investigated further. Molecular docking should be carried out to dock GTP into TG6 based on a TG3 template and GTPγS into TG6 based on a TG2 template. This in turn will increase the accuracy of the predictions on the impact of TG6 mutations and also identify differences in nucleotide binding without the structural characterisation of TG6 using X-ray crystallography.

Furthermore, it is still not known whether TG6 is capable of hydrolysing GTP. Additional models (generated with Prime homology) with different nucleotides e.g. GDP and GMP, could aid in the elucidation of this mechanism (Table 4.23). In these models the changes in local conformation and the interactions involved in nucleotide stabilisation, following step-wise hydrolysis of the phosphate groups can provide a model mechanism. In addition, these models could help to predict if the TG6 mutants can act to influence the predicted GTPase activity.

the existing knowledge of TG6 substrate selectivity (Table 4.23). Firstly, the difference in the substrate selectivity of WT TG6 can be assessed by priming the glutamine substrate pocket against a library of peptide substrates and using information from the literature (Stamnaes *et al.* 2010). Similar molecular docking experiments have been carried out for TG3 (Ahvazi and Steinert 2003). Furthermore, this method could be used to assess the optimal substrate of lysine binding pocket. This information in turn could be tested biochemically. With regards to disease, increased data could be used to assess the substrate selectivity of the TG6 mutants, to identify the various mechanism in which these mutants induce neuronal cell death.

Table 4.23: Additional TG6 models.

Method	Model	Ions	Template	TG	Reason
Homology modelling	TG6•GTP	N/a	4PYG	TG2	GTP binding
	TG6•GTPγS	2x Ca^{2+} 1x Mg^{2+}	1RLE	TG3	GTPγS binding and substrate tunnel
	TG6•GDP	3x Ca^{2+}	1VJJ	TG3	GDP binding and substrate tunnel
	TG6•GDP	2x Ca^{2+} 1x Mg^{2+}	1VJJ	TG3	GDP binding and substrate tunnel
	TG6•GMP	2x Ca^{2+} 1x Mg^{2+}	1SGX	TG3	GMP binding and substrate tunnel
	TG6 active	3x Ca^{2+}	1L9N	TG3	Activation
	TG6 inactive	1x Ca^{2+}	1L9M	TG3	Activation
	TG6 open	N/a	2Q3Z	TG2	Substrate binding
Molecular docking	TG6•GTP	N/a		TG3	GTP binding
	TG6•GTPγS	N/a		TG2	GTPγS binding
	TG6•GMP	N/a		TG2	GMP binding
	TG6 open	N/a	2Q3Z	TG2	Various substrate binding

Models made are highlighted in green.

TG6 N137S, D327G, V391M, T426N and L502Q were grouped in to 'functional' mutants. These mutants showed extensive changes in the interactions involved in the stabilisation of Ca^{2+} and nucleotides (GTP not GTPγS). In addition, these mutants showed no change/ little or increased stability (FoldX) and no change or decreased aggregation propensity (A3D). The environment of the mutation showing a loss or gain of interactions can indicate a functional mutant depending on its location. For example, L502Q may cause altered of GTP binding and D327G can impact Ca^{2+} binding.

In some cases, the TG6 mutants showed changes which indicated they had structural and functional implications. This included the TG6 mutants: P359L, R412C, P493L, L502Q Q652dup, G508D, L517W, V527E, E574del and Q652dup. These mutants showed decreased stability and decreased solubility (FoldX and A3D) and homology modelling of the local environment commonly shows a loss of interactions. In addition, significant impact is observed with regards to the interactions present Ca^{2+} and/or nucleotide binding sites. The combination of structural and functional predictions are likely to be primarily due to structural consequences which has a secondary impact on the function of TG6. However, additional work is required to identify the specific underlying mechanisms for this group.

One TG6 mutant was predicted to be benign (TG6 E406K) as no predictions were made in favour of a structural or functional consequence. Additionally, despite the number of bioinformatic tools used and homology modelling approaches the predicted impact of TG6 D510H was unable to be predicted. As a result it remains categorised as probably damaging but of unknown mechanism. As only one TG6 mutant was not able to be grouped into predicted impact, this shows the combination of bioinformatic analysis and

homology modelling was able to categorise TG6 mutants and prime these findings against subsequent bioinformatic analysis.

4.4.9 Conclusion

In order to validate the findings of the bioinformatic and homology method used, the impact of the TG6 mutations on structure and function need to be assessed biochemically. Following biochemical testing, the bioinformatic method can be further improved.

Table 4.24: Summary of Bioinformatic and homology modelling results.

	Bioinformatics										Homology modelling						
TG6	**Pathogenicity**			**Stability**		**Solubility**					**Local Environment**		**Ca^{2+} binding**				
				FoldX			A3D score		A3D PDB				Fig. 4.1A			Fig 4.1	
	Conserved?	PolyPhen-2	SIFT	Closed	Open	AGGRESCAN	Closed	Open	Closed	Open	Closed	Open	Site 1	Site 2	Site 3	Site 1	
R111C	N	Y	Y	-	-	-	-	=	-	-	-	-	=	=	+	=	=
N137S	Y	Y	Y	=	+	-	=	=	+	+	-	-	+	-	-	+	+
Q181H	N	Y	Y	-	+	=	=	=	=	=	-	-	+	=	=	=	=
D327G	Y	Y	Y	+	=	-	=	=	=	-	-	=	+	-	-	-	=
P359L	Y	Y	Y	-	-	-	-	=	=	-	=	=	+	+	+	-	+
V391M	Y	Y	Y	+	=	+	=	=	=	=	=	+	+	=	=	=	+
E406K	N	N	N	+	=	-	=	=	=	=	=	=	+	=	=	=	-
R412C	N	Y	Y	+	=	-	-	-	-	-	-	=	+	-	=	=	+
T426N	Y	Y	Y	=	-	+	=	=	=	=	=	-	+	-	+	-	+
Y441C	Y	Y	Y	-	-	+	-	=	=	=	-	-	=	=	=	=	=
R448W	N	Y	Y	=	=	+	=	-	-	-	=	=	=	=	-	=	=
P493L	N	Y	N	=	=	-	-	=	=	+	-	+	+	-	+	-	+
L502Q	N	Y	N	-	=	+	=	+	+	+	+	+	=	=	+	=	=
L502Q/ Q652dup	N/Y	N/a	N/a	+	-	N/a	N/a	N/a	N/a	N/a	N/a	N/a	=	=	+	-	+

G508D	Y	Y	Y	-	-	+	=	=	+	=	=	-	+	-	+	-	=
D510H	Y	Y	Y	=	=	-	=	+	=	-	=	+	=	=	=	=	=
L517W	N	Y	Y	-	-	=	-	=	=	=	=	=	=	=	=	-	=
V527E	N	Y	Y	-	-	+	=	=	=	=	=	=	+	-	+	-	=
E574del	N	N/a	N/a	-	+	-	N/a	N/a	N/a	N/a	=	-	=	-	=	-	=
Q652dup	Y	N/a	N/a	-	-	=	N/a	N/a	N/a	N/a	-	+	=	=	+	-	+

Highlighted are the results supporting either a structural (yellow) or functional (red) mechanism. Pathogenicity a
(Y) or no (N) and N/a if the tool analysis missense mutations only. The predicted differences are denoted as: +: in
compared with WT, =: equal to WT. The grouping is highlighted as follows: structural (S, yellow), functional
orange), benign (B, green) and unknown (Unk, blue).

CHAPTER 5

5. GENERATION OF TG6 MUTANTS AND BIOCHEMICAL ANALYSIS OF TG6 MUTANTS

5.1 INTRODUCTION

Following the bioinformatics and molecular modelling analysis, the generated predictions were assessed. First, the TG6 mutant baculovirus was produced and was used to generate TG6 mutant protein. Two mutants from each of the categories (structural, functional and structural and functional, established in Chapter 4, was selected for functional analysis using the isopeptidase and MDC assay (described in Chapter 2). However, TG6 D510H (predicted with unknown consequence) and E406K (predicted as benign) were not generated in time for analysis. In addition, the impact of the mutation on the binding of GTP and Ca^{2+} was investigated. The selected TG6 mutants are listed in Table 5.1.

Table 5.1: TG6 mutant protein analysis.

TG6 Mutant	Category	Analysis
R111C	Structural	Biochemical and cellular
Q181H	Structural	Biochemical and cellular
D327G	Functional	Biochemical and cellular
V391M	Functional	Cellular
E406K	Benign	Cellular
Y441C	Structural	Cellular
R448W	Structural	Cellular
L502Q	Functional	Biochemical and cellular
L502Q Q652dup	Structural and functional	Biochemical and cellular
D510H	Unknown	Cellular
L717W	Structural	Biochemical and cellular
E574del	Structural and functional	Biochemical and cellular
Q652dup	Structural and functional	Cellular

There is existing biochemical data for the TG6 mutants: R111C, D327G, L517W and E574del, (outlined in Section 4.1). In short, the transamidation activity was significantly reduced in HEK293 cells transfected with TG6 R111C, D327G, L517W and E574del and all had a significantly reduced half-life when compared with WT TG6 (Guan *et al.* 2013; Guo *et al.* 2014). In addition, TG6 D327G and L517W expressing cells (NIH3T3 and HEK293 cells) had increased sensitivity to apoptosis (Guan *et al.* 2013).

Results obtained for TG6 R111C, D327G, L517W and E574del allowed direct comparison to the literature. It also allowed those mutants with no previous biochemical data (Q181H, L502Q and L502Q Q652dup) to be compared to their category counterparts. This could assess whether bioinformatic predictions are sufficient to

categorise the TG6 mutants which helps establish a mechanism for neuronal cell death observed in SCA patients. The molecular mechanism of the TG6 mutants (Table 5.1) were also investigated in a cellular context. Together, the biochemical and cellular data could highlight potential pathogenic mechanisms and aid the understanding of the role in TG6 in neuronal cells. With this in mind the aims for this chapter have been formulated.

Aims of this chapter:

1. Recombinantly express TG6 mutants using the baculovirus *Sf9* cell expression system.
    a. Generate pFastBac plasmids, containing the mutated TG6 cDNA for the generations of the recombinant bacmid.
    b. Produce of baculovirus harbouring mutated TG6 for expression in *Sf9* cells.
    c. Express and purify TG6 mutant proteins.
2. Investigate the biochemical properties of the TG6 mutants using recombinantly expressed TG6.
    a. Assess the function of the TG6 mutants using the isopeptidase and MDC assay.
    b. Investigate the regulation of the TG6 mutants by measuring the binding of Ca^{2+} and GTP.
3. Evaluate the effect of the TG6 mutations in a mammalian cells.

5.2 METHODS

5.2.1 Cloning of TG6 mutants in to pFastBac

To generate recombinant TG6 mutants, TG6 baculovirus must be produced. This involved inserting the DNA of the mutants into the pFastBac plasmid. pFastBac with WT TG6 was donated by Zedira GmbH and pcDNA3 plasmids harbouring mutant TG6 (R111C, Q181H, V391M, Y441C, R448W, L502Q, L502Q Q652dup, D510H, and E574del) were gifted by M. Basso (Centre for Integrative Biology, University of Trento,). The plasmids pFastBac D327G and L517W were previously generated in our lab (A.Heil, Zedira GmbH). 10 µL of pcDNA3 and 1 µL of pFastBac plasmids were incubated with 20 µL DH5α Max efficiency competent cells (Invitrogen, 21034013) on ice for 30 minutes. The cells were heat shocked at 42 °C for 45 seconds and incubated on ice for a further 5 minutes. 200 µL of pre-warmed (37 °C) S.O.C media (Invitrogen, 15544034) was added to each tube and incubated at 37 °C for 1.5 hours. The cells were then grown on carbenicillin (100 µg/mL, Fisher Bioreagents, BP2648-5) agar plate at 37 °C overnight. 5 mL of LB media (with 100 µg/mL carbenicillin) was inoculated with positive bacterial colony and incubated overnight, shaking at 37 °C. Plasmids were purified from *E. coli* using Plasmid Miniprep kit (Qiagen, 27104) and DNA concentration of was measured using a Nanovue™4282 (GE healthcare).

5.2.2 Subcloning of TG6 mutants

Subcloning approaches were used to take the mutated DNA from the pcDNA3 TG6 mutant plasmid and clone them into pFastBac. Different restriction enzyme combinations were selected based on the mutation required (Table 5.2). The plasmids pcDNA3 and pFastBac plasmids were digested using respective restriction enzymes overnight at 37 °C (1 µg DNA, 1 µL restriction enzyme (NEB or Promega), 2 µL of restriction enzyme buffer (containing bovine serum albumin (BSA, 10 µg/µL), NEB and Promega), H_2O added to 20 µL). The digested DNA was analysed using a 1% agarose gel electrophoresis. The bands containing the respective mutation and plasmid backbone were excised from the gel (Table 5.2). The DNA was gel purified using a gel extraction kit (Qiagen, 28704) and the pFastBac plasmid backbone was treated with calf intestine alkaline phosphatase (Promega, M1821), following manufactures protocol. The DNA was purified again using the gel extraction kit (Qiagen, 28704). The purified DNA was analysed using a 1% agarose gel electrophoresis before ligation.

The reactions for the ligation of the insert into the plasmid backbone ((1:3, backbone: insert) 1 µL T4 DNA Ligase (Promega, M1801), 2 µL ligase buffer and H_2O up to 20µL) were incubated at 20 °C overnight. The resulting DNA was transformed into DH5α Max efficiency competent cells as above (Invitrogen, 21034013). Plasmids were purified as before and sequenced (Eurofins).

Table 5.2: Restriction digest of pcDNA3 TG6 mutants.

TG6 mutant	Restriction enzyme(s)	Band required (bp)
R111C	*XmaI* (NEB, R0180) or *Bsu36I* (NEB, 0524)	1380 or 1132
Q181H		
V391M		
Y441C		
R448W		
D510H		
L502Q	*PvuII* (Promega, R6331 and *BstEII* (Promega, R6641)	1448
L502Q Q652dup		
E547del		

TG6 R412C and Q652dup were generated using site directed mutagenesis primers (Table 5.3, Eurofins) and pFastBac WT TG6 was used as the template.

Table 5.3: TG6 site directed mutagenesis primers.

TG6 mutant	Primer (5'-3')
R412C	F: CCGGGAGTGTGTATACTCAAACACGAAGAAG
	R: GAGTATACACACTCCCGGCTCTCATCCTCGTG
Q652dup	F: CCAGGAACAGCAGCTCAGCATCGACGTG
	R: CGTGAGCTGCTGTTCCTGGAGAAGACCGCTG

The forward (F) and reverse (R) primers used to introduce site-specific mutations.

Gradient PCR (Table 5.4) established the best annealing temperature for each set of primers. Reactions were set up accordingly: 10 µL 5x GC Phusion buffer (ThermoFisher, F530S), 1 µL Phusion (ThermoFisher, F530S), 0.3 µg of template DNA (pFastBac WT TG6) and 2 µM of forward and reverse primers (Table 5.3). The amplified DNA was analysed using DNA electrophoresis (1% agarose). The PCR reactions were digested at 37 °C, overnight with *DpnI* (Promega, R6231) to remove the template DNA. The DNA was PCR purified using the gel extraction kit (Qiagen, 28704). *E. coli* transformations were carried out next, as before (Section 5.2.1), using 1µL of purified plasmid and 25 µL of DH5αMax efficiency competent cells (Invitrogen, 21034013). Overnight cultures were set up for each of the clones and the purified plasmid DNA was sequenced by Eurofins.

Table 5.4: Site-directed mutagenesis PCR conditions.

PCR	Time (s)	Temperature (°C)	Cycles
Initial denaturing	120	94	1x
Denaturing	60	94	20x
Annealing	45	60-72	
Extension	420	72	
Final extension	600	72	1x
Store	∞	4	-

5.2.3 Recombinant TG6 mutant bacmid preparation for Baculovirus generation.

pFastBac containing WT TG6 or a TG6 mutant (1 µL) was added 100 µL of Max Efficiency® DH10Bac cells (Invitrogen, 10361012) and transformed as described above (Section 5.2.1). 1 mL of prewarmed S.O.C medium (20 °C, Invitrogen, 15544034) was added to the cells and the mixture was incubated at 37 °C for >3 hours. Cells (100-400 µL) were spread onto agar plates containing: 50 µg/mL kanamycin (Fisher Bioreagents, BP906), 10 µg/mL tetracycline (Fisher Bioreagents, BP912), 7 µg/mL gentamycin (Fisher Bioreagents, BP918), 40 µg/mL IPTG (Fisher Bioreagents, BP1755) and 100 µg/mL X-Gal (Promega, V3941) and the plates were incubated at 37 °C for 48 hours. Following blue/white colony selection, white colonies were picked for overnight growth in LB medium (containing: 50 µg/mL kanamycin, 10 µg/mL tetracycline, and 7 µg/mL gentamycin at 37 °C). The bacmid was extracted using Purelink™ Hipure plasmid miniprep kit (Invitrogen, 210002) as per manufacturer's instructions.

The successful integration of pFastBac TG6 into the bacmid was investigated with PCR. (100 ng Bacmid, 4 µL dNTPs (10 mM, dATP, 10 mM dCTP, 10 mM dGTP and 10 mM dTTP, Promega) , 10 µL Phusion HF buffer (ThermoFisher, F530S), 0.5 µL M13 forward primer (5'-GTTTTCCCAGTCACGAC-3') (Eurofins), 0.5µL M13 reverse primer (5'-CAGGAAACAGCTATGAC-3') (Eurofins), 1 µL Phusion polymerase (ThermoFisher, F530S), up to 50 µL dH_2O). The PCR conditions followed those suggested by the Bac to Bac Baculovirus expression system kit (Invitrogen, 10359016) (Table 5.5). 10 µL was removed from the reaction and analysed by agarose gel electrophoresis (1%). The bacmid alone is 300 bp whereas the bacmid containing TG6 is ~5000 bp long.

Table 5.5: Bacmid PCR conditions.

PCR	Time (s)	Temperature (°C)	Cycles
Initial denaturing	180	93	1x
Denaturing	45	95	30x
Annealing	45	55	
Extension	300	72	
Final extension	420	72	1x
Store	∞	4	-

5.2.4 TG6 mutant Baculovirus generation (P1, P2 and P3).

Sf9 cells were seeded at $5x10^5$ cells/mL (>70% viability) in Sf-900™ II SFM media (Gibco, 10902088) without antibiotic for 1 hour at 27 °C in a 6 well plate (Starstedt, 83.3920.0005). 100 µL Sf-900™ II SFM media (Gibco, 10902088) with 6 µL rBacmid and 100 µL media with 6 uL TransIT®-LT1 transfection reagent (Mirus, MIR 2304) were mixed together and incubated (45 minutes at 20 °C). An additional 800 µL of Sf-900™ II SFM media (Gibco, 10902088) was added to the transfection reaction, the media was removed from the cells and the 1 mL mixture was added to each well containing the seeded cells (P0 virus). The plate was incubated at 27 °C for 4 days. The supernatant was retained (P1 Virus) and centrifuged (5 minutes, 500 xg,). For the P2-virus, *Sf9* cells were seeded at a density of $1.5x10^6$ cells/mL and infected with P1 virus. The culture was incubated for 5 days at 27 °C and 90 rpm. Cells were counted, and viability checked throughout. The remaining culture was centrifuged (5 minutes, 500 xg) and the supernatant was collected. For the P3 virus, the steps for the P2 virus were repeated but the culture was expanded.

5.2.5 Determining success TG6 expression in *Sf9* cells using MDC assay

Following each passage of the baculovirus, the infected cells were collected and centrifuged (5 minutes, 500 xg). The supernatant was removed and the cell pellet was lysed using lysis buffer (20 mM Tris-HCl, pH 7.5, 200 mM NaCl, 1 mM EDTA, 1% Triton-X-100 and 1 mM DTT). The lysate was incubated on ice for 10 minutes and centrifuged (5 minutes, 18000 xg). The activity of the supernatant was measured using the MDC activity assay (as described in Section 2.2.11), 10 µL of lysate and 80 µL of MDC assay buffer and 2 mM $CaCl_2$). The negative control represents bacmid containing no TG6 cDNA. The resulting data was blank corrected and averaged over three replicates.

5.2.6 Analysis of *Sf*9 cell lysates

Sf9 cells were seeded at a density of $5x10^5$ in a 6 well plate (Starstedt) and were incubated over overnight at 27 °C. The amount of virus added was based on its calculated virus titre (Equation 3.1) and incubated for a further 72 hours. Post-infection the supernatant was removed and the cells were lysed with lysis buffer (20 mM Tris-HCl, pH 7.5, 200 mM NaCl, 1 mM EDTA, 1% Triton X100 and 1 mM DTT). The lysate was centrifuged (5 minutes, 18000 xg). 10 µL of *Sf9* cell lysates was analysed using SDS-PAGE (4-20% Tris glycine, Invitrogen) and Western blot analysis (as described in Section 2.2.4 and 2.2.5).

5.2.7 Measuring the viral titre of the TG6 mutant P3 Baculovirus

The titre of the virus was measured used the BacPAK Baculovirus Rapid Titre Kit (Takara, 631406) as follows. *Sf9* cells were seeded at a density of 6.5 $x10^4$ cells/well in a 96 well plate (cell culture grade, Starstedt, 83.3926) and incubated at 27 °C for 1 hour and the medium was aspirated. The virus was diluted (10^{-3}, 10^{-4} and 10^{-5}) in *Sf9* medium and 25 µL of each of the virus dilutions was added to the appropriate wells and incubated at RT for 1 hour. The medium was aspirated and 50 µL methyl cellulose overlay was added before the microtitre plate was incubated at 27 °C for 43-47 hours. Formyl buffered acetone (150 µL) was added to each well and incubated for 10 minutes at RT and then removed. Each well was washed with 200 µL PBS plus 0.05% Tween 20 (PBS-T) for 5 minutes (3x) and 50 µL normal goat serum (diluted 1 in 4) was added for 5 minutes. 25 µL of mouse gp64 antibody (1:200) was added to each well and incubated at 37 °C for 30 minutes. The wells were washed with PBS-T for 5 minutes (2x) and subsequently incubated with 50 µL of goat anti-mouse antibody/ HRP conjugate (1:250) for 25 minutes at 37 °C. The wells were washed with PBS-T and 50 µL blue peroxidase substrate was added and developed for 3 hours. The blue foci were counted using a light microscope in each of the dilutions and used to calculate the virus titre with the following equation.

Equation 5.1: Calculation of viral titre.

$$Virus\ titre\ (IFU/ml) = average\ number\ of foci\ per\ well \times dilution\ factor \times 40$$

5.2.8 Expression and purification of His-tagged mutant TG6 from *Sf9* cells

Mutant TG6 proteins were expressed and purified as described in Section 3.2.13 and 3.2.14.

5.2.9 SDS-PAGE analysis of purified TG6 mutants

Due to the low concentration of the various TG6 mutants, 200 µL aliquots were lyophilised and resuspended in 8 M urea. 0.5 µg of each TG6 mutant was diluted in equivalent 2X sample buffer. SDS-PAGE was carried out as described in Section 2.2.4

5.2.10 Enzymatic activity of TG6 mutants

TG6 proteins were kinetically analysed using the MDC assay and the TG isopeptidase activity assay (as described in Section 2.2.10 and 2.2.11) with WT TG6. Data was baseline and blank corrected and averaged over three replicates with ± S.D.

5.2.11 Binding of allosteric regulators to TG6 mutants

The GTP binding and Ca^{2+} binding was assessed in the context of the isopeptidase activity assay. The activation of TG6 isopeptidase activity was measured using various concentrations of $CaCl_2$ and the inhibition of TG6 isopeptidase activity with various concentrations of GTP (as detailed in Section 2.2.15 and 3.2.18).

The following methods were used by the co-authors of the publication by Tripathy *et al*. 2017, on which I am a co-author.

5.2.12 Immunocytochemistry

Subcellular localisation of TG6 was investigated using immunocytochemistry. COS-7 cells were grown on poly-D-Lysine coated coverslips. Cells were then fixed with 4% paraformaldehyde for 15–20 minutes, at RT, 48 hours post transfection. Cells were then permeabilised with 0.1% Triton X-100 (in phosphate buffered saline, PBS) and subsequently blocked with PBS containing 10% FBS and 0.05% Triton X-100 for 45 minutes at RT. Fixed COS-7 cells were incubated with TG6 polyclonal antibody (Zedira #A017, 1:250) at 4 °C overnight. Washes (3x) with PBS followed and the slides were incubated with an Alexa-Fluor conjugated goat-anti rabbit secondary antibody (ThermoFisher scientific, dilution 1:1000) for 1 hour at RT in the dark. The coverslips were then washed with PBS and mounted on glass slides with DAPI containing Prolong

Gold antifade mounting media (Life technologies #P36961). Slides were imaged with Zeiss AxioObserver.Z1 microscope.

Confocal imaging was performed using a laser-scanning motorised confocal system (Nikon A1) equipped with an Eclipse Ti-E inverted microscope and four laser lines (405, 488, 561 and 638 nm). Z-series images were taken with an inter-stack interval of 0.5 µm using 60x oil objective. Laser intensities and camera settings were maintained identically within the same experiment to allow comparison of different experimental groups. Cell (n=100) were analysed from three independent experiments (N=3)

5.2.13 *In vitro* transamidase activity

Transamidase activity was measured using a biotinylated substrate peptide (0.5 mM, Biotinyl-Thr-Val-Gln-Gln-Glu-Leu-OH), in the presence of 5 mM DTT and 5 mM $CaCl_2$, with 25 µg of protein from cell lysates overexpressing TG6 for 15 minutes at 30 °C. Samples were analysed using SDS-PAGE and Western blot analysis. The nitrocellulose membrane was incubated with anti-streptavidin (IRDye 800CW #926-322301:10000) antibody for 1 hour at RT in the dark. Detection was carried out using an Odyssey infrared imagining system.

5.2.14 Western blotting of HEK293T cell lysates

HEK293T cells were lysed post transfection (25 mM Tris-HCl pH 7.4, 100 mM NaCl, 1 mM EGTA, 1% Triton-X-100, protease and phosphatase inhibitors). 25-50 µg of proteins were separated and analysed using SDS-PAGE. Detection of TG6 via the TG6 antibody (Zedira #A017, 1:1500) and the V5 tag antibody (ThermoFisher 1:1500) required the InfraRed dye conjugated antibody (1:10000), the signal was captured by an Odyssey infrared imaging system (LI-COR Biosciences).

5.2.15 Analysing protein turnover of TG6 mutants

Post-treatment with CHX treatment (0.1 mg/mL, 6 hours incubation) on HEK293T cells (overexpressing WT TG6 and TG6 mutants), the cells were collected in PBS and centrifuged (500 xg for 5 minutes, 4 °C). Radioimmunoprecipitation assay (RIPA) buffer (1% NP-40, 0.1% SDS, 50 mM Tris–HCl pH 7.4, 150 mM NaCl, 0.5% sodium deoxycholate, 1 mM EDTA) was used to lyse the cells, which were then sonicated and centrifuged (5000 xg, 10 minutes, and 4 °C). The supernatant was removed and

centrifuged (100000 xg, 30 minutes), this represents the soluble fraction. The insoluble fraction was collected by washing the remaining cell pellet in RIPA lysis buffer and centrifugation (100000 xg, 30 minutes). The supernatant was discarded and the pellet was resuspended in a buffer containing: 8 M Urea, 4% CHAPS, 50 mM DTT, 40 mM Tris-HCL pH 7.4, 2.5 mM EDTA. In experiments using MG132 (10 μM, proteasome inhibitor) or bafilomycin (200 nM, autophagy inhibitor), the cells were lysed after 6 hours. Expression was assessed in the cell lysates using Western blotting.

5.2.16 Statistics

A combination of statistical analysis was carried out. This included the one-way ANOVA (followed by Tukey's *post hoc* test), the two-way ANOVA (followed by Bonferroni's *post hoc* test) and the Student's *t*-test. The selected analysis was dependent on the specific experiment. Statistically significant results were defined as follows: $*P < 0.05$; $**P < 0.01$; $***P < 0.001$ and $****P < 0.0001$.

5.3 RESULTS

5.3.1 TG6 mutant cloning into pFastBac.

13 out of the 23 identified TG6 mutants were selected for biochemical investigation (TG6 R111C, Q181H, D327G, V391M, R412C, Y441C, R448W, L502Q, L502Q Q652dup, D510H, L517W, E574del and Q652dup) as these were the only identified mutants where the plasmids were readily available. The pFastBac *TGM6* and the mutated pcDNA3 *TGM6* were digested and the DNA fragment of interest was isolated for ligation into the pFastBac backbone (Fig. 5.1). Nine out of 11 of the TG6 mutants being generated via subcloning were successfully cloned. Cloning for the remaining two TG6 mutants (TG6 V391M and R448W) is still ongoing.

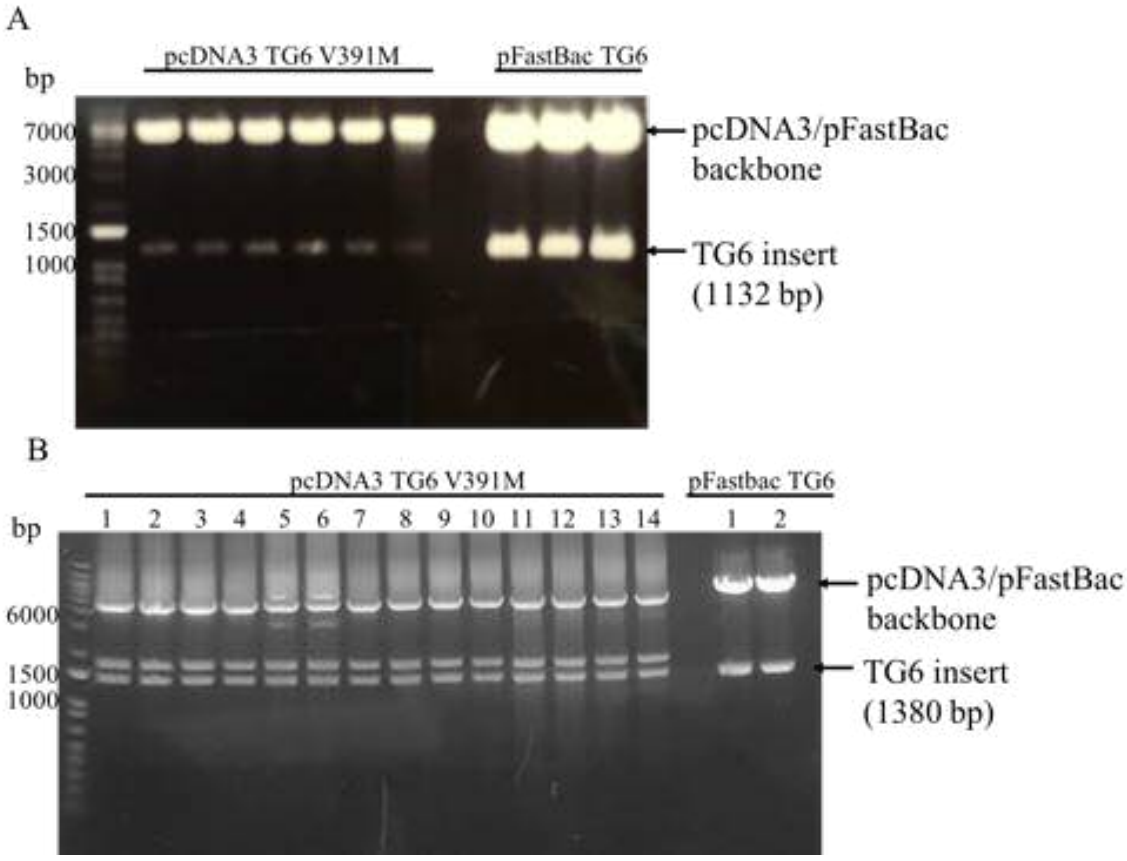


Figure 5.1: Subcloning of TG6 mutants.
TGM6 mutant DNA was excised from pcDNA backbone using selected restriction enzymes (Table 5.2). Following separation of DNA fragments using agarose gel electrophoresis, fragments were gel purified, for subcloning into pFastbac. A: pcDNA TG6 V391M and pFastBac TG6 was digested using *Bsu36I*. B: The pcDNA TG6 V391M and pFastBac were digested using *XmaI.*

TG6 R412C and Q652dup were generated by site-directed mutagenesis (Fig. 5.2). Different annealing temperatures were used to establish the optimal conditions of mutagenesis. The optimal annealing temperature for TG6 R412C was 68.2 °C and TG6

Q652dup 61.8 °C. TG6 Q652dup was successfully introduced into TG6 cDNA, whereas attempts to introduce R412C were unsuccessful i.e. colonies were screened but the mutation was not introduced when sequenced by Eurofins.

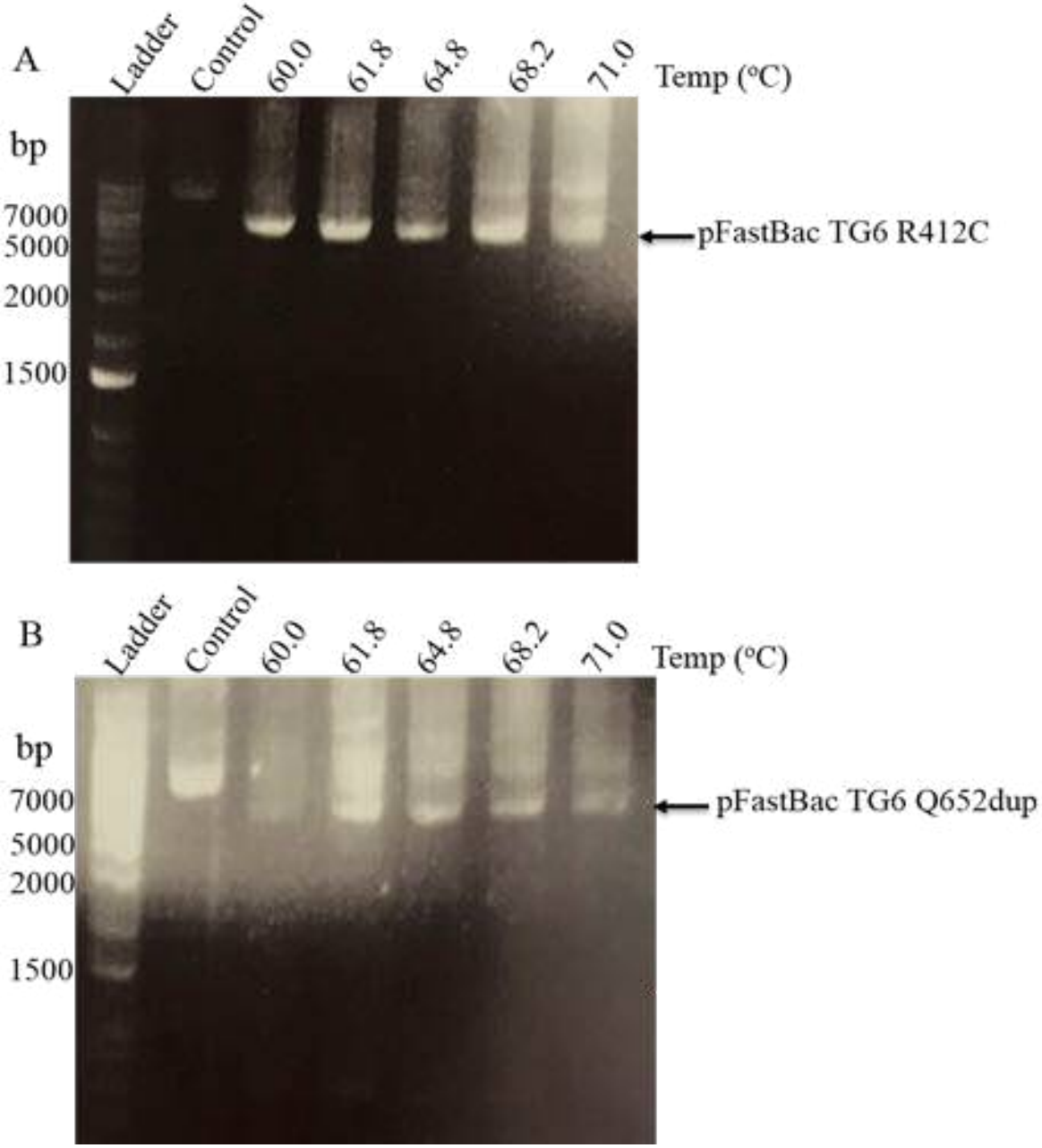


Figure 5.2: Site directed mutagenesis of TG6 cDNA.
Site-directed mutagenesis was carried out to generate TG6 R412C and Q652dup and pFastBac TG6 was used as the template. PCR negative control was unamplified pFastBac TG6 of equivalent amount added to the reaction. A gradient of annealing temperatures (60-71 °C) were used to identify the best amplification. A: PCR reaction for pFastBac TG6 R412C generation. B: PCR reaction for pFastBac TG6 Q652dup generation. Data is representative of one independent experiment (N=1).

5.3.2 Production of TG6 mutant baculovirus

Bacmid PCR was carried out and generated a PCR product of ~5000 bp, (pFastBac (~2300 bp) plus TG6 cDNA (~2100 bp) (Fig. 5.3). A bacmid with no TG6 cDNA was used as a negative control which produced a ~300 bp band. This was carried out for all 10 TG6 mutants and a positive clone (successful integration of mutated TG6 into the bacmid) was identified for WT TG6, R111C, Q181H, D327G, D510H and E574del (Fig. 5.3). Positive clones were also identified for Y441C, L502Q, L502Q Q652dup, L517W and Q652dup (data not shown).

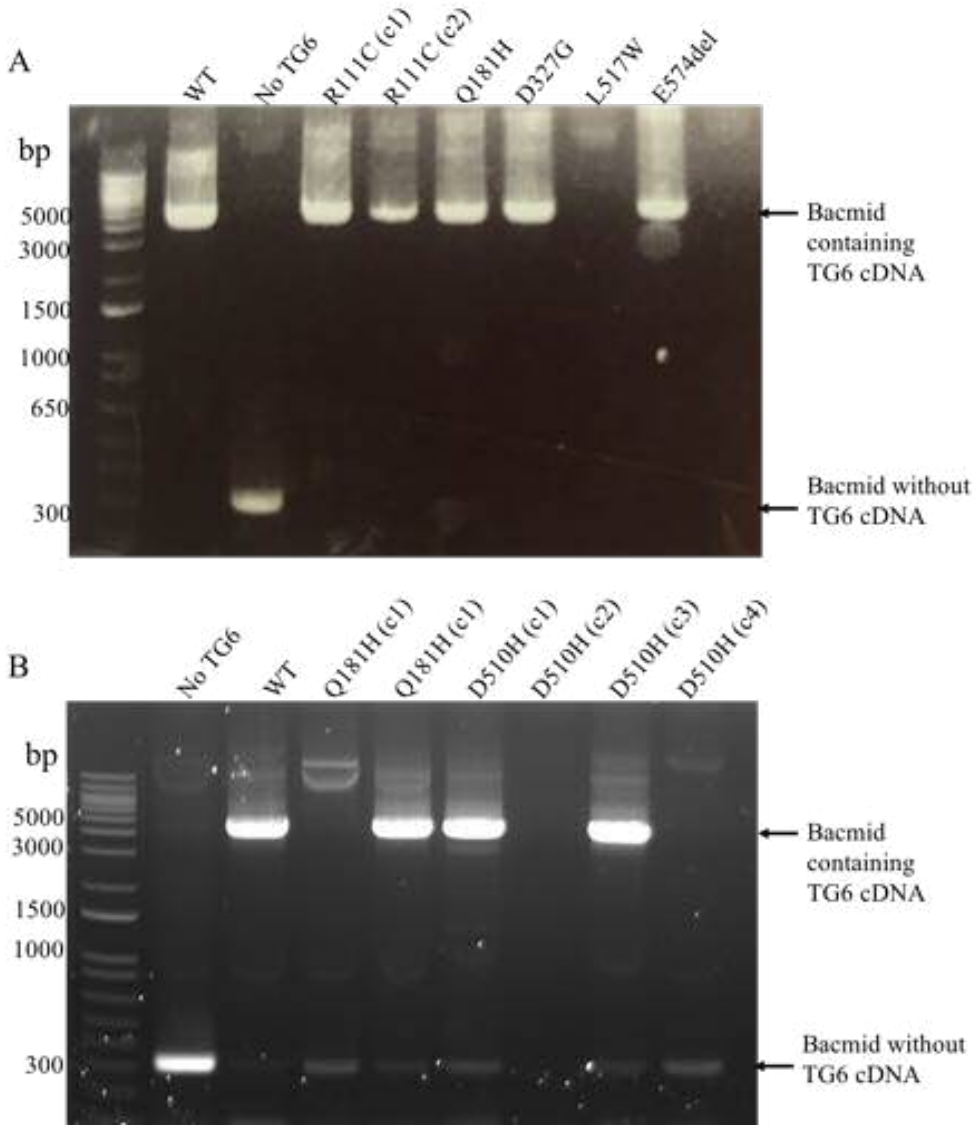


Figure 5.3: PCR of TG6 mutant bacmid.
The integration of TG6 cDNA into bacmid DNA was determined by PCR. Primers recognised the flanking region of the bacmid recombination site and generated a band of ~5000 bp with the TG6 cDNA and without generates a ~300 bp band. Multiple clones were screened per TG6 mutant. Representative examples are shown. A: Positive integration TG6 cDNA for R111C, Q181H, D327G and E574del. B: Positive integration of Q181H and D510H.

The lysates from virally infected *Sf9* cells, used to generate P1, P2 and P3 virus, were analysed for transamidation using the MDC assay (assay buffer containing 2mM $CaCl_2$). This was carried out for the TG6 mutants: R111C, Q181H, D327G, L502Q, L502Q Q652dup, L517W and E574del (Fig 5.4). Transamidation activity was observed for L502Q and L502Q Q652dup with the P1 virus (Fig. 5.4A). Reduced activity was observed for Q181H, D327G and E574del whilst R111C and L517W was comparable to the negative control. Similar results were observed for the P2 virus with regards to R111C, L502Q, L502Q Q652dup, and L517W (Fig 5.4B). In contrast, the activity for Q181H, D327G and E574del showed activity comparable to the negative control on this occasion. All mutants show comparable activity in P3 virus lysates, indicating protein has not been expressed or has been degraded when making P3 virus.

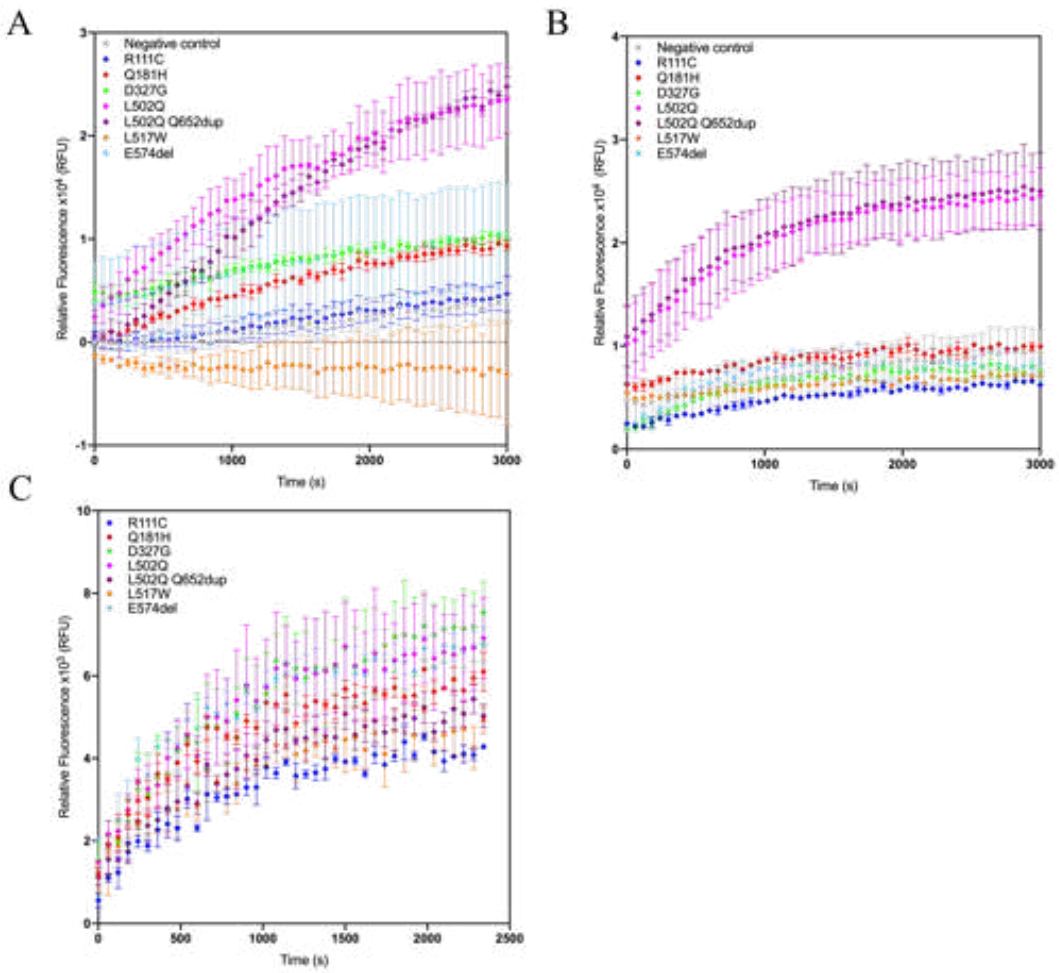


Figure 5.4: Baculovirus generation from recombinant bacmid.
The MDC assay was carried out on *Sf9* cell lysates to establish successful expression of TG6. Measurements are blank corrected and averaged over three replicates with ± S.D (n=3) is representative of 1 independent experiment (N=1). A: Passage 1 virus. B: Passage 2 virus. C: Passage 3 virus.

The TG6 activity from cell lysates confirmed successful production of TG6 baculovirus and indicated that the protein correctly folded. The protein expression in cell lysate was assessed, by using Western blot lysate post infection (72 hours), to support the successful production of baculovirus (Fig. 5.5). In this case the TG6 mutants Q181H, D327G, L502Q, L502Q Q652dup and E574del have been successfully generated but on this occasion WT TG6 protein was not detected. Both TG6 R111C and L571W had no expression and therefore the generation of baculovirus was repeated.

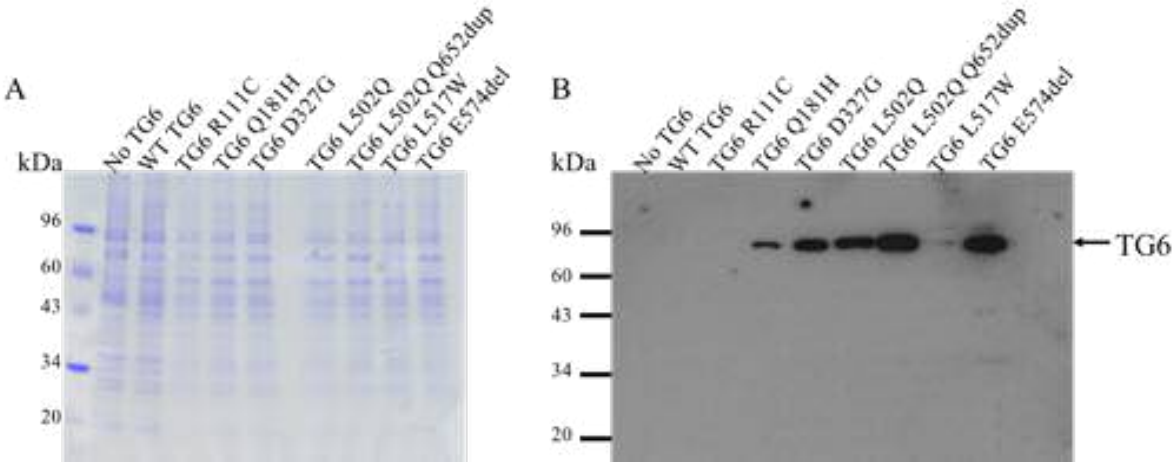


Figure 5.5: SDS-PAGE and Western blot of TG6 baculovirus infected Sf9 cells.
Sf9 cells were infected the appropriate amount of virus and was incubated for 72 hours. Cells were collected and lysed. 10 µL of cell lysate, for each TG6 mutant. A: SDS-PAGE Coomassie R staining analysis of *Sf9* cell lysates. B: Western blot analysis of *Sf9* cell lysates. The membrane was stained using polyclonal anti-TG6. Data is representative of one independent experiment.

The TG6 mutant baculovirus was titred and the measured titres are shown in Table 5.6 from these the volume of supernatant required for 1 L of cell culture for protein expression was calculated (Table 5.6).

Table 5.6: Measured viral titres of TG6 mutant baculovirus.

TG6	Average foci (in viral dilution 10^{-5})	Titre x 10^{7}(IFU/mL)	Volume (uL) per 1L culture	Protein concentration (mg/mL)
WT	38.5	15	650	1.0
R111C	31.0	12	800	0.1
Q181H	10.5	4.3	2330	0.3
D327G	11.5	4.6	2170	1.1
L502Q	14.3	5.7	1750	1.1
L502Q Q652dup	12.7	5.1	2000	0.9
L517W	35.3	14	700	<0.1
E574del	13.0	4.5	2200	0.4

5.3.3 Expression and purification of TG6 mutants in *Sf9* cells

Following infection with TG6 baculovirus, the viability and density of the *Sf9* cells was calculated before purification. WT TG6, TG6 Q181H, D327G, L502Q, L502Q Q652dup and E574del had a cell viability of 50-60% and a cell density of ~4.0x10^6 cells/mL. The desired viability of the *Sf9* cells post-infection is >70% and a cell density of 4.0x10^6 cells/mL (starting cell density of 2.0x10^6 cells/mL) for the greatest yield of protein. Despite sub-optimal cell viability, WT TG6, TG6 D327G, L502Q and L502Q Q652dup yielded comparable concentrations of purified TG6 (Table 5.6). TG6 Q181H and E574del purification resulted in significantly reduced yields of protein compared with WT TG6 (0.25 and 0.35 mg/mL, respectively). In contrast, the TG6 R111C and L517W mutants consistently had a cell viability of <10% and an increase in the cell density was not observed. Both TG6 R111C and L517W cultures produced <0.1 mg/mL of purified protein, even when cell culture volume was expanded (0.5 L to 1 L). The consistently low yields of these proteins could indicate that they are very unstable and could be toxic to the cells. As a result, the *Sf9* cells could be undergoing premature cell death leading to lower protein yields when compared with WT and the other TG6 mutants. TG6 Y441C, D510H and Q652dup were not expressed and purified as the virus was not available at this time.

The purity of the TG6 mutants were analysed using SDS-PAGE analysis with Coomassie blue R staining (Fig. 5.6). A single band was observed for WT TG6 and all TG6 mutants, except TG6 L517W, indicating their purification was successful. TG6 L517W showed four bands, one band was migrated with the same apparent size of WT TG6 and the remaining three bands could be attributed to contaminating proteins or other TG6 products. Following successful purification of the TG6 mutants, biochemical investigations to establish the impact of the mutant on the structure and function of TG6, were carried out.

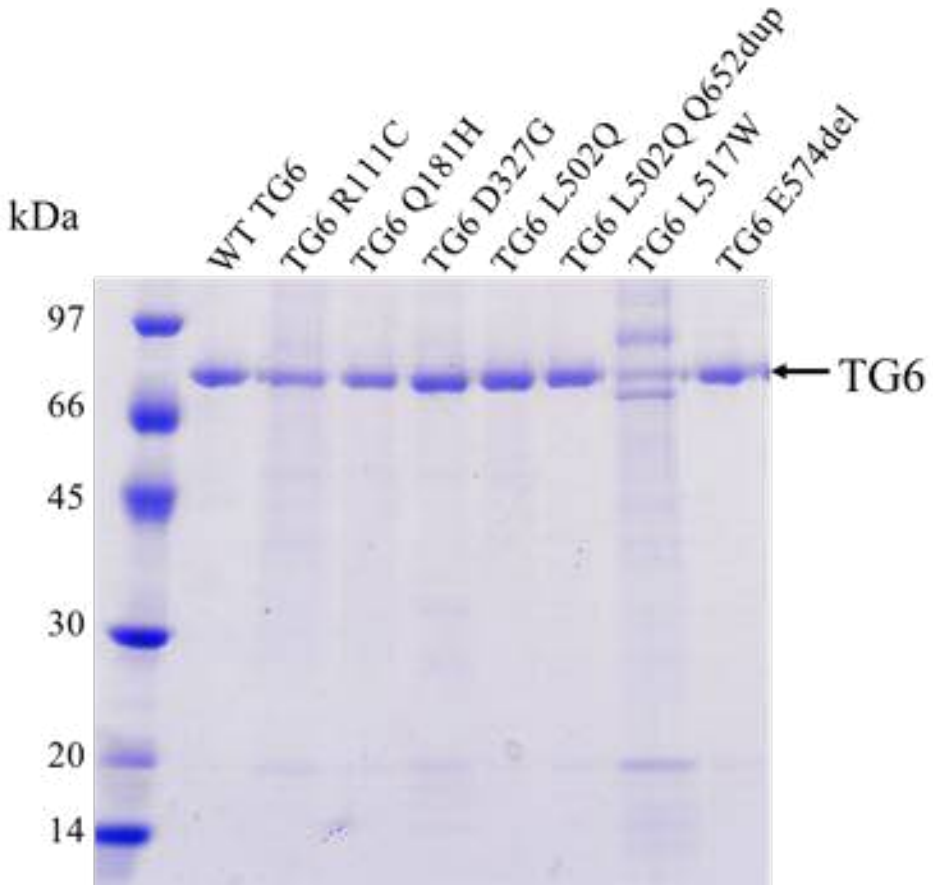


Figure 5.6: SDS-PAGE analysis of purified TG6 mutants.
TG6 was lyophilised and resolved in 8 M urea. 0.5 µg of WT TG6 and TG6 mutant samples, in equivalent 2x sample buffer, were run on 4-20% SDS-PAGE gel under reducing conditions, Coomassie brilliant blue R was used to stain and visualise the protein (Gel by A.Schmitt-Thomee). Data is representative of 1 independent experiment.

5.3.4 TG6 mutant activity and allosteric regulation

The activity of the TG6 mutants were investigated using the isopeptidase activity assay and MDC assay (Fig. 5.7). The isopeptidase activity (Fig. 5.7A) and the transamidation activity was measured by the MDC assay (Fig. 5.7C). In each case, the rate of reaction for each of the TG6 mutants was calculated (Table 5.7). The results from the isopeptidase activity and the MDC were normalised to percentage activity where TG6 WT was 100%, when comparing the activity of the TG6 mutants in this way the differences are apparent (Fig 5.7B and D). This showed the activity of all the TG6 mutants was hindered when compared to WT TG6. TG6 R111C, Q181H and E574del have greater isopeptidase activity compared with the transamidation activity, whereas TG6 D327G, L502Q, L502Q Q652dup and L571W have increased transamidation activity compared to their isopeptidase activity. This is most apparent with TG6 D327G where the difference between in activity is 22.3% in favour of the transamidation activity.

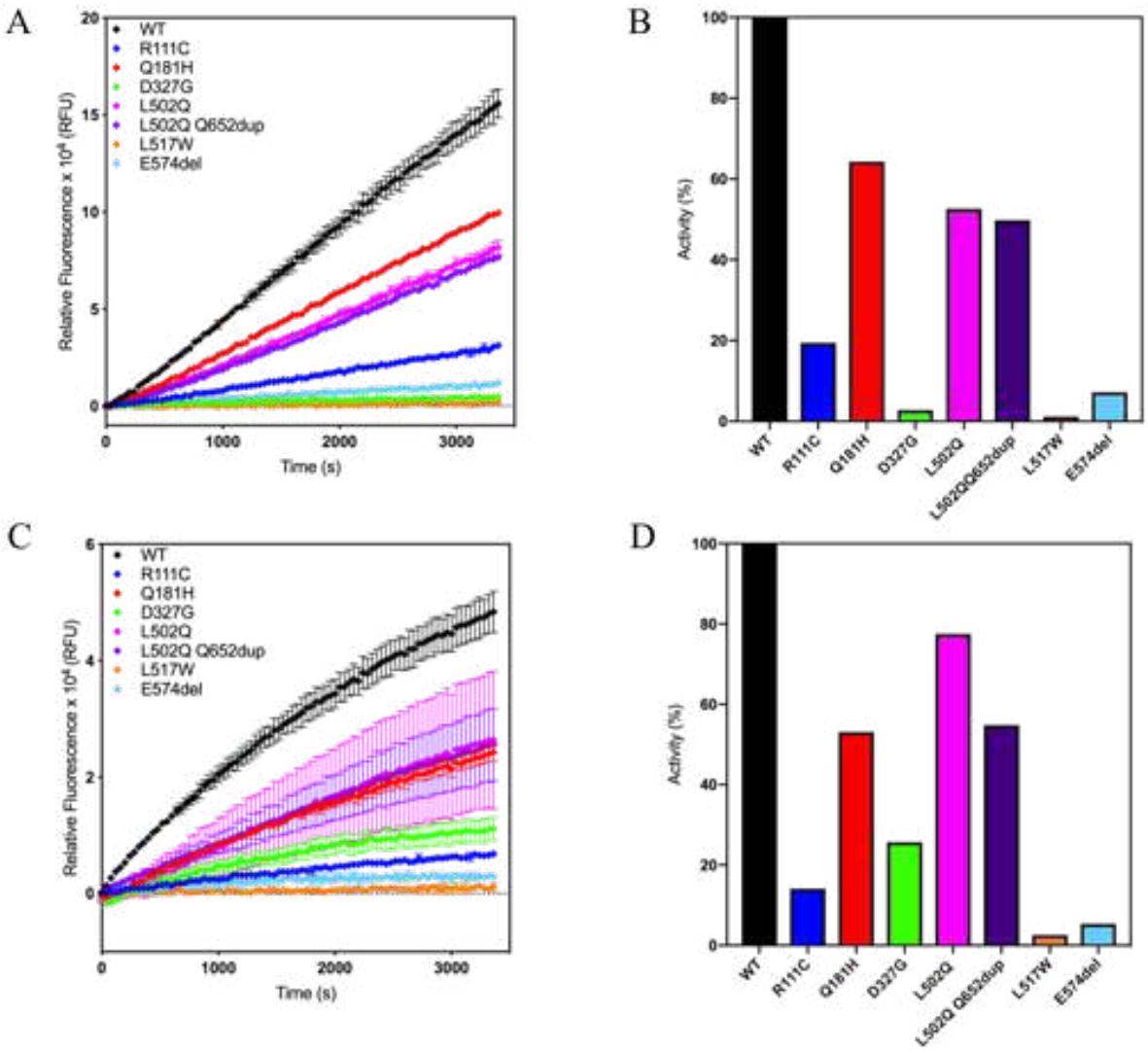


Figure 5.7: Isopeptidase and amine incorporation activity of TG6 mutants.
The activity of purified TG6 (20 µg /mL) was assessed using the isopeptidase activity assay and the MDC assay. Data is representative of the mean of 2 wells ± SD (n=2) and representative of two independent experiments (N=2). A: The isopeptidase activity measured as in Fig 2.6, over 1 hour. B: The rate of isopeptidase activity was determined, by linear regression, for each of the TG6 mutants and compared to WT. C: Transamidation activity was measured using the MDC assay over 1 hour. D: Rate of transamidation activity of the TG6 mutants were compared with WT.

The Ca^{2+} binding and GTP binding was assessed in the context of the isopeptidase activity for the TG6 mutants. Ca^{2+} binding/activation was measured using increasing concentrations (0.15 µM to 32 mM) of $CaCl_2$ to activate the isopeptidase activity (Fig. 5.8A). The K_D for Ca^{2+} binding for TG6 R111C, L502Q and L502Q Q652dup Ca^{2+} is comparable to WT TG6 (Table 5.7). In contrast, TG6 Q181H, D327G and E574del were less sensitive to Ca^{2+} and require higher concentrations for activation. The TG6 L517W

mutant was completely insensitive to Ca^{2+} and isopeptidase activity could not be activated with high $CaCl_2$ concentrations (32 mM, data not shown).

The inhibition of isopeptidase activity was measured using increasing concentrations of GTP (0-8 mM) (Fig. 5.8B). The K_D of GTP for each of the TG6 mutants were calculated (half maximal activity of normalised data, Table 5.7). The inhibition of the isopeptidase activity of TG6 L517W with GTP, could not be assessed as its isopeptidase activity was severely limited in the absence of GTP. The WT TG6 had a K_D of 240 µM for GTP and TG6 Q181H was the only TG6 mutant with a similar K_D for GTP at 220 µM. TG6 R111C, D327G and L502Q all showed decreased sensitivity to inhibition with GTP when compared with WT TG6 (Table 5.7). Conversely, TG6 L502Q Q652dup and E574del have increased sensitivity to inhibition with GTP.

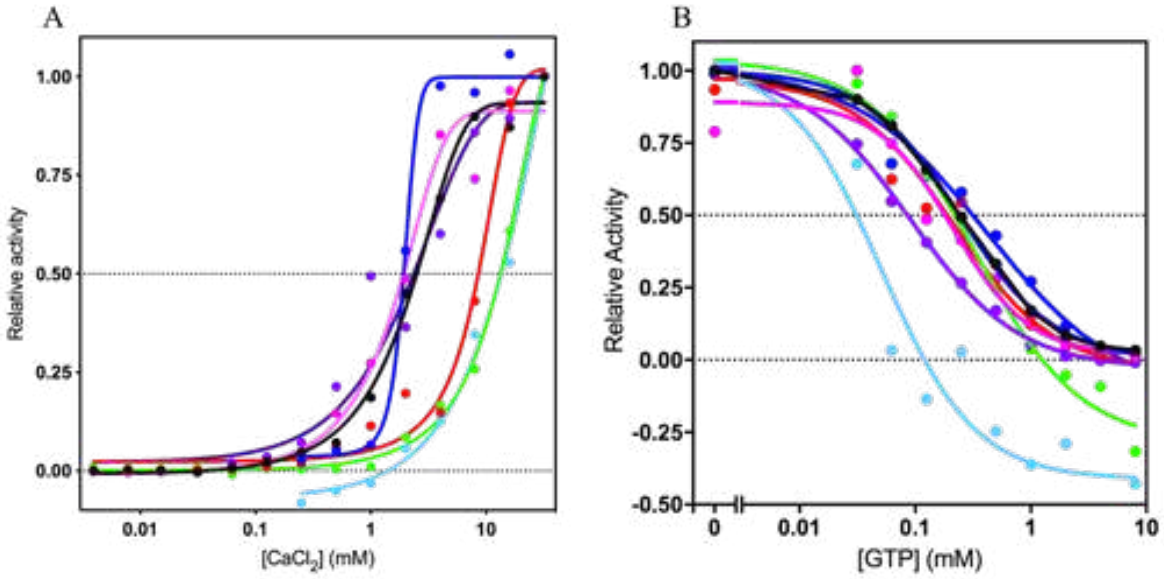


Figure 5.8: Binding of Ca^{2+} and GTP by TG6 mutants.
Ca^{2+} and GTP binding were assessed using the isopeptidase activity assay. The change in fluorescence was measured for 20 µg/ml of TG6 (WT TG6 (black), TG6 R111C (blue), Q181H (red), D327G (green), L502Q (pink), L502Q Q652dup (purple) and E574del (light blue)) for different concentrations of Ca^{2+} or GTP over 1 hour. Data was blank and baseline corrected and are the mean of 2 wells (n=2) and representative of 1 experiment (N=1). The rate of reaction was measured with linear regression and was normalised (0 to 1) and then fitted to a sigmoidal dose response curve. The apparent K_D was calculated from half maximal inhibition/activation. A: $CaCl_2$ (15.6 µM to 32 mM) was used to activate TG6. B: Inhibition of isopeptidase activity by GTP (31.3 µM to 8 mM) was measured. Work aided by Dr R. Griffiths.

From the data obtained, it is apparent that some of the TG6 mutations impact the activation of TG6. For example Q181H, D327G and E574del as they require higher Ca^{2+} concentrations for the activation. Alternatively, the TG6 mutants L502Q Q652dup and E574del have higher affinity for GTP and R111C and D327G have decreased sensitivity

to GTP, suggesting TG6 regulation (transition from open to close conformation) may be affected. Consequently, the impact on the regulation, i.e. Ca^{2+} and GTP binding, of TG6, upon introduction of the mutation, could be important the in the pathogenic mechanism.

Table 5.7: Kinetic analysis of TG6 mutants.

TG6	Isopeptidase activity rate (RFU/s)	SEM	Isopeptidase Activity (%)	MDC activity rate (RFU/s)	SEM	MDC activity (%)	Ca^{2+} K_D (mM)	GTP K_D (µM)
WT	47.7	0.254	100.0	13.9	0.190	100.0	2.4	240
R111C	9.3	0.062	19.4	2.0	0.028	14.1	2.0	420
Q181H	30.6	0.063	64.3	7.4	0.067	53.1	9.1	220
D327G	1.3	0.030	2.8	3.6	0.102	25.7	13.4	330
L502Q	25.1	0.089	52.6	8.5	0.211	61.2	2.1	210
L502Q Q652dup	23.7	0.026	49.8	7.6	0.049	54.8	3.2	90
L517W	0.5	0.051	1.1	0.4	0.029	2.5	n/a	n/a
E574del	3.4	0.122	7.2	0.7	0.423	5.4	12.3	50

The following results are the work of the co-authors of the publication by Tripathy *et al*. 2017, on which I am a co-author.

5.3.5 Investigating the subcellular distribution of TG6 mutants

In addition to biochemical analysis, WT TG6 and the TG6 mutants were investigated in a cellular context to establish the cellular mechanisms involving the TG6 mutants. The expression and subcellular localisation was investigated in mouse cerebellum, primary mouse cerebellar and cortical neurones. TG6 expression was observed in the granular layer of the cerebellar cortex and the Purkinje cells. WT TG6 was predominantly localised to the nucleus and the perinuclear regions. Similar localisation was observed in COS-7 cells, with 70% localised to nucleus (Fig. 5.9).

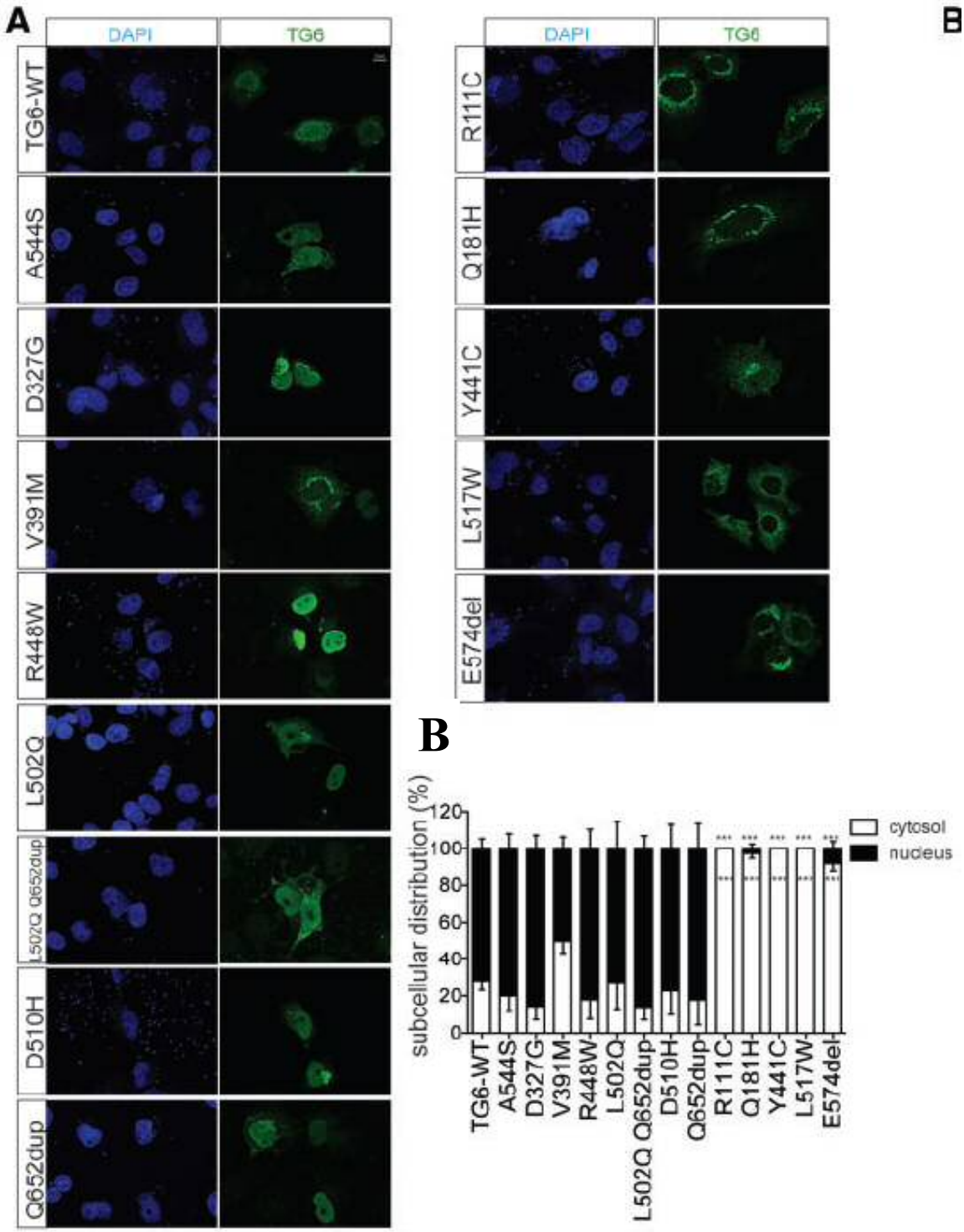


Figure 5.9: Subcellular localisation of TG6 mutants in COS-7 cells.
A: Immunocytochemical analysis for WT TG6 and the TG6 mutants in COS-7 cells. The images are representative of three independent experiments. B: Quantification of the subcellular localisation of WT TG6 and the TG6 mutants. Graph shows mean ± SEM, n=100 cells from three independent experiments. Statistical analysis using two-way ANOVA with Bonferroni's post hoc test (***P <0.001) (Tripathy *et al.* 2017).

WT TG6 and the TG6 mutants: A544S, D327G, V391M, R448W, L502Q, L502Q Q652dup, D510H, Q652dup were localised to the nucleus. In contrast, TG6 R111C, Q181H, Y441C, L517W and E574del were predominantly localised to the perinuclear space and the cytoplasm. Experiments with calnexin (ER marker protein) and WT TG6 showed 13% of WT TG6 was co-localised with calnexin compared with 45% of TG6 R111C (Fig. 5.10). This indicates that TG6 Q181H, Y441C, L517W and E574del all result in a loss of TG6 localisation in the nucleus and increased accumulation in the ER, as with TG6 R111C.

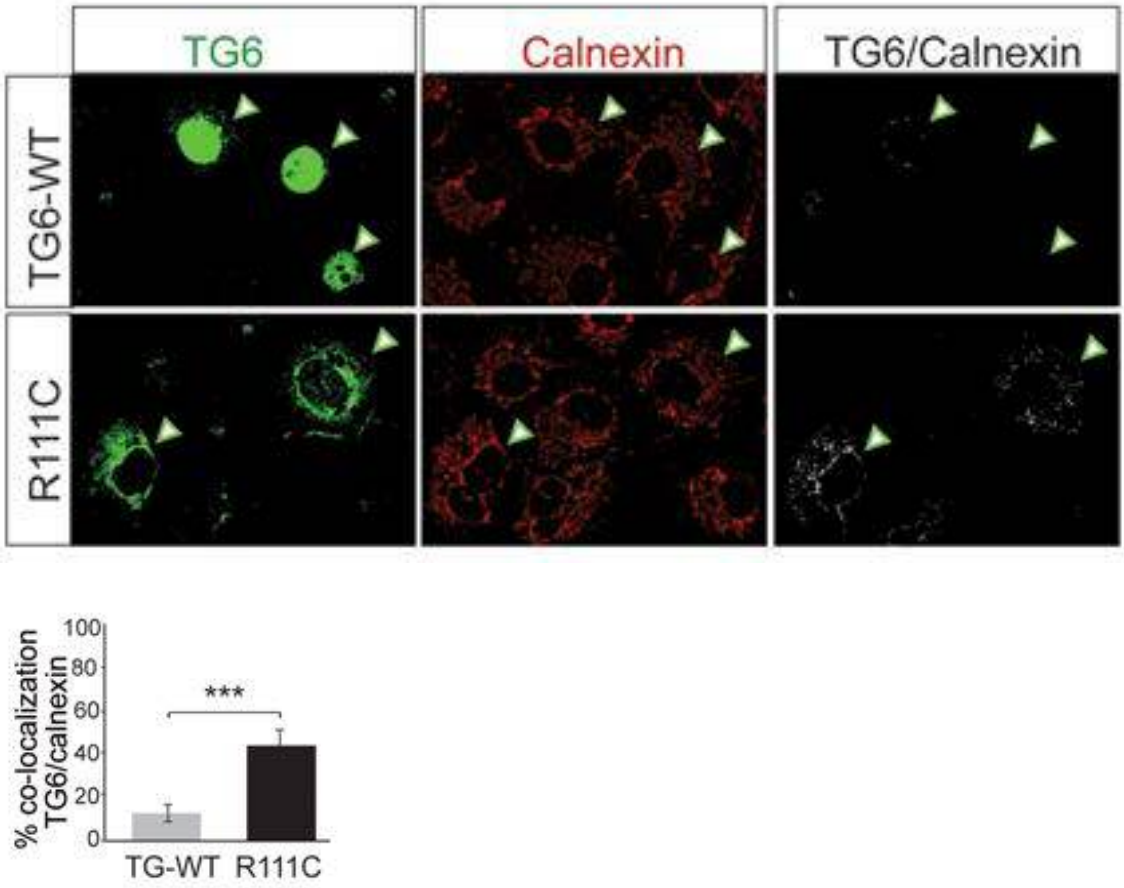


Figure 5.10: Co-localisation of TG6 mutants with calnexin.
Immunocytochemical analysis of the co-localisation studies of TG6 with calnexin in COS-7 cells expressing WT TG6 and TG6 R111C. The co-localisation of TG6 and calnexin are highlighted with arrowheads. Co-localisation of TG6 and calnexin is represented as percent co-localisation ±SEM. Statistical significance: ***P<0.001 (Students t test).

5.3.6 Transamidase activity of TG6 mutants in HEK293T cell lysates

Transamidase activity was measured for each of the mutants using a biotinylated peptide that irreversibly cross-links to TG proteins when in an environment containing high levels of Ca^{2+} (5 mM, Fig. 5.11) (Tripathy *et al.* 2017). The enzymatic activity of TG6 mutants: R111C (86%), Q181H (60%), Y441C (77%), L517W (90%), E754del (80%), D327G

(54%) and V391M (77%) was significantly decreased when compared with WT whilst the transamidase activity was marginally impaired for the remaining TG6 mutants. Those TG6 mutants showing abnormal subcellular localisation also have a major reduction in transamidase activity. This indicates the importance of subcellular localisation with regards to activity.

5.3.7 TG6 mutants with abnormal localisation induce UPR

To investigate protein turnover TG6 levels were measured following treatment with CHX after 6 hours (Tripathy *et al.* 2017). A 23% decrease in TG6 accumulation was seen for WT, 27% for TG6 D327G and major reduction of 53% was found for TG6 R111C (Fig. 5.12). Additional analysis confirmed this was not due to reduced mRNA levels and thus reduced protein expression.

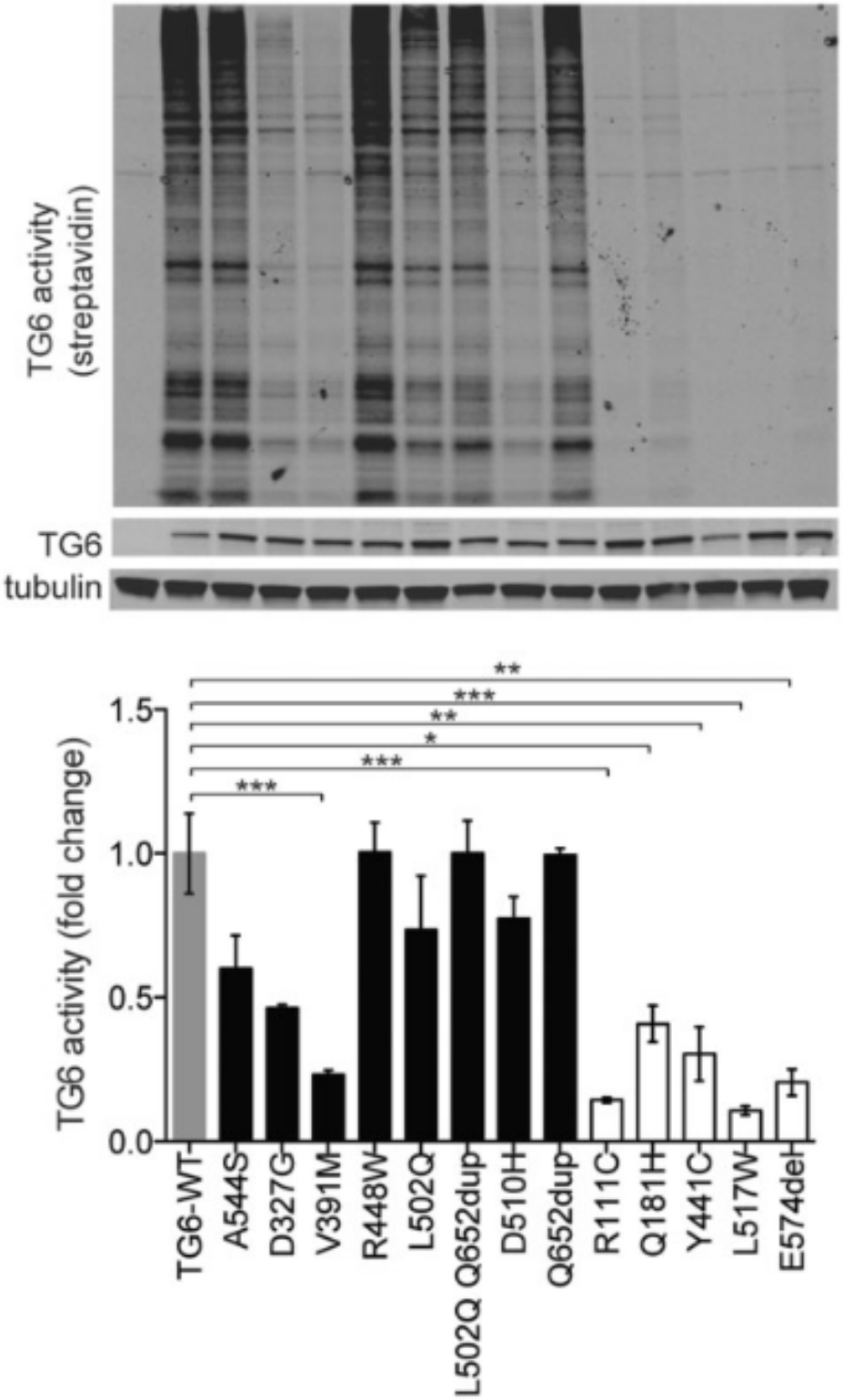


Figure 5.11: Transamidase activity of TG6 mutants.
The transamidase activity of the TG6 mutants was measured by the irreversible crosslinking of biotinylated glutamine substrates by the TG6 mutants (in HEK293T cells). A: Western blotting with streptavidin conjugated fluorophore was used to detect crosslinked substrates of various molecular weight. B: Densitometric analysis of western blot to quantify transamidase activity (shown as fold change) of WT TG6 and the TG6 mutants. Graph shows mean SEM and statistical analysis with one-way ANOVA with Tukey's *post hoc* test (*P <0.05; **P <0.01; ***P <0.001) (Tripathy *et al.* 2017).

To establish if degradation was due to autophagy or initiation of the UPR, cells were treated with CHX and bafilomycin or MG132 (Tripathy *et al.* 2017). Degradation of TG6 R111C was inhibited by MG132 indicating a UPR response was involved. In addition, under normal conditions, TG6 R111C was found distributed equally between in soluble and insoluble fraction whereas WT and TG6 D327G were found predominantly in the soluble fraction following cell lysis (data not shown). When in the presence of MG132, TG6 R111C was found predominantly in the insoluble fraction indicating TG6 R111C more readily aggregates and is degraded by the proteasome.

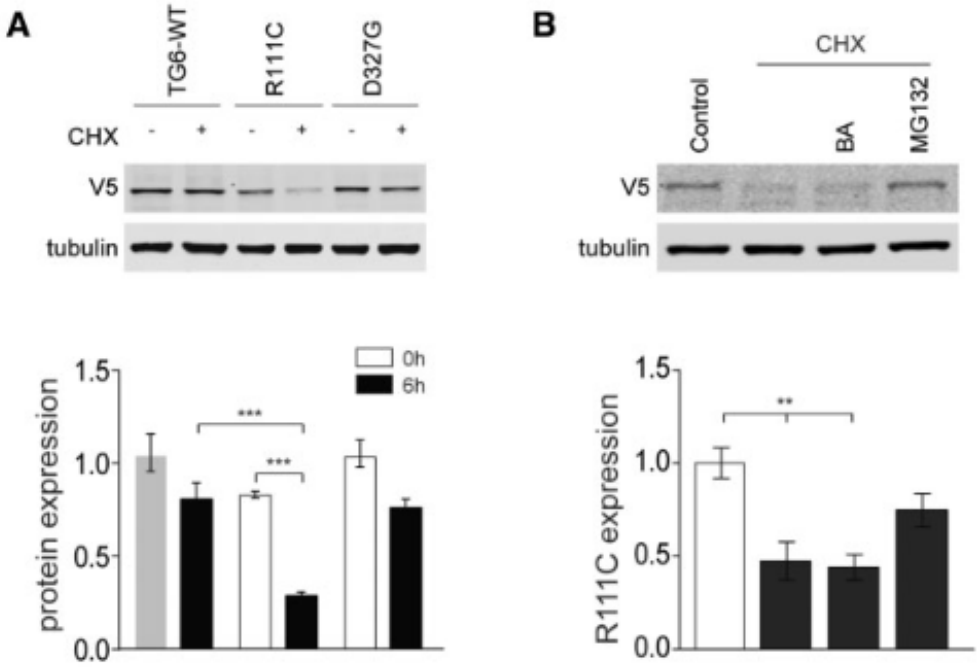


Figure 5.12: Analysis the turnover of WT and TG6 mutants in HEK293T cells. Protein turnover was assessed for WT TG6, TG6 R111C and D327G upon CHX treatment. A: Western blot analysis was used to assess TG6 expression level overtime from HEK293T cell lysates. Densitometric analysis quantified changes expression over time (0-6 hours). Graph shows mean SEM and was statistically analysed using two-way ANOVA with Bonferroni's *post hoc* test (***P <0.001). B: Western blot analysis was used to assess the TG6 expression levels following treatment with CHX and bafilomycin or MG132. Graph shows mean SEM and statistical analysis of one-way ANOVA with Tukey's *post hoc* test (**P <0.01) (Tripathy *et al.* 2017).

5.4 DISCUSSION

Many of the TG6 mutants lacked biochemical characterisation and therefore predictions made by the bioinformatic and homology method were used to predict the outcome of enzyme functional experiments.

5.4.1 Stability of TG6 mutants

Most TG6 mutants have not been previously expressed and purified using this method (excluding D372G and L517W, which has previously been expressed by our lab). Consequently, the amount of soluble TG6 mutant protein purified in comparison to WT TG6 gave an insight into the stability of the protein. Purification of TG6 D327G, L502Q and L502Q Q652dup produced yields comparable to WT TG6 (1 mg/mL, Table 5.6) suggesting their stability maybe similar to WT. On the other hand, TG6 R111C, Q181H, L517W and E574del purified at concentrations between 0.1-0.35 mg/mL. One explanation for reduced concentrations could be the reduced stability of these four TG6 mutants. This in most cases can be supported by the bioinformatics analysis carried out in Chapter 4 (summary in Table 4.24). For example, in closed conformation, TG6 R111C, Q181H, L517W and E574del were predicted to be unstable (destabilising to highly destabilising) and TG6 D327G (highly stabilising). However, L502Q and L502Q Q652dup were predicted to be unstable (L502Q destabilising and Q652dup highly destabilising) using FoldX (Table 5.8) but was actually shown to express and purify similar to WT (Table 5.8). In addition, A3D predicts a decreased solubility for TG6 R111C and TG6 L517W. Data obtained from FoldX and A3D, could explain why the extracted protein concentrations were significantly lower for these mutants. In addition, the protein turnover of WT TG6, TG6 R111C and D327G was investigated in HEK293T cells following CHX treatment. This showed WT TG6 and D327G had similar protein turnover whereas TG6 R111C had reduced stability. This is linked to the mutant proteins location in the cell, whereby those localised to the perinuclear regions are predicted to be misfolded. Furthermore, the TG6 R111C had reduced stability compared with WT measured, is consistent with the literature (Guo *et al.* 2014). In contrast, TG6 D327G showed reduced stability in the literature, whereas here little difference in stability was detected when compared with WT.

Furthermore, the viability of *Sf9* cells was 0-10% for TG6 R111C and L517W, which is significantly lower than the other TG6 mutants with a viability of 50-60%. During the 72

hours the cells divide once before viral infection halts mitosis and WT TG6 and TG6 D327G, L502Q and L502Q Q652dup had a cell density of ~4.0x10^6 cells/mL when starting from a density of 2.0x10^6 indicating the ratio of virus to cells was correct for successful protein expression. In contrast, TG6 R111C, Q181H, L517W and E574del did not achieve a density of 4.0x10^6 cells/mL when starting from 2.0x10^6 cells/mL thus indicating premature death of the *Sf9* cells, which in turn impacted protein expression. This supports the hypothesis that TG6 R111C, Q181H, L517W and E574del are unstable, likely due to misfolding, which in turn may be toxic to the cell. This is supported by increased toxicity observed in *Drosophila*, where WT TG6 and TG6 D327G have comparable toxicity but TG6 R111C shows significantly increased toxicity (Tripathy *et al.* 2017). Alternatively, the TG6 mutants may have reduced expression levels compared to WT (Fig. 5.5). This is unlikely as no changes in expression were noted in cellular studies (Tripathy *et al.* 2017).

Table 5.8: Comparison of stability of TG6 mutants.

TG6 Mutant	Concentration (mg/mL)	FoldX (closed conformation)	AGGRESCAN3D (closed conformation)	Cellular localisation
WT	1.0	N/a	N/a	Nucleus
R111C	0.1	Destabilised	Decreased solubility	Peri-nuclear
Q181H	0.3	Highly destabilised	No change	Peri-nuclear
D327G	1.1	Stable	No change	Nucleus
L502Q	1.1	Destabilised	Increased solubility	Nucleus
L502Q Q652dup	0.9	Highly destabilised	N/a	Nucleus
L517W	<0.1	Highly destabilised	Decreased solubility	Peri-nuclear
E574del	0.4	Highly destabilised	N/a	Peri-nuclear

Tripathy *et al*. showed TG6 R111C initiates the UPR and is targeted to the proteasome. *Sf9* cells contain a 26S proteasome which functions in a similar way to the human proteasome. Proteasome activity is required for baculovirus life cycle progression and also be involved in the degradation of cellular proteins (Lyupina *et al.* 2016). It has been shown that TG6 R111C and it is predicted that TG6 Q181H, L517W and E574del initiate the UPR in human cells. This mechanism may also be taking place in *Sf9* cells and would hence explain the drop in the concentration of these mutants extracted. In order to address this hypothesis, the transfection of the baculovirus for TG6 R111C, Q181H, L517W and

E574del needs to be further investigated. For example the optimal harvest time needs to be improved and the transfection efficiency of the virus can be assessed perhaps by measuring the cell size, cell density, enzymatic function and cell viability over a time course (Ustun-Aytekin *et al.* 2014). This will allow further insight to whether the transfection of the virus has impacted the amount of protein purified or if the UPR is initiated in *Sf9* cells and the resulting protein degraded.

If sufficient protein can be isolated for TG6 R111C, Q181H, L517W and E574del, structural analysis of the TG6 mutants can assess the impact of the mutation on TG6 structure. As both size exclusion chromatography and CD spectroscopy were used for TG2 GTP binding variants (Chapter 2), the structure of TG6 mutants can be assessed using these techniques. This would also support the pathogenic mechanism elucidated for TG6 R111C i.e. targeting by the unfolded protein response for proteasomal degradation (Tripathy *et al.* 2017). In addition, this information would identify if altered subcellular localisation is linked with structural deficits of TG6 Q181H, Y441C, L517W and E574del.

5.4.2 Function and regulation of TG6 mutants

The impact of the TG6 mutants on the function of TG6 was assessed using the assays developed in Chapter 2. The MDC assay investigates the incorporation of MDC into the glutamine substrate (β-casein), by monitoring the increase in fluorescence of MDC (Lorand *et al.* 1992). The isopeptidase activity assay investigates the binding of a quenched glutamine substrate (A102, Zedira) by the Cys273, which forms part of the catalytic triad (Adamczyk *et al*. 2013). As such each assay monitors the binding of the first substrate (isopeptidase activity assay) and second substrate (MDC assay) (Fig. 1.4). The activity of the TG6 mutants differed with respect to the activity assay used when compared with WT TG6 (Fig. 5.7). TG6 R111C and L517W had increased isopeptidase activity compared with the transamidase activity. However, L517W is likely to be inaccurate as the measured fluorescence was minimal. The remaining TG6 mutants had greater transamidation activity for which the highest difference was observed with TG6 D327G. TG6 D327G showed transamidation activity was significantly higher (difference of 22.3%) than isopeptidase activity. This suggests the mutants have altered substrate selectivity and/or have a preference for one enzymatic function over the other. Transamidase activity was also investigated in HEK293T cell lysates using a biotinylated glutamine substrate. This showed similar results to the MDC assay. For example, the

change in activity was as follows: TG6 L517W< R111C< E574del <Q181H < D327G <L502Q <L502Q Q652dup < WT. The main difference was in the activity of TG6 Q181H and D327G for example in the MDC assay TG6 Q181H had higher activity than TG6 D327G and in the transamidase activity from the cell lysate, the opposite was true. In combination, these results showed those TG6 mutants predicted to be misfolded (TG6 R111C, Q181H, L517W and E574del) had significantly reduced activity when compared with WT. This decrease in function is likely due structural consequences but this requires confirmation by structural analysis. Interestingly, D327G which is involved in the coordination of Ca^{2+} at Ca^{2+} binding site 3, was active in both MDC assay but had limited isopeptidase activity, which could be important in disease progression.

Table 5.9: Comparison of Ca^{2+} and GTP binding.

TG6 mutant	Ca^{2+} binding	GTP binding	Ca^{2+} binding	GTP binding
	Biochemistry		Bioinformatics	
R111C	Increased	Reduced	No impact	No impact
Q181H	Reduced	No change	No impact	No impact
D327G	Reduced	Reduced	Probable impact	Probable impact
L502Q	No change	No change	No impact	Probable impact
L502Q Q652dup	Reduced	Increased	Probable impact	Probable impact
E574del	Reduced	Increased	No impact	Probable impact

The regulation of isopeptidase activity was assessed by titration of Ca^{2+} or GTP into the reaction. TG6 D327G, Q181H, L502Q Q652dup and E574del had reduced affinity for Ca^{2+} when compared to WT TG6 and TG6 R111C had increased affinity for Ca^{2+} (Table 5.7, 5.9 and Fig. 5.8). However, D327G and L502Q Q652dup were predicted by homology modelling methods to have an impact on Ca^{2+} binding (Table 5.9). Furthermore, the GTP binding was reduced for TG6 R111C and D327G, whereas increased affinity was measured for L502Q Q652dup and E574del (Table 5.9). The impact of the mutations on GTP binding for TG6 D327G, L502Q, L502Q Q652dup and E574del consistent with the homology modelling results. TG6 L517W had too little activity in the isopeptidase to measure the GTP inhibition and isopeptidase activity could not be activated in excessive Ca^{2+} concentrations (32 mM). It is unclear on this occasion if the lack of signal is due to the lack of function from structural implications caused by

the mutation, as predicted by bioinformatics and in existing literature or due to the low concentrations purified (Table 5.6) and/or the contamination of proteins in the sample (Fig. 5.6). For the other TG6 mutants, the altered affinities for the Ca^{2+} and GTP in the TG6 mutants can indicate altered regulation of activity for the TG6 mutants.

To further identify the altered allosteric regulation of TG6, a method of directly measuring the binding of GTP and Ca^{2+} could increase the sensitivity and accuracy of the measured affinities which in turn can help to differentiate the different mechanisms of SCA35 pathogenesis. These additional experiments could support the data presented here (Fig. 5.8). GTP binding can be measured using BODIPY GTP/ GTPγS (as described in Chapter 2). However, the affinity of GTP for TG6 is much lower than TG2 (200 μM compared with 2 μM) and therefore higher concentrations of BODIPY GTP is required. Furthermore, when preliminary investigations were carried out, the signal to noise ratio was much higher and little signal was observed at the concentrations used. Alternatively, surface plasmon resonance (SPR) can accurately measure binding affinities, this would be suitable for TG6 as only low amounts of protein (50 nM) are required (Applications reviewed here (Nguyen *et al.* 2015)). In addition, alternative ways to investigate Ca^{2+} binding will allow more accurate affinities to be calculated. This would aid in understanding if Ca^{2+} binding is directly affected or the activation is affected for other reasons. This can measure the different affinities of the three Ca^{2+} binding sites and the cooperative binding nature of the Ca^{2+} binding in TG's, which was absent when measuring isopeptidase activation. This has previously been carried out using ITC with TG2 and TG3 (Ahvazi *et al.* 2003; Ahvazi *et al.* 2004a; Begg *et al.* 2006a; Begg *et al.* 2006b; Kiraly *et al.* 2009) but requires large amount of protein (applications reviewed here (Falconer 2016)). ITC can monitor multiple binding events and would be suitable for measuring the binding of Ca^{2+} at site 2 and 3. Ca^{2+} binding at site 1 is important in the stabilisation of the structure and therefore presents more of a challenge.

In addition, to identify if the TG6 mutants shows further preference to selective substrates or preference toward one enzymatic function, the deamidation activity of WT TG6 and the TG6 mutants should be assessed. An assay to determine deamidation activity has been developed in collaboration with Zedira. However, only preliminary data is available and the assay requires optimisation to identify the accuracy of the results. Moreover, it is still unknown if TG6 has GTPase activity. This function could be important in differentiating the different mechanisms between those mutants capable of the binding GTP. The

GTPase activity could be investigated with the GTPase Glo assay optimised in Chapter 2. However, this assay showed unexpected results with TG2 which are yet to be understood.

5.4.3 Conclusion

The novel data presented in this chapter has shown the successful production of the TG6 mutant baculovirus and the following expression and purification. Differences in the TG6 mutants stability was apparent and needs to be investigated further. Structural analysis is required to confirm selected TG6 mutants are misfolded and also to indicate whether they have been correctly grouped by the bioinformatic analysis. The TG6 mutants also have varied function and regulation which supports bioinformatics data in most cases. To fully identify the different mechanisms of neuronal cell death and assess the accuracy of the bioinformatic data further mutants need to be screened.

CHAPTER 6

6. GENERAL DISCUSSION

The work carried out here has investigated the mechanist aspects of TG2 externalisation to identify the role of GTP, development of an *E. coli* expression system for TG6, the development of a bioinformatic method to assess the impact of TG6 mutants and the biochemical characterisation of WT TG6 and TG6 mutants.

6.1 TG2 externalisation

Here, the role of GTP in the externalisation of TG2 was assessed. This work was completed with two aims in mind. 1) To assess the role of GTP in the secretion of TG2. 2) To develop assays to assess the regulation of other TG's and TG mutants by GTP. This work was in aid of supporting one or both of the two hypothetical mechanisms formed by our group. The first was, GTP induces the closed compact conformation which is required for its secretion. The second hypothesis is GTP hydrolysis is required to generate energy required for secretion by a phospho-transfer mechanism. Previous work (Fig. 2.3, R. Griffiths, 2017) detected the presence of WT, K173N and to a lesser degree K173L in the conditioned media. No externalisation was detected for K173N/F174D and R580A suggesting they either are not in the correct conformation for externalisation or are unable to hydrolyse GTP and thus generate the energy required for externalisation.

The TG2 GTP binding site mutants (K173N, K173L, K173N/F174D and R580A) were used to probe the mechanistic role of GTP in the context of conformation and hydrolysis. The conformation of the TG2 GTP binding mutants was assessed using CD spectroscopy (with and without GTP), size exclusion chromatography and native PAGE. The CD spectroscopy data presented here indicates WT TG2 and the TG2 GTP binding mutants adopt a conformation between open and closed conformation in the absence of GTP. WT, K173N and K173L represent nucleotide bound TG2 (GTP or ATP), TG2 K173N/F174D and R580A closely resemble open TG2 (Table 2.4). In agreement, native PAGE in the presence of GTP showed K173N/F174D and R580A were unable to adopt closed compact conformation (Fig. 2.11). Previously, TG2 R580A was shown using different methods e.g. sensitivity to trypsin and μ-calpain, native PAGE and sedimentation velocity experiments to have altered conformation but no published data is available for K173N/F174D for comparison (Begg *et al.* 2006a).

GTP binding experiments previously carried out for TG2 K173N, K173L and R580A (summarised in Table 6.1, (Iismaa *et al.* 2000; Begg *et al.* 2006a; Begg *et al.* 2006b; Datta *et al.* 2007; Ruan *et al.* 2008; Singh *et al.* 2016) were consistent with data presented here (GTP inhibition, BOIDPY GTP binding and native PAGE, Fig. 2.10, 2.11 and 2.17 and R. Griffiths 2017) and confirm R580A is unable to bind GTP and assume compact conformation, whilst TG2 K173N and K173L can. Differences in the binding of GTP is observed when introducing the mutants into TG2•GTP structure (Fig. 2.4). TG2 K173N/F174D showed a loss of hydrophobic contacts between F174 and the guanine ring of GTP and instead N173 and D174 interacts with the hydroxyl group on guanine. This mutation is predicted to have lower binding due to loss of hydrophobic stacking. R580A showed a loss of three interactions with the α and β phosphates, predicting a loss of affinity of TG2 for GTP. This is consistent with data obtained with respected to GTP binding.

GTPγS binding was measured for WT TG2 and all the TG2 GTP binding mutants (all above FXIIIA) which contradicts previously published data, where GTPγS binding was abolished/reduced for R580A (Fig. 2.16 and 2.17 and Table 6.1). Inconsistency between GTP and GTPγS binding and affinity when investigating the GTP binding mutants has not previously been measured suggesting different nucleotide binding modes. The different interactions in nucleotide (GTP and GTPγS) stabilisation are detailed in Fig. 2.20 and Table 2.9, with respect to WT TG2. Consistent with the hypothesis that GTPγS has different binding modes molecular docking experiments showed GTPγS adopts two distinct conformations. This could explain the TG2 mutant GTPγS binding capabilities. It is evident GTP and GTPγS cannot be interchangeably used when assessing GTP binding, but more investigation is required.

Furthermore, the role of GTP hydrolysis was assessed using two assays: BODIPY GTP turnover and GTPase-Glo™ assay. The BODIPY GTP turnover is reliant on BODIPY GTP binding and is therefore bias and showed similar results to BODIPY GTP binding (Fig. 2.22). In contrast, the GTPase Glo™ assay showed GTP hydrolysis rates R580A> K173N/F174D> K173L> WT> K173N (Fig. 2.27). This does not support the GTP binding/ hydrolysis data presented here or published in the literature (Iismaa *et al.* 2000; Begg *et al.* 2006a; Begg *et al.* 2006b). GTP hydrolysis for TG2 and the TG2 GTP binding mutants has previously been measured using the charcoal method (release of the γ-phosphate (P^{32}) measured by scintillation counting). This method has not been used to

assess GTP hydrolysis by K173N/F174D and R580A, however TG2 R580K and R580L showed high GTP hydrolysis activity when compared with WT TG2, using this method (Datta *et al.* 2007). TG2 R580K and R580L also had no detectable GTP binding (Begg *et al.* 2006a; Begg *et al.* 2006b; Datta *et al.* 2007), Table 6.1). This is unexpected result is also seen with TG2 K173N/F174D and R580A.

Table 6.1: Summary of GTP binding/ hydrolysis of TG2 mutants in literature.

Method	Nucleotide	Mutant	Binding Result	Ref
GTP photolabelling	GTP	K173N K173L R580A R580K/L	Yes Yes No No	(Iismaa *et al.* 2000; Begg *et al.* 2006b; Datta *et al.* 2007; Ruan *et al.* 2008)
GTPγS filter binding assays	GTPγS	K173N K173L R580A	No No N/a	(Iismaa *et al.* 2000)
GTP inhibition of TGase activity	GTPγS	K173N K173L R580A (R579A) R580K/L	Reduced Reduced No No	(Iismaa *et al.* 2000; Begg *et al.* 2006b; Datta *et al.* 2007)
GTPase (charcoal method)	GTP	K173N K173L R580K/L	Reduced No Yes	(Iismaa *et al.* 2000; Datta *et al.* 2007)
ITC	GTPγS	K173N K173L R580A (R579A)	N/a N/a Reduced	(Begg *et al.* 2006b)
Native PAGE	GTP	K173N K173L R580A (R579A) R580K	N/a N/a No No	(Begg *et al.* 2006a)
BODIPY GTP binding	GTPγS	K173N K173L R580K/L	N/a Yes No	(Datta *et al.* 2007; Singh *et al.* 2016)

In further support, previous data has shown the catalytic core of TG2 alone is capable of hydrolysing GTP at a higher rate when compared with FL TG2 (Iismaa *et al.* 1997). This suggests the closed conformation is not required for GTPase activity. As both K173N/ F173D and R580A have altered conformation this can explain the higher GTPase activity of these mutants compared to WT TG2.

In addition, I have shown the binding affinity of ATP is unexpectedly not affected by the TG2 GTP binding mutants (K173N and K173N/F173D). TG2 K173N/F174D was

insensitive to inhibition by GTP (Fig. 2.10), whilst inhibition by ATP is comparable to WT TG2 (Fig. 2.28). The TG2•ATP structure gave no explanation as to why the ATP affinity remains the same and shows ATP binds in the nucleotide pocket similarly to GTP (Fig. 2.20 B(i) and Table 2.9) (Han *et al.* 2010). This suggests K173 and F174 are not essential for ATP binding but are important for GTP binding. It has previously been suggested there are two different pockets for binding of ATP and GTP but this idea was rejected following the publishing of the solved TG2•ATP structure but this could explain the lack of affinity change (Lai *et al.* 1998).

Together, this data suggests 1) The conformation of TG2 is important in TG2 export. WT TG2, K173N and K173L adopt the closed compact conformation and are externalised from the cell. Whereas both K173N/F174D and R580A are not exported and have been identified in an alternative conformation. 2) GTP hydrolysis is not involved in TG2 export. In our hands both TG2 K173N/F174D and R580A are capable of GTP hydrolysis at greater rates to WT TG2 and are not exported from the cell. This suggesting GTP hydrolysis is not required. However, as all the TG2 GTP binding mutants are capable of GTP hydrolysis it is impossible to rule out its involvement in the TG2 secretion mechanism. 3) Measurable GTP binding and the compact conformation is not essential for GTP hydrolysis. GTP hydrolysis was measured for all GTP binding mutants irrespective of their adopted conformation. With regards to the BODIPY GTP turnover assay the hydrolysis activity could be too fast to be captured for TG2 K173N/F174D and R580A.

With the conformation of TG2 being the most likely factor influencing TG2 export, further investigation into this aspect of the mechanism of TG2 secretion is required. This in turn can lead to therapeutically targeting this pathway to aid in the treatment of related pathological conditions. These results have shown the developed GTP binding and GTP hydrolysis assays are capable of measuring small changes in the conformation, enzymatic function and regulation of TG2, using the TG2 GTP binding mutants. This allows other TG's to be assessed in the same way to measure changes caused by pathogenic mutation such as those found in TG6.

6.2 Expression of TG6 in *E.coli*

I have developed and optimised and *E.coli* expression system for TG6 expression and purification and have investigated the introduction of proteins tags which aid in the

solubility, folding and expression of TG6. The protein tags MBP, NusA and Trx have been previously shown to aid in the folding and solubility of their fused proteins and SUMO-intein aids in protein expression (Terpe 2003; Nallamsetty and Waugh 2006). Despite attempts to optimise expression and purification of the TG6 fusion proteins, contaminating proteins and cleavage products were consistently observed in the purified sample. Data showed the FL TG6 fusion proteins were unstable and cleavage products were detected despite the presence of protease inhibitors (Fig. 3.7). Furthermore, the relative activity of the TG6 fusion proteins was low when compared with baculovirus/*Sf*9 cell generated TG6 (Fig. 3.11). It was evident the expression in *E.coli* system was not capable of expressing intact TG6 which was correctly folded and consequently, the *E.coli* system was abandoned. Subsequent generation of recombinant TG6 was generated using the baculovirus/*Sf9* cell system.

6.3 Biochemical analysis of WT TG6

As biochemical data on purified WT TG6 is limited, first work was carried out to biochemically assess WT TG6. I have shown TG6 can be successfully expressed and purified in the baculovirus/ *Sf*9 cell expression system, with minor losses (Fig. 3.12). WT TG6 showed dose-dependent activation with Ca^{2+} and dose dependent inhibition with GTP and binding constants (K_D) were determined. This showed WT TG6 has similar affinity for Ca^{2+} and 10x lower affinity for GTP when compared with WT TG2 (Fig. 3.14), which is consistent with published data (Thomas *et al.* 2013). Furthermore, these experiments provided a basis on which to test the identified TG6 mutants.

TG6 has previously been shown as stable at 4 °C over seven days (D. Aeschlimann, personal communication). This was consistent with data presented here when samples were measured post-centrifugation (Fig. 3.14A and B). In contrast, there is progressive aggregation with no centrifugation suggesting TG6 stability is hugely decreased when precipitates are not removed regularly (Fig. 3.14C and D). When in a cellular environment, TG6 is stable (Fig. 5.11) which indicates TG6 is only unstable when isolated and therefore may require intracellular chaperones to maintain its structure and solubility (Martinez-Torrecuadrada *et al.* 2005; Shuo-shuo *et al.* 2011). When purified TG2 (expressed in *E. coli)* and TG3 (express in baculovirus/ *Sf*9 expression system) are stable, which highlights differences between the transglutaminases and implies TG6 may be processed differently and could harbour post-translational modifications processed differently by *Sf9* cells (Marchal *et al.* 2001). However, no post-translational

modifications such as glycosylation are yet to be detected for TG6 (D. Aeschlimann, personal communication). Structural analysis of WT TG6 has previously been carried out using size exclusion chromatography and CD spectroscopy (D. Aeschlimann personal communication), however when repeated on this occasion no meaningful data was obtained.

6.4 Bioinformatic method to predict impact of TG6 mutants

Here, I have developed a novel bioinformatic method for predicting the impact of TG6 mutation, using a combination of tools. The TG6 mutants fell into five main groups (structural, functional, structural and functional, benign and unknown) (Table. 4.24). This showed despite extensive bioinformatic analysis the impact of one TG6 mutant (D510H) was unable to be predicted. Mutations showing decreased stability (FoldX) and solubility (A3D) were predicted to have structural consequences, whereas mutations that showed alteration in the functional binding sites (homology modelling) were predicted to impact function. Mutations with decreased stability and solubility and altered interactions in the functional binding sites were characterised as 'structural and function' consequence. Benign mutations showed similar results to WT TG6.

Bioinformatic approaches have been used to assess mutations in TG2 (Thangaraju *et al.* 2017). This predominantly involved the use of sequence based tools but homology models were used to assess the stability of TG2 and investigate the position of the mutation. Furthermore, the published TG2 mutations were biochemically assessed via their isopeptidase and transamidation activity, GTP binding and fibronectin binding to support bioinformatic predictions. In this case, no structural analysis was carried out for these TG2 mutants. This is similar to the method used to the assess the TG6 mutants (Chapter 4 and 5).

A similar approach has been used to predict the outcome of clinical mutations in various diseases such as cancer, primary congenital glaucoma, hearing loss, cystic fibrosis (Dorfman *et al.* 2010; Yilmaz 2015; Firasat *et al.* 2018). In these studies they were unable to rely on computational methods to aid in predicting the pathogenicity of mutations. The most commonly tools used are: SIFT, PolyPhen-2 and PANTHER (Thomas *et al.* 2003; Adzhubei *et al.* 2010; Sim *et al.* 2012). Here, I have found SIFT and PolyPhen-2 are not sufficient to predict pathogenicity or predict the impact of the mutation and analysis with PANTHER was not included in my bioinformatic method. It is apparent predictions made

based on the homology modelling alone are considered more informative, depending on the validity of the model (Pang *et al.* 2013; Fazel-Najafabadi *et al.* 2015; Badior *et al.* 2017; Dong *et al.* 2017). In the case of TG6 there is no structure available and as a result the results obtained from homology modelling have their limitations. None of the generated homology models showed high validity and quality but this could be due to the low sequence homology when compared with TG2 and TG3 (Table 4.1). Consequently, a combination of sequence based tools and 3D tools such as homology modelling is most informative. Supporting evidence from biochemical studies for of all the identified TG6 mutants allowed the developed a bioinformatic method to be tested and compared.

This method can be used to assess additional TG6 mutants and mutations in other proteins to aid understanding of the mechanism of genetic variants in disease pathogenesis.

6.5 Biochemical characterisation of TG6 mutants

Following the bioinformatic predictions, TG6 mutants were generated for biochemically analysis using the baculovirus/*Sf*9 cell expression system. Here I show analysis of at least two mutants from each of the predicted groups (structural, functional, structural and functional, Table 6.2), which has not previously been carried out. However, the unknown (D510H) and benign (E406K) groups were not represented as they were not generated at the time.

TG6 R111C, Q181H, L517W and E574del, were unstable during purification indicating these mutations impact protein architecture (Table 5.6). Furthermore, these mutations showed the greatest functional deficits (except Q181H) which is likely related to misfolding (Fig. 5.7). In addition functional deficits were measure for all remaining mutants when compared with WT TG6. TG6 D327G, L502Q and L502Q Q652dup had impaired activity but were stably expressed and purified indicating a functional impact (Fig. 5.7 and Table 5.6). Furthermore differences in activity between isopeptidase and transamidation was apparent for the TG6 mutants when compared with WT. This suggests possible changes in substrate specificity due to local changes in the substrate binding pockets despite the mutation not being located nearby. For example the isopeptidase activity or TG6 D327G is barely detectable (Fig 5.7A and B), whilst transamidation activity is measurable (Fig. 5.7C and D). The difference between isopeptidase and transamidation activity was also observed in Chapter 2 with TG2 K173L and R580A (Fig. 2.6). This indicates changes in substrate selectivity are not TG6 specific.

Furthermore, this suggests altered substrate selectivity could be a possible pathogenic pathway caused by TG6 mutants.

Data regarding the regulation of TG6 by GTP and Ca^{2+} identified at least three possible mechanisms. Where by TG6 R111C could be constitutively active due to insensitivity to GTP, whilst D327G is insensitive to physiological levels of GTP and Ca^{2+} indicating under resting cellular conditions it is inactive. L502Q Q652dup and E574del have enhanced GTP binding, which could indicate these TG6 mutants are inactive under physiological conditions but could have higher GTPase activity, but GTPase activity has not yet been identified for TG6. TG6 L502Q behaved similar to WT TG6 and therefore could be benign or function via an unknown mechanism.

Table 6.2: Grouped TG6 mutants.

TG6	Group	Progress
R111C	Structural	Protein
Q181H	Structural	Protein
D327G	Functional	Protein
V391M	Functional	Not generated
E406K	Benign	Not generated
R412C	Structural and functional	Not generated
Y441C	Structural	Bacmid
R448W	Structural	Not generated
L502Q	Functional	Protein
L502Q Q652dup	Structural and functional	Protein
D510H	Unknown	Bacmid
L517W	Structural and functional	Protein
E574del	Structural and functional	Protein
Q652dup	Structural and functional	Bacmid

This data is further supported by work carried out by Tripathy *et al.* which showed the TG6 mutants induce neuronal cell death through distinct mechanisms. The main mechanism investigated was the induction of the UPR consistent with abnormal subcellular localisation and low transamidase activity (Fig. 5.9 and 10). However, as all the mutants did not abide by this criteria, it suggests an alternative mechanism can be undertaken. Those mutants suggested to be targeted to the proteasome for degradation, via the UPR, are: TG6 R111C, Q181H, Y441C, L517W and E574del. TG6 V319M and D327G are localised to the nucleus but have reduced transamidase activity suggesting alternative mechanisms. Only R111C (structurally compromised) and D327G

(functionally compromised) was analysed in the context of the UPR, but the other TG6 mutants require analysis to determine if they are targeted to the UPR.

At this time it is difficult to comment on the accuracy of the bioinformatic predictions without the structural characterisation of the TG6 mutants and the analysis of more TG6 mutants. However, from the biochemical experiments carried out, five possible pathogenic mechanisms for the TG6 mutants can be interpreted from this set of mutants. 1) protein misfolding 2) altered substrate selectivity 3) inactive TG6 4) constitutive activity and 5) enhanced GTP hydrolysis. This data shows the grouping of the TG6 mutants is not straightforward and in fact covers multiple mechanisms to cause neuronal cell death. Further elucidating the pathways involved, can highlight possible therapeutic targets to prevent pathogenesis in SCA35 patients.

In addition, this may be relevant for other neurological diseases which involve *TGM6* mutations and this work can aid in their treatment. For example episodic ataxia and other sporadic ataxia groups with unknown diagnosis, PD and TG6 L517W has been identified in AML (Wang *et al.* 2010; Fogel *et al.* 2012; Pan *et al.* 2015; Westenberger *et al.* 2015; Choi *et al.* 2017). Furthermore, TG6 antibodies have been identified in GA and MS (Hadjivassiliou *et al.* 2008a; Cristofanilli *et al.* 2017). The biochemical work here carried out to further understand of WT TG6 can aid in investigating how TG6 positive cells are recognised by the immune system.

6.6 Additional identified TG6 mutations

A single nucleotide polymorphism in *TGM6*, in an intronic region between exon 5 and 6, was recently identified as a susceptible loci for Tuberculosis (TB) by genome wide sequencing (Zheng *et al.* 2018). TB patients had significantly reduced *TGM6* gene expression in peripheral blood mononuclear cells compared with healthy controls. Tgm6-deficient mice (CRISPR/Cas-9 mediated) showed greater infiltration of immune cells and tissue damage was observed with homozygotes (Zheng *et al.* 2018). Also, *TGM6* was identified as biomarker of colon cancer metastasis and with a panel other genes which predicts poor prognosis (Zhou *et al.* 2019). TG2 has been extensively linked to cancer and therefore it is unsurprising TG6 may also have a role (Eckert *et al.* 2014). This further suggests the role of TG6 in the pathogenesis of a number of different diseases and emphasises the importance of this work.

Furthermore, a recent paper suggests the prevalence of *TGM6* mutations in autosomal dominant SCA is inflated and indicates a high risk of misdiagnosis of patients harbouring *TGM6* mutations (Fung *et al.* 2019). Four previously published TG6 mutants (Table 6.3) and four novel mutations were identified in (R211C, P347L, A385V and F670S) in this disease cohort (University of Hong Kong). Furthermore, the mutation TG6 S39C which was identified in a family based exome sequencing study to identify undiagnosed conditions, was also identified here (Farwell *et al.* 2015). It was suggested exon 6 (codes for part of the catalytic core) is a hotspot for *TGM6* mutations associated with SCA35 and have a calculated inflation of up to 111 fold (Fung *et al.* 2019). Despite the suggestion *TGM6* mutations are inflated in the diagnosis of SCA35, it is evident from the data presented here and other published data the *TGM6* mutations play a role in causing disease, which may not be explicitly related to SCA35 (Table 6.3).

Table 6.3: Identification of known TG6 mutations in exome screening cohort.

Mutation	Disease
c.7+1G>T	Congenital myopathy
V391M	Mitochondrial disorder Maturity onset diabetes of the young Noonan syndrome Infantile hypertrophic cardiomyopathy
D510H	Dystonia Developmental delay
L517W	Spastic paraplegia Febrile convulsions

When analysing the phenotype/ genotype relationship of the TG6 mutants (Table 1.5 and Table 4.3), it is not apparent that mutations in any particular part domain is more detrimental then others. However, earlier onset of the disease is generally observed in the *N* terminal end of the protein compared with the *C* terminal, with R412C having the earliest mortality (patient aged 9). Common symptoms are seen with all mutants e.g. Dysarthria, gait, tremor, saccade slowing, myoclonus and cerebellar ataxia. Whilst other symptoms such as epilepsy Ocular dysmetria, Pseudobullar palsy and nystagmus are seen with certain mutations this is also not domain specific within the TG6 protein. As a result it is difficult to confirm the presence of a SCA35 ‘hotspot’ as identified by Fung et al. 2019.

6.7 Future work

6.6.1 Stabilisation of TG6 and TG6 mutants

An important issue with expressing and purifying TG6, was its stability and solubility. Attempts were made to address stability by introducing fusion tags (e.g. Trx, MBP and NusA) to TG6 as shown in Chapter 3 but were unsuccessful in the in an *E. coli* expression system (Terpe 2003; Nallamsetty and Waugh 2006; Costa *et al.* 2013). An alternative approach would include the co-expression of molecular chaperones which will aid in protein folding (Shuo-shuo *et al.* 2011).

In addition, the polymorphic state of TG6 could be investigated to identify if the 'aggregates' observed are due to misfolding during the expression of TG6 or if TG6 is more stable when polymorphic. This in turn can identify whether its polymeric state is in dynamic equilibrium with the monomer. Furthermore, the ideal storage conditions need to be established to allow sufficient time for enough protein to be generated for crystallisation trials. Storage at -20 °C was not sufficient to prevent aggregation and lyophilisation results in losses of TG6. High concentrations (<3.5 mg/mL) are required for crystallisation so improving stability and storage would aid in this process. Crystallisation trays should also be set up immediately following large scale purification to reduce the risk of aggregation if no storage conditions are suitable. Consequently, it may be worthwhile to streamline the process by automation such as an Akta system instead of centrifugation washes. If an X-ray crystal structure can be obtained, then the bioinformatic method can be validated and the method accuracy can be improved.

6.6.2 Structural analysis of TG6 mutants

To identify different mechanisms used by the TG6 mutants to induce neuronal cells death, structural analysis of TG6 is required. This would also support the bioinformatic data collected on the TG6 mutants to confirm their accuracy. Both analytic size exclusion chromatography and circular dichroism could give an indication into whether the TG6 mutants are correctly folded. This has previously been carried out for WT TG6 by our lab and therefore similar conditions can be used to for the TG6 mutants. Again, this may need to be carried out immediately after purification to give an accurate presentation of protein conformation without storage.

6.6.3 Additional functional analysis

The functional analysis carried here out needs repeated with a new batch of TG6 mutant proteins to ensure the data is accurate. Furthermore, GTP binding and Ca^{2+} binding needs to be assessed directly. For example BODIPY GTP/GTPγS or ITC can be used to assess GTP binding. Ca^{2+} binding had previously been measured for TG3 and TG2 by equilibrium dialysis and ITC. These methods could be adopted to measure Ca^{2+} binding for WT TG6 and the TG6 mutants (Bergamini *et al.* 1987; Iismaa *et al.* 2000; Ahvazi *et al.* 2003). In addition, biochemical analysis of the preliminary TG6 mutants set needs to be completed, this includes the analysis of TG6 E406K and D510H. This will allow assessment of the accuracy of the grouping predicted by the bioinformatic method.

6.6.4 Substrate selectivity

As both transamidation and isopeptidase activity have been measured, it is important to identify if other enzymatic functions are compromised by the TG6 mutants. This would detect if there are changes in the substrate binding sites, which lead to altered substrate selectivity. Assays to establish the deamidation activity of TG6 and the TG6 mutants could show mutants have a preference various substrates which may be important in identifying the mechanism by which the initiate neuronal cell death. Furthermore, the developed assay can be used to assess the deamidation activity of TG2. This could be relevant information for the role of both TG2 and TG6 in gluten-related diseases and could be further investigated with peptide docking experiments to probe both the lysine and glutamine substrate pockets.

6.6.5 GTPase activity of TG6

GTPase activity has been measured for TG2 and TG3 but it is still unclear if TG6 has GTPase activity (Bergamini *et al.* 1987; Iismaa *et al.* 2000; Ahvazi *et al.* 2004a). This can be tested using the GTPase Glo™ assay, where the incubation times may need to be extended to compensate for reduced affinity. TG2 has been shown to have a role in signal transduction but a physiological role of the GTPase activity of TG3 is yet to be established (Iismaa *et al.* 2000). This is relevant if TG6 has GTPase activity and could be important for its role in the nervous system. As a result it could be another mechanism by which the TG6 mutants cause neuronal cell death.

6.6.6 Externalisation of TG6 and TG6 mutants

As TG6 is capable of binding GTP, it may be externalised from the cells. It would be of interest to use the existing system for TG2 to identify if WT TG6 is externalised and it would also help to identify if the TG6 mutants are also deficient in externalisation (Griffiths, 2017). This could also be relevant the role of TG6 in autoimmune diseases such as GA and will aid in understanding the capabilities of TG6.

6.7 Conclusion

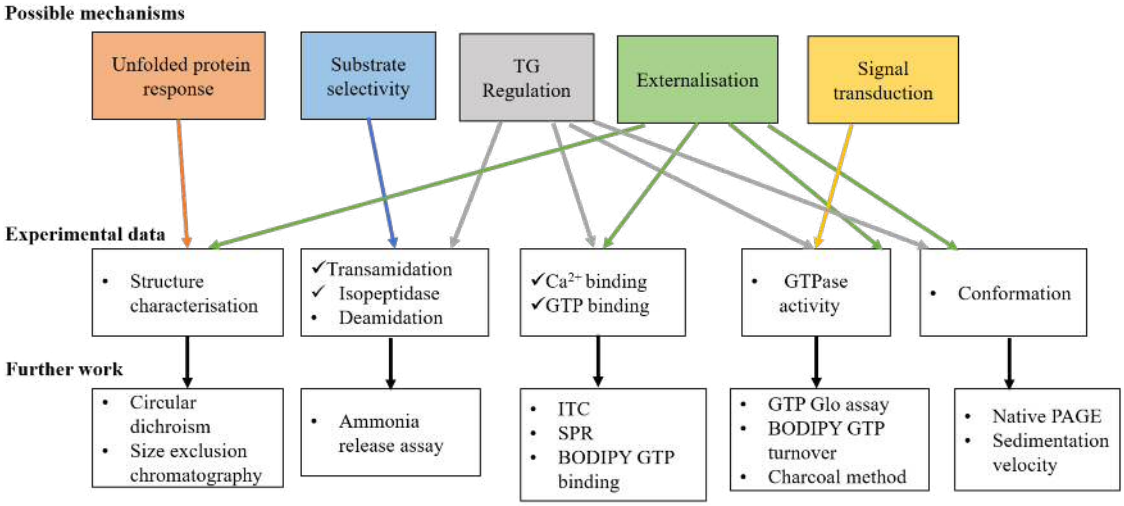


Figure 6.1: Summary of TG6 mutant mechanisms.
Summary of data collected on the TG6 mutants and the possible mechanism involved in neuronal cell death. Further work required has also been summarised.

In combination, this would allow us to identify the TG6 mutants which have impacted the structure and function of TG6. The different possible mechanism by which the TG6 mutants can cause neuronal cell death have been summarised in Fig. 6.1. Preliminary data has been collected to start differentiating the possible mechanisms of TG6 mutants however further work is required to identify exactly the mechanisms involved and in turn can be of therapeutic value. Furthermore, this data will aid in testing the accuracy of bioinformatic predictions which can be used in combination or in isolation when determining the impact of novel TG6 mutant.

7. REFERENCES

Achyuthan, K. E. and Greenberg, C. S. (1987). Identification of a guanosine triphosphate-binding site on guinea pig liver transglutaminase. Role of GTP and calcium ions in modulating activity. *J Biol Chem* **262**(4):1901-1906.

Adamczyk, M., (2013). Biomarkers for arthritis: regulation of extracellular transglutaminase activity by non-conventional export. *PhD Thesis*, Cardiff University.

Adamczyk, M., Heil, A. and Aeschlimann, D. (2013). Real-time fluorescence assay for monitoring transglutaminase activity. *BMG LABTECH.* Application note 234

Adamczyk, M., Griffiths, R., Dewitt, S., Knauper, V. and Aeschlimann, D. (2015). P2X7 receptor activation regulates rapid unconventional export of transglutaminase-2. *J Cell Sci* **128**(24):4615-4628.

Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., . . . Sunyaev, S. R. (2010). A method and server for predicting damaging missense mutations. In: *Nat Methods*. Vol. 7. United States, pp. 248-249.

Aeschlimann, D. and Knauper, V. (2017). P2X7 receptor-mediated TG2 externalization: a link to inflammatory arthritis? *Amino Acids* **49**(3):453-460.

Aeschlimann, D., Koeller, M. K., Allen-Hoffmann, B. L. and Mosher, D. F. (1998). Isolation of a cDNA encoding a novel member of the transglutaminase gene family from human keratinocytes. Detection and identification of transglutaminase gene products based on reverse transcription-polymerase chain reaction with degenerate primers. *J Biol Chem* **273**(6):3452-3460.

Aeschlimann, D. and Paulsson, M. (1994). Transglutaminases: protein cross-linking enzymes in tissues and body fluids. *Thromb Haemost* **71**(4):402-415.

Aeschlimann, D. and Thomazy, V. (2000). Protein crosslinking in assembly and remodelling of extracellular matrices: the role of transglutaminases. *Connect Tissue Res* **41**(1):1-27.

Ahvazi, B., Boeshans, K. M., Idler, W., Baxa, U. and Steinert, P. M. (2003). Roles of calcium ions in the activation and activity of the transglutaminase 3 enzyme. *J Biol Chem* **278**(26):23834-23841.

Ahvazi, B., Boeshans, K. M., Idler, W., Baxa, U., Steinert, P. M. and Rastinejad, F. (2004a). Structural basis for the coordinated regulation of transglutaminase 3 by guanine nucleotides and calcium/magnesium. *J Biol Chem* **279**(8):7180-7192.

Ahvazi, B., Boeshans, K. M. and Steinert, P. M. (2004b). Crystal structure of transglutaminase 3 in complex with GMP: structural basis for nucleotide specificity. *J Biol Chem* **279**(25):26716-26725.

Ahvazi, B., Kim, H. C., Kee, S. H., Nemes, Z. and Steinert, P. M. (2002). Three-dimensional structure of the human transglutaminase 3 enzyme: binding of calcium ions changes structure for activation. *Embo j* **21**(9):2055-2067.

Ahvazi, B. and Steinert, P. M. (2003). A model for the reaction mechanism of the transglutaminase 3 enzyme. *Exp Mol Med* **35**(4):228-242.

Anwar, R., Stewart, A. D., Miloszewski, K. J., Losowsky, M. S. and Markham, A. F. (1995). Molecular basis of inherited factor XIII deficiency: identification of multiple mutations provides insights into protein function. *Br J Haematol* **91**(3):728-735.

Arentz-Hansen, H., Korner, R., Molberg, O., Quarsten, H., Vader, W., Kooy, Y. M., . . . McAdam, S. N. (2000). The intestinal T cell response to alpha-gliadin in adult celiac disease is focused on a single deamidated glutamine targeted by tissue transglutaminase. *J Exp Med* **191**(4):603-612.

Badior, K. E., Alka, K. and Casey, J. R. (2017). SLC4A11 Three-Dimensional Homology Model Rationalizes Corneal Dystrophy-Causing Mutations. *Hum Mutat* **38**(3):279-288.

Baldwin, R. L. (1986). Temperature dependence of the hydrophobic interaction in protein folding. *Proc Natl Acad Sci U S A* **83**(21):8069-8072.

Balklava, Z., Verderio, E., Collighan, R., Gross, S., Adams, J. and Griffin, M. (2002). Analysis of tissue transglutaminase function in the migration of Swiss 3T3 fibroblasts: the active-state conformation of the enzyme does not affect cell motility but is important for its secretion. *J Biol Chem* **277**(19):16567-16575.

Basmanav, F., Cau, L., Tafazzoli, A., Mechin, M. C., Wolf, S., Romano, M. T., . . . Betz, R. C. (2016). Mutations in Three Genes Encoding Proteins Involved in Hair Shaft Formation Cause Uncombable Hair Syndrome. *Am J Hum Genet* **99**(6):1292-1304.

Baumgartner, W., Golenhofen, N., Weth, A., Hiiragi, T., Saint, R., Griffin, M. and Drenckhahn, D. (2004). Role of transglutaminase 1 in stabilisation of intercellular junctions of the vascular endothelium. *Histochem Cell Biol* **122**(1):17-25.

Begg, G. E., Carrington, L., Stokes, P. H., Matthews, J. M., Wouters, M. A., Husain, A., . . . Graham, R. M. (2006a). Mechanism of allosteric regulation of transglutaminase 2 by GTP. *Proc Natl Acad Sci U S A* **103**(52):19683-19688.

Begg, G. E., Holman, S. R., Stokes, P. H., Matthews, J. M., Graham, R. M. and Iismaa, S. E. (2006b). Mutation of a critical arginine in the GTP-binding site of transglutaminase 2 disinhibits intracellular cross-linking activity. *J Biol Chem* **281**(18):12603-12609.

Beissbarth, T., Tye-Din, J. A., Smyth, G. K., Speed, T. P. and Anderson, R. P. (2005). A systematic approach for comprehensive T-cell epitope discovery using peptide libraries. *Bioinformatics* **21 Suppl 1**:i29-37.

Benkert, P., Biasini, M. and Schwede, T. (2011). Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics* **27**(3):343-350.

Bergamini, C. M., Signorini, M. and Poltronieri, L. (1987). Inhibition of erythrocyte transglutaminase by GTP. *Biochim Biophys Acta* **916**(1):149-151.

Bertoni, M., Kiefer, F., Biasini, M., Bordoli, L. and Schwede, T. (2017). Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology. *Sci Rep* **7**(1):10480.

Bienert, S., Waterhouse, A., de Beer, T. A., Tauriello, G., Studer, G., Bordoli, L. and Schwede, T. (2017). The SWISS-MODEL Repository-new features and functionality. *Nucleic Acids Res* **45**(D1):D313-d319.

Boscolo, S., Lorenzon, A., Sblattero, D., Florian, F., Stebel, M., Marzari, R., . . . Tongiorgi, E. (2010). Anti transglutaminase antibodies cause ataxia in mice. *PLoS One* **5**(3):e9698.

Bourne, H. R. (1997). G proteins. The arginine finger strikes again. *Nature* **389**(6652):673-674.

Candi, E., Oddi, S., Terrinoni, A., Paradisi, A., Ranalli, M., Finazzi-Agro, A. and Melino, G. (2001). Transglutaminase 5 cross-links loricrin, involucrin, and small proline-rich proteins in vitro. *J Biol Chem* **276**(37):35014-35023.

Candi, E., Paradisi, A., Terrinoni, A., Pietroni, V., Oddi, S., Cadot, B., . . . Melino, G. (2004). Transglutaminase 5 is regulated by guanine-adenine nucleotides. *Biochem J* **381**(Pt 1):313-319.

Carija, A., Navarro, S. and Ventura, S. (2016). Data on correlation between Abeta42 structural aggregation propensity and toxicity in bacteria. *Data Brief* **7**:143-147.

Carrio, M. M. and Villaverde, A. (2002). Construction and deconstruction of bacterial inclusion bodies. *J Biotechnol* **96**(1):3-12.

Cassidy, A. J., van Steensel, M. A., Steijlen, P. M., van Geel, M., van der Velden, J., Morley, S. M., . . . McLean, W. H. (2005). A homozygous missense mutation in TGM5 abolishes epidermal transglutaminase 5 activity and causes acral peeling skin syndrome. *Am J Hum Genet* **77**(6):909-917.

Chester, N. and Marshak, D. R. (1993). Dimethyl sulfoxide-mediated primer Tm reduction: a method for analyzing the role of renaturation temperature in the polymerase chain reaction. *Anal Biochem* **209**(2):284-290.

Cho, S. Y., Choi, K., Jeon, J. H., Kim, C. W., Shin, D. M., Lee, J. B., . . . Kim, I. G. (2010). Differential alternative splicing of human transglutaminase 4 in benign prostate hyperplasia and prostate cancer. *Exp Mol Med* **42**(4):310-318.

Choi, K. D., Kim, J. S., Kim, H. J., Jung, I., Jeong, S. H., Lee, S. H., . . . Choi, J. H. (2017). Genetic Variants Associated with Episodic Ataxia in Korea. *Sci Rep* **7**(1):13855.

Conchillo-Sole, O., de Groot, N. S., Aviles, F. X., Vendrell, J., Daura, X. and Ventura, S. (2007). AGGRESCAN: a server for the prediction and evaluation of "hot spots" of aggregation in polypeptides. *BMC Bioinformatics* **8**:65.

Costa, S. J., Almeida, A., Castro, A., Domingues, L. and Besir, H. (2013). The novel Fh8 and H fusion partners for soluble protein expression in Escherichia coli: a

comparison with the traditional gene fusion technology. *Appl Microbiol Biotechnol* **97**(15):6779-6791.

Cristofanilli, M., Gratch, D., Pagano, B., McDermott, K., Huang, J., Jian, J., . . . Sadiq, S. A. (2017). Transglutaminase-6 is an autoantigen in progressive multiple sclerosis and is upregulated in reactive astrocytes. *Mult Scler* **23**(13):1707-1715.

Datta, S., Antonyak, M. A. and Cerione, R. A. (2006). Importance of Ca(2+)-dependent transamidation activity in the protection afforded by tissue transglutaminase against doxorubicin-induced apoptosis. *Biochemistry* **45**(44):13163-13174.

Datta, S., Antonyak, M. A. and Cerione, R. A. (2007). GTP-BINDING-DEFECTIVE FORMS OF TISSUE TRANSGLUTAMINASE TRIGGER CELL DEATH†. *Biochemistry* **46**(51):14819-14829.

de Marco, A., Deuerling, E., Mogk, A., Tomoyasu, T. and Bukau, B. (2007). Chaperone-based procedure to increase yields of soluble recombinant proteins produced in E. coli. In: *BMC Biotechnol*. Vol. 7. p. 32.

Dong, X., Su, X., Yu, J., Liu, J., Shi, X., Pan, Q., . . . Cao, H. (2017). Homology modeling and molecular dynamics simulation of the HIF2alpha degradation-related HIF2alpha-VHL complex. *J Mol Graph Model* **71**:116-123.

Dorfman, R., Nalpathamkalam, T., Taylor, C., Gonska, T., Keenan, K., Yuan, X. W., . . . Durie, P. (2010). Do common in silico tools predict the clinical consequences of amino-acid substitutions in the CFTR gene? *Clin Genet* **77**(5):464-473.

Dubbink, H. J., de Waal, L., van Haperen, R., Verkaik, N. S., Trapman, J. and Romijn, J. C. (1998). The human prostate-specific transglutaminase gene (TGM4): genomic organization, tissue-specific expression, and promoter characterization. *Genomics* **51**(3):434-444.

Eckert, R. L., Kaartinen, M. T., Nurminskaya, M., Belkin, A. M., Colak, G., Johnson, G. V. and Mehta, K. (2014). Transglutaminase regulation of cell function. *Physiol Rev* **94**(2):383-417.

Fadouloglou, V. E., Kokkinidis, M. and Glykos, N. M. (2008). Determination of protein oligomerization state: two approaches based on glutaraldehyde crosslinking. *Anal Biochem* **373**(2):404-406.

Fagerberg, L., Hallstrom, B. M., Oksvold, P., Kampf, C., Djureinovic, D., Odeberg, J., . . . Uhlen, M. (2014). Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol Cell Proteomics* **13**(2):397-406.

Falconer, R. J. (2016). Applications of isothermal titration calorimetry - the research and technical developments from 2011 to 2015. *J Mol Recognit* **29**(10):504-515.

Farwell, K. D., Shahmirzadi, L., El-Khechen, D., Powis, Z., Chao, E. C., Tippin Davis, B., . . . Tang, S. (2015). Enhanced utility of family-centered diagnostic exome sequencing with inheritance model-based analysis: results from 500 unselected families with undiagnosed genetic conditions. *Genet Med* **17**(7):578-586.

Fazel-Najafabadi, E., Vahdat Ahar, E., Fattahpour, S. and Sedghi, M. (2015). Structural and functional impact of missense mutations in TPMT: An integrated computational approach. *Comput Biol Chem* **59 Pt A**:48-55.

Firasat, S., Kaul, H., Ashfaq, U. A. and Idrees, S. (2018). In silico analysis of five missense mutations in CYP1B1 gene in Pakistani families affected with primary congenital glaucoma. *Int Ophthalmol* **38**(2):807-814.

Fleckenstein, B., Qiao, S. W., Larsen, M. R., Jung, G., Roepstorff, P. and Sollid, L. M. (2004). Molecular characterization of covalent complexes between tissue transglutaminase and gliadin peptides. *J Biol Chem* **279**(17):17607-17616.

Fogel, B. L., Lee, J. Y., Lane, J., Wahnich, A., Chan, S., Huang, A., . . . Coppola, G. (2012). Mutations in rare ataxia genes are uncommon causes of sporadic cerebellar ataxia. *Mov Disord* **27**(3):442-446.

Fraij, B. M. and Gonzales, R. A. (1997). Organization and structure of the human tissue transglutaminase gene. *Biochim Biophys Acta* **1354**(1):65-71.

Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., . . . Shenkin, P. S. (2004). Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem* **47**(7):1739-1749.

Friesner, R. A., Murphy, R. B., Repasky, M. P., Frye, L. L., Greenwood, J. R., Halgren, T. A., . . . Mainz, D. T. (2006). Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J Med Chem* **49**(21):6177-6196.

Fukui, M., Kuramoto, K., Yamasaki, R., Shimizu, Y., Itoh, M., Kawamoto, T. and Hitomi, K. (2013). Identification of a highly reactive substrate peptide for transglutaminase 6 and its use in detecting transglutaminase activity in the skin epidermis. *Febs j* **280**(6):1420-1429.

Fung, J. L. F., Tsang, M. H. Y., Leung, G. K. C., Yeung, K. S., Mak, C. C. Y., Fung, C. W., . . . Chung, B. H. Y. (2019). A significant inflation in TGM6 genetic risk casts doubt in its causation in spinocerebellar ataxia type 35. *Parkinsonism Relat Disord*.

Gil-Garcia, M., Bano-Polo, M., Varejao, N., Jamroz, M., Kuriata, A., Diaz-Caballero, M., . . . Ventura, S. (2018). Combining Structural Aggregation Propensity and Stability Predictions To Redesign Protein Solubility. *Mol Pharm* **15**(9):3846-3859.

Goncharova, I., Novotna, J. and Urbanova, M. (2012). Stacked and continuous helical self-assemblies of guanosine monophosphates detected by vibrational circular dichroism. *Anal Bioanal Chem* **403**(9):2635-2644.

Greenberg, C. S., Birckbichler, P. J. and Rice, R. H. (1991). Transglutaminases: multifunctional cross-linking enzymes that stabilize tissues. *Faseb j* **5**(15):3071-3077.

Grenard, P., Bates, M. K. and Aeschlimann, D. (2001). Evolution of transglutaminase genes: identification of a transglutaminase gene cluster on human chromosome 15q15.

Structure of the gene encoding transglutaminase X and a novel gene family member, transglutaminase Z. *J Biol Chem* **276**(35):33066-33078.

Griffiths, R., (2017). Novel biomarkers for arthritis: The role of P2X7 receptor in transglutaminase 2 export and activation. *PhD Thesis*, Cardiff University.

Guan, W. J., Wang, J. L., Liu, Y. T., Ma, Y. T., Zhou, Y., Jiang, H., . . . Tang, B. S. (2013). Spinocerebellar ataxia type 35 (SCA35)-associated transglutaminase 6 mutants sensitize cells to apoptosis. *Biochem Biophys Res Commun* **430**(2):780-786.

Guerois, R., Nielsen, J. E. and 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):369-387.

Guex, N., Peitsch, M. C. and Schwede, T. (2009). Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: a historical perspective. *Electrophoresis* **30 Suppl 1**:S162-173.

Guo, Y. C., Lin, J. J., Liao, Y. C., Tsai, P. C., Lee, Y. C. and Soong, B. W. (2014). Spinocerebellar ataxia 35: novel mutations in TGM6 with clinical and genetic characterization. *Neurology* **83**(17):1554-1561.

Hadjivassiliou, M., Aeschlimann, P., Sanders, D. S., Maki, M., Kaukinen, K., Grunewald, R. A., . . . Aeschlimann, D. P. (2013). Transglutaminase 6 antibodies in the diagnosis of gluten ataxia. *Neurology* **80**(19):1740-1745.

Hadjivassiliou, M., Aeschlimann, P., Strigun, A., Sanders, D. S., Woodroofe, N. and Aeschlimann, D. (2008a). Autoantibodies in gluten ataxia recognize a novel neuronal transglutaminase. *Ann Neurol* **64**(3):332-343.

Hadjivassiliou, M., Gibson, A., Davies-Jones, G. A., Lobo, A. J., Stephenson, T. J. and Milford-Ward, A. (1996). Does cryptic gluten sensitivity play a part in neurological illness? *Lancet* **347**(8998):369-371.

Hadjivassiliou, M., Sanders, D. S., Woodroofe, N., Williamson, C. and Grunewald, R. A. (2008b). Gluten ataxia. *Cerebellum* **7**(3):494-498.

Halgren, T. A., Murphy, R. B., Friesner, R. A., Beard, H. S., Frye, L. L., Pollard, W. T. and Banks, J. L. (2004). Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J Med Chem* **47**(7):1750-1759.

Han, B. G., Cho, J. W., Cho, Y. D., Jeong, K. C., Kim, S. Y. and Lee, B. I. (2010). Crystal structure of human transglutaminase 2 in complex with adenosine triphosphate. *Int J Biol Macromol* **47**(2):190-195.

Hartley, D. L. and Kane, J. F. (1988). Properties of inclusion bodies from recombinant Escherichia coli. *Biochem Soc Trans* **16**(2):101-102.

Heil, A., (2015). Development of a novel 3D human cartilage model system to investigate changes in cartilage associated with osteoarthritis. *PhD Thesis*, Cardiff University.

Hong, P., Koza, S. and Bouvier, E. S. (2012). Size-Exclusion Chromatography for the Analysis of Protein Biotherapeutics and their Aggregates. *J Liq Chromatogr Relat Technol* **35**(20):2923-2950.

Ichinose, A. and Davie, E. W. (1988). Characterization of the gene for the a subunit of human factor XIII (plasma transglutaminase), a blood coagulation factor. *Proc Natl Acad Sci U S A* **85**(16):5829-5833.

Iismaa, S. E. (2016). The prostate-specific protein, transglutaminase 4 (TG4), is an autoantigen associated with male subfertility. *Ann Transl Med* **4**(Suppl 1):S35.

Iismaa, S. E., Chung, L., Wu, M. J., Teller, D. C., Yee, V. C. and Graham, R. M. (1997). The core domain of the tissue transglutaminase Gh hydrolyzes GTP and ATP. *Biochemistry* **36**(39):11655-11664.

Iismaa, S. E., Mearns, B. M., Lorand, L. and Graham, R. M. (2009). Transglutaminases and disease: lessons from genetically engineered mouse models and inherited disorders. *Physiol Rev* **89**(3):991-1023.

Iismaa, S. E., Wu, M. J., Nanda, N., Church, W. B. and Graham, R. M. (2000). GTP binding and signaling by Gh/transglutaminase II involves distinct residues in a unique GTP-binding pocket. *J Biol Chem* **275**(24):18259-18265.

Im, M. J., Riek, R. P. and Graham, R. M. (1990). A novel guanine nucleotide-binding protein coupled to the alpha 1-adrenergic receptor. II. Purification, characterization, and reconstitution. *J Biol Chem* **265**(31):18952-18960.

Jacobson, M. P., Friesner, R. A., Xiang, Z. and Honig, B. (2002). On the role of the crystal environment in determining protein side-chain conformations. *J Mol Biol* **320**(3):597-608.

Jacobson, M. P., Pincus, D. L., Rapp, C. S., Day, T. J., Honig, B., Shaw, D. E. and Friesner, R. A. (2004). A hierarchical approach to all-atom protein loop prediction. *Proteins* **55**(2):351-367.

Jang, T. H., Lee, D. S., Choi, K., Jeong, E. M., Kim, I. G., Kim, Y. W., . . . Park, H. H. (2014). Crystal structure of transglutaminase 2 with GTP complex and amino acid sequence evidence of evolution of GTP binding site. *PLoS One* **9**(9):e107005.

Jayadev, S. and Bird, T. D. (2013). Hereditary ataxias: overview. *Genet Med* **15**(9):673-683.

Jeitner, T. M., Muma, N. A., Battaile, K. P. and Cooper, A. J. (2009). Transglutaminase activation in neurodegenerative diseases. *Future Neurol* **4**(4):449-467.

John, S., Thiebach, L., Frie, C., Mokkapati, S., Bechtel, M., Nischt, R., . . . Smyth, N. (2012). Epidermal transglutaminase (TGase 3) is required for proper hair development, but not the formation of the epidermal barrier. *PLoS One* **7**(4):e34252.

Johnson, K. A. and Terkeltaub, R. A. (2005). External GTP-bound transglutaminase 2 is a molecular switch for chondrocyte hypertrophic differentiation and calcification. *J Biol Chem* **280**(15):15004-15012.

Johnson, W. C. (1999). Analyzing protein circular dichroism spectra for accurate secondary structures. *Proteins* **35**(3):307-312.

Kapust, R. B. and Waugh, D. S. (1999). Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Sci* **8**(8):1668-1674.

Kim, H. C., Lewis, M. S., Gorman, J. J., Park, S. C., Girard, J. E., Folk, J. E. and Chung, S. I. (1990). Protransglutaminase E from guinea pig skin. Isolation and partial characterization. *J Biol Chem* **265**(35):21971-21978.

Kim, I. G., Gorman, J. J., Park, S. C., Chung, S. I. and Steinert, P. M. (1993). The deduced sequence of the novel protransglutaminase E (TGase3) of human and mouse. *J Biol Chem* **268**(17):12682-12690.

Kim, I. G., Lee, S. C., Lee, J. H., Yang, J. M., Chung, S. I. and Steinert, P. M. (1994). Structure and organization of the human transglutaminase 3 gene: evolutionary relationship to the transglutaminase family. *J Invest Dermatol* **103**(2):137-142.

Kim, I. G., McBride, O. W., Wang, M., Kim, S. Y., Idler, W. W. and Steinert, P. M. (1992). Structure and organization of the human transglutaminase 1 gene. *J Biol Chem* **267**(11):7710-7717.

Kim, S. Y., Grant, P., Lee, J. H., Pant, H. C. and Steinert, P. M. (1999). Differential expression of multiple transglutaminases in human brain. Increased expression and cross-linking by transglutaminases 1 and 2 in Alzheimer's disease. *J Biol Chem* **274**(43):30715-30721.

Kimple, M. E., Brill, A. L. and Pasker, R. L. (2013). Overview of Affinity Tags for Protein Purification. *Curr Protoc Protein Sci* **73**:Unit-9 9.

Kiraly, R., Csosz, E., Kurtan, T., Antus, S., Szigeti, K., Simon-Vecsei, Z., . . . Fesus, L. (2009). Functional significance of five noncanonical Ca2+-binding sites of human transglutaminase 2 characterized by site-directed mutagenesis. *Febs j* **276**(23):7083-7096.

Korsgren, C. and Cohen, C. M. (1991). Organization of the gene for human erythrocyte membrane protein 4.2: structural similarities with the gene for the a subunit of factor XIII. *Proc Natl Acad Sci U S A* **88**(11):4840-4844.

Kranz, T. M., Harroch, S., Manor, O., Lichtenberg, P., Friedlander, Y., Seandel, M., . . . Malaspina, D. (2015). De novo mutations from sporadic schizophrenia cases highlight important signaling genes in an independent sample. *Schizophr Res* **166**(1-3):119-124.

Lai, T. S., Slaughter, T. F., Peoples, K. A., Hettasch, J. M. and Greenberg, C. S. (1998). Regulation of human tissue transglutaminase function by magnesium-nucleotide complexes. Identification of distinct binding sites for Mg-GTP and Mg-ATP. *J Biol Chem* **273**(3):1776-1781.

Landegren, N., Sharon, D., Shum, A. K., Khan, I. S., Fasano, K. J., Hallgren, A., . . . Kampe, O. (2015). Transglutaminase 4 as a prostate autoantigen in male subfertility. *Sci Transl Med* **7**(292):292ra101.

Lebendiker, M. and Danieli, T. (2014). Production of prone-to-aggregate proteins. *FEBS Lett* **588**(2):236-246.

Lee, B. and Richards, F. M. (1971). The interpretation of protein structures: estimation of static accessibility. *J Mol Biol* **55**(3):379-400.

Lek, M., Karczewski, K. J., Minikel, E. V., Samocha, K. E., Banks, E., Fennell, T., . . . MacArthur, D. G. (2016). Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**(7616):285-291.

Li, M., Pang, S. Y., Song, Y., Kung, M. H., Ho, S. L. and Sham, P. C. (2013). Whole exome sequencing identifies a novel mutation in the transglutaminase 6 gene for spinocerebellar ataxia in a Chinese family. *Clin Genet* **83**(3):269-273.

Li, Z., Kessler, W., van den Heuvel, J. and Rinas, U. (2011). Simple defined autoinduction medium for high-level recombinant protein production using T7-based Escherichia coli expression systems. *Appl Microbiol Biotechnol* **91**(4):1203-1213.

Lin, C. C., Gan, S. R., Gupta, D., Alaedini, A., Green, P. H. and Kuo, S. H. (2018). Hispanic Spinocerebellar Ataxia Type 35 (SCA35) with a Novel Frameshift Mutation. *Cerebellum*.

Lindemann, I., Böttcher, J., Oertel, K., Weber, J., Hils, H., Pasternack, R., Linne, U., Heine, A. and Klebe, G. (2012). Inhibitors of transglutaminase 2: a therapeutic option in celiac disease. *Zedira newsletter.* https://zedira.com/data/newsletter/newsletter_20080902.pdf. Accessed September 17, 2019

Lindfors, K., Kaukinen, K. and Maki, M. (2009). A role for anti-transglutaminase 2 autoantibodies in the pathogenesis of coeliac disease? *Amino Acids* **36**(4):685-691.

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

Liu, Y. T., Tang, B. S., Lan, W., Song, N. N., Huang, Y., Zhang, L., . . . Wang, J. L. (2013). Distribution of transglutaminase 6 in the central nervous system of adult mice. *Anat Rec (Hoboken)* **296**(10):1576-1587.

Lorand, L. and Graham, R. M. (2003). Transglutaminases: crosslinking enzymes with pleiotropic functions. *Nat Rev Mol Cell Biol* **4**(2):140-156.

Lorand, L., Murthy, S. N., Parameswaran, K. N., Velasco, P. T. and Wilson, J. (1992). Amide bond cleavage monitored continuously through detection of a dansylcadaverine leaving group. *Biochem Biophys Res Commun* **186**(1):334-341.

Lyupina, Y. V., Zatsepina, O. G., Serebryakova, M. V., Erokhov, P. A., Abaturova, S. B., Kravchuk, O. I., . . . Mikhailov, V. S. (2016). Proteomics of the 26S proteasome in

Spodoptera frugiperda cells infected with the nucleopolyhedrovirus, AcMNPV. *Biochim Biophys Acta* **1864**(6):738-746.

Maiuri, L., Ciacci, C., Auricchio, S., Brown, V., Quaratino, S. and Londei, M. (2000). Interleukin 15 mediates epithelial changes in celiac disease. *Gastroenterology* **119**(4):996-1006.

Marchal, I., Jarvis, D. L., Cacan, R. and Verbert, A. (2001). Glycoproteins from insect cells: sialylated or not? *Biol Chem* **382**(2):151-159.

Martinez-Alonso, M., Gonzalez-Montalban, N., Garcia-Fruitos, E. and Villaverde, A. (2008). The Functional quality of soluble recombinant polypeptides produced in Escherichia coli is defined by a wide conformational spectrum. *Appl Environ Microbiol* **74**(23):7431-7433.

Martinez-Torrecuadrada, J. L., Romero, S., Nunez, A., Alfonso, P., Sanchez-Cespedes, M. and Casal, J. I. (2005). An efficient expression system for the production of functionally active human LKB1. *J Biotechnol* **115**(1):23-34.

Mondal, S., Hsiao, K. and Goueli, S. A. (2015). A Homogenous Bioluminescent System for Measuring GTPase, GTPase Activating Protein, and Guanine Nucleotide Exchange Factor Activities. In: *Assay Drug Dev Technol*. Vol. 13. pp. 444-455.

Murthy, S. N., Iismaa, S., Begg, G., Freymann, D. M., Graham, R. M. and Lorand, L. (2002). Conserved tryptophan in the core domain of transglutaminase is essential for catalytic activity. *Proc Natl Acad Sci U S A* **99**(5):2738-2742.

Nallamsetty, S. and Waugh, D. S. (2006). Solubility-enhancing proteins MBP and NusA play a passive role in the folding of their fusion partners. *Protein Expr Purif* **45**(1):175-182.

Nguyen, H. H., Park, J., Kang, S. and Kim, M. (2015). Surface plasmon resonance: a versatile technique for biosensor applications. *Sensors (Basel)* **15**(5):10481-10510.

Nibbeling, E. A. R., Duarri, A., Verschuuren-Bemelmans, C. C., Fokkens, M. R., Karjalainen, J. M., Smeets, C., . . . Verbeek, D. S. (2017). Exome sequencing and network analysis identifies shared mechanisms underlying spinocerebellar ataxia. *Brain* **140**(11):2860-2878.

Pan, L. L., Huang, Y. M., Wang, M., Zhuang, X. E., Luo, D. F., Guo, S. C., . . . Wang, S. Y. (2015). Positional cloning and next-generation sequencing identified a TGM6 mutation in a large Chinese pedigree with acute myeloid leukaemia. *Eur J Hum Genet* **23**(2):218-223.

Pang, C., Cao, T., Li, J., Jia, M., Zhang, S., Ren, S., . . . Zhan, Y. (2013). Combining fragment homology modeling with molecular dynamics aims at prediction of Ca(2)(+) binding sites in CaBPs. *J Comput Aided Mol Des* **27**(8):697-705.

Pinkas, D. M., Strop, P., Brunger, A. T. and Khosla, C. (2007). Transglutaminase 2 undergoes a large conformational change upon activation. *PLoS Biol* **5**(12):e327.

Porzio, O., Massa, O., Cunsolo, V., Colombo, C., Malaponti, M., Bertuzzi, F., . . . Barbetti, F. (2007). Missense mutations in the TGM2 gene encoding transglutaminase 2 are found in patients with early-onset type 2 diabetes. Mutation in brief no. 982. Online. *Hum Mutat* **28**(11):1150.

Rosano, G. L. and Ceccarelli, E. A. (2014). Recombinant protein expression in Escherichia coli: advances and challenges. *Front Microbiol* **5**:172.

Ruan, Q., Tucholski, J., Gundemir, S. and Johnson Voll, G. V. (2008). The Differential Effects of R580A Mutation on Transamidation and GTP Binding Activity of Rat and Human Type 2 Transglutaminase. *Int J Clin Exp Med* **1**(3):248-259.

Ruiz-Lopez, M., Freitas, M. E. and Fox, S. H. (2017). Late onset cerebellar ataxia SCA 35 in a Caucasian sibling pair. *Journal of the Neurological Sciences* **381**:892-893.

Sardy, M., Karpati, S., Merkl, B., Paulsson, M. and Smyth, N. (2002). Epidermal transglutaminase (TGase 3) is the autoantigen of dermatitis herpetiformis. *J Exp Med* **195**(6):747-757.

Sastry, G. M., Adzhigirey, M., Day, T., Annabhimoju, R. and Sherman, W. (2013). Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J Comput Aided Mol Des* **27**(3):221-234.

Schein, C. H. and Noteborn, M. H. M. (2019). Formation of Soluble Recombinant Proteins in Escherichia Coli is Favored by Lower Growth Temperature. *Bio/Technology* **6**(3):291.

Schuppan, D., Dieterich, W., Ehnis, T., Bauer, M., Donner, P., Volta, U. and Riecken, E. O. (1998). Identification of the autoantigen of celiac disease. *Ann N Y Acad Sci* **859**:121-126.

Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F. and Serrano, L. (2005). The FoldX web server: an online force field. *Nucleic Acids Res* **33**(Web Server issue):W382-388.

Sezonov, G., Joseleau-Petit, D. and D'Ari, R. (2007). Escherichia coli physiology in Luria-Bertani broth. *J Bacteriol* **189**(23):8746-8749.

Shanmugarajah, P. D., Hoggard, N., Currie, S., Aeschlimann, D. P., Aeschlimann, P. C., Gleeson, D. C., . . . Hadjivassiliou, M. (2016). Alcohol-related cerebellar degeneration: not all down to toxicity? *Cerebellum Ataxias* **3**:17.

Shuo-shuo, C., Xue-zheng, L. and Ji-hong, S. (2011). Effects of co-expression of molecular chaperones on heterologous soluble expression of the cold-active lipase Lip-948. *Protein Expr Purif* **77**(2):166-172.

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., . . . Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**:539.

Sim, N. L., Kumar, P., Hu, J., Henikoff, S., Schneider, G. and Ng, P. C. (2012). SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res* **40**(Web Server issue):W452-457.

Singh, G., Zhang, J., Ma, Y., Cerione, R. A. and Antonyak, M. A. (2016). The Different Conformational States of Tissue Transglutaminase Have Opposing Affects on Cell Viability. *J Biol Chem* **291**(17):9119-9132.

Sjostrom, H., Lundin, K. E., Molberg, O., Korner, R., McAdam, S. N., Anthonsen, D., . . . Sollid, L. M. (1998). Identification of a gliadin T-cell epitope in coeliac disease: general importance of gliadin deamidation for intestinal T-cell recognition. *Scand J Immunol* **48**(2):111-115.

Sollid, L. M. (2002). Coeliac disease: dissecting a complex inflammatory disorder. *Nat Rev Immunol* **2**(9):647-655.

Sollid, L. M. and Jabri, B. (2013). Triggers and drivers of autoimmunity: lessons from coeliac disease. In: *Nat Rev Immunol*. Vol. 13. England, pp. 294-302.

Spina, A. M., Esposito, C., Pagano, M., Chiosi, E., Mariniello, L., Cozzolino, A., . . . Illiano, G. (1999). GTPase and transglutaminase are associated in the secretion of the rat anterior prostate. *Biochem Biophys Res Commun* **260**(2):351-356.

Sreerama, N. and Woody, R. W. (2000). Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Anal Biochem* **287**(2):252-260.

Stamnaes, J., Dorum, S., Fleckenstein, B., Aeschlimann, D. and Sollid, L. M. (2010). Gluten T cell epitope targeting by TG3 and TG6; implications for dermatitis herpetiformis and gluten ataxia. *Amino Acids* **39**(5):1183-1191.

Staub, A., Guillarme, D., Schappler, J., Veuthey, J. L. and Rudaz, S. (2011). Intact protein analysis in the biopharmaceutical field. *J Pharm Biomed Anal* **55**(4):810-822.

Steinert, P. M. and Marekov, L. N. (1995). The proteins elafin, filaggrin, keratin intermediate filaments, loricrin, and small proline-rich proteins 1 and 2 are isodipeptide cross-linked components of the human epidermal cornified cell envelope. *J Biol Chem* **270**(30):17702-17711.

Steinert, P. M., Parry, D. A. and Marekov, L. N. (2003). Trichohyalin mechanically strengthens the hair follicle: multiple cross-bridging roles in the inner root sheath. *J Biol Chem* **278**(42):41409-41419.

Stenberg, R., Hadjivassiliou, M., Aeschlimann, P., Hoggard, N. and Aeschlimann, D. (2014). Anti-transglutaminase 6 antibodies in children and young adults with cerebral palsy. *Autoimmune Dis* **2014**:237107.

Stieler, M., Weber, J., Hils, M., Kolb, P., Heine, A., Buchold, C., . . . Klebe, G. (2013). Structure of active coagulation factor XIII triggered by calcium binding: basis for the design of next-generation anticoagulants. *Angew Chem Int Ed Engl* **52**(45):11930-11934.

Studer, R. A., Christin, P. A., Williams, M. A. and Orengo, C. A. (2014). Stability-activity tradeoffs constrain the adaptive evolution of RubisCO. *Proc Natl Acad Sci U S A* **111**(6):2223-2228.

Terpe, K. (2003). Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Appl Microbiol Biotechnol* **60**(5):523-533.

Thangaraju, K., Kiraly, R., Demeny, M. A., Andras Motyan, J., Fuxreiter, M. and Fesus, L. (2017). Genomic variants reveal differential evolutionary constraints on human transglutaminases and point towards unrecognized significance of transglutaminase 2. *PLoS One* **12**(3):e0172189.

Thomas, H., (2004). Characterisation of a novel transglutaminase, TGase 6, and determination of its expression pattern. *PhD Thesis*, Cardiff University.

Thomas, H., Beck, K., Adamczyk, M., Aeschlimann, P., Langley, M., Oita, R. C., . . . Aeschlimann, D. (2013). Transglutaminase 6: a protein associated with central nervous system development and motor function. *Amino Acids* **44**(1):161-177.

Thomas, P. D., Kejariwal, A., Campbell, M. J., Mi, H., Diemer, K., Guo, N., . . . Doremieux, O. (2003). PANTHER: a browsable database of gene products organized by biological function, using curated protein family and subfamily classification. *Nucleic Acids Res* **31**(1):334-341.

Tripathy, D., Vignoli, B., Ramesh, N., Polanco, M. J., Coutelier, M., Stephen, C. D., . . . Basso, M. (2017). Mutations in TGM6 induce the unfolded protein response in SCA35. *Hum Mol Genet* **26**(19):3749-3762.

Ustun-Aytekin, O., Gurhan, I. D., Ohura, K., Imai, T. and Ongen, G. (2014). Monitoring of the effects of transfection with baculovirus on Sf9 cell line and expression of human dipeptidyl peptidase IV. *Cytotechnology* **66**(1):159-168.

Wang, J. L., Yang, X., Xia, K., Hu, Z. M., Weng, L., Jin, X., . . . Tang, B. S. (2010). TGM6 identified as a novel causative gene of spinocerebellar ataxias using exome sequencing. *Brain* **133**(Pt 12):3510-3518.

Wang, Z., Li, N., Wang, Y., Wu, Y., Mu, T., Zheng, Y., . . . Fang, X. (2012). Ubiquitin-intein and SUMO2-intein fusion systems for enhanced protein production and purification. *Protein Expr Purif* **82**(1):174-178.

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., . . . Schwede, T. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res* **46**(W1):W296-w303.

Westenberger, A., Svetel, M., Dragaševic, N., Braenne, I., Dobricic, V., Hicks, A. A., Tomic, A., Kresojevic, N., Pawlack, H., KGrütz, K., Domingo, A., Erdmann, J., V.S. Kostic, V. S. and Klein C. (2016). Do mutations in the TGM6 (SCA35) gene cause early-onset Parkinson's disease? [abstract]. *Mov Disord.* 31 (suppl 2). https://www.mdsabstracts.org/abstract/do-mutations-in-the-tgm6-sca35-gene-cause-early-onset-parkinsons-disease/. Accessed September 17, 2019.

Whitmore, L. and Wallace, B. A. (2004). DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res* **32**(Web Server issue):W668-673.

Whitmore, L. and Wallace, B. A. (2008). Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. *Biopolymers* **89**(5):392-400.

Willard, F. S., Kimple, A. J., Johnston, C. A. and Siderovski, D. P. (2005). A direct fluorescence-based assay for RGS domain GTPase accelerating activity. *Anal Biochem* **340**(2):341-351.

Yang, Z. H., Shi, M. M., Liu, Y. T., Wang, Y. L., Luo, H. Y., Wang, Z. L., . . . Xu, Y. M. (2018). TGM6 gene mutations in undiagnosed cerebellar ataxia patients. *Parkinsonism Relat Disord* **46**:84-86.

Yilmaz, A. (2015). Bioinformatic Analysis of GJB2 Gene Missense Mutations. *Cell Biochem Biophys* **71**(3):1623-1642.

Zambrano, R., Jamroz, M., Szczasiuk, A., Pujols, J., Kmiecik, S. and Ventura, S. (2015). AGGRESCAN3D (A3D): server for prediction of aggregation properties of protein structures. *Nucleic Acids Res* **43**(W1):W306-313.

Zhang, H., Zheng, X., Kwok, R. T. K., Wang, J., Leung, N. L. C., Shi, L., . . . Tang, B. Z. (2018). In situ monitoring of molecular aggregation using circular dichroism. *Nat Commun* **9**(1):4961.

Zheng, R., Li, Z., He, F., Liu, H., Chen, J., Xie, X., . . . Ge, B. (2018). Genome-wide association study identifies two risk loci for tuberculosis in Han Chinese. *Nat Commun* **9**(1):4072.

Zhou, Y., Lu, Z., Wang, X., Selvaraj, J. N. and Zhang, G. (2018). Genetic engineering modification and fermentation optimization for extracellular production of recombinant proteins using Escherichia coli. *Appl Microbiol Biotechnol* **102**(4):1545-1556.

Zhou, Y., Zang, Y., Yang, Y., Xiang, J. and Chen, Z. (2019). Candidate genes involved in metastasis of colon cancer identified by integrated analysis. *Cancer Med* **8**(5):2338-2347.

8. SUPPLEMENTRAY INFORMATION

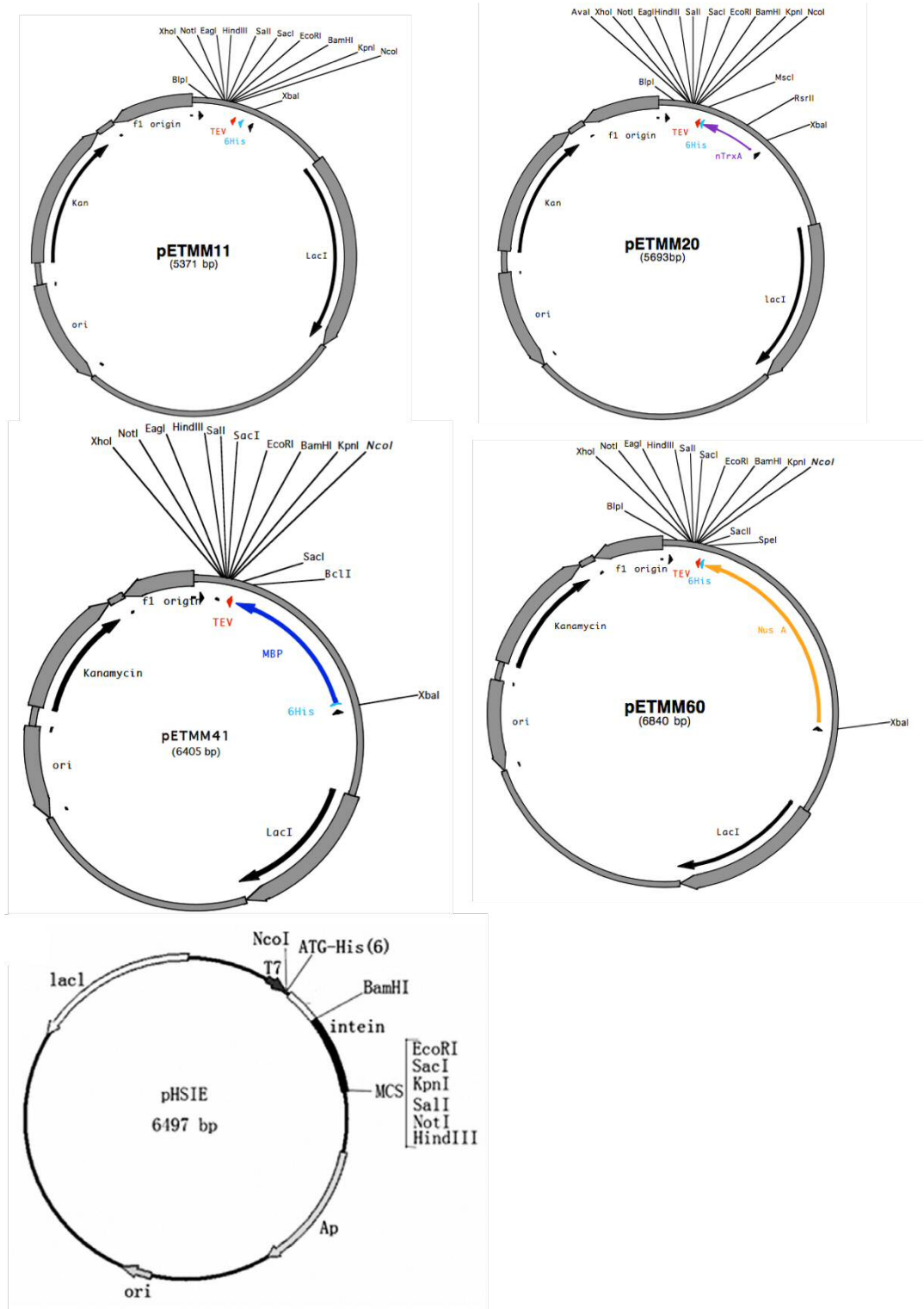


Figure S1: Vectors selected for TG6 cDNA cloning.
All five constructs have a *N*-terminal His-tag to aid in TG6 purification. pETMM 20, 41 and 60 have an additional *N*-terminal protein tag (Thioredoxin (Trx), Maltose binding protein (MBP) and *N*-utilisation substance A (NusA) respectively) which aids in the solubilisation of TG6. pHSIE contains a SuMO-intein tag which helps to increase expression.

Table S1: TG6 mutations linked to SCA35 patients, identified in different population samples.

		GnomAD			ExAC		
Protein/DNA	**Population**	**Allele count**	**Allele number**	**Allele Frequency**	**Allele count**	**Allele number**	**Allele Frequency**
c.7+1G>T	African	0	23976	0.00E+00	0	8210	0
	Ashkenazi Jewish	0	9968	0.00E+00	-	-	-
	East Asian	37	19388	2.02E-03	15	6782	2.21E-03
	European (Finnish)	0	23696	0.00E+00	0	3752	0.00E+00
	European (non-Finnish)	0	123970	0.00E+00	0	0	0.00E+00
	Latino	0	34294	0.00E+00	0	0	0.00E+00
	South Asian	0	28484	0.00E+00	0	0	0.00E+00
	Other	0	6892	0.00E+00	0	0	0.00E+00
	Total	**37**	**270668**	**1.40E-04**	**15**	**18744**	**1.61E-04**
R111C	African	2	24970	8.01E-05	0	10404	0.00E+00
	Ashkenazi Jewish	0	10370	0.00E+00	-	-	-
	East Asian	4	19952	2.01E-04	1	8654	1.16E-04
	European (Finnish)	0	25124	0.00E+00	0	6614	0.00E+00
	European (non-Finnish)	1	129194	0.00E+00	0	66734	0.00E+00
	Latino	0	35440	0.00E+00	0	11570	0.00E+00
	South Asian	0	30616	0.00E+00	0	16512	0.00E+00
	Other	0	7228	0.00E+00	0	908	0.00E+00
	Total	**7**	**282894**	**7.89E-06**	**1**	**121396**	**8.23E-06**
N137S	-	-	-	-	-	-	-
Q181H	-	-	-	-	-	-	-
G279ter	African	0	16120	0.00E+00	0	10042	0.00E+00
	Ashkenazi Jewish	0	10068	0.00E+00	-	-	-
	East Asian	0	18362	0.00E+00	0	8540	0.00E+00

	European (Finnish)	0	21488	0.00E+00	0	6276	0.00E+00
	European (non-Finnish)	3	113152	2.65E-05	1	65540	1.53E-05
	Latino	0	34512	0.00E+00	0	11382	0.00E+00
	South Asian	2	30596	6.54E-05	0	16444	
	Other	0	6098	6.50E-05	0	880	0.00E+00
	Total	**5**	**250396**	**2.04E-05**	**1**	**119104**	**8.39E-06**
D327G	African	0	16256	2.98E-05	0	10400	0.00E+00
	Ashkenazi Jewish	0	10074	0.00E+00	-	-	-
	East Asian	0	18390	0.00E+00	0	8632	0.00E+00
	European (Finnish)	0	21604	0.00E+00	0	6602	0.00E+00
	European (non-Finnish)	1	113180	8.84E-06	0	66620	0.00E+00
	Latino	1	34590	2.89E-05	1	11568	8.65E-05
	South Asian	22	30616	7.19E-04	11	16512	6.66E-04
	Other	0	6128	0.00E+00	0	906	0.00E+00
	Total	**24**	**250838**	**9.57E-05**	**12**	**121240**	**9.90E-05**
P359L	African	0	16248	0.00E+00	0	10396	0.00E+00
	Ashkenazi Jewish	0	10080	0.00E+00	-	-	-
	East Asian	0	18394	0.00E+00	0	8650	0.00E+00
	European (Finnish)	0	21648	0.00E+00	0	6614	0.00E+00
	European (non-Finnish)	13	113752	1.14E-04	3	66712	4.50E-05
	Latino	5	34590	1.45E-04	1	11576	8.64E-05
	South Asian	13	30616	4.25E-04	3	16512	1.82E-04
	Other	0	6140	0.00E+00	2	906	2.21E-03
	Total	**31**	**251468**	**1.23E-04**	**9**	**121366**	**7.42E-05**
V391M	African	0	24964	0.00E+00	0	10400	0.00E+00
	Ashkenazi Jewish	0	10370	0.00E+00	-	-	-

	East Asian	188	19954	9.42E-03	97	8642	1.12E-02
	European (Finnish)	0	25118	0.00E+00	**0**	**6614**	**0.00E+00**
	European (non-Finnish)	13	129162	1.01E-04	3	66702	4.50E-05
	Latino	7	35440	1.98E-04	3	11576	2.59E-04
	South Asian	10	30616	3.27E-04	6	16512	3.63E-04
	Other	3	7226	4.15E-04	0	906	0.00E+00
	Total	**221**	**282850**	**7.81E-04**	**109**	**121352**	**8.98E-04**
E406K	African	8	24962	3.21E-04	1	10394	9.62E-05
	Ashkenazi Jewish	0	10370	0.00E+00	-	-	-
	East Asian	1	19950	5.01E-05	0	8650	0.00E+00
	European (Finnish)	56	25122	2.23E-03	9	6614	1.36E-03
	European (non-Finnish)	263	129162	2.04E-03	139	66702	2.08E-03
	Latino	35	35440	9.88E-04	8	11578	6.91E-04
	South Asian	0	30616	0.00E+00	1	16512	6.06E-05
	Other	12	7222	1.66E-03	2	906	2.21E-03
	Total	375	282844	1.33E-03	160	121356	1.32E-03
R412C	African	0	24960	0.00E+00	**0**	**10386**	**0.00E+00**
	Ashkenazi Jewish	0	10370	0.00E+00	-	-	-
	East Asian	1	19952	5.01E-05	0	8650	0.00E+00
	European (Finnish)	0	25122	0.00E+00	0	6614	0.00E+00
	European (non-Finnish)	7	129168	5.42E-05	5	66682	7.50E-05
	Latino	0	35440	0.00E+00	0	11578	0.00E+00
	South Asian	0	30616	0.00E+00	0	16510	0.00E+00
	Other	0	7224	0.00E+00	0	906	0.00E+00
	Total	**8**	**282852**	**2.83E-05**	**5**	**121326**	**4.12E-05**
T426N	-	-	-	-	-	-	-
Y441C	African	1	16230	6.16E-05	-	-	-

	Ashkenazi Jewish	0	10062	0.00E+00	-	-	-
	East Asian	0	18392	0.00E+00	-	-	-
	European (Finnish)	0	21648	0.00E+00	-	-	-
	European (non-Finnish)	2	113360	1.76E-05	-	-	-
	Latino	0	34582	0.00E+00	-	-	-
	South Asian	0	30614	0.00E+00	-	-	-
	Other	0	6134	0.00E+00	-	-	-
	Total	**3**	**251022**	**1.20E-05**	-	-	-
R448W	African	69	22628	3.05E-03	33	4592	7.19E-03
	Ashkenazi Jewish	63	9848	6.40E-03	-	-	-
	East Asian	23	18332	1.26E-03	10	2642	3.79E-03
	European (Finnish)	745	22450	3.32E-02	131	1564	8.38E-02
	European (non-Finnish)	2235	114170	1.96E-02	938	24990	3.75E-02
	Latino	199	32780	6.07E-03	41	2432	1.69E-02
	South Asian	295	27914	1.06E-02	139	9516	1.46E-02
	Other	100	6712	1.49E-02	14	368	3.80E-02
	Total	**3729**	**254834**	**1.46E-02**	**1306**	**46104**	**2.83E-02**
P493L	African	0	16208	0.00E+00	0	10230	0.00E+00
	Ashkenazi Jewish	0	10072	0.00E+00	-	-	-
	East Asian	19	18386	1.03E-03	12	8562	1.40E-03
	European (Finnish)	0	21264	0.00E+00	0	6280	8.73E-05
	European (non-Finnish)	0	113354	0.00E+00	0	65894	0.00E+00
	Latino	1	34568	2.89E-05	1	11460	0.00E+00
	South Asian	0	30604	0.00E+00	0	16374	0.00E+00
	Other	0	6114	0.00E+00	0	878	0.00E+00
	Total	**20**	**250570**	**7.98E-05**	**13**	**119678**	**1.09E-04**

L502Q	African	0	24922	0.00E+00	-	-	-
	Ashkenazi Jewish	0	10362	0.00E+00	-	-	-
	East Asian	0	19940	0.00E+00	-	-	-
	European (Finnish)	0	24512	0.00E+00	-	-	-
	European (non-Finnish)	2	128868	1.55E-05	-	-	-
	Latino	0	35420	0.00E+00	-	-	-
	South Asian	0	30608	0.00E+00	-	-	-
	Other	0	7212	0.00E+00	-	-	-
	Total	**2**	**281844**	**7.10E-06**	-	-	-
G508D	African	7	24908	2.81E-03	1	10084	9.92E-05
	Ashkenazi Jewish	0	10360	0.00E+00	-	-	-
	East Asian	0	19928	0.00E+00	0	8524	0.00E+00
	European (Finnish)	58	24002	2.42E-03	15	6266	2.39E-03
	European (non-Finnish)	333	128662	2.58E-03	150	65614	2.29E-03
	Latino	4	35408	1.13E-04	0	11416	0.00E+00
	South Asian	0	30604	0.00E+00	0	16328	0.00E+00
	Other	9	7206	1.25E-03	0	870	0.00E+00
	Total	**411**	**281078**	**1.46E-03**	**166**	**119102**	**1.39E-03**
D510H	African	0	24914	0.00E+00	0	10068	0.00E+00
	Ashkenazi Jewish	0	10356	0.00E+00	-	-	-
	East Asian	42	19916	2.11E-03	18	8496	2.12E-03
	European (Finnish)	0	23928	0.00E+00	0	6262	0.00E+00
	European (non-Finnish)	0	128654	0.00E+00	0	65550	0.00E+00
	Latino	0	35398	0.00E+00	0	11402	0.00E+00
	South Asian	0	30598	0.00E+00	0	16314	0.00E+00
	Other	0	7202	0.00E+00	0	864	0.00E+00

	Total	**42**	**280966**	**1.50E-04**	**18**	**118956**	**1.51E-04**
L517W	African	0	24852	0.00E+00	0	9922	0.00E+00
	Ashkenazi Jewish	0	10336	0.00E+00	-	-	-
	East Asian	40	19888	2.01E-03	13	8438	1.54E-03
	European (Finnish)	0	23180	0.00E+00	0	6220	0.00E+00
	European (non-Finnish)	0	128288	0.00E+00	0	65060	0.00E+00
	Latino	0	35374	0.00E+00	0	11342	0.00E+00
	South Asian	0	30562	0.00E+00	0	16204	0.00E+00
	Other	1	7188	1.39E-04	**0**	**854**	**0.00E+00**
	Total	**41**	**279668**	**1.47E-04**	**13**	**118040**	**1.10E-04**
V527E	African	273	24680	1.11E-02	99	9588	1.03E-02
	Ashkenazi Jewish	47	10298	4.56E-03	-	-	-
	East Asian	0	19800	0.00E+00	0	8314	0.00E+00
	European (Finnish)	0	21562	0.00E+00	0	6078	0.00E+00
	European (non-Finnish)	54	127002	4.25E-04	36	63808	5.64E-04
	Latino	52	35328	1.47E-03	15	11186	1.34E-03
	South Asian	2	30486	6.56E-05	2	15990	1.25E-04
	Other	11	7144	1.54E-03	1	842	1.19E-03
	Total	**439**	**276300**	**1.59E-03**	**153**	**115806**	**1.32E-03**
E574del	African	0	16256	0.00E+00	-	-	-
	Ashkenazi Jewish	0	10078	0.00E+00	-	-	-
	East Asian	1	18394	5.44E-05	-	-	-
	European (Finnish)	0	21640	0.00E+00	-	-	-
	European (non-Finnish)	0	113720	0.00E+00	-	-	-
	Latino	0	34590	0.00E+00	-	-	-
	South Asian	0	30616	0.00E+00	-	-	-

	Other	0	6134	0.00E+00	-	-	-
	Total	**1**	**251428**	**3.98E-06**	-	-	-
Q652dup	African	13	24948	5.21E-04	8	10404	7.69E-04
	Ashkenazi Jewish	1	10352	9.66E-05	-	-	-
	East Asian	0	19950	0.00E+00	0	8648	0.00E+00
	European (Finnish)	37	25078	1.48E-03	9	6612	1.36E-03
	European (non-Finnish)	300	128930	2.33E-03	144	66704	2.16E-03
	Latino	2	35430	5.65E-05	0	11576	0.00E+00
	South Asian	0	30614	0.00E+00	0	16512	0.00E+00
	Other	8	7208	1.11E-03	1	906	1.10E-03
	Total	**361**	**282510**	**1.28E-03**	**162**	**121362**	**1.34E-03**

Population frequencies for Mutations in TGM6 identified in patients. Data as obtained from genome databases (GnomAD and ExAc)