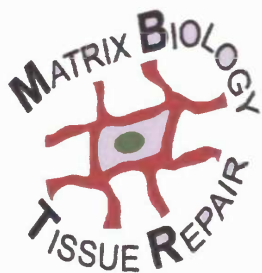


BINDING SERVICES
Tel +44 (0)29 2087 4949
Fax +44 (0)29 20371921
e-mail bindery@cardiff.ac.uk

**Characterisation of a novel transglutaminase,
TGase 6, and determination of its expression
pattern**

Helen Ann Thomas

2004



**University of Wales, College of Medicine,
Cardiff**

UMI Number: U583962

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI U583962

Published by ProQuest LLC 2013. Copyright in the Dissertation held by the Author.
Microform Edition © ProQuest LLC.

All rights reserved. This work is protected against
unauthorized copying under Title 17, United States Code.



ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Summary

Transglutaminases represent a family of enzymes capable of stabilising protein assemblies by γ -glutamyl- ϵ -lysine crosslinks. These enzymes catalyse a calcium-dependent transferase reaction between the γ -carboxamide group of a peptide-bound glutamine residue and primary amines, most commonly the ϵ -amino group of lysine residues. Eight different transglutaminase gene products have previously been characterised in man, and found to have specialised in the crosslinking of proteins in different biological processes. Identification of a gene cluster on mouse chromosome 2 and analysis of the homologous chromosomal sequences on human chromosome 20q11 revealed the presence of a novel gene, TGM6, located upstream from the gene encoding transglutaminase 3. A corresponding gene product, transglutaminase 6 could be identified in a small lung carcinoma cell line, called H69, demonstrating that this was a functional gene. The gene product in human and mouse was subsequently characterised by determining its primary structure and deduced amino acid sequence. This revealed the existence of several splice variants. Analysis of its expression on a RNA and protein level using *in situ* hybridisation and immunohistochemistry revealed that transglutaminase 6 is expressed predominantly in the central nervous system but also in a number of other tissues, for example, in the epidermis of skin, in the developing vertebrae and other long bones, and in the retinal cells of the eye. Expression was also detected in the brain, particularly in the neuronal cells of the cerebral cortex and the Purkinje cells of the cerebellum. The expression in the cerebral cortex correlated with neurogenesis during development. Transglutaminase activity has been implicated in the formation of aberrant protein complexes in the central nervous system resulting in nerve cell degeneration, for example, in Alzheimer's disease and Huntington's disease. Based on its expression pattern, transglutaminase 6 is a possible candidate to generate the underlying transglutaminase-related pathological changes.

Abstract

Transglutaminases (TGases) represent a family of enzymes capable of stabilising protein assemblies by γ -glutamyl- ϵ -lysine crosslinks. These enzymes catalyse a calcium-dependent transferase reaction between the γ -carboxamide group of a peptide-bound glutamine residue and primary amines, most commonly the ϵ -amino group of lysine residues. Eight different TGase gene products have previously been characterised in man, and found to have specialised in the crosslinking of proteins in different biological processes. Identification of a gene cluster on mouse chromosome 2 and analysis of the homologous chromosomal sequences on human chromosome 20q11 revealed the presence of a novel gene, TGM6, located upstream from the gene encoding TGase 3. A corresponding gene product, TGase 6 could be identified in a small lung carcinoma cell line, called H69, demonstrating that this was a functional gene. The gene product in human and mouse was subsequently characterised by determining its primary structure and deduced amino acid sequence. This revealed the existence of several splice variants. Analysis of its expression on a RNA and protein level using *in situ* hybridisation and immunohistochemistry revealed that TGase 6 is expressed predominantly in the central nervous system but also in a number of other tissues, for example, in the epidermis of skin, in the developing vertebrae and other long bones, and in the retinal cells of the eye. Expression was also detected in the brain, particularly in the neuronal cells of the cerebral cortex and the Purkinje cells of the cerebellum. The expression in the cerebral cortex correlated with neurogenesis during development. TGase activity has been implicated in the formation of aberrant protein complexes in the central nervous system resulting in nerve cell degeneration, for example, in Alzheimer's disease and Huntington's disease. Based on its expression pattern, TGase 6 is a possible candidate to generate the underlying TGase-related pathological changes.

Acknowledgements

I would like to express my gratitude to my supervisor, Professor Daniel Aeschlimann, for his guidance and encouragement during this research project. I would also like to thank Dr. Pascale Grenard, Dr. Konrad Beck and Mrs Pascale Aeschlimann for their assistance with aspects of this research.

I would like to acknowledge Kate Silverthorne and Sally Rosser-Davies for their continuous help, support and friendship throughout the last few years, and a special thank you to Martin Langley and Dr. Stuart Jones, who made the transition from CTBL to the Dental School a lot easier. Thanks to the four of you for always being able to turn my tears into laughter.

I must acknowledge the continual support and encouragement given by my parents, Raymond and Margaret (and a special thank you to my Dad for allowing me to “borrow” his laptop and printer). Most of all, I would like to express my sincerest thanks and appreciation to my boyfriend David, whose love, understanding, patience and sense of humour have helped me to survive the last 4 years.

Abbreviations

AAP	abridged anchor primer
A β	amyloid β -protein
AD	Alzheimer's disease
AEC	3-amino-9-ethyl-carbazole
APP	amyloid precursor protein
BCA	bicinchoninic acid
BCIP	5-bromo-4-chloro-3-indolyl phosphate
bp	base pairs
Ca ²⁺	calcium ion
CE	cornified cell envelope
CIAP	calf intestinal alkaline phosphatase
DAG	diacylglycerol
DEPC	diethyl pyrocarbonate
DIG	digoxigenin
dATP	2'-deoxyadenosine 5'-triphosphate
dCTP	2'-deoxycytidine 5'-triphosphate
dGTP	2'-deoxyguanosine 5'-triphosphate
dTTP	2'-deoxythymidine 5'-triphosphate
dNTP	2'-deoxynucleotide 5'-triphosphate
DDT	dithiothreitol
ECM	extracellular matrix
EDC	1-ethyl-3-(3-dimethylaminopropylcarbodiimide-HCl)
EDTA	diaminoethanetetra-acetic acid
ELISA	enzyme linked immunosorbent assay

FAK	focal adhesion kinase
FN	fibronectin
fP	forward primer
GH β	GTP-binding β -subunit
GST	glutathione-S-transferase
GTP	guanosine 5-triphosphate
HAP	huntingtin associated protein
HCl	hydrogen chloride
HD	Huntington,s disease
IP ₃	inositol triphosphate
IPTG	isopropyl β D-thiogalactoside
KCl	potassium chloride
KSCN	potassium thiocyanate
LB	luria-bertia
LI	lamellar ichthyosis
Mg ²⁺	magnesium ion
MgCl ₂	magnesium chloride
MgSO ₄	magnesium sulphate
MOPS	morpholinopropanesulfonic acid
3-NP	3-nitropropionic acid
NAC	non-A β component of amyloid plaque
NaCl	sodium chloride
NBT	nitroblue tetrazolium salt
NEM	N-ethylmaleimide
NF4	nuclear factor 4

NFκB	nuclear factor-κB
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PD	Parkinson's disease
PDI	protein disulphide isomerase
PHF	paired helical filament
PIP₂	phosphatidylinositol diphosphate
PLC	phospholipase C
PLG	heavy phase lock gel
PMSF	phenylmethylsulfonylfluoride
PSP	progressive supranuclear palsy
RACE	rapid amplification of 5' mRNA end
rP	reverse primer
RT	room temperature
SDS	sodium dodecyl sulphate
SDS-PAGE	sodium dodecyl sulphate-polyacrylamide gel electrophoresis
STR	short random repeat
TBS	tris buffered saline
TEA	triethanolamine
TF	transcription factor
TGase	transglutaminase
TGFβ	transforming growth factor-β
UAP	universal amplification primer

Table of contents

1.0 Introduction	1
1.1 Introduction.....	1
1.1.1 Enzyme mechanism.....	1
1.1.2 The Transglutaminase family	3
1.1.3 The Transglutaminase gene family.....	5
1.1.4 Domain organisation.....	7
1.1.5 Regulation of expression	8
1.1.5.1 Transglutaminase 2.....	9
1.1.5.2 Transglutaminase 1.....	10
1.1.5.3 Transglutaminase 3.....	11
1.1.5.4 Factor XIII	11
1.1.5.5 Transglutaminase 4.....	12
1.1.5.6 Erythrocyte protein band 4.2	13
1.1.5.7 Transglutaminase 5.....	14
1.1.5.8 Conclusion	14
1.1.6.1 Factor XIII	16
1.1.6.1.1 The a subunit.....	16
1.1.6.1.2 The active site.....	17
1.1.6.1.3 The b subunit	17
1.1.6.1.4 Activation of factor XIII.....	18
1.1.6.1.5 Function of factor XIII a subunit	19
1.1.6.1.7 Factor XIII deficiency.....	20
1.1.6.2 Tissue transglutaminase.....	21
1.1.6.2.1 Regulation of transamidation activity.....	22
1.1.6.2.2 Functions of TGase 2.....	23
1.1.6.2.2.1 Organising the extracellular matrix	24
1.1.6.2.2.2 Programmed cell death	26
1.1.6.2.2.3 TGase 2 as a signal transducing G protein	27
1.1.6.2.2.4 Pathogenic role of TGase 2 in disease.....	27
1.1.6.2.2.5 Wound healing, inflammation and fibrosis	27
1.1.6.2.3 Neurodegenerative diseases.....	28
1.1.6.2.4 Coeliac disease.....	28

1.1.6.2.5 Consequences of TGase 2 deletion.....	29
1.1.6.3 Keratinocyte transglutaminase.....	29
1.1.6.4 Epidermal transglutaminase.....	32
1.1.6.5 Prostate transglutaminase	33
1.1.6.6 Erythrocyte protein band 4.2	34
1.1.6.7 Transglutaminase 5.....	35
1.1.6.8 Transglutaminase 6.....	36
1.1.6.9 Transglutaminase 7.....	36
1.1.6.10 Summary of functions.....	37
1.1.7 Non-mammalian Transglutaminases	38
1.1.8.1 Food production.....	40
1.1.8.2 Medical applications.....	41
1.1.8.2.1 Enzyme substitution therapy.....	41
1.1.8.2.2 Diagnostics	41
1.1.8.2.3 Inhibition of excessive crosslinking	41
1.1.8.3 Cosmetics.....	42
1.1.8.4 Textiles	42
1.1.8.5 Conclusion.....	42
1.2 Neurodegenerative diseases.....	44
1.2.1 Expression of TGases in normal brain.....	44
1.2.2 Alzheimer's disease.....	46
1.2.3 Parkinson's disease.....	51
1.2.4 Progressive supranuclear palsy.....	51
1.2.5 Huntington's disease.....	52
1.2.5.1 Mouse studies	56
1.2.6 Conclusion	58
1.3 Aim of thesis.....	60
Aim 1	60
Aim 2	60
Aim 3	60
Aim 4	60
Aim 5	61

2.0 Characterisation of a novel transglutaminase, TGase 6.....	62
2.1 Introduction.....	62
2.2 Materials and Methods	63
2.2.2 RNA extraction using TRIzol for Northern blotting	63
2.2.3 RNA extraction using TRI reagent.....	64
2.2.4 Conversion of total RNA to mRNA	64
2.2.5 cDNA synthesis	65
2.2.5.1 cDNA synthesis using the cDNA Cycle kit (Invitrogen)	66
2.2.6 Amplification of TGase 6 from various cell lines	67
2.2.7 Agarose gel electrophoresis.....	67
2.2.8 Cloning of TGase 6 by anchored PCR	68
2.2.9 QIAquick gel extraction	68
2.2.10 Sequencing.....	69
2.2.10.2 Sequencing performed using the Thermo Sequenase™ Cy5/Cy5.5 Dye Terminator Sequencing Kit (Amersham Pharmacia) and a Long-Read Tower (Visible Genetics).....	70
2.2.10.2.1 Sequencing reactions using Cy5 dye-labelled ddNTP terminators	70
2.2.10.2.2 Sequencing reactions using Cy5.5 dye-labelled ddNTP terminators	71
2.2.10.3 Purification of sequencing reactions using ethanol precipitation	71
2.2.11 Rapid amplification of 5'- mRNA end.....	72
2.2.12 PCR amplification of 5' end.....	73
2.2.14 PCR amplification from mouse brain cDNA/genomic DNA	74
2.2.15.1 Preparation of probe	75
2.2.15.2 Determination of cpm/ml of labelled probe.....	76
2.2.15.3 Hybridisation and detection.....	76
2.2.16 Northern blotting using DIG-labelling	76
2.2.16.1 Preparation of probe	77
2.2.16.1.1 TA-cloning.....	77
2.2.16.1.3 Transformation	77
2.2.16.1.4 Preparation of X-Gal/LB agar plates for transformation.....	78
2.2.16.1.5 Isolation of plasmid DNA.....	78
2.2.16.1.6 DIG labelling of probe.....	79
2.2.16.2 Hybridisation and detection.....	80
2.2.17 Primer extension analysis	81

2.3 Results.....	82
2.3.1 Detection of TGase 6 transcripts in various cell lines	82
2.3.2 Determination of the cDNA sequence of TGase 6	82
2.3.3 Determination of transcript size.....	95
2.3.4 Primer extension analysis	96
2.3.5 3'-untranslated region.....	99
2.3.6 Sequencing of mouse TGase 6	100
2.3.7 Proximal promoter sequence	106
2.4 Discussion.....	108
3.0 RNA expression of TGase 6	115
3.1 Introduction.....	115
3.2 Materials and Methods	116
3.2.1 Northern Hybridisation.....	116
3.2.2 In situ hybridisation.....	116
3.2.2.1 Preparation of mouse TGase 6 RNA probes.....	116
3.2.2.2 Preparation of mouse TGase 5 RNA probes.....	119
3.2.2.3 Preparation of mouse tissue sections	120
3.2.2.4 Hybridisation	120
3.2.3 Luxol Fast Blue/Cresyl violet stain	121
3.2.4 RNA samples	121
3.2.4.1 cDNA synthesis from RNA.....	122
3.2.4.2. TGase screening.....	122
3.3 Results.....	123
3.3.1 Northern blotting	123
3.3.2 In situ hybridisation.....	125
3.3.2.1 TGase 6 expression in newborn mouse	125
3.3.2.2 TGase 5 expression in newborn mouse	130
3.3.2.3 TGase 6 expression in mouse embryo development	131
3.3.2.3.1 11 day mouse embryo.....	132
3.3.2.3.2 13 day mouse embryo.....	133
3.3.2.3.3 16 day mouse embryo.....	134
3.3.2.4 TGase 5 expression in mouse embryo development	137
3.3.3 Amplification of TGases from mouse brain cDNA.....	138

3.4 Discussion.....	139
4.0 Antibody generation and characterisation of protein expression of TGase 6.....	141
4.1 Introduction.....	141
4.2 Materials and Methods	142
4.2.1 Expression of the b barrel 1 and 2 domain of human TGase 6.....	142
4.2.1.1 Isolation of DNA fragment encoding C-terminal b barrel 1 and 2 domain of human TGase 6	142
4.2.1.2 Construction of expression vector	142
4.2.1.3 Expression of the GST-fusion protein	143
4.2.1.4 Purification of GST-fusion protein by affinity chromatography.....	144
4.2.2 SDS-PAGE	144
4.2.3 Western blotting.....	144
4.2.4 Membrane stripping.....	145
4.2.5 Peptides.....	145
4.2.6 Peptide coupling for TGase 6S	145
4.2.6.1 Peptide coupling for TGase 6L.....	146
4.2.6.2 Preparation of control conjugate.....	146
4.2.6.3 Protein concentration determination.....	147
4.2.7 Production of antisera.....	147
4.2.8 ELISA (Enzyme Linked ImmunoSorbent Assay).....	147
4.2.9 Analysis of mouse brain tissue	148
4.2.10 Affinity purification of specific antibodies from TGase 6S antiserum	148
4.2.10.1 Affinity purification of specific antibodies from TGase 6L antiserum	149
4.2.11 Immunohistochemistry	149
4.3 Results.....	151
4.3.1 Expression of the b barrel 1 and 2 domain of human TGase 6.....	151
4.3.1.1 Generation of the expression construct for the b barrel 1 and 2 domain of TGase 6	151
4.3.1.2 Expression and purification of the GST-fusion protein.....	151
4.3.2 Peptide design and synthesis	153
4.3.3 Peptide coupling	156
4.3.4 Production of antisera.....	160
4.3.5 Characterisation of TGase 6 antisera.....	161

4.3.6 Determination of TGase 6 expression in mouse brain.....	165
4.3.7 Antibody purification.....	167
4.3.8 Immunohistochemistry	168
4.4 Discussion.....	174
5.0 Generation of full-length TGase 6 cDNA	176
5.1 Introduction.....	176
5.2 Materials and Methods	177
5.2.1 PCR amplification	177
5.2.2 Gel extraction.....	177
5.2.2.1 Gel extraction using a QIAquick gel extraction kit.....	177
5.2.2.2 Gel extraction using a GeneClean kit.....	178
5.2.3 Sequencing.....	178
5.2.4 Ligation of PCR product into vector	178
5.2.5 Transformation	179
5.2.6 Isolation of plasmid DNA.....	179
5.2.7 Restriction endonuclease digests	179
5.3 Results.....	181
5.3.1 Amplification of TGase 6 cDNA.....	181
5.4 Discussion.....	191
6.0 Analysis of patient sera.....	193
6.1 Introduction.....	193
6.2 Materials and Methods	194
6.2.1 Cell extracts	194
6.2.2 SDS-PAGE	194
6.2.3 Western blotting.....	194
6.3 Results.....	196
6.3.1 Analysis of patient sera samples.....	196
6.4 Discussion.....	200
7.0 Discussion.....	202
7.1 Determination of TGase 6 primary structure.....	202
7.2 Three-dimensional structure of TGase 6	202

7.2.1 Active site structure	204
7.2.2 Function of cis peptide bonds	207
7.2.3 Calcium binding sites	210
7.2.3.1 Ca ²⁺ ion binding site 1	210
7.2.3.2 Ca ²⁺ ion binding site 2	211
7.2.3.3 Ca ²⁺ ion binding site 3	212
7.2.4 Transamidase activation mechanism	215
7.2.4.1 Glutamine pocket	215
7.2.4.2 Lysine pocket	217
7.2.5 GTP binding	220
7.2.6 Conclusion	224
7.3 Expression of TGase 6	224
7.4 Future directions	226
References	228

Appendices

List of figures

Figure No.	Description
Figure 1.1	Reaction scheme of the TGase-catalysed transfer reaction
Figure 1.2	Schematic of the organisation of the identified TGase gene clusters
Figure 1.3	Phylogenetic tree of the TGase gene family
Figure 1.4	Genomic organisation and protein domains of TGase
Figure 1.5	Schematic of 5' flanking regulatory sequences involved in gene regulation of TGases
Figure 1.6	Cellular distribution and possible functions of various TGases
Figure 1.7	Possible role of TGase 2 in AD
Figure 1.8	Possible role of TGase 2 in HD
Figure 2.1	PCR strategy for amplification of cDNA sequences of TGase 6
Figure 2.2	Amplification of TGase 6 from H69 cells
Figure 2.3	Nucleotide and deduced amino acid sequence for the long form of TGase 6
Figure 2.4	Nucleotide and deduced amino acid for the short form of TGase 6.
Figure 2.5	PCR strategy for amplification of cDNA sequences of TGase 6 at the 5' end
Figure 2.6	Amplification of the 5'-end of TGase 6
Figure 2.7	Amplification of the 5'-end of TGase 6
Figure 2.8	Schematic of identified TGase 6 gene product
Figure 2.9	Amplification of the 5'-end of TGase 6
Figure 2.10	Amplification of the 5'-end of TGase 6
Figure 2.11	Size of transcripts of TGase 6 in H69 cells
Figure 2.12	Primer extension analysis of poly (A ⁺) RNA isolated from H69 cells.
Figure 2.13	Human TGase 6 sequence showing the major transcription start site
Figure 2.14	Amplification of the 5'end of human TGase 6
Figure 2.15	Sequence alignment of human and mouse TGase 6 gene in exon I
Figure 2.16	Structure of 3'-untranslated region of human TGase 6 gene
Figure 2.17	PCR strategy for amplification of cDNA sequences of mouse TGase 6
Figure 2.18	Amplification of TGase 6 from mouse brain using human primers.
Figure 2.19	Amplification of mouse TGase 6 from mouse brain

- Figure 2.20** Sequence alignment for human and mouse TGase 6
- Figure 2.21** Nucleotide sequence of proximal 5' region of mouse TGase 6 gene
- Figure 2.22** Comparison of amino acid sequence of human TGase 6 with the other members of the TGase family.
- Figure 3.1** Schematic of TGase 6 *in situ* probe
- Figure 3.2** Determination of generated DIG-labelled RNA probe
- Figure 3.3** Schematic of TGase 5 *in situ* probe
- Figure 3.4** Representation of tissue expression pattern for TGase 6
- Figure 3.5** *In situ* hybridisation analysis for TGase 6 on sagittal newborn mouse sections
- Figure 3.6** *In situ* hybridisation analysis for TGase 6 on sagittal newborn mouse sections
- Figure 3.7** Comparison of *in situ* hybridisation analysis with a luxol fast blue/cresyl violet stain
- Figure 3.8** *In situ* hybridisation analysis for TGase 5 on sagittal newborn mouse sections
- Figure 3.9** *In situ* hybridisation analysis for TGase 5 on sagittal newborn mouse sections
- Figure 3.10** *In situ* hybridisation analysis for TGase 6 on sagittal 11 day mouse embryo
- Figure 3.11** *In situ* hybridisation analysis for TGase 6 on sagittal 13 day mouse embryo
- Figure 3.12** *In situ* hybridisation analysis for TGase 6 on sagittal 16 day mouse embryo
- Figure 3.13** Overview of TGase 6 expression in mouse development
- Figure 3.14** Overview of *in situ* hybridisation analysis on the cerebrum
- Figure 3.15** Comparison of *in situ* hybridisation analysis for TGase 5 and 6 on mouse embryo tissue sections
- Figure 3.16** Amplification of TGases from mouse brain cDNA
- Figure 4.1** Sequence of TGase 6 β barrel domains ligated in pGEX2T
- Figure 4.2** Expression of C-terminal β barrel 1 and 2 domain of hTGase 6 in *E. coli* BL21

- Figure 4.3 Affinity purification of GST-fusion protein on glutathione-Sepharose 4B
- Figure 4.4 Image of the predicted structure of TGase 6_L
- Figure 4.5 Image of the predicted structure of TGase 6_S
- Figure 4.6 Comparison of peptide sequences
- Figure 4.7 Analysis of fractions for TGase 6_S peptide-hemocyanin conjugate
- Figure 4.8 Analysis of fluorescence of fractions for control experiment
- Figure 4.9 Analysis of fractions of TGase 6_L
- Figure 4.10 Analysis of fractions for TGase 6_L peptide-hemocyanin conjugate
- Figure 4.11 Characterisation of sera against peptide conjugates
- Figure 4.12 Characterisation of sera against GST-fusion proteins for TGase 6
- Figure 4.13 ELISA analysis to determine specificity of TGase 6_S sera
- Figure 4.14 ELISA analysis to determine specificity of TGase 6_L sera
- Figure 4.15 Detection of TGase 6 in extracted mouse brain tissue
- Figure 4.16 Analysis of affinity purified TGase 6_L antibodies
- Figure 4.17 Immunostaining of sagittal newborn mouse tissue sections with purified antibodies to TGase 6_L and TGase 5
- Figure 4.18 Immunostaining of sagittal mouse embryo sections with purified antibodies to TGase 6_L
- Figure 5.1 Amplification of TGase 6 from H69 cDNA
- Figure 5.2 Strategy for subcloning of PCR products to obtain full-length TGase 6 cDNA
- Figure 5.3 EcoRI endonuclease digestions on plasmid DNA containing insert 1, 2 and 3
- Figure 5.4 EcoRI endonuclease digestion on plasmid DNA containing insert 1.1.
- Figure 5.5 Schematic showing the cleavage sites for NcoI and XbaI endonucleases on insert 1.1
- Figure 5.6 Endonuclease digestion using NcoI and XbaI on plasmid DNA containing inserts 1.1 and 2
- Figure 5.7 Schematic showing the cleavage sites for NcoI and XbaI on insert 2.
- Figure 5.8 Schematic showing strategy for the joining of insert 1.1 and insert 2
- Figure 5.9 Schematic showing the cleavage sites for NcoI and XbaI on insert 2.
- Figure 5.10 Schematic showing cloning of insert 1.1 into the pCRII vector containing insert 2

- Figure 5.11 Endonuclease digestion of plasmid containing insert 2 with NcoI and XbaI
- Figure 5.12 Schematic showing the cleavage sites for EcoRI and BspEI on pCRII containing insert 1.1 and insert 2
- Figure 5.13 Schematic showing the cleavage sites for EcoRI and BspEI in the pCRII vector containing insert 3
- Figure 5.14 Schematic showing the cloning of insert 3 into the pCRII vector containing insert 1.1 and 2
- Figure 6.1 Western blotting analysis of patient sera samples
- Figure 7.1 Sequence comparison of flexible solvent exposed loop joining α -helical region of catalytic core domain to first β -strand of barrel 1
- Figure 7.2 Ribbon image of the predicted structure of TGase 6
- Figure 7.3 Reaction pathway and a proposed mechanism for TGase 6-catalysed transamidations
- Figure 7.4 View of key residues involved in the active site of TGase 6
- Figure 7.5 Superposition of TGase 6 and TGase 3 residues involved in *cis* peptide formation
- Figure 7.6 Superposition of key TGase 6 and TGase 3 residues involved in Ca^{2+} binding.
- Figure 7.7 View of key residues involved in Ca^{2+} ion binding sites and active site of TGase 6.
- Figure 7.8 View of key residues of TGase 6 located in the glutamine pocket
- Figure 7.9 View of key residues of TGase 6 located in the lysine pocket
- Figure 7.10 Superposition of key residues found in guanine nucleotide-binding site pocket in TGase 6 and TGase 3 complex

List of tables

Table No.	Description
Table 1.1	TGase terminology
Table 1.2	Summary of parts of the brain affected by neurodegenerative diseases
Table 2.1	Nucleotide sequence of exon/intron/exon junctions for human TGase 6
Table 2.2	DNA structure of human TGases
Table 2.3	Nucleotide sequence of exon/intron/exon junctions for mouse TGase 6
Table 7.1	Sequence alignment of human and mouse TGase 6 with other members of TGase family in the regions of the <i>cis</i> peptide bonds.
Table 7.2	Tabulated drawing of key residues involved in two parts of the TGase reaction
Table 7.3	Amino acid sequence alignment of TGases around the guanine nucleotide-binding site pocket

1.0 Introduction

1.1 Introduction

Transglutaminases (TGases) are known as a widely distributed group of enzymes which catalyse a post-translational modification of proteins through the formation of isopeptide bonds. This transpires either by protein crosslinking via ϵ -(γ -glutamyl)lysine bonds or by incorporation of primary amines into specific glutamine residues.

Proteins exhibiting TGase activity have been discovered in organisms ranging from mammals to invertebrates. However, mammalian TGases are the most well characterised. Similarly to other significant cellular functions found in mammalian cells, TGases need calcium binding for their activity (Burgoyne and Weiss, 2001), which essentially signifies that TGases are effectively inactive under normal conditions. The crosslinked products formed by the action of TGases have been found to accumulate in several tissues and organ systems including skin, bone, and wound healing. TGases have also been shown to contribute to several human diseases such as neurodegenerative disease (Lesort et al., 2000), autoimmune diseases, for example, celiac disease (Dietrich et al., 1997) and diseases involving the epidermis of skin.

1.1.1 Enzyme mechanism

TGases catalyse the calcium-dependent acyl transferase reaction (EC2.3.2.13), which results in the formation of an isopeptide bond between the γ -carboxamide group of a peptide-bound glutamine residue and various primary amines (Folk and Finlayson, 1977, Lorand and Conrad, 1984). A glutamine residue fulfils the role of acyl donor and the most typical acyl-acceptors are ϵ -amino groups of peptide-bound lysine residues or primary amino groups of polyamines (Fig. 1.1). The reaction is a multiple stage process whereby the first stage involves the active site cysteine reacting with the γ -carboxamide group of the glutamine residue to form a γ -glutamylthioester (known as the acyl enzyme intermediate). This results in the release of ammonia (Pedersen et al., 1994), (Fig. 1.1 A). In the second stage, the acyl enzyme intermediate then reacts with a nucleophilic primary amine which attacks the thioester bond resulting in the

formation of an isopeptide bond. This regenerates the active-centre cysteine residue of the enzyme to its original form, thereby allowing it to participate in further cycles of catalysis. The rate-limiting step in these reactions is thought to be the formation of the covalent acyl enzyme intermediate (Curtis et al., 1974).

Most commonly, γ -glutamyl- ϵ -lysine crosslinks are generated in or between proteins by reaction with the ϵ -amino group of lysine residues (Fig. 1.1 B1). This results in the formation of covalently crosslinked, generally insoluble structures, such as the fibrin clot formed in haemostasis, skin cornified envelopes and a variety of extracellular matrix structures (Aeschlimann and Thomazy, 2000). The reaction with polyamines (Fig. 1.1 B2) generates protein modifications which possibly have an effect on the biological activity or turnover of the target protein. In an environment which lacks appropriate amines for crosslinking, TGases can also react with water (Fig. 1.1 B3) to hydrolyse glutamine to glutamate (Folk and Finlayson, 1977). This reaction is thought to be of biological importance with regard to celiac disease (Anderson et al., 2000).

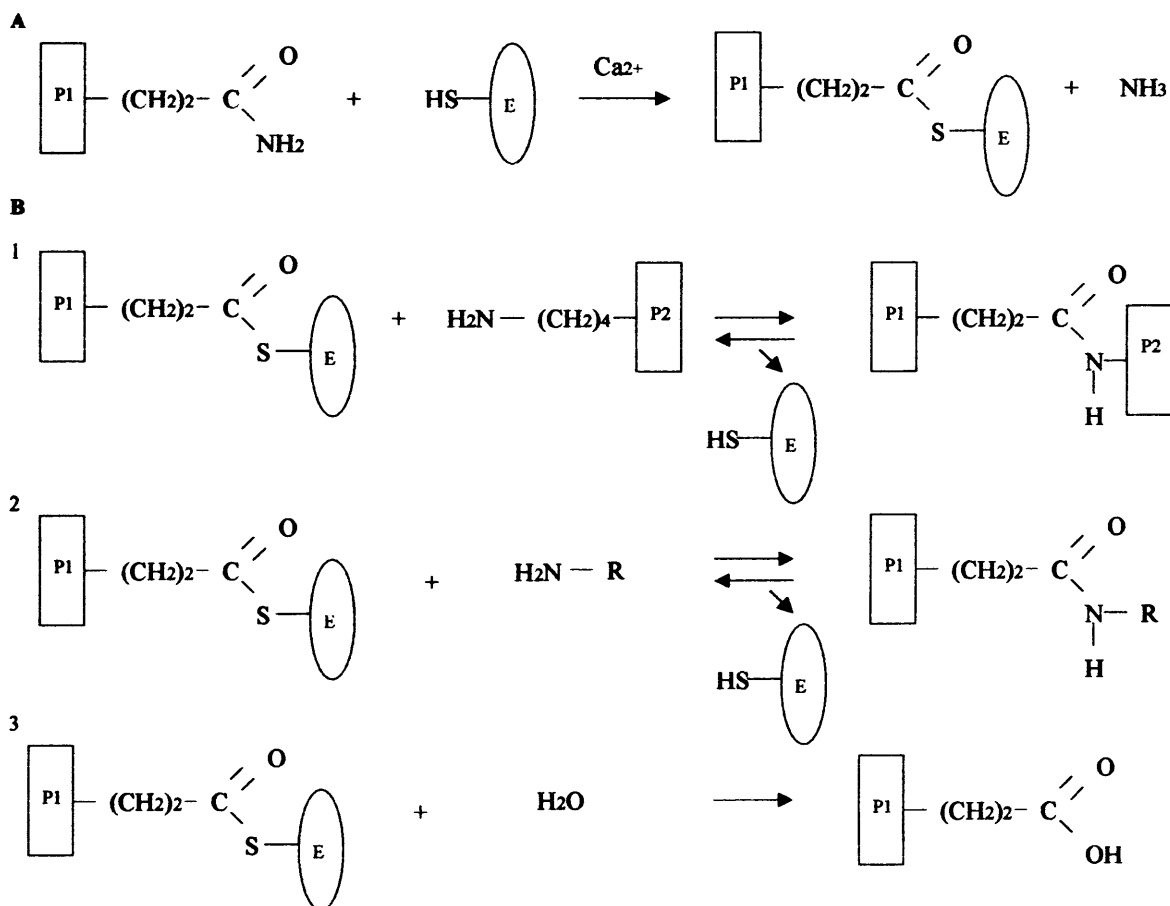


Fig. 1.1 Reaction scheme for the TGase-catalysed transfer reaction. E represents the TGase enzyme and P represents the protein.

1.1.2 The Transglutaminase family

Due to their unique physical and biochemical properties, six different TGases were initially identified in tissue extracts and purified using classical biochemical separation techniques (Aeschlimann and Paulsson, 1994). The circulating zymogen factor XIII which is modified by a thrombin-dependent proteolysis into the active factor XIIIa has a role in the stabilisation of fibrin clots and in wound healing. Keratinocyte TGase (TGase 1), which occurs in membrane-bound and soluble forms, is activated by proteolysis and is associated with terminal differentiation of keratinocytes. Tissue TGase (TGase 2) is found in a number of cell types and tissues in the body. However, regardless of its abundance and the large number of studies carried out, its function remains elusive. It has been implicated in a number of diverse processes such as stabilisation of the extracellular matrix in development and wound healing, GTP binding and hydrolysis, in signalling, and intracellular crosslinking in programmed cell death (Aeschlimann and Thomazy, 2000). Epidermal TGase (TGase 3) also needs proteolysis to become active, and similar to TGase 1, has a role in the terminal differentiation of keratinocytes. Prostate TGase (TGase 4) is an androgen-regulated protein which has a role in semen coagulation. Band 4.2 protein is a membrane cytoskeleton component expressed at a high level in erythroid cells. It is a unique member of the TGase family as it is the only one to have lost its enzymatic activity and to adopt a purely structural role.

An assay for detection and identification of TGase gene products based on RT-PCR with degenerate primers has been developed by our lab. Using this technique, the gene product of the TGM5 gene was discovered in keratinocytes (Aeschlimann et al., 1998). Through the use of restriction mapping and PCR analysis, it was revealed that the TGM5 gene is located ~11kb upstream of the EPB42 gene on chromosome 15q15 (Grenard et al., 2001). From additional sequence analysis of this sequence locus, another novel TGase gene product, TGase 7 was revealed and the respective gene product was isolated from human prostate carcinoma cells. This novel gene, TGM7, is located ~9kb upstream of the TGM5 gene. The corresponding region is located on chromosome 2, 2F1-F3 in mouse, and using radiation hybrid mapping, it was shown that orthologous genes for all three human genes are located at this region of mouse chromosome 2. Of interest is the fact that the syntenic region of human chromosome

20q11 which harbours the genes for TGase 2 and TGase 3, is positioned adjacent to the identified TGase gene cluster on distal mouse chromosome 2. Based on comparative analysis of amino acid sequence and gene structure, the genes encoding TGase 2 and TGase 3 are more closely related to the TGM5, TGM7 and EPB42 genes in comparison to the other TGase genes. Analysis of genomic sequences also identified a third gene, TGM6, which is adjacent to the TGM3 gene in the locus on chromosome 20q11, thereby identifying a second cluster comprising of TGM3, TGM6 and TGM2. Using radiation hybrid mapping, the chromosomal location of the mouse *tgm2* and *tgm3* genes have been determined, and positions were identified approximately 20cM and 5cM distal to the *tgm7/tgm5/epb42* gene cluster, respectively (Grenard et al., 2001). This positions six TGase genes on distal mouse chromosome 2, indicating that initially a single TGase gene locally duplicated, and that this was succeeded by a larger section of the chromosomal region duplicating, resulting in the organisation of the genes identified in mouse. It is thought that in humans, these chromosomal regions redistributed to two different chromosomes (Fig. 1.2).

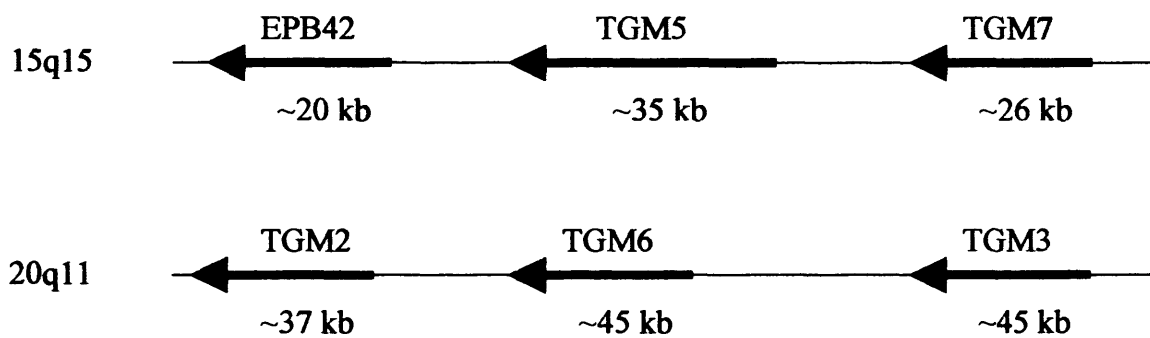


Fig. 1.2 Schematic of the organisation of the identified TGase gene clusters in the human genome.

Therefore, there are currently nine different TGase genes that have been identified in the human genome. Various terms have been used to describe individual TGase gene products so a list of the genes and alternative names for the gene products is given in Table 1.1.

Table 1.1 TGase terminology.

Gene product	Alternative names	Gene name
Factor XIIIa	plasma TGase, fibrin stabilising factor, fibrinolygase	F13A1
TGase 1	TG _K , keratinocyte TGase, TGase type 1	TGM1
TGase 2	TG _C , tTG, cytosolic TGase, endothelial TGase, liver TGase, tissue TGase, TGase type 2, Guh	TGM2
TGase 3	TG _E , epidermal TGase, hair follicle TGase, TGase type 3	TGM3
TGase 4	TG _P , prostate TGase, TGase type 4, androgen regulated secretory protein	TGM4
TGase 5	TG _X , TGase type 5	TGM5
TGase 6	TG _Y , TGase type 6	TGM6
TGase 7	TG _Z , TGase type 7	TGM7
Band 4.2	erythrocyte protein band 4.2	EPB42

1.1.3 The transglutaminase gene family

All vertebrate TGases appear to be encoded by a family of closely related genes. Alignment of the gene products reveals a high degree of sequence similarity and all family members exhibit a similar gene organisation, with conservation of intron distribution and intron splice sites. Comparison of the structure of the individual TGase genes shows that they may be divided into two subclasses (Aeshlimann and Paulsson, 1994), wherein the genes encoding TGase 2 (Fraij and Gonzales 1997), TGase 3 (Kim et al., 1994), band 4.2 protein (Korsgren and Cohen, 1991), TGase 4 (Dubbink et al., 1998), TGase 5, TGase 6 and TGase 7 (Grenard et al., 2001), contain 13 exons and the genes encoding factor XIIIa-subunit (Ichinose and Davie, 1998) and TGase 1 (Kim et al., 1992) contain 15 exons. Exon IX of the former group is separated into two exons (X and XI) in TGase 1 and factor XIIIa, and the amino-

terminal extensions of TGase 1 and factor XIIIa are comprised of an additional exon. Phylogenetic analysis also indicated that an early gene duplication event gave rise to two different lineages, one comprising TGase 2, TGase 3 and band 4.2 protein, and the other, factor XIII a-subunit and TGase 1.

More recently, a more detailed analysis of the evolutionary relationship between the TGase genes was carried out, based on new data. The amino acid similarity based on the sequence alignment was calculated, and the evolutionary distances derived using different algorithms. All algorithms predicted a close relationship between factor XIIIa-subunit and TGase 1 (lineage 1), and TGase 5, TGase 3, band 4.2 protein and TGase 2 (lineage 2), respectively (Fig. 1.3) (Grenard et al., 2001). The genomic organisation and gene structure also supports the close relationship of the genes and corroborates the placement of these in a separate phylogenetic branch. The exact relationship of TGase 4 to these two lineages is less certain, but is likely to have branched off from lineage 2 approximately at the same time as factor XIIIa and TGase 1 diverged. Only a single TGase gene has been identified in genomes of invertebrate species, and it is therefore likely that the separation of TGase into four branches is likely to have occurred after divergence of invertebrates from proto-vertebrates. A similar relationship between invertebrate and vertebrates has been found for many orthologous genes, and it has recently been proposed that octaploidy in early vertebrates may have resulted in two successive genome duplications in early vertebrates (Spring, 1997). The phylogenetic analysis by Grenard and colleagues (Grenard et al., 2001) is consistent with this model. It is thought that one branch of lineage 2 has subsequently undergone multiple duplications locally to generate clusters of TGase genes on mouse chromosome 2 containing a total of six genes.

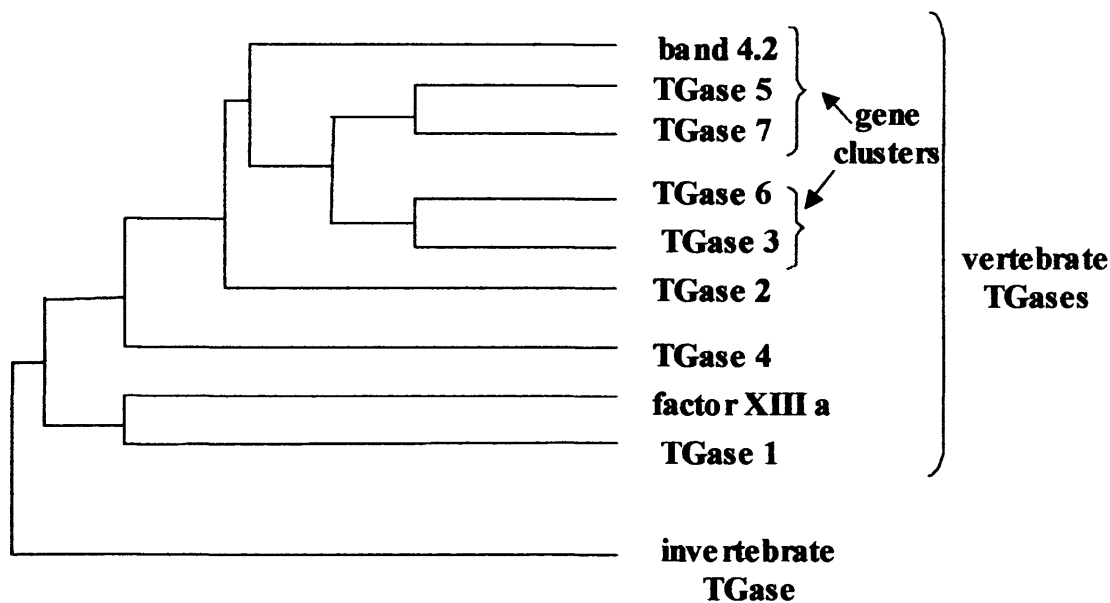


Fig. 1.3 Phylogenetic tree of the TGase gene family. A hypothetical pedigree for the gene family is given that is consistent with the data on the sequence relationship of the individual gene products as well as with the data on gene structure and genomic organisation. Adapted from (Grenard et al., 2001).

1.1.4 Domain organisation

The initial structural studies on TGases were carried out on the A subunit zymogen of factor XIII which is proteolytically processed by thrombin to produce the active enzyme (Yee et al., 1994, Weiss et al., 1998). Studies showed that each factor XIII subunit consists of four domains designated the N-terminal β -sandwich, the core domain (containing the catalytic and the regulatory sites) and the C-terminal β -barrels 1 and 2, and that two monomers form into the native dimer via the surfaces in domain 1 and 2. The core domain is highly conserved among the different members of the TGase family, likely based on the conservation of function during evolution. The N- and C-terminal domains show some conservation, while the most variation is seen in the region (about 30 amino acids in length) linking the core domain to the β -barrel domains. This is an unstructured region which is vital in the regulation of catalytic activity (Casadio et al., 1999). Additionally, tail extensions are found in some of the enzymes, highlighting that the specialised functions of the individual TGases are based on adaptation of N- and C-terminal regions.

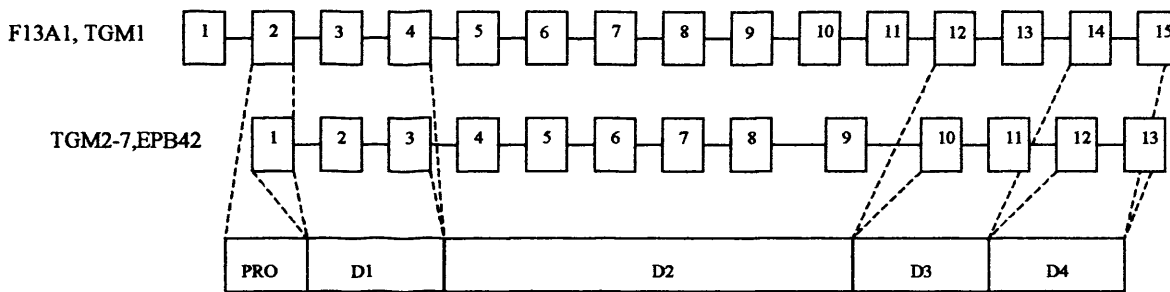


Fig. 1.4. Genomic organisation and protein domains of transglutaminase. The genes of factor XIII (F13A1) and TGase 1 (TGM1) contain 15 exons (numbered) and 14 introns, whereas the genes encoding TGase 2-7 (TGM2-7) and band 4.2 (EPB42) contain 13 exons (numbered) and 12 introns. In F13A1 and TGM1, the exon 9 equivalent is divided by an extra intron to give exons 10 and 11. Exon 1 is non-coding in these genes, and exon 2 encodes an amino-terminal propeptide. The exons encoding the four domains (d1-d4) of the mature protein are indicated by broken lines. Adapted from (Lorand and Graham, 2003).

All TGase members lack classical glycosylation and disulphide bonds, regardless of the presence of several potential sites for N-linked glycosylation, and a number of cysteine residues, respectively. In addition, TGases (even those that are secreted, e.g. factor XIII and TGase 2), have no amino-terminal hydrophobic leader sequence which is common of secreted proteins. The majority are post-translationally modified by the removal of the initiator methionine residue and N-acylation of the penultimate amino acid residue. Importantly, all members of the vertebrate gene family require calcium for activation which is discussed further in chapter 7 (except possibly plant and bacterial TGases, but this is under further investigation).

1.1.5 Regulation of expression

The genes for the TGase family show a significant similarity in their organisation, however, in the 5' flanking regions, there is very little homology between their nucleotide sequences and the mechanisms each gene has adapted to control its expression differs between the different TGase genes.

1.1.5.1 Transglutaminase 2

The promoter of the TGM2 gene is complex which could be a reflection of the various roles the enzyme plays in different biological situations. It appears that TGase 2 has a typical type II promoter structure (Lu et al., 1995): it contains a well-defined TATA box element situated 24 nucleotides upstream from a canonical cap site. An extensive GC-rich region upstream from the TATA box element which encodes at least two consensus Sp1-binding sites is also incorporated in the promoter, and a canonical CAAT box element is located further upstream.

Several important morphogens, including retinoids and members of the TGF β gene family, have been shown to have a role in the regulation of TGase 2 expression. Retinoids have been found to induce TGase 2 expression in several different cell types and due to the cloning of the promoter region of TGase 2, it has been revealed that consensus motifs of the oestrogen receptor-type binding site are present which may have a function as retinoic acid response elements. The nuclear receptors for trans-retinoic acid RAR- β and RAR- γ , as well as *cis*-retinoic acid, RXRs, have been shown to be the mediators of retinoic acid-induced TGase 2 expression (Gentile et al., 1992). Several *in vitro* studies with receptor selective agonists and antagonists have implicated that various retinoid receptor combinations are involved in the up-regulation of TGase 2, which suggests that different retinoid receptor combinations may direct regulation of TGase 2 in different biological contexts.

In the mouse TGase 2 gene promoter, a response element for members of the transforming growth factor- β (TGF- β) protein family has been identified 868 nucleotides upstream of the transcriptional start site (Ritter and Davies, 1998). It has been shown in various cells *in vitro* that this element can direct the bone morphogenetic protein-2 (BMP-2) and BMP-4-dependent repression of promoter activity. Using the same motif, TGF- β 1 has a role in either inducing or repressing TGase 2 expression in a cell type-specific manner (Nunes et al., 1997), and since the activation of TGF- β 1 is induced by TGase 2, it is possible that an amplification loop is initiated in which TGF- β 1 induces synthesis of TGase 2, which subsequently converts latent to active TGF- β 1.

Pro-inflammatory cytokines such as interleukin-6 (IL-6) also induce TGase 2 expression. It has been demonstrated that IL-6 plays a role in the up-regulation of TGase 2 in hepatocytes on a mRNA and protein level (Suto et al., 1993), and the recent cloning of the promoter region of TGase 2 revealed potential responsive elements for IL-6 (Suto et al., 1993a).

1.1.5.2 Transglutaminase 1

TGase 1 also contains a TATA box-like motif (ATAAA) upstream of the cap site, however, in contrast to TGase 2, it lacks a putative CAAT box element (Medveder et al., 1999). Deletion analysis positioned the minimal promoter at -37/+60 and identified two major sites that are of importance in the transcriptional control of TGase 1: a palindromic sequence located from -490 to -470bp (TGpal) containing a CREB/AP-1 like site in the centre, and the region from bp -53 to -37 containing a functional Sp1 site. Positioned just upstream from the TATA-like box is a GC-rich region incorporating the Sp1 consensus element. Another consensus Sp1 sequence is located at bp-194 but no evidence has been found to indicate that this site is of any importance in regulating the TGase 1 gene. It has been indicated that Sp1 elements have a role in the regulation of several genes during squamous differentiation. For, optimal transcriptional activation of the involucrin gene, the Sp1 site is required and the protein complexes which bind to this site include Sp1.

The functionally important CREB/AP-1 like element TGATGTCA was found contained in a 22bp palindrome. It has been demonstrated that several transcription factors, including members of the CREB and Fos/Jun families can bind to the CREB/AP-1 like element. It has also been suggested that CREB/AP-1 like elements have a role in the transcriptional regulation of several other squamous markers, including involucrin. It is probable that different CREB/AP-1 protein complexes are involved in controlling the regulation of squamous-specific genes induced at various stages during differentiation. Studies have shown that PKC activation, which regulates the expression of Jun/Fos family members, is an important factor in signalling pathways controlling squamous differentiation and TGase 1. This further supports a role for CREB/AP-1 like elements. However, the role of specific CREB/AP-1 complexes in squamous differentiation is still not well defined.

1.1.5.3 Transglutaminase 3

The correct transcription of the human TGase 3 gene has been found to depend on the simultaneous effect of elements located in both the proximal promoter and in regions distal from it (Lee et al., 1996). The distal elements are necessary for late epidermal differentiation-specific transcription. The proximal promoter of the TGase 3 gene however, contains the information to direct expression in stratified squamous epithelia. This region has been mapped to sequences between –126 and –91, which contains Sp1 and Ets-like binding sites. The overall activity of the proximal promoter region is a result of the co-operative interactions between these positively acting motifs.

The Sp1 transcription factor is important for the regulation of several epidermally expressed genes. However, the TGase 3 gene is the first epidermally expressed gene in which the Ets transcription factors have been shown to be of importance. These transcription factors have been implicated in the regulation of gene expression during several biological processes, including growth control and developmental or transformation programmes. Interestingly, in the TGase 3 gene proximal promoter region, the interactions not only involved Sp1, but also additional proteins, indicating that the relationship between the Ets and Sp1 factors is modulated by interactions with additional unidentified nuclear proteins (Lee et al., 1996).

1.1.5.4 Factor XIII

In contrast to the other TGase genes which have their translation start sites inside exon 1 adjacent to the transcription start site, exon 1 of the factor XIIIa gene encodes only a 5'-untranslated region, and its translation start site (ATG) is present in exon II (Kida et al., 1999). The factor XIIIa subunit promoter is a TATA box-less promoter. Studies have revealed several *cis*-elements and *trans*-acting factors involved in the expression of factor XIIIa (Kida et al., 1999). A potential NF-1 element was located at bp-13 to –39 and studies revealed that protein binding to this element was abundant. However, it was demonstrated that the NF-1 element alone was not sufficient to drive reporter gene expression. The MZF-1/Sp-1 element found at bp-92 to –73 appears to be most essential for the expression of factor XIIIa as removal or mutation of this site

abolished promoter activity (Kida et al., 1999). *cis*-acting elements located upstream from the proximal promoter region also play a role in modulating the transcriptional activity. Two MZF-1/Sp-1 elements were found located adjacent to one another at bp-260 to -223 (Kida et al., 1999). Studies revealed that the nuclear protein which bound to the first MZF-1/Sp-1 element (at bp-233) showed low affinity for MZF-1 consensus oligonucleotides regardless of the fact that the MZF-1 sequence at bp-223 is exactly the same as the sequence at bp-260. It has therefore been suggested that the MZF-1-like protein which binds to the element at bp-223 may differ from those that bind at the other two MZF/Sp-1 sites, and it may be a novel MZF-1 protein.

Previously, it has been found that GATA-1 and Ets proteins have important roles in the expression of several genes in the megakaryotic lineage. It has been revealed that a GATA-1 and two Ets-1 motifs are located in an enhancer region of factor XIIIa, and that nuclear proteins bind to these sites (Kida et al., 1999). Studies using MEG-01 cells revealed that mutation of the GATA-1, but not of Ets elements resulted in a significant decrease in promoter activity. Therefore it has been concluded that the GATA-1 motif plays a role in the megakaryocytic expression of the factor XIIIa gene. Interestingly, a short tandem repeat (STR) polymorphism is present in the 5'-flanking region of factor XIIIa. It is believed that since the GATA-1 site and STR polymorphism are adjacent to each other, the STR polymorphism may be involved in activation of expression by the GATA proteins and/or modifying their binding affinity to DNA. In conclusion, the possible interaction of these proteins (MZF-1-like protein, Sp1 and GATA-1 and GATA-2) with factor XIIIa promoter elements likely direct factor XIIIa expression, resulting in a complex pattern of tissue- and cell type-specific gene regulation.

1.1.5.5 Transglutaminase 4

Sequence analysis has revealed that the TGase 4 proximal promoter contains a TATA-box like CATAA motif, located 49bp upstream from the major transcription start site, but lacks a typical CCAAT box. The region between positions -113 and -61 was found to be essential for core promoter activity (Dubbink et al., 1998). Further analysis demonstrated the functional significance of an Sp1 binding motif, 5'-

ACCCCGCCCC-3' located at position -96 to -87. This sequence was shown to bind the ubiquitous transcription factors Sp1 and Sp3 (Dubbink et al., 1999).

A processed cyclophilin A pseudogene (PPIP) was found located at position -1276 to -563, which has a 94% similarity to cyclophilin cDNA. Due to the absence of a correct start codon (position -1263) and the disruption of the open reading frame by deletions and insertions, this gene cannot produce a functional protein product. Functional analysis of the TGase 4 gene promoter demonstrated that basal activity is located within the proximal 500bp of the TGase 4 promoter and that no extra activity is comprised by the larger constructs which include the PPIP pseudogene, and more distal promoter information (Dubbink et al., 1998). However, the incorporation of the PPIP pseudogene and repetitive elements in the TGase 4 promoter do not necessarily determine the 5' boundary of the regulatory region, as, for example, the prostate-specific antigen (PSA) promoter includes an upstream enhancer responsible for androgen-regulated and tissue-specific expression which is distanced from the proximal promoter by several Alu-type repeats.

1.1.5.6 Erythrocyte protein band 4.2

In the band 4.2 gene, the transcription start site was found to be located at nucleotide -226 (Korsgren and Cohen, 1991). A comparison of the band 4.2 sequence with the genomic sequences for two other erythrocyte proteins, β -globin and porphobilinogen deaminase enabled the identification of the five possible regulatory elements: (1) a possible TATA element located -27 to -21nt upstream from the start site; (2) a short GC rich domain which could represent an Sp1 binding site (-41 to -47nt upstream); (3) a possible CAAT box (-58 to -66nt); (4) a CAAC box (-79 to -88nt) and (5) two GF-1 binding domains, located at -23 to -28 and -173 to -178. These elements are spaced a similar distance from the transcription start site and have a comparable spacing and order in all of these genes.

1.1.5.7 Transglutaminase 5

Primer extension experiments revealed that the major transcription initiation site used in keratinocytes is located 157nt upstream of the translation start codon (Grenard et al., 2001). Computer analysis was carried out on the proximal promoter region in order to determine potential binding sites of transcription factors. A classical TATA-box sequence was not found in the TGase 5 gene but other potential transcription factor binding sites were identified, suggesting that the TGase 5 promoter is a TATA-less promoter. A number of TATA-less genes have shown interaction of CCAAT/enhancer-binding protein, NF-1 and upstream stimulatory factor (USF) to form a core proximal promoter, and this may also be the case for TGase 5. Ap1, Ets and Sp1 elements are all typically found in keratinocyte-specific genes as discussed for TGase 1 and TGase 3, and may regulate transcription in keratinocytes. In TGase 5, several Ap1 sites are located within 2.5kb of upstream sequence and may interact with the proximal Ap1 factor for activation.

Unfortunately no experimental information is available for the promoter regions of TGase 5, nor for TGase 6 and TGase 7 at this time.

1.1.5.8 Conclusion

In contrast to the significant similarity the TGases exhibit in their gene organisation, their 5' nucleotide sequence differs, and the mechanisms for their gene regulation seems to be quite diverse. For example, the factor XIII gene is regulated mainly by transcription factors MZF, NF-1 and Sp1. To compare TGase 1, TGase 2 and TGase 3, all three contain consensus Sp1 recognition motifs near the transcription start site. The Sp1 transcription factor is important for the regulation of several epidermally expressed genes. Interestingly, the TGase 3 gene has also been found to contain Ets-like motifs. This suggests that members of the Ets family of transcription factors may play an important role in the regulation of late differentiation of genes in the epidermis (Lee et al., 1996). Due to the variation in transcription factors found within certain genes but also the similarity within others which are expressed in similar tissues, it is possible that the transcription factors play a major role in the cell type-

specific expression of each TGase. A schematic of the known 5' regulatory sequences for TGases is given in Fig. 1.5.

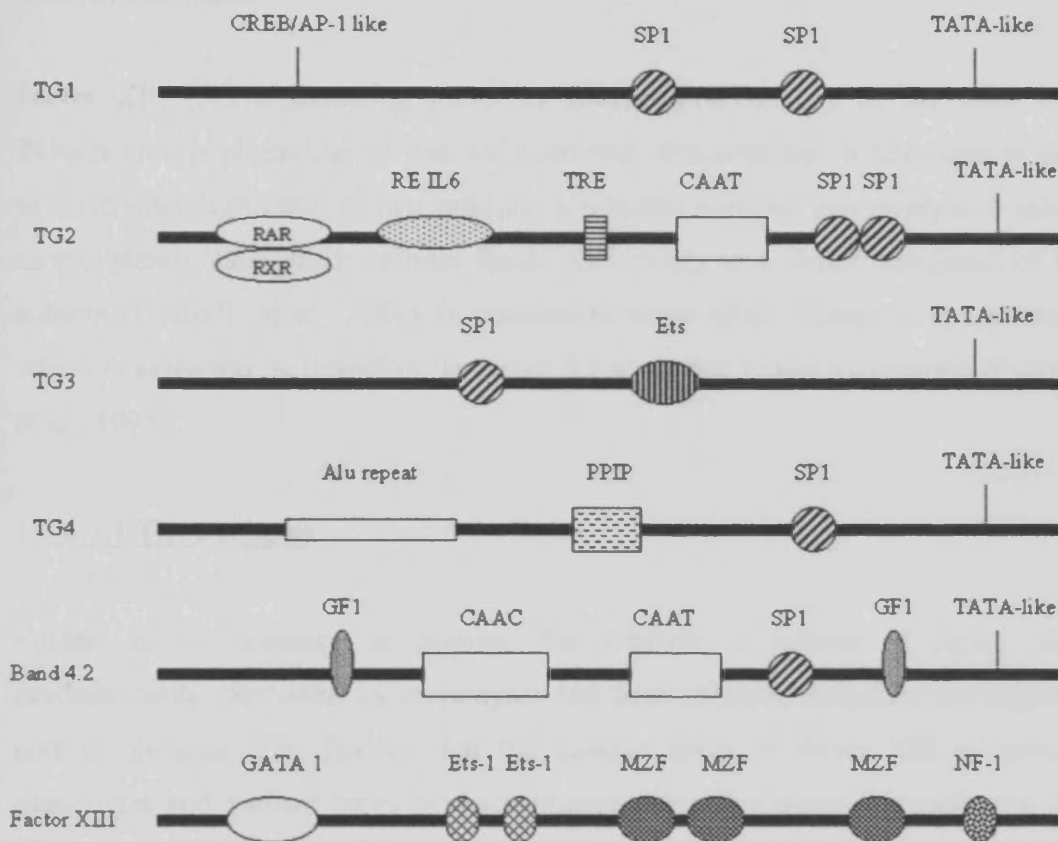


Fig. 1.5 Schematic of the 5' flanking regulatory sequences involved in the gene regulation of TGases. Only the TGases whose sequence of transcription factors has been experimentally determined are shown in the schematic.

1.1.6 Members of the TGase family and their functions

1.1.6.1 Factor XIII

Factor XIII (fibrin-stabilising factor or fibrinoligase) is one of the most studied TGases and its physiological role has been well characterised. It circulates in plasma as a tetramer composed of two catalytic α subunits and two non-catalytic β subunits, noncovalently associated: cellular factor XIII exists as a dimer composed of two α subunits (Carrell, et al., 1989). In contrast to many other TGases, it is a proenzyme which is activated by thrombin (to factor XIIIa) during blood coagulation (Greenberg et al., 1991).

1.1.6.1.1 The α subunit

Further to its presence in plasma, the catalytic α subunit of factor XIII is predominantly expressed by monocytes and macrophages, megakaryocytes/platelets and in placenta. The finding that the cellular form of factor XIII is present in monocytes and various types of macrophages, including tissue macrophages, shows that it is present in most of the organs and tissues of the body (Kida et al., 1999).

The complete amino acid sequence of the human placental factor XIII α subunit was determined by a combination of cDNA cloning and protein sequencing techniques (Ichinose et al., 1990). It consists of 731 amino acid residues with a molecular mass of ~83kDA and is identical with the primary structure of cellular factor XIIIa. Six potential glycosylation sites are present, but no carbohydrate modification was detected. The cDNA sequence reveals that the enzyme does not contain a typical hydrophobic leader sequence for secretion. This could explain the presence of the protein in the cytoplasm and is consistent with the lack of carbohydrate or disulphide bonds in the protein.

The gene for the α subunit is located on chromosome 6p24-25. The gene consists of 15 exons and 14 introns which spans about 180kb (Ichinose et al., 1990). The thrombin cleavage sites (exons 2 and 12), potential calcium binding sites (exons 6 and

11) and active centre sequence (exon7) are encoded by separate exons, indicating that the introns separate the a subunit into specific structural and functional domains (Greenberg et al., 1991).

1.1.6.1.2 The active site

In the crystal structure, the Cys314 residue is located at the amino terminus of the longest helix in the subunit. Cys314 is known to be an essential amino acid for factor XIII activity. The Cys314 side-chain is hydrogen bonded to His373, which then interacts with the side chain of Asp396. All three residues are conserved among the members of the TGase family which are enzymatically active, and result in a catalytic triad which is comparable to the Cys-His-Asn triad found in the papain family of cysteine proteases. The active site helix containing the nucleophilic cysteine residue and an adjacent four-stranded β sheet containing the His and Asp/Asn residues are found in both factor XIII and the cysteine proteases, indicating that the structure of the proteins in the active site area is also conserved. The order of the catalytic triad residues and the order and direction of the helix and β strands are conserved. Due to these findings, it has been proposed that factor XIII and cysteine proteases are evolutionarily related and have diverged from a common ancestral protein (Pedersen et al., 1994). However, even though the overall organisation of the cysteine proteases and factor XIII catalytic domains are comparable, protein folding within the two structures is very different. Therefore, the structural similarity between the two is limited (Yee et al., 1996).

1.1.6.1.3 The b subunit

The primary structure of the b subunit of factor XIII was derived by cDNA cloning (Ichinose et al., 1986) and was found to consist of 641 amino acids and the addition of 8.5% carbohydrate gives a molecular mass of ~ 80kDa (Ichinose et al., 1990). The b subunit is synthesised in the liver (Sakata and Aoki, 1982). It consists of ten tandem repeats of about 60 amino acids each, which are called GP-I structures as they were first identified in β 2-glycoprotein I (Lozier et al., 1984), or Sushi domains because of their shape (Ichinose et al., 1990). The gene for the b subunit spans ~28kb and was

localised to chromosome 1q32-32.1 (Webb et al., 1989). It consists of 12 exons interrupted by 11 introns. The first exon encodes a 20 amino acid leader sequence and the last exon codes for the COOH-terminal region of the protein. The remaining 10 exons code for the 10 sushi structures, indicating that each sushi structure is encoded by a single exon in the gene (Greenberg et al., 1991). Similar repeats have been found in a total of 53 genes by the human genome project (Venter et al., 2001) and these include β 2-glycoprotein I, complement protein H and decay accelerating factor (Greenberg et al., 1991). As their physiological function most often involves binding to other proteins, it is likely that the sushi domains in the b subunit serve as a protein binding module.

1.1.6.1.4 Activation of factor XIII

In contrast to many TGases, factor XIII is a proenzyme which is converted in blood to the active enzyme by thrombin, in the presence of fibrin. First, thrombin cleaves the Arg37-Gly38 peptide bond in factor XIIIa, resulting in an activation peptide consisting of 37 amino acid residues being separated from the amino terminus of the a subunit. It has recently been shown that after being cleaved by thrombin, the activation peptide stayed attached to the molecule at its original position obstructing the active site cavity (Yee et al., 1995). In the presence of calcium ions, the a subunits dissociate from the b subunits, unmasking the active site cysteine within the catalytic core domain. Fibrinogen is responsible for the reduction in calcium concentration needed for b subunit dissociation to levels that are present in plasma (Credo et al., 1978). Polymerising fibrin molecules provide binding sites for factor XIII and thrombin, thereby increasing the rate of thrombin cleavage (Lorand et al., 1993). The ability of fibrin to contribute to factor XIII activation ensures that factor XIIIa is produced after fibrin and restricts the production of factor XIIIa in plasma. Plasma comprises of ~1 factor XIII molecule to every 100 fibrinogen molecules. It is necessary for factor XIIIa to interact with several sites on the fibrin molecule, in order for a significant number of isopeptide bonds in fibrin to be catalysed during blood clot formation. Specific substrate binding sites have been determined for factor XIII and factor XIIIa on fibrinogen and fibrin (Achyuthan et al., 1996).

1.1.6.1.5 Function of factor XIII a subunit

The active form of factor XIIIa, plays a role in the normal hemostasis by producing a mechanically stronger, more rigid and more elastic clot which is less susceptible to fibrinolysis. Following tissue injury, formation of a blood clot is needed to restore vascular integrity and to promote a provisional matrix for the initiation of wound repair (Mosesson, 1992). Fibrin and plasma fibronectin (FN), the main components of the clot, are crucial to these functions. Clot polymerisation commences when soluble fibrinogen is transformed to fibrin by thrombin. The fibrin monomers are then able to assemble into polymers (Furie and Furie, 1998). Following this, soluble FN is incorporated into the clot with fibrin. As the clot matures, intermolecular crosslinking occurs between fibrin molecules and between FN and fibrin molecules, catalysed by factor XIIIa (Mosher, 1976). Covalent crosslinking between fibrin molecules is crucial for the structural stability of the clot and the presence of FN is essential as it provides additional binding sites for cell attachment and migration into the wound. Another important reaction catalysed by factor XIIIa is the crosslinking of α_2 -antiplasmin to fibrin (Sakata and Aoki, 1982), which increases the resistance of the clot to plasmin.

Factor XIII is also known to have a role in maintaining pregnancy as women with homozygous factor XIII deficiency are not able to carry a pregnancy to term and lose their fetus by habitual abortion. During pregnancy, factor XIII is found within the placenta. Factor XIII acts as an adhesive protein during the cytotrophoblast invasion of the endometrium and is vital shortly after 4-5 weeks of gestation (Asahina et al., 1998). If the concentration of factor XIIIa is low in the placental bed, the cytotrophoblastic shell does not properly form and increases the risk of abortion (Asahina et al., 2000).

1.1.6.1.6 Function of factor XIII b subunit

The literature available so far does not provide a clear answer as to the function of the b subunit of factor XIII. However, the most conclusive results suggest that it functions as a stabiliser to the a subunit (Karges and Metzner, 1996). It has also been suggested

that subunit b: 1) protects the a subunit from proteolytic degradation (Cook, 1994); 2) binds plasma factor XIII to a sub-population of fibrinogen in the plasma (Siebenlist et al., 1996); and 3) inhibits the contact activation of blood coagulation (Saito et al., 1990).

1.1.6.1.7 Factor XIII deficiency

Congenital factor XIII deficiency is reported to be in the range of 1 in 5 million in the United Kingdom and Japan. It affects all races and both sexes equally, and is inherited in an autosomal recessive manner (Ichinose et al., 1996). As a result of the determination of the genomic organisation and sequence for both subunits, it is now possible to classify factor XIII deficiency at the DNA level (Anwar and Miloszewski, 1999): the a subunit deficiency (former type II deficiency); the b subunit deficiency (former type I deficiency) and a possible combined deficiency of both the a and b subunits.

The first sign of bleeding in affected individuals is typically from the umbilical cord after birth (Sakata and Aoki, 1982). The major cause of death is intracranial haemorrhage which arises in one fourth of patients. Superficial bruising and haematomas in subcutaneous tissue and muscle are common, and patients may also bleed around the joints after trauma. Deficiency leads to “delayed bleeding” after trauma, which is caused by the premature lysis of haemostatic clots. Due to the lack of crosslinking between α 2-antiplasmin and fibrin, the clots have a reduced resistance to fibrinolysis.

Beck and colleagues (Beck et al., 1961) investigated the effect of factor XIII on the critical events of wound healing (i.e. on fibroblast proliferation, on collagen formation and on the attachment of cells to their substrata). In factor XIII deficient patients, cell contact was lost, previously elongated cells became irregular in shape and the production of collagen fiber could not be identified, resulting in abnormal wound healing.

Acquired factor XIII deficiency has also been reported in connection with various diseases (Egbring et al., 1996). Factor XIII deficiency can be caused by impaired synthesis of factor XIII b subunit (liver disease) and/or increased turnover and consumption of factor XIIIa (systemic haematologic diseases and septicemia).

1.1.6.2 Tissue transglutaminase

Tissue transglutaminase (TGase 2, TG_C), was the first TGase to be identified by Heinrich Waelsch over 40 years ago as a liver enzyme incorporating amines into proteins (Sarker et al., 1957). TGase 2 is a unique member of the TGase family as in addition to being a transamidating enzyme, it also functions as a signal transducing G protein (Takahashi et al., 1986). Like other TGases, TGase 2 lacks carbohydrate modification, irrespective of the presence of multiple potential N-linked glycosylation sites (Ikura et al., 1988), it lacks an N-terminal hydrophobic leader sequence for secretion, and it lacks disulfide bonds.

TGase 2 is a ubiquitous enzyme which is constitutively expressed in endothelium, smooth muscle cells and fibroblasts accounting for its widespread organ distribution (Thomazy and Fesus, 1989). At the cellular level, TGase 2 is present both intracellularly, and extracellularly where it localises to the cell surface and the extracellular matrix. Transport into the nucleus is thought to be via the interaction with the nuclear transport protein, importin- α_3 (Peng et al., 1999), while transportation to the cell surface is still by an unknown mechanism.

The human TGase 2 gene has been mapped to chromosome 20q11 (Gentile et al., 1994) and in mouse, *tgm2* is on the syntenic region located on the distal arm of chromosome 2 (Grenard et al., 2001). The gene for human TGase 2 is reported to be 32.5kb comprising of 13 exons and 12 introns (Gentile et al., 1994). Full-length TGase 2 is 687 amino acids and has a predicted molecular mass of ~77kDa (Greenberg et al., 1991). Different forms of varying sizes as a result of alternative splicing have been reported (Fraij and Gonzales, 1997). Most recently a short alternative splice variant of TGase 2 (s-TGase 2, 652 amino acids, ~73kDa and an mRNA of 2.4kb) has been identified in astrocytes treated with cytokines and in brain of Alzheimer's diseased patients (Citron et al., 2001). This form differs from wild-

type TGase 2 found in untreated astrocytes, normal brain and other tissues in both the 3' coding (exon 13, β -barrel 2) and non-coding regions.

The three-dimensional structure of TGase 2 was derived from that of factor XIII (Yee et al., 1994). Studies have shown that the N-terminal domain in TGase 2 contains the fibronectin binding site, the catalytic core contains the catalytic triad for acyl-transfer reactions and a conserved Trp vital for this catalytic activity (Murthy et al., 2002), and the second C-terminal β -barrel domain contains a phospholipase C-binding segment (Hwang et al., 1995).

1.1.6.2.1 Regulation of transamidation activity

Like other TGases, calcium is required for TGase 2 to assume a catalytically active conformation (Achyuthan and Greenberg, 1987). Investigations of the structural arrangement for the activation of TGase 2 have indicated that initialising transamidating (i.e. crosslinking) activity requires movement of protein domains to enable accessibility of substrates to the active site (Mariani et al., 2000, Casadio et al., 1999). It has been well documented that Cys277 is the essential nucleophile required for transamidation (Yee et al., 1994). In TGase 2, Cys277 is found in the centre of a groove located within the catalytic domain (domain 2), and it is part of the catalytic triad, Cys277-His335-Asp358. Under latent conditions this active centre is concealed by the overlying domains 3 and 4. It has been suggested, based on site-directed mutagenesis (Murthy et al., 2002) and crystallography (Liu et al., 2002), that during activation, calcium binds at its main binding site situated in domain 2. Following binding, the region unfolds, disrupting the structure of the adjacent loop 455-478, which links domains 2 and 3, and domains 3 and 4 move away from each other and from domain 2. This results in a conformation allowing access to the active site for the transamidating activity. Other Ca^{2+} sites are believed to be present, but have not as yet been identified (Ambrus et al., 2001). In TGase 2, a small number of tryptophan residues have been found to be essential (Murthy et al., 2002). For example, Trp241, which is located near to the newly formed substrate channel is thought to have a role in the stabilisation of the transition state intermediate.

In addition to its transamidating activity, TGase 2 binds and hydrolyses guanosine 5-triphosphate (GTP). A distinctive guanine nucleotide-binding site, that had not been identified in any other protein until recently, is situated in a cleft between the catalytic core and the first β -barrel (Liu et al, 2002). This region of the TGase 2 gene, shows little sequence homology with the same region in the other TGases. This binding of GTP to TGase 2 causes a conformational change that results in inhibition of TGase activity as well as decreased calcium binding (Bergamini, 1998). The GTP binding pocket is found in a β -structure segment, which is not close to the active site Cys277 in the primary sequence, but is very near in the spatial location, so it may be possible that the GTP has a direct action on the reactivity of cysteine (Iismaa et al., 2000). As a result, in the GTP-bound form of TGase 2, access to the transamidation active site is prevented by two loops and the active site cysteine is hydrogen-bonded to a Tyr residue (Liu et al., 2002). It has been determined that the binding of GTP to TGase 2 protects against proteolysis by trypsin, and GTP was also able to protect TGase 2 from proteolysis by the calcium-dependent protease calpain (Zhang et al, 1998). This indicates that the conformational change in TGase 2 as a result of GTP-binding makes it resistant to proteolysis in general. Interestingly, the addition of calcium reverses the inhibitory effect of GTP, indicating that the respective calcium and GTP concentrations may play a role in the regulation of the two catalytic functions of TGase 2 (Kanaji et al., 1993).

1.1.6.2.2 Functions of TGase 2

Surprisingly, the function of TGase 2 has remained rather ambiguous. Regardless of its abundance and the extensive number of studies carried out, the physiological function of TGase 2 remains largely speculative, and maybe diverse depending on the tissue or the biological event. TGase 2 has been associated with processes as diverse as stabilisation of the extracellular matrix and in wound healing, GTP binding, and programmed cell death. TGase 2 has also been associated with several pathologies including tissue fibrosis in various organs, atherosclerosis, neurodegenerative diseases and celiac disease.

1.1.6.2.2.1 Organising the extracellular matrix

TGase 2 plays a significant role in the organisation of the ECM by mediating cell matrix interactions that influence cell spreading and migration (Balklava et al, 2002, Akimov et al., 2000), by advancing tissue mineralisation (Aeschlimann et al., 1995), and stabilising dermo-epidermal junctions (Qian and Glanville, 1997, Raghunath et al., 1996). However, some of its functions in the ECM may be independent from its transamidating activity.

Studies have shown that TGase 2 binds fibronectin with high affinity (LeMosy et al., 1992). The formation of the TGase 2-fibronectin complex takes place independently of the catalytic and GTP-related activities of TGase 2. The fibronectin binding site of TGase 2 is located in the 28kDa amino-terminal sandwich region of the protein (Jeong et al., 1995), whereas the TGase 2-binding domain of fibronectin is in a 42kDa gelatin-binding fragment (Radek et al., 1993). This complex has the ability to form ternary complexes with collagen (Turner and Lorand, 1989), and integrins (β_1 and β_3) (Akimov et al., 2000) facilitating adhesion, spreading and motility of cells. Fibronectin, or the fibronectin-TGase 2 complex, also initiate the phosphorylation of focal adhesion kinase (FAK) (Verderio et al., 2002). As well as being able to interact with integrins through a complex with fibronectin, it has also been demonstrated that TGase 2 can associate directly with fibronectin receptors (Akimov et al., 2000). Another role for TGase 2 has also been suggested whereby the enzyme can mediate cell adhesion independently of integrin (Balklava et al., 2000). It is therefore a possibility that the accumulation of TGase 2 in the ECM provides a reservoir of the enzyme which could provide instant and increased protein crosslinking as a result of tissue insult, thereby maintaining tissue integrity and contributing to cell stability by increased cell adhesion.

TGase 2 is also thought to contribute to angiogenesis and wound healing (Greenberg et al., 1987). This notion is supported by the presence of crosslinked fibrin α/γ -chain hybrids characteristic of TGase 2 crosslinking in vascular lesions (Shainoff et al., 1991). Various substrate proteins for TGase 2 such as fibronectin, vitronectin and collagen type V are present in blood plasma and may contribute to the stability of

platelet-fibrin-endothelium and to the process of wound healing. Crosslinked complexes containing fibronectin, vitronectin, or plasminogen have been shown in endothelial cell cultures (Martinez et al., 1994) in support of a role for TGase crosslinking in assembly of the subendothelial extracellular matrix and in wound healing.

It has become apparent that the key structures forming the dermo-epidermal junction are crosslinked by TGases and that this process is important for providing stability at the tissue interface (Raghunath et al., 1996). Several basement membrane components including osteonectin, fibronectin, and the major constituent of anchoring fibrils, collagen VII, have been shown to be substrates for TGase 2 and to occur in crosslinked complexes in tissues. TGase 2 is expressed by basal keratinocytes and dermal fibroblasts which co-operate in the synthesis of this tissue interface and TGase 2 is thought to be the enzyme primarily involved in this process (Aeschlimann and Paulsson, 1991).

The crosslinking of calcium-binding matrix proteins catalysed by TGase 2 is vital in bone formation (Aeshlimann et al., 1995). As chondrocytes go through the different phases of differentiation, they express and externalise TGase 2, as well as matrix proteins, for example, osteopontin and osteonectin, which, in addition to fibronectin and collagen II and XI, have a potential role as crosslinking substrates (Aeschlimann et al., 1995, Wozniak et al., 2000). It has been suggested that the formation of crosslinked protein clusters at the cell surface leads to nucleation and growth of hydroxyapatite crystals. N- ϵ (γ -glutamyl)lysine crosslinks are plentiful in the bone matrix and also appear to correspond with ectopic mineralisation in pathological processes such as atherosclerosis (Bowness et al., 1994), and osteoarthritis (Johnson et al., 2001).

It has been indicated by *in vitro* experiments that TGase 2 may have a role in the activation of the latent form of TGF- β . TGF- β is a pivotal cytokine, secreted by several types of cells as an inactive trimeric complex comprising of the cytokine, its cleaved pro-peptide and a second gene product, the latent TGF- β binding protein (LTBP) which is a constituent of microfibrils in skin (Raghunath et al., 1998). It has been demonstrated that TGase 2 is needed for the protease-mediated pathway of TGF-

β activation (Crawford et al., 1998). Antibodies to TGase 2 prevent the activation of latent TGF- β indicating that LTBP is a TGase 2 substrate, and via this subunit, the complex is crosslinked to the ECM (Nunes et al., 1997). Due to this, the cytokine accumulates in the ECM, from which it is released by proteolysis.

1.1.6.2.2.2 Programmed cell death

Programmed cell death (apoptosis) plays a vital role in biological processes which are essential for morphogenesis and hormone-induced tissue remodelling. Apoptosis is a cell-autonomous process of cell death which can take place either in association with cell-cycle arrest or as the last phase of terminal differentiation, seemingly independent of the cell cycle. TGase 2 has been implicated as a mediator of apoptosis (Aeschlimann and Thomazy, 2000, Fesus et al., 1987, Melino et al., 1994), however, the evidence is conflicting. It was suggested that TGase 2 had an important role in preventing the loss of intracellular components prior to clearance by phagocytosis by stabilising the apoptotic cells by intracellular crosslinking (Knight et al, 1991). Some studies have shown that overexpression of TGase 2 stimulates apoptosis, in contrast to its inhibition by antisense constructs which protects against cell death (De Laurenzi and Melino, 2001). However, it is also becoming evident that the incidence of apoptosis and TGase 2 expression do not necessarily correspond. There is evidence that apoptosis can occur without the induction of TGase 2 expression (Monczak et al., 1997), and moreover, overexpression may even induce cell survival by inhibiting programmed cell death (Benedetti et al., 1996). However, recent studies have suggested that TGase 2 may have a role as a sensitiser of death stimuli, through a process thought to include hyperpolarisation of the mitochondrial membrane (Piacentini et al., 2002). Verification that TGase 2 is not required for the apoptotic mechanism came from studies of TGase 2-knockout mice, which do not display any overt developmental phenotype, indicating that apoptosis is not affected in a major way due to the loss of TGase 2 (De Laurenzi and Melino, 2001, Nanda et al., 2001). However, due to its ability to crosslink cytoskeletal, intermediate filament and nuclear proteins into insoluble aggregates, it may have a role in preventing inflammatory responses that would result from disruption of dying cells and the disbanding of their contents. Thus, TGase 2 may have a role in the later stages of some forms of apoptosis.

1.1.6.2.2.3 TGase 2 as a signal transducing G protein

TGase 2 is a unique member of the TGase family as in addition to being a transamidating enzyme, it also functions as a signal transducing G protein in hormone receptor signalling (Lesort et al., 2000). In this case, TGase 2 was shown to be identical to a previously characterised high-molecular weight G-protein known as G α which regulates the activation of phospholipase C (PLC) (Nakaoka et al., 1994). It has been demonstrated that TGase 2/G α can activate the non-classical δ_1 isoform of PLC (Das et al., 1993, Feng et al., 1996), which can then hydrolyse phosphatidylinositol diphosphate (PIP₂) to produce second messengers inositol triphosphate (IP₃) and diacylglycerol (DAG). IP₃ is a strong inducer for the release of calcium from intracellular stores (Gill et al., 1989) whilst DAG plays a role in the activation of protein kinase C (Lee and Severson, 1994).

1.1.6.2.2.4 Pathogenic role of TGase 2 in disease

TGase 2 has been implicated in a number of chronic diseases, particularly in inflammatory diseases including chronic wound healing, and chronic degenerative diseases, for example, arthritis and neurodegenerative pathologies. The prevailing role of TGase 2 in these disorders seems to be associated with its interaction with, and stabilisation of, the ECM.

1.1.6.2.2.5 Wound healing, inflammation and fibrosis

TGase 2, found in vascular walls is most likely involved in the angiogenic stage of wound repair. In addition to its interaction with and stabilisation of the ECM, this function is probably mediated by its role as a cell-adhesion protein or as an integrin co-receptor (Greenberg et al., 1987, Fromm et al., 1995). Increased TGase activity and TGase 2-specific crosslinking products have been found in atherosclerotic plaques (Bowness et al., 1994) and TGase 2 may have a role in the initiation of tissue mineralisation. A link between TGase 2 and the pathogenesis of the chronic inflammatory diseases of the joints, including rheumatoid arthritis and osteoarthritis has also been reported (Johnson et al., 2001). Fibrogenesis is the normal follow-up of inflammatory repair, and crosslinking by TGase 2 was found abundantly in fibrotic

diseases such as renal and lung fibrosis (Richards et al., 1991, Johnson et al., 1999), liver cirrhosis and liver fibrosis (Mirza et al., 1997, Grenard et al., 2001a). An important role of TGase 2 in many of these disorders is possibly associated with its involvement in the activation of inflammatory mediators such as TGF β (Johnson et al., 2001) Abnormal cytokine activity can cause mineralisation of diseased joints and activation of growth factors such as TGF β ₁ and cytokines, e.g. interleukin-6, can cause increased induction and expression of TGase 2, resulting in an autocrine activation loop (Kunio et al., 1998). Therefore, it may be assumed that TGase 2 crosslinking contributes to these pathological conditions in several ways: by increasing the resistance of the deposited ECM to turnover by regulating cell motility; and by the activation of cytokines, and therefore may have a regulatory role in the process.

1.1.6.2.3 Neurodegenerative diseases

TGases, particularly TGase 2 have been associated with various degenerative neurological diseases, for example, Huntington's disease, Alzheimer's disease, Parkinson's disease and progressive suprabulbar palsy. The disorders are characterised by the accumulation of highly crosslinked insoluble protein aggregates which are thought to promote progressive neuronal death (Kim et al., 2002, Lesort et al., 2000). Due to the crosslinking function of TGases, the expression of TGases (such as TGase 2) in the central nervous system, and that TGase activity is increased in diseased brain tissue, TGases and particularly TGase 2 have been implicated in the production of intracellular aggregates, and as a result, the pathogenesis of these diseases. This will be discussed further in section 1.2.

1.1.6.2.4 Coeliac disease

Coeliac disease is a humoral immune response to the wheat protein gluten (gliadin) and to the intestinal epithelium which results in malnutrition caused by inflammation and almost complete loss of the villi of the intestinal epithelia. TGase 2 is thought to have a role in the progression of this disease in susceptible individuals in two ways: 1) TGase 2 is involved in producing T cell stimulatory gluten peptides through deamidation of specific glutamine residues in gliadin (Anderson et al., 2000,

Willemijn et al., 2002). The TGase 2-formed disease-inducing epitopes trigger a pathological immune response that destroys the epithelia; 2) The formation of heteromeric gliadin-TGase 2 complexes stimulates T-cell mediated initiation of IgA type antibody production by TGase 2-specific B cells (Dietrich et al., 1997).

1.1.6.2.5 Consequences of TGase 2 deletion

Recently, two groups generated TGase 2 knockout mice in order to determine the role of TGase 2. Nanda and colleagues (Nanda et al., 2001) used the Cre/*loxP* site-specific recombinase approach to develop a mouse line in which TGase 2 expression was ubiquitously inactivated (deletion from exon 6 to exon 8). Despite the complete absence of TGase 2, the mice were viable, fertile, phenotypically normal and were born at the expected Mendelian frequency. Melino and colleagues (De Laurenzi and Melino, 2001), generated mice with a deletion of exon 5 in the *tgm2* gene (which includes the active site) by homologous recombination techniques. Again, despite the absence of TGase 2, the mice were viable and fertile and showed no developmental abnormalities.

A possible explanation for the apparently normal phenotype is that other TGases, (for example, TGase 5 and TGase 7) in mammalian tissues can compensate for the loss of TGase 2 (De Laurenzi and Melino, 2001, Nanda et al., 2001). However, preliminary evidence indicates reduced adherence of primary fibroblasts and impaired response in dermal wound healing (Mearns et al., 2002). TGase 2 deficient mice also display mild glucose intolerance and hyperglycaemia due to reduced insulin secretion (Bernassola et al., 2002).

1.1.6.3 Keratinocyte transglutaminase

Keratinocyte transglutaminase (TG_K, TGase 1) is involved in the formation of a cornified cell envelope in terminally differentiating epidermal keratinocytes. The gene spans 14.3kb and contains 15 exons. It encodes a 92kDa protein consisting of 816 amino acid residues and has been mapped to chromosome 14q11.2 (Yamanish et al., 1992). In the epidermis the enzyme is first expressed at a low level in the proliferative basal layer, in a slightly higher amount in the suprabasal cells assigned to

differentiation, and then in greater amounts in the granular layer as terminal differentiation in the epidermis and cornified cell envelope assembly take place (Steinert et al., 1996). The enzyme is post-translationally modified by fatty acid acylation and phosphorylation (Rice et al., 1996). A unique characteristic of TGase 1 is that it is membrane bound and it has been shown that a large proportion of the enzyme pool is attached to the membrane through myristate and palmitate chains (Steinert et al., 1996a). It has been demonstrated that in epidermal keratinocytes, the enzyme exists in several forms in the cytosolic and membrane fractions. The cytosolic enzyme occurs as a full-length form of 106kDa, which has a low specific activity, and two proteolytically processed forms, one of 67kDa and another consisting of 67kDa and 33kDa fragments which are non-covalently associated. These two processed forms have specific activities 5-10 fold higher. These three isoforms contribute 5% to 35% of the total TGase 1 enzyme activity, respectively, in proliferating or terminally differentiating cells (Kim et al., 1995). The membrane bound enzyme exists in two forms: a full-length inactive zymogen: or a 67/33/10kDa chain complex that is highly active (Steinert et al., 1996).

One important function of stratified squamous epithelia is to provide a physical barrier against the environment and protection for the tissues internal to them. Most of this barrier function is provided by a cornified cell envelope (CE), which is a specialised structure formed on the inside of the plasma membrane as the cells differentiate (Nemes and Steinert, 1999). In all stratified squamous epithelia, the CE commonly consists of a macromolecular assembly of highly insoluble proteins built from some or many of the following components (Steinert et al., 1996a, Steinert and Marekov, 1997): annexin I, filaggrin, involucrin, loricrin, type II keratin intermediate filament proteins, small proline-rich proteins and various cell junction proteins including desmoplakin. These proteins become crosslinked together by the action of one or more TGase enzymes, including TGase 1 (Nemes and Steinert, 1999, Greenberg et al., 1991). CEs from the various tissue sources are uniformly about 10nm thick (Jarnik et al., 1998) and have a low density which indicates that the structure is quite porous and has no or very little water barrier function. However, in the epidermis where water barrier function is crucial for survival, this function is provided by the covalent attachment of a monomolecular layer of lipids on the exterior surface of the protein, specifically to involucrin, envoplakin and periplakin

(Marekov and Steinert, 1998). It is proposed that the attachment occurs by a transesterification of ceramides from their linoleic esters onto glutamine residues of the CE scaffold proteins, a reaction catalysed by TGase 1.

Autosomal recessive lamellar ichthyosis (LI) is a severe congenital scaling skin disorder. LI is commonly evident at birth as a “collodion baby” due to a collodion-like translucent membrane encasing the body (Roop, 1995). After the membrane is shed, the skin develops severe ichthyosis with large brown plate-like scales. Patients may also suffer from scarring alopecia, ectropion, eclabium and decreased sweating. It has been demonstrated that LI is caused by a lack of, or severely diminished TGase 1 activity, as a result of mutations in the TGM1 gene (Russell et al., 1995). Several mutations have now been identified, for example, point mutations resulting in amino acid substitutions, large deletions of essential exon sequences and interruptions of intron-exon boundaries resulting in nonsense proteins or premature chain termination events (Yang et al., 2001). The mutations may directly affect the correct folding of the protein or the subsequent post-synthetic processing of the protein into its active functional forms. It is noteworthy that only external epithelia, i.e. the epidermis and hair are affected by LI; internal epithelia, such as buccal mucosa and the oesophagus, are not affected. Therefore, it has been suggested that the ceramide and other lipid esterification events which exclusively take place in the epidermis (and hair cuticle layers of the developing hair fiber) can only be accomplished by TGase 1, and other TGases cannot compensate for its role (Ahvazi et al., 2002).

However, LI is not caused by mutations in the TGM1 gene in all cases; the disease has been mapped to at least three different genetic loci (Parmentier et al., 1996), suggesting that TGM1 mutations are not the cause in all cases. Nonetheless, TGase 1 activity was severely reduced in some of the cases. Therefore, it is conceivable that mutations may be present in other genes whose products are responsible for processing TGase 1 into its active and functional form. Such products may include proteases, phosphatases or lipid metabolising enzymes.

Furthermore to its role in LI, diminished TGase 1 activity has been implicated in other autosomal ichthyoses (Ahvazi et al., 2002). These include: X-linked ichthyosis, which is a disease known to be due to loss-of-function mutations in the steroid sulfatase

gene; and Sjogren-Larsson syndrome, a disease caused by loss of function mutations in the fatty aldehyde dehydrogenase gene, leading to an accumulation of long-chain fatty alcohols in skin and other organs. Preliminary data suggests that these mutations result in an interference with the production of active TGase enzymes.

1.1.6.4 Epidermal transglutaminase

Epidermal transglutaminase (TG_E, TGase 3) is also required in the formation of the cornified cell envelope, where it crosslinks various structural proteins in the late stages of keratinocyte differentiation (Yamanish et al., 1992). Additionally, TGase 3 is thought to be needed for the cross-linking of the structure protein trichohyalin and the keratin intermediate filaments to assemble a rigid structure within the inner root sheath cells (Steinert et al., 1971), and therefore play a role in the shape determination of the developing hair cortical cells internal to the sheath structure. TGase 3 expression has also been identified in the brain (Kim et al, 1999). The native TGase 3 protein molecule consists of 629 amino acid residues and has a molecular mass of 77kDa (Kim et al., 1993). The TGM3 gene contains 13 exons interrupted by 12 introns and has been mapped to human chromosome 20q11. Activation of the zymogen form of TGase 3 requires proteolytic cleavage at Ser469 (Kim et al., 1990), located in a hinge region that separates the catalytic core and β -barrel 1 domains. The resultant 50kDa fragment, containing the catalytically essential SH-group and 27kDA carboxy terminal fragment, remain non-covalently associated to each other, whilst the cleavage induces a conformational change that results in the exposure of the active thiol. The zymogen can be activated by several different proteases including trypsin, thrombin, and proteinase K. However, it is more likely that the proenzyme is cleaved by a member of the calcium-dependent cysteine proteinases, the calpains, in the epidermis.

It is known that the inactive zymogen possesses only one tightly-bound calcium ion located at a site encompassing residues Asn224-Asn229. (Ahvazi et al., 2002). However, crystallographic studies have recently demonstrated that when the zymogen is proteolytically cleaved, a further two calcium ions are bound at two other sites, Asn430-Glu448 and Asp320-Ser325 respectively, and the enzyme becomes highly

active (Fig. 7.6 and 7.7). Specifically, the H-bonding of residue Ser323 in the catalytic core domain with Arg570 in the barrel 1 domain is disrupted to allow Asp324 to coordinate with the calcium ion at the third site. This results in the opening of a channel that passes through the enzyme. It is proposed that the channel will allow both glutamine and lysine substrates to enter into the active site (Ahvazi et al., 2002a). Also, it has recently been determined that TGase 3, like TGase 2, can bind and hydrolyse GTP (Ahvazi et al., 2004). These findings are discussed further in chapter 7.

1.1.6.5 Prostate transglutaminase

Prostate transglutaminase (TG_p, TGase 4), is an androgen-regulated protein involved in semen coagulation and its expression is restricted to the prostate (Aeschlimann and Thomazy, 2000).

The TGM4 gene spans approximately 35kb and consists of 13 exons and 12 introns. It has been localised to human chromosome 3p21.33-p22 (Gentile et al., 1994). The cDNA encodes a protein of 684 amino acids which has a molecular mass of 77kDa. The deduced amino acid sequence shows a significant similarity with the other members of the TGase family, and the overall identity of human TGase 4 with the other TGases is approximately 30%, with the highest similarity in the active site region (Dubbink et al., 1996).

Rodent TGase 4 is involved in the formation of copulatory plugs in the female genital tract after coitus (Folk and Finlayson, 1997). In contrast, no physiological function has been determined for the human enzyme (Dubbink et al., 1996). Despite the fact that both enzymes are synthesised in the same organ, various differences occur between the rat and the human TGase 4: 1) the human enzyme has a much lower level of expression than the abundantly expressed rat enzyme; 2) the human and rat TGase 4 amino acid sequences share an identity of only 53%; and 3) human TGase 4 is only produced in the secretory epithelial cells of the human prostate (Dubbink et al., 1998). These variations may reflect a difference in the function of the respective enzymes. While a role in semen coagulation cannot be excluded for human TGase 4, other functions related to the appropriate sperm transfer to the female have been suggested.

It may also play a role in suppressing the sperm surface antigenicity, resulting in the prevention of an immunological response in the female genital tract (Dubbink et al., 1996).

1.1.6.6 Erythrocyte protein band 4.2

Band 4.2 is a component of the cytoskeletal network underlying the erythrocyte plasma membrane, but it is also found in other cells and tissues such as platelets, kidney and brain (Aeschlimann and Paulsson, 1994). The gene for human erythrocyte band 4.2 is approximately 20kb in size and consists of 13 exons and 12 introns. The gene has been localised to human chromosome 15q15 (Cohen et al., 1993). The protein comprises of 691 amino acids and has a molecular mass of 77kDa. Band 4.2 has significant homology to known TGases, but contains a cysteine to alanine substitution in the TGase active site (Korsgren et al., 1990). With the loss of its enzymatic activity, band 4.2 has become a purely structural protein. An isoform of band 4.2, called B4.2L has been identified which contains a 30 amino acid insert after Gln-3 at the NH₂-terminus of the protein (Sung et al., 1992). A short form of band 4.2, with exon III spliced out has also been reported, however, the difference in function of the isoforms is presently unknown.

Band 4.2 associates with the cytoplasmic domain of the anion exchanger band 3 and may also interact with ankyrin and band 4.1 protein. It is likely that band 4.2 still retains the TGase protein-binding site, but is unable to covalently crosslink the bound proteins. This indicates that TGases may also have a role as structural components (Aeschlimann and Paulsson, 1994), and one function of band 4.2 might be to prevent excessive crosslinking of the cytoskeletal network of intact cells by competing with TGase 2 for the same binding sites.

Band 4.2 is clearly important for normal erythrocyte function since inherited band 4.2 deficiency, called spherocytosis, elliptocytosis or ovalostomatocytosis causes accelerated erythrocyte destruction and abnormally shaped, fragile erythrocytes (Korsgren and Cohen, 1991). Affected individuals of this rare genetic disorder are predominantly found in Japan and suffer from haemolytic anaemia, which may be mild to severe.

1.1.6.7 Transglutaminase 5

A new member of the TGase family, TGase 5 (TG_x), has recently been identified in differentiating keratinocytes (Aeschlimann et al., 1998). The gene has been mapped to the 15q15.2 region of chromosome 15 by fluorescence *in situ* hybridisation. The gene consists of 13 exons separated by 12 introns and spans about 35kb. TGase 5 is expressed in four different isoforms (Candi et al, 2001), full length, $\Delta 3$ (deletion of exon 3), $\Delta 13$ (deletion of exon 13), and $\Delta 3\Delta 13$ (deletion of exon 3 and 13). Western blot analysis of keratinocyte cells revealed a band of 84kDa corresponding to the full length TGase 5 enzyme as well as other bands of lower molecular weight corresponding to the splice variants. However, the only active isoforms are the full length and $\Delta 13$.

TGase 5 enzyme expression is not completely confined to the epidermis where it is abundantly expressed (Grenard et al, 2001). However, TGase 5 appears to have a role in keratinocyte differentiation and cornified envelope (CE) assembly. TGase 5 has been identified in the upper layers of normal human epidermis, showing a progression of expression, extending from the spinous layer to the horny layer. It is induced during the early phases of keratinocyte differentiation *in vitro*. It has also been proven that TGase 5 acts on specific epidermal substrates such as loricrin, involucrin and SPR3 (Candi et al., 2001). Accordingly, it seems that TGase 5 is present and active in the very early stages of epidermal differentiation.

The expression of TGase 5 has been analysed in pathological conditions such as psoriasis, ichthyosis (lamellar and vulgaris) and Darier's disease. The data showed that TGase 5 is expressed in differentiating keratinocytes *in vivo* (Candi et al., 2002), and that in many of the conditions, its expression is up-regulated, and has a contributory role in the hyperkeratotic phenotype observed in ichthyosis.

So far, TGase 2 and TGase 3 are the only members of the TGase family for which regulation by guanine nucleotides has been demonstrated. Comparative analysis of amino acid alignments and a homology-derived three-dimensional model has led to the identification of a putative GTP-binding pocket (Candi et al., 2003). It is reported

that TGase 5 is regulated by guanine/adenine nucleotides. GTP and ATP were found to inhibit the crosslinking activity of TGase 5 *in vitro*, in a dose-dependent manner and that Ca^{2+} was able to reverse the inhibition. Also, TGase 5 can hydrolyse GTP

Therefore, based on its expression TGase 5 is likely to play an important role in keratinocyte differentiation and CE assembly and it is possible that increased expression of TGase 5 in certain diseases contributes to the hyperkeratotic phenotype, maybe as a compensatory mechanism. However, the expression of TGase 5 is not restricted to the epidermis and its capability to crosslink non-epidermal substrates and hydrolyse GTP, indicates that other roles for TGase 5 need to be considered.

1.1.6.8 Transglutaminase 6

The TGase 6 gene product was identified in a small lung carcinoma cell line, called H69. This cell line has neuronal cell characteristics such as the expression of neuron-specific enolase and brain isozyme of creatine kinase, which suggests that TGase 6 expression may be specific to neuronal cells. TGase 6 and TGase 2 are part of the same subgroup of gene products which are closely related and may therefore have comparable properties and substrate specificities. TGase activity and in particular TGase 2 activity, has been implicated in the formation of aberrant protein complexes in the central nervous system, for example, in Alzheimer's and Huntington's diseases. Based on its close relation to TGase 2, TGase 6 is a logical candidate for a function in the central nervous system to potentially contribute to the underlying TGase-related pathological changes. The aim of this PhD is therefore to characterise the novel TGase, TGase 6, and to determine its expression pattern.

1.1.6.9 Transglutaminase 7

A full-length cDNA sequence for another novel TGase, TGase 7 (TG_Z), has recently been determined in our lab (Grenard et al., 2001). The deduced protein consists of 710 amino acids and has a calculated molecular mass of 80,065Da. Based on the preservation of critical residues for enzyme function and domain folding, and the extensive similarity of TGase 7 to the other members of the TGase protein family, it can be predicted that the characterised cDNA encodes an active TGase. Initial studies

on mRNA expression have found that it is highly expressed in testis and lung. RT-PCR analysis confirmed that TGase 7 is expressed in testis and lung, as well as demonstrating expression in brain, skin, kidney, heart, pancreas placenta, skeletal muscle and prostate.

1.1.6.10 Summary of functions

TGases are associated with bone and cartilage remodelling and bone ossification, development of lung, heart, and the central and peripheral nervous systems. Factor XIII is vital for blood coagulation. Also, TGase 1 and band 4.2 are needed for the formation of the cornified cell envelope and for the integrity of erythrocyte membrane, respectively. TGase 3 is also essential for skin integrity. In contrast, the biological functions of other TGase members, for example TGase 2, are more diverse with roles in wound healing, stabilisation of the extracellular matrix and programmed cell death. The functions of TGases in other species have been studied less, but it is thought that orthologues in other mammals probably have similar functions to those in humans.

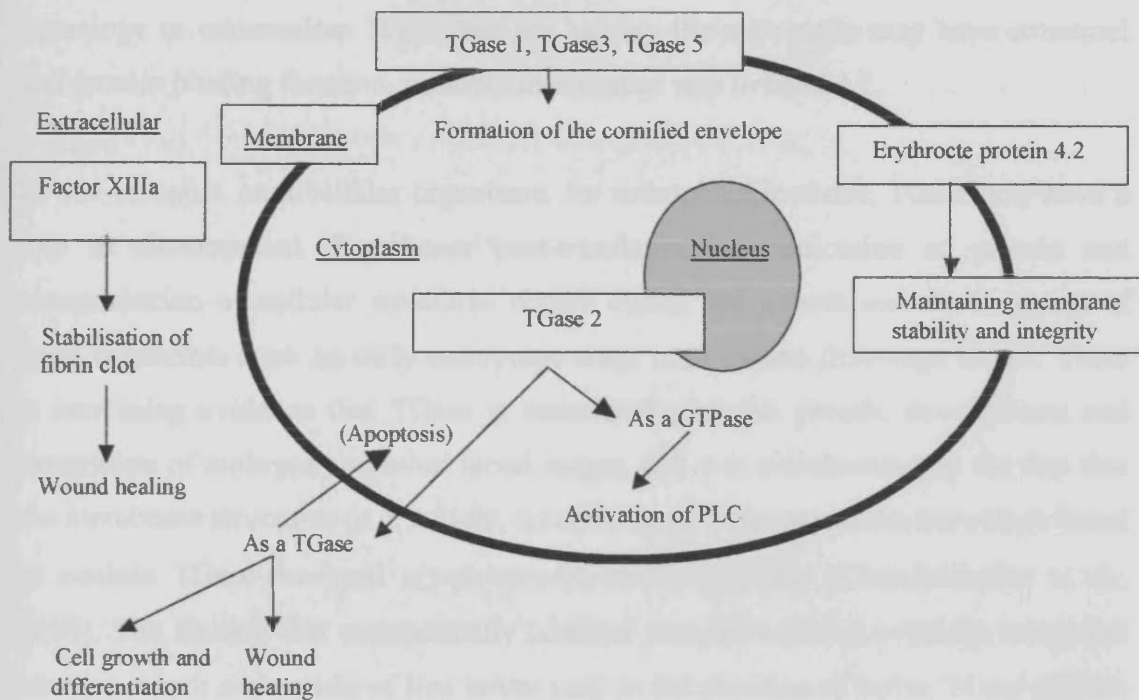


Fig. 1.6 Cellular distribution and possible functions of various TGases. The members of the TGase enzyme family can be found as extracellular zymogen (e.g. factor XIII), membrane-bound (e.g. TGase 1), cytosolic (TGase 2), and even as nucleus associated (e.g. TGase 2). Depending on their localisation, TGases may modulate specific cell functions.

1.1.7 Non-mammalian Transglutaminases

TGase-like activity has also been identified in plants, early eukaryotes and bacteria. Some of the TGase genes found in these organisms have been successfully cloned. Even though a number of TGases identified in these organisms have preserved a certain degree of homology to the known TGases, some show no structural or functional similarities (Makarova et al., 1999, Schmidt et al., 1998). For example, a TGase cloned from the bacteria *Streptovorticillium spp.* shows no homology to any of the known TGases (Makarova et al., 1999). Also, the necrotizing factor 1 from *Escherichia coli*, which deaminates a specific glutamine residue in Rho GTPases, contains the catalytic cysteine and histidine residues and is able to catalyse the TGase reaction (Schmidt et al., 1998) but shows no similarity to mammalian TGases. In contrast, several proteins were identified in *Mycobacteria spp.* (Makarova et al., 1999), which are homologous to mammalian TGases. Also, proteins which have a

homology to mammalian TGase but are lacking the active site may have structural and protein binding function, possibly in a similar way to band 4.2.

In non-complex multicellular organisms, for example nematodes, TGase may have a role in development. Significant post-translational modification of protein and reorganisation of cellular structures occurs during the growth and development of these organisms from an early embryonic stage until mature first-stage larvae. There is increasing evidence that TGase is associated with the growth, development and maturation of embryos and other larval stages, and it is corroborated by the fact that the membrane structures (e.g. cuticle, epicuticle), of these organisms have been found to contain TGase-catalysed $\epsilon(\gamma\text{-glutamyl})\text{lysine}$ isopeptides (Chandrashekar et al., 1998). The finding that metabolically labelled proteins were successfully integrated into the sheath and cuticle of live larvae only in the presence of active TGase (Mehta et al., 1992) adds further support that the TGase-catalysed protein crosslinking reactions in growth and development are of importance.

Calcium-dependent TGase activity has been detected in the free-living nematode *Caenorhabditis elegans* (Mehta et al., 1992). However, identification of a gene with a notable homology to vertebrate or bacterial TGases has not been possible. This indicates that either the TGase homologue is one of the gene products that has not yet been sequenced or that the TGase found in the *C. elegans* nematode shows no similarity to other members of the TGase superfamily. However, a potential nematode TGase does show significant homology to a protein disulphide isomerase (PDI)-related endoplasmic reticulum protein Erp 60 that has been shown to have TGase activity *in vitro* (Chandrashekar et al., 1998). Similar TGase homologues have also been cloned from other filarial nematodes (Natsuka et al., 2002).

It is possible that TGase-catalysed reactions contribute to the protection of free-living microorganisms from environmental stresses and parasitic microorganisms from the immune attack of their hosts. It has been shown that endogenous TGase has the capability to use nematode proteins as substrates (Mehta et al., 1996). Many parasites can coat themselves with hosts' proteins thereby avoiding the hosts' immune attack. Recently, a protein (p68) from a rodent host was identified covalently integrated onto

the surface of young developing larvae in a TGase-dependent manner. These findings indicate that TGase-catalysed reactions may also have a role in increasing the survival of these organisms under certain non-physiologic conditions.

The awareness that TGase has a vital role in the growth, development and survival of nematodes may provide biochemical targets to monitor conditions induced by these organisms in humans and animals.

1.1.8 Biotechnological applications of TGases

The use of TGases in the biomedical and biotechnological areas of research is most likely one of the fastest-growing areas of TGase research. TGases are being used in a broad range of commercial sectors varying from the textile industry to the cosmetic industry. Due to the possible involvement of TGases in a number of human diseases, their role as therapeutic targets or as identifiable markers for certain diseases has resulted in interest from the diagnostics and pharmaceutical industries. Below is a summary of just some of the applications of TGases.

1.1.8.1 Food production

In 1981, it was realised that TGase could be utilised for the adaptation of proteinaceous food products and TGase 2 was used initially (Matheis and Whitaker, 1987). TGase from the microbe *Streptoverticillium sp.* remains the only commercially available source of microbial TGase, and it is widely used within the industry. Some examples of the use of TGase in this industry include: 1) application of TGase for the restructuring of many different meats and fish into forms which are visually and texturally more satisfactory for the consumer (Kuraishi et al., 1997); 2) in the dairy industry, TGase treatment of milk before fermentation in the production of yoghurt yields a product with increased firmness and decreased water separation properties (Kuraishi et al., 2001), TGase treatment of milk is also a helpful step in the manufacture of cheese products and another benefit of the application of TGase is in the manufacture of low calorie ice-cream which is softer and smoother (Kuraishi et al., 2001); 3) in the baking industry, TGase has been used to improve the properties of dough. Dough treated with TGase has greater resistance to stretching, and baked

bread has a larger volume. Cake products show similar improvements and oil-fried dough products absorb less fat when treated with TGase, resulting in an improvement in quality and a reduction in fat content.

1.1.8.2 Medical applications

1.1.8.2.1 Enzyme substitution therapy

A fibrin sealant consisting of thrombin and a concentrated fibrinogen solution enriched in factor XIII was developed during the 1980s. Since then, fibrin sealants have been applied in many aspects of surgery for haemostasis and the sealing of organs, and generally as a tissue glue. One of the early therapeutic applications of TGases was the use of substitutive therapy with plasma-derived factor XIII (Jurgensen et al., 1997) to assist blood clotting in patients with factor XIII deficiency and also associated with major surgery.

1.1.8.2.2 Diagnostics

The established technique of diagnosis for celiac disease is a biopsy of the small intestine. However, due to the finding that the predominant endomysial autoantigen is TGase 2 (Dietrich et al., 1997), ELISA methods have been developed using recombinant human TGase 2 (Sardy et al., 1999). The ELISA is also used for checking the dietary compliance of celiac disease sufferers and the ELISA is now used worldwide.

1.1.8.2.3 Inhibition of excessive crosslinking

After surgery or serious burns, patients may suffer from hypertrophic scarring, which is failure of normal remodelling of scar tissue due to an increased level of TGase. A topical cream has been developed which contains putrescine which competes for TGase crosslinking (Collighan et al., 2002). Inhibition of the action of TGase to prevent excessive collagen III deposition as opposed to collagen I, enables correct remodelling of collagen during healing, resulting in decreased contracture and an improved cosmetic result. This cream is currently in clinical trials.

1.1.8.3 Cosmetics

The formation of corneocytes is dependent on the crosslinking of different cornified envelope proteins, a process mediated by TGases. From a commercial perspective, modification of the skin surface can be achieved by the covalent inclusion of various primary amine-containing compounds using TGase. For example, TGase can be used to incorporate antimicrobials, hair and skin conditioning agents, anti-inflammatories and colouring agents (Collighan et al., 2002). Topical preparations comprising of TGase and one or more of the corneocyte proteins, for example, involucrin, or loricrin, are thought to produce a protective layer on the surface of hair, skin and nails. By using the essential plant oil pulegone in topical skin treatments, an increase in the expression of TGase 1 and the production of ceramides has been observed. The appearance of dryness and wrinkles in aging skin is thought to be caused partially by thinning of the dermis and by reduction in barrier properties of the epidermis, and therefore an increase in TGase 1 and ceramides is suggested to decrease these effects.

1.1.8.4 Textiles

A more environmentally friendly method to coat leather uses TGase. In this process TGase and casein, are applied to the leather as a coating. This abolishes the need for toxic hardening agents (Collighan et al., 2002).

It has been demonstrated that TGase treatment of textiles improves the durability of finished garments. Combined with an appropriate protease, TGase treatment of wool leads to, for example, increased shrink resistance, reduced pilling, improved softness, and improved dye uptake. In addition, fibrous textiles can be covalently modified by the incorporation of primary amine containing compounds such as dyes, softeners and water repellents using TGase.

1.1.8.5 Conclusion

Various TGase-catalysed reactions have been employed in the covalent modification of proteinaceous materials in commercial processes. The food industry currently uses the most substantial amount of enzyme, where TGase has been vital for the generation

of novel foods and food production. However, with increased research, their potential applications within a variety of fields will undoubtedly increase further.

1.2 Neurodegenerative diseases

A diverse range of neurodegenerative disorders are characterised by damage to neurones, which is possibly caused by the formation of toxic, insoluble proteins. The processes involved in the formation of abnormal intra-neuronal inclusion bodies or extracellular aggregates in neurodegenerative diseases are uncertain. However, it is predicted that processes common to many neurodegenerative diseases as well as disease-specific processes exist. Many studies have suggested that one of the processes responsible for the generation of abnormal protein aggregates in neurodegenerative diseases is TGase crosslinking. Several reports have identified TGase activity in regional and subcellular tissue locations affected by the disease and the presence of TGase substrates in the protein aggregates (Kaytor and Warren, 1999). Studies have revealed the expression of TGase 1, 2, and 3 in human brain tissue (Kim et al., 1999) and it is possible that more TGases may be present but as yet have not been identified. A role for TGase-crosslinking in neurodegenerative diseases is substantiated by the presence of TGase-catalysed isodipeptide bonds in abnormal inclusions in Parkinson's disease, Alzheimer's disease, progressive supranuclear palsy and Huntington's disease.

1.2.1 Expression of TGases in normal brain

Several reports have described the presence of TGase activity in numerous different neural tissues including brain, spinal cord, central and peripheral nervous system (Gilad et al., 1985, Gilad and Varon, 1985), and that it is developmentally regulated. There is evidence that TGase activities are located in cytosolic and/or membrane compartments (Tetzlaff et al., 1998) and that membrane-associated brain TGases contain latent activity which may be activated by thrombin, proteases and physical agents (Hand et al., 1993). It has been demonstrated that TGase activity is enriched in synaptosomal fractions in comparison to other subcellular fractions (Gilad and Varon, 1985) and TGase 2 has been immunolocalised to axons in cultured cerebellar granule neurons. It has been suggested that neuronal TGases modulate synaptic plasticity and neurotransmitter release (Facchiano et al., 1993, Pastuszko et al., 1986) and play a role in axonal regeneration (Ando et al., 1993, Gilad et al., 1985). Another study demonstrated that a high degree of crosslinking of endogenous protein substrates took

place in synaptic membranes (Perry et al., 1995). It has also been shown that in isolated superior cervical and nodose ganglia, intracellular TGase activities can be induced and activated by agents such as extracellular potassium and acetylcholine (Slife et al., 1985), which directly cause an influx of calcium.

Even though TGase 2 is mainly a cytosolic protein, it also has been found in particulate fractions (Slife et al., 1985, Hand et al., 1993) and in the extracellular matrix (Aeschlimann and Paulsson, 1994). Nuclear localisation of TGase 2 has been reported (Lesort et al., 1998) but the level of TGase 2 in the nucleus appeared quite low in resting cells. However, an increase in intracellular calcium concentrations resulted in a translocation of TGase 2 into the nucleus and there is evidence to suggest that the nuclear transport of TGase 2 may be mediated by the nuclear transport protein importin α 3 (Peng et al., 1999).

TGase 1 and 3 enzymes are abundantly expressed in terminally differentiating stratified squamous epithelia. The ~77kDa TGase 3 is normally located in the cytosol, but is proteolytically processed during calcium-induced terminal differentiation into a functional 50/27kDa complex (Kim et al., 1993, Kim et al., 1990). The majority of the TGase 1 exists in epithelial cells as a ~100kDa membrane-bound protein of low specific activity, but some may be activated by proteolytic processing during terminal differentiation into a 67/23/10kDa form with very high specific activity and some intact low specific activity or higher specific activity processed forms are released into the cytosol (Kim et al., 1995). Therefore, depending on the extent of proteolytic activation, cytosolic or membrane fractions of epithelial cells may contain large amounts of TGase 1 protein of either low or high activity. However, in contrast, the TGase 2 enzyme is disrupted by proteolysis (Folk and Chung, 1985, Greenberg et al., 1991). Therefore, proteolytic activation of TGase 1 can result in considerable alterations in total TGase activity in cells. Due to this, it is thought that significant increases in total TGase activity could occur in normal or abnormal neuronal cells as a result of proteolytic activation of TGase 1 and/or 2.

RT-PCR and immunohistochemical studies by Kim *et al.* (Kim et al., 1999) determined that the mRNA and protein for TGase 2 is abundantly expressed in the four regions of human brain studied; amygdala, cerebellum, corpus callosum, and

cortex. This finding is in agreement with previously published reports. The TGase 3 enzyme was observed to a minor degree in all the tissues tested, but was found more significantly in the amygdala. A significant amount of TGase 1 mRNA expression was observed in cerebellum and corpus callosum. In contrast to epithelial tissues where most TGase 1 is membrane bound, it was found that only about 20% of the total TGase activity was associated with membranes from the cortex and cerebellum, all of which was attributable to the TGase 1 protein. Therefore, up to 80% of the TGase 1 enzyme was present in the cytosolic fraction. Due to the finding that TGase 1 had relatively low activity levels (Kim et al., 1990) it was proposed that most of it remained in its intact low specific activity (unprocessed) form.

1.2.2 Alzheimer's disease

The most common age-related neurodegenerative disorder, Alzheimer's disease (AD), is associated with the selective damage of brain areas and neural circuits, including neurons in the neocortex, hippocampus, and amygdala. Impaired memory, thinking and behaviour are a consequence of dysfunction and loss of neurons in these neural circuits. Two defining characteristics of AD pathology are the presence of intraneuronal neurofibrillary tangles and extracellular senile plaques. Neurofibrillary tangles are composed predominantly of paired helical filaments (PHF) (Delacourte and Defossez, 1986) and the main constituent of the PHFs is the hyperphosphorylated form of the microtubule associated protein tau (Goedert et al., 1992). Senile plaques are composed of the amyloid β -protein ($A\beta$), a 39-42 amino acid peptide which is proteolytically derived from a transmembrane glycoprotein, the amyloid precursor protein (APP) (Kang et al., 1987). Two main APP processing pathways have been identified, the α -pathway and the β pathway (Selkoe, 1998). APP processing via the α pathway involves a cleavage in the $A\beta$ region by α -secretase which prevents the formation of $A\beta$ (Sisoda et al., 1990). Proteolysis of APP by the β pathway involves cleavage of APP by β -secretase at the N-terminus of $A\beta$, and an N-terminal soluble APP fragment (β -APPs) is released. The C-terminal fragment of APP is then cleaved by γ -secretase to generate $A\beta$ (Haass et al., 1992). It is now known that $A\beta$ is produced under physiological conditions (Haass et al., 1992) and is found in the cerebrospinal fluid and plasma in healthy subjects (Seubert et al., 1992). The full length $A\beta$ peptide contains two lysine and one glutamine residue. It has been shown

that all three residues are utilised by TGase 2 in the formation of insoluble polymers (Rasmussen et al., 1994). It has been proposed that a TGase, possibly TGase 2, might have a role in the pathogenesis of AD by contributing to the formation of one or both of the insoluble protein aggregates. Immunocytochemical staining showed the presence of TGase in senile plaques in tissue sections and also in isolated amyloid plaque cores obtained from AD brains (Zhang et al., 1998a). For example, two groups have demonstrated that β -amyloid is a substrate for TGase (Rasmussen et al., 1994, Ikura et al., 1993) and using SDS-PAGE have shown that crosslinking of A β by TGase 2 forms dimers, trimers, tetramers and other high molecular weight oligomers that correspond to protein aggregates extracted from the brains of AD patients. Through incorporation of site specific probes followed by enzymatic digestion, it was revealed that the two lysine residues in the full-length A β are modified by TGase 2 and contributed in the TGase-catalysed formation of polymers (Rasmussen et al., 1994). As well as A β -peptide, another amyloid plaque constituent known as NAC (non-A β component of amyloid plaque in AD) is also a good substrate for TGases (Jensen et al., 1995) In addition, both kinetic experiments and biochemical analysis demonstrated that tau is a substrate for TGase (Miller and Johnson, 1995) and immunocytochemistry showed that TGase 2 co-localises with a hyperphosphorylated form of tau that is present in the neurofibrillary tangles of AD brains (Appelt et al., 1996). These findings have also been confirmed by another group through the use of immunoprecipitation and western blotting (Singer et al., 2002). These findings add support to the hypotheses that TGases play a role in the formation of plaques and tangles in AD brain. Therefore if plaques and tangles play a causative role in the progression of AD, then the activation of TGases may be a vital factor in the disease.

Several TGases, including TGase 2, are expressed within the membrane, cytosol and the nucleus of neurons (Kim et al., 1995). Johnson *et al.* have demonstrated that TGase activity in AD brain was significantly increased in comparison to age- and postmortem-interval matched controls (Johnson et al., 1997). In the prefrontal cortex of AD specimens, a region where the disease is highly prevalent, TGase activity was significantly greater than the control samples. In contrast, there were no significant differences in TGase activity between controls and AD patients in the cerebellum, a region of the brain where there is virtually no pathological changes. These findings were confirmed by Kim *et al.* who also demonstrated that total TGase activity is

elevated in the affected areas of AD brains (Kim et al., 1999a). It was also shown that the number of $\epsilon(\gamma\text{-glutamyl})\text{lysine}$ linkages in insoluble proteins from AD brains is significantly greater than the number found in normal brain tissues. The increase in TGase activity is thought to be due to the elevated expression of TGase 1 and 2 mRNAs and proteins. Histochemical analysis using TGase isoform-specific antibodies have further demonstrated increased staining for TGase 1 and 2 in the areas of the brain affected by AD when compared to normal individuals.

There are three features of TGases which support the possibility that they have a role in neurodegeneration in AD; it is activated by calcium, its activity is increased during apoptosis, and oxidation of proteins increases their ability to be crosslinked by TGases (Groenen et al., 1993). TGase is a calcium activated enzyme (Folk and Finlayson, 1977), and therefore impairment or alteration to calcium homeostasis is likely to cause inappropriate activation of the enzyme. In AD, there is evidence indicating that the concentration of intracellular calcium is altered, which might contribute to the neurodegeneration in AD brain (Mattson et al., 2001). Fibroblasts from AD patients have been found to contain increased concentrations of bound, cytosolic and free calcium (Peterson and Goldman, 1986). Further studies have shown that alterations in cerebral calcium metabolism occurs in AD (Gibson et al., 1996). It has been demonstrated that $A\beta$, an obligatory component of AD pathology, disrupts calcium homeostasis in neuronal cultures (Mattson et al., 1993). It has also been shown that influxes of calcium lead to increased immunoreactivity with the Alz-50 antibody, which specifically recognises a particular pathological conformation of tau found in AD brain (Carmel et al., 1996). Further studies have shown that mutations in presilin 1, which cause early onset inherited forms of AD also perturb calcium-signalling processes (Mattson et al., 2001). An increase in intracellular calcium concentration is likely to cause increased TGase transamidating activity, thereby disrupting normal cell processes and contributing to cellular damage.

Oxidative stress and free radical production may also have a role in the pathology of AD (Beal et al., 1993) by altering intracellular calcium levels through several different processes (Mattson, 2000). For example, $A\beta$ instigates calcium-dependent neurotoxicity via a process which may involve increased production of reactive oxygen species (Mark et al., 1996), and it has also been demonstrated that oxidation

of proteins results in an increase in their ability to be crosslinked by TGase (Groenen et al., 1993). In addition, it has been shown that markers of protein oxidation are increased in particular areas affected in AD brains, such as the cortex and hippocampus, in comparison to control brains (Hensley et al., 1995). Oxidative stress has also been shown to increase the expression of the inducible transcription activator nuclear factor- κ B (NF κ B) (Baeuerle, 1991), a transcription factor which can up-regulate TGase 2 gene transcription (Mirza et al., 1997). As a consequence of increased intracellular calcium concentrations, there may be a TGase 2-mediated increase of phospholipase A₂ activity (Cordella-Miele et al., 1993), resulting in further free radical production via the release of arachidonic acid and other polyunsaturated fatty acids (Chan and Fishman, 1980). Studies have shown that neuroinflammation occurs in AD and that inflammatory cytokines, including TNF- α and interleukin-6 are up-regulated in AD brain (Rogers et al., 1996). As inflammatory cytokines can induce TGase 2 expression (Suto et al., 1993) it has been suggested that TGase 2 levels and activity could be increased in AD.

There is increasing evidence to suggest that apoptosis may have a role in the neurodegenerative process of AD (Mattson, 2000). For example, overexpression of mutant presenilin 1 in cultured neuronal cells considerably enhanced apoptosis induced by A β (Mattson et al., 2001). DNA fragmentation characteristic to apoptosis was identified in tissue sections from the hippocampus of AD cases, but not in controls (Lassman et al., 1995). It has been suggested that if an apoptotic process is involved in the progress of the disease, it is likely that TGase 2 levels and activity would be up-regulated under such conditions.

Two possible TGase substrates which may play a role in AD are glyceraldehyde 3-phosphate dehydrogenase and the α -ketoglutarate dehydrogenase complex. In the presence of glutamine donors, both of these proteins are inactivated by TGase 2 (Gentile et al., 1998). This is a significant finding as a decrease in cerebral metabolism precedes the onset of brain atrophy and cognitive impairment due to a loss of α -ketoglutarate dehydrogenase activity (Gibson et al., 2000). Therefore, by inactivating α -ketoglutarate dehydrogenase and certain other enzymes involved in energy metabolism, for example, glyceraldehydes 3-phosphate dehydrogenase, TGases may contribute to AD.

In summary, there is circumstantial/indirect evidence that TGases, particularly TGase 2 may be a contributing factor in AD. The increase of TGase 2 levels specifically in brain regions affected by AD indicates that the enzyme may have a compensatory or pathogenic role in the progression of the disease and Fig. 1.7 provides a summary diagram of its possible roles. However, further research is required to confirm or reject this hypothesis. Investigations into a possible role of the newly characterised TGases must be carried out before a definitive role for this enzyme or other TGases in AD can be established.

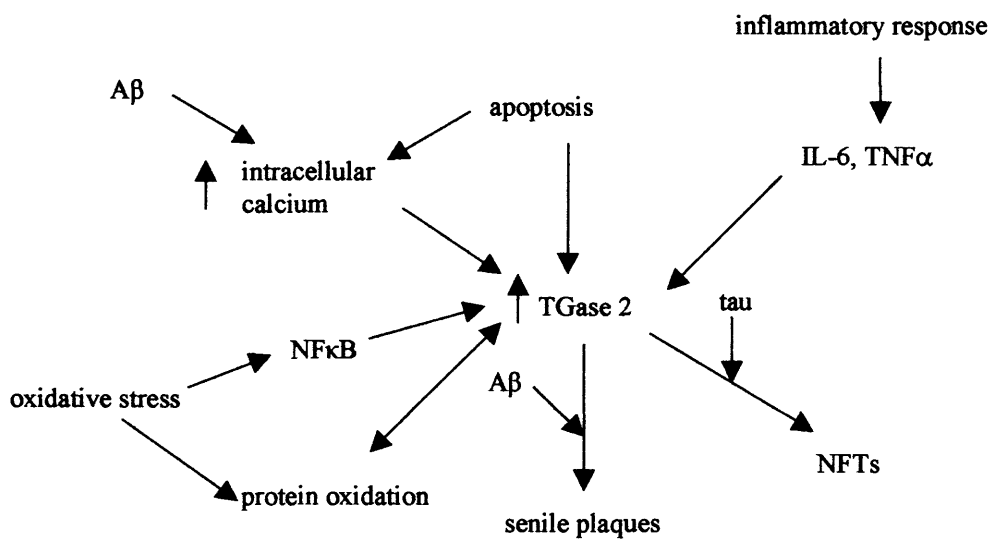


Fig. 1.7 Possible role of TGase 2 in AD. Apoptotic cell death of neurons is a characteristic feature of AD. Apoptosis is often linked to an increase in TGase 2 levels and activity. One possible factor which contributes to premature neuronal death in AD is the amyloid- β ($A\beta$) peptide. $A\beta$ cytotoxicity is linked to an increase in intracellular calcium concentration, which may lead to inappropriate activation of TGase 2. It is also thought that oxidative stress has a contributory role in the pathology of AD. By activating NF κ B or by initiating protein oxidation, oxidative stress may be a factor in the increase of TGase 2 expression and activity in AD. The inflammatory response which is linked with the pathology may also lead to an increase in TGase 2 levels in AD brains. As $A\beta$ and tau are proven substrates for TGases, an increase in TGase 2 activity and expression may be a contributory factor in the formation of senile plaques and/or neurofibrillary tangles, which are the aggregates characteristic of AD. Adapted from (Lesort et al., 2002).

1.2.3 Parkinson's disease

Parkinson's disease (PD) is the second most common neurodegenerative disorder after AD and is characterised by tremor, muscular rigidity, bradykinesia, and postural instability (Riess et al., 1998). The symptoms are thought to result from abnormal dopamine levels or degeneration of dopaminergic neurons in affected areas of the central nervous system. PD is also characterised by the widespread distribution of intracytoplasmic inclusion bodies (Lewy bodies), which are between 5 and 25nm in diameter (Kim et al., 1999a). It has been proposed that Lewy bodies have a causative role in neurodegeneration, as these aggregates have also been identified in certain cases of AD (Mezey et al., 1998).

The main constituent of Lewy bodies is α -synuclein (Spillantini et al., 1997) and mutations within α -synuclein have been observed in cases of familial PD (Riess et al., 1998). α -synuclein is a 140kDa protein that is homologous to two other gene products, β - and γ -synuclein (Clayton and George, 1998). The functions of all three proteins remains to be determined. A portion of α -synuclein known as the non-amyloid component (NAC), has been identified in the Lewy bodies of PD patients (Arima et al., 1998) and in the neuritic plaques of AD patients (Wirhth et al., 2000). This is an interesting finding as NAC acts as a substrate for TGases (Jensen et al., 1995) and is neurotoxic to primary dopaminergic neurons (Forloni et al., 2000). NAC contains glutamine and lysine residues that function as carboxamide and amine donors, respectively, for $\epsilon(\gamma$ -glutamyl)lysine formation mediated by TGase, and this results in the development of NAC polymers. Aggregated forms of NAC are more toxic than the monomeric form (El-Agnaf et al., 1998, Forloni et al., 2000), so TGases may have a role in the pathology of PD (and AD) by catalysing the formation of toxic NAC aggregates.

1.2.4 Progressive supranuclear palsy

Progressive supranuclear palsy (PSP), is a neurodegenerative disorder with clinical symptoms such as postural instability, vertical gaze palsy, axial rigidity and dementia (Golbe et al., 1988). Pathologically PSP is characterised by neuronal loss, gliosis,

neuropil threads, tau immunoreactive astrocytes, and the formation of neurofibrillary tangles (NFT) (Feany and Dickson, 1996, Bereron et al., 1997, Komori, 1999).

Neurofibrillary tangles occur in the affected areas of PSP patients (Flament et al., 1991). Similar aggregates are found in AD and it has been postulated that TGases may have a role in the formation of neurofibrillary tangles in PSP based on the finding that: 1) TGase activity is increased in AD brains and; 2) tau (the main component of the tangles) is a substrate for TGase. It has been demonstrated that tau is a major constituent of the insoluble proteins isolated from the brain of PSP patients and that these tau aggregates include $\epsilon(\gamma\text{-glutamyl})\text{lysine}$ crosslinks (Zemaitaitis et al., 2000), furthermore, it was found that the levels of $\epsilon(\gamma\text{-glutamyl})\text{lysine}$ crosslinks in the vulnerable brain regions globus pallidus and pons was remarkably greater in PSP patients than in controls. Recently, it has been demonstrated that TGase 1 and 2 message and protein are expressed selectively in brain regions of PSP and that TGase 1 is significantly upregulated in the globus pallidus and cerebellum in PSP, regions rich in NFT pathology (Zemaitaitis et al., 2003). Future studies need to determine that TGase isoforms 1 and 2 have a role in the development and progression of PSP.

1.2.5 Huntington's disease

Expansion of CAG trinucleotide repeats encoding polyglutamine stretches has been recognised as a common mutation for eight neurodegenerative diseases (Cummings and Zoghbi, 2000) including Huntington's disease (HD), spinobulbar muscular atrophy, dentatorubral pallidolusian atrophy and spinocerebellar ataxias types 1, 2, 3, 6, and 7. All eight disorders are inherited in an autosomal dominant fashion, and are characterised by a gradual decrease in coordination and/or motor neuron degeneration and are linked with different mental syndromes (Vonsattel and DiFiglia, 1998). The CAG triplet expansion arises in a different gene in each of the diseases, which alters a protein that is specific to the disease. In all these genes, the CAG trinucleotide repeats are located in the coding region and are translated into a stretch of polyglutamine in the expressed proteins. The polyglutamine stretches are the only common homologous domains found within the proteins (Trottier et al., 1995). Of these disorders, HD is the most prevalent.

Huntington's disease (HD) is a progressive neurodegenerative disorder with a characteristic pattern of neuronal loss, and consequent chorea, psychiatric alterations and intellectual decline (Duyao et al., 1993). It is inherited in an autosomal dominant fashion and affects ~1 in 10,000 individuals in most populations of European origin (Huntington's Disease Collaborative research Group, 1993). The onset of HD generally occurs in adults in mid-life, with a long-term duration of 15-20 years (Mende-Mueller et al., 2001). Occasionally, HD is expressed in juveniles, typically exhibiting more severe symptoms including rigidity and a more rapid course. Juvenile onset of HD is associated with a preponderance of paternal transmission of the disease allele. Currently, there is no cure for HD and so no therapeutic approach to delay the onset of symptoms (Huntington's Disease Collaborative research Group, 1993).

1993 marked a major breakthrough in research into HD with the discovery that the disease-causing mutation is the expression of a variable stretch of CAG triplets in the first exon of the IT15 gene that encodes a widely expressed 348kDa cytoplasmic protein named huntingtin (Huntington's Disease Collaborative research Group, 1993). Normal individuals contain <35 CAG repeats, whereas individuals with adult-onset HD possess an expansion of 38/39-55 CAG repeats. Expansions of 70 or more repeats are associated with juvenile-onset HD. HD brains display characteristic neuropathological alterations, graded from 0 to 4, with grade 4 representing severe brain atrophy. Advanced grades show a reduction in striatum, cerebral cortex, as well as hippocampus, amygdala, and thalamus brain tissues. Neuronal loss is especially severe in striatum (Mende-Mueller et al., 2001). In spite of early recognition that normal huntingtin function is required for embryonic development, evidence of its function in brain cells over the whole lifetime is lacking.

Several investigators have proposed that HD could be caused in part by abnormal protein-protein interactions related to the elongated polyglutamine stretch in huntingtin (Lesort et al., 2000). A pathological hallmark of HD brain is the presence of cytoplasmic and nuclear aggregates in specific neuronal populations that contain the NH₂-terminal region of mutant huntingtin (DiFiglia et al., 1997), although it is not clear whether they are beneficial (Saudou et al., 1998), or harmful (Davies et al., 1997). Two potential mechanisms have been proposed to explain how expanded polyglutamine domains form insoluble aggregates. It has been hypothesised that

polyglutamine peptides self associate, forming a noncovalent 'polar zipper' involving a beta sheet held together by hydrogen bonds between the main chain and the side chain amines (Perutz et al., 1994). The second proposed mechanism is that an enzymatic reaction, possibly in association with a polar zipper mechanism, is the contributing factor to the protein-protein interactions. It has been hypothesised that, because of the role glutamine plays in the reaction catalysed by TGase, an increase in the number of glutamines beyond a threshold may result in a protein becoming a TGase substrate (Green, 1993), and therefore contribute to aggregate formation in HD brain.

In support of the hypothesis that TGase is involved are the findings that total TGase activity and TGase 2 levels are increased in HD brain (Karpuj et al., 1999, Lesort et al., 1999) in a grade and region dependent manner compared to control brain. In fact, it has been shown that the mRNAs for TGase 1, 2 and 3 are present in human brains (Kim et al., 1999), and all three mRNAs are elevated in HD patient brains compared to age matched controls (Kim et al., 2002). Also, *in vitro* mutant huntingtin is a substrate for TGase 2 (Kahlem et al., 1998). In contrast, it has been demonstrated that TGase 2 selectively associates with truncated, but not full-length huntingtin, and that in intact cells neither wild type nor mutant huntingtin is a TGase 2 substrate. However, there are proteins that are TGase 2 substrates which selectively associate with mutant truncated huntingtin (Chun et al., 2001). These findings indicate that it is unlikely that TGase 2 has a causative role in aggregate formation in HD brain but that it may contribute to the etiology of HD by modifying proteins which preferentially associate with truncated mutant huntingtin.

A possible role for TGase 2 in aggregate formation has been proposed to occur in a series of events as follows (Cooper et al., 2002): 1) an expanded polyglutamine domain results in structural changes which cause a tendency towards misfolding of the mutated protein; 2) misfolding leads to cleavage of the mutated protein by caspases resulting in a truncated form which is then able to enter the nucleus; 3) TGase 2 can associate with wild type and mutant truncated huntingtin, but only modifies huntingtin associated proteins (HAPs) that bind to truncated mutant huntingtin; 4) modification of these HAPs (which may be transcription factor (TF) regulators) by TGase 2 might change their function resulting in increased

transcription of cell death genes and/or reduced transcription of cell survival genes contributing to the cell death process in HD.

Studies clearly demonstrated that neither mutant nor wild type huntingtin (full-length, truncated or highly truncated) is a TGase 2 substrate *in situ* (Lesort et al., 2002), however, proteins that are associated specifically with the truncated mutant but not wild type huntingtin are selectively polyaminated by TGase 2 (Chun et al., 2001). In addition, several of the proteins that selectively associate with mutant polyglutamine proteins contain polyglutamine or Gln-rich domains and therefore may be substrates for TGase 2 (Lesort et al., 2002). For example, the transcriptional co-activators CREB binding proteins and TAFII₁₃₀ contain polyglutamine and/or Gln-rich regions and selectively associate with mutant polyglutamine proteins. Previous studies have demonstrated that proteins can be polyaminated by TGases *in vitro* (Lesort et al., 2000) and *in vivo* (Piacentini et al., 1998), and it has been shown that inclusion of polyamines into particular protein substrates can alter their function and/or metabolism. Therefore, it has been proposed that the TGase 2 catalysed polyamination of particular proteins that are specifically associated with the mutant truncated huntingtin alters the function and/or localisation of these proteins and therefore may be a factor in the etiology of HD (Lesort et al., 2000).

Even though the pathological mechanism by which the CAG repeats cause neuronal death are still not known, one hypothesis is that the mutation may lead to disruption of energy metabolism resulting in slow excitotoxic neuronal death (Beal, 1998). There is significant evidence to support this hypothesis. Early ultrastructural studies of cortical biopsies obtained from patients with either juvenile or adult onset of HD showed abnormal mitochondria (Goebel et al., 1978). HD brain has also provided evidence of impaired mitochondrial enzyme activity. For example, early studies showed deficiencies in pyruvate dehydrogenase and succinate dehydrogenase activities in the striatum of HD cases relative to controls (Butterworth et al., 1985). Also in support of the hypothesis, changes in the respiratory chain enzyme activities in HD post-mortem brain clearly show the presence of defects in mitochondrial complexes II, III and IV (Browne et al., 1999). Dysfunction of the mitochondria leads to a decrease in cellular ATP production and alterations in ionic-ATPase-pumps function. This can lead to a partial membrane depolarisation, removal of the voltage-dependent Mg²⁺ blockage of

N-Methyl-D-Aspartate receptor-associated channels permitting their activation by endogenous glutamate resulting in calcium influx into the neurons (Riepe et al., 1995). A rise in the intracellular concentration of calcium has been implicated in having a role in the increased production of reactive oxygen species and oxidative damage. All these factors are thought to constitute an appropriate environment to activate TGase for the following reasons: (1) TGase is a calcium-dependent enzyme, so an increase in intracellular calcium concentration may result in an abnormal TGase activation in the nucleus (Lesort et al., 1998); (2) It has previously been reported that, *in vitro*, oxidation can modify specific proteins making them better substrates for TGase 2, and this could contribute to the increased TGase 2-modification of protein in response to oxidative stress *in situ* (Groenen et al., 1993).

1.2.5.1 Mouse studies

Evidence of the association of mitochondrial dysfunction with HD has become apparent due to animal studies using mitochondrial toxins such as 3-nitropropionic acid (3-NP). 3NP is a plant toxin and mycotoxin that has been associated with neurological illness in humans and animals and is an irreversible inhibitor of succinate dehydrogenase and therefore inhibits the Krebs cycle and mitochondria complex II activity of the electron transport chain (Alston et al., 1977, Coles et al., 1979). Administration of 3-NP to non-human primates resulted in neuropathological and biochemical changes that were similar to those observed in HD. For example, administration of 3-NP resulted in the presence of striatal lesions that were similar to the neurodegeneration pattern that arises in HD (Brouillet and Hantraye, 1995, Brouillet et al., 1995). 3-NP treatment also caused various dyskinetic movements and dystonic postures similar to the clinical features of HD. These findings suggest that mitochondrial dysfunction may contribute to the pathogenesis of HD.

Several types of transgenic mice have been generated in an attempt to model the pathogenesis of HD. All types of transgenic mice expressing an abnormal polyglutamine (polyQ) sequence demonstrate a progressive HD-like pathology (Mangiarini et al., 1996, Hodgson et al., 1999, Guidetti et al., 2001). One of the most widely characterised models is the Bates R6 line (Mangiarini et al., 1996) (which expresses exon 1 of the huntingtin gene with a 150 CAG expansion), which shows

neuropathological changes consistent with HD (Davies et al., 1997, DiFiglia et al., 1997).

Two groups have studied how the administration of the TGase inhibitor cystamine affects HD transgenic mice. One group demonstrated that the TGase inhibitor prolongs the lives of HD transgenic mice and reduces abnormal movements (Dedeoglu et al., 2002). The other group also showed that the transgenic mice had an increased lifespan and in addition, demonstrated that the administration of cystamine slowed aggregate formation (Karpuj et al., 2002). These preliminary results suggest that TGase inhibitors may be a therapeutic tool for the neurodegenerative diseases.

Studies revealed that in contrast to HD, TGase activity was not elevated in the brains of the R6/2 mice (Johnson et al., 1997). But although TGase was not elevated in the brain, knocking out TGase 2 significantly improved rotorod performance and increased the lifespan of the R6/2 mice. Mastroberardino *et al.* created a new animal model by mating R6/1 mice transgenic for exon 1 of the human huntingtin gene (Mangiarini et al., 1996) and knockout mice with a disrupted TGase 2 gene (De Laurenzi and Melino, 2001). Studies revealed a reduction in neuronal cell death and prolonged survival occur in R6/1, TGase 2^{-/-} mice compared to R6/1 mice (Mastroberardino et al., 2002), which is in agreement with the results obtained by Johnson *et al.* It was also found that the transgene expression resulted in a marked increase in TGase 2-catalysed ε(γ-glutamyl)lysine bonds compared to control mice indicating that the polyglutamine tracts are TGase 2 substrates *in vivo*. Correspondingly, depletion of TGase 2 in HD transgenic mice resulted in a significant reduction in crosslink formation. However, as the crosslinks were not completely abolished and some remaining ability of brain homogenates to catalyse crosslinking was observed, it suggests that other TGases are active in HD mouse brain. Interestingly however, an increase in neuronal intranuclear inclusions (NII) was found in R6/1, TGase 2^{-/-} mice compared to control R6/1 mice (Mastroberardino et al., 2002). This finding gives support to the previous suggestion by some authors (Kuemmerle et al., 1999, Saudou et al., 1998) that aggregates are not necessarily toxic. Cell death in the cortex and striatum of HD transgenic mice is characterised by condensed neurons lacking the typical apoptotic features (Turmaine et al., 2000, Iannicola et al., 2000). Studies revealed a substantial reduction in the number of dying

cells in the neocortex and striatum of HD transgenic/TGase 2 knockout brains, as compared to HD transgenic animals. This finding taken together with the finding that there was a significant improvement in motor performances and survival of R6/1, TGase 2^{-/-} mice in comparison to R6/1 mice indicates an important role for TGase 2 in the regulation of neuronal cell death occurring in HD and it is unlikely that TGase 2 plays a vital role in the formation of aggregates in HD. Fig. 1.8 provides a summary diagram indicating how TGase 2 may have a role in HD.

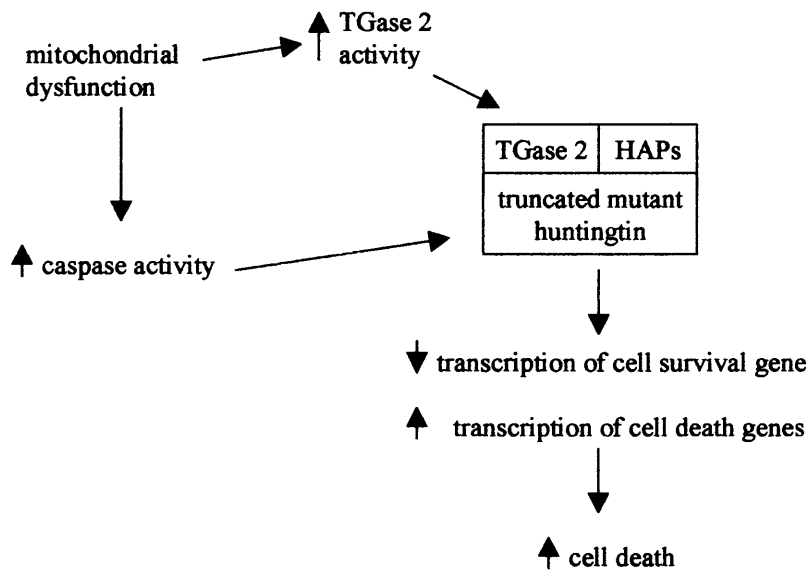


Fig. 1.8 Possible role of TGase 2 in HD. In HD, mitochondrial function is impaired which could result in an increase in TGase 2 activity, as well as an increase in caspase activation. Activation of caspase could lead to proteolysis of huntingtin resulting in a cleaved huntingtin product that is able to associate with TGase 2. TGase 2 is only able to modify huntingtin associated proteins (HAPs) that bind to truncated mutant huntingtin. TGase 2 modification of HAPs may result in a change in their function causing an increase in transcription of cell death genes and/or a decrease in transcription of cell survival genes contributing to the cell death process in HD. Adapted from (Lesort et al., 2002).

1.2.6 Conclusion

There is compelling evidence that TGases have specific roles in the etiology of neurodegenerative disease. The following factors determine the extent to which TGases contribute to a disease: 1) accessibility of potentially toxic substrates; 2) the process of enzyme activation; and 3) which isoform of TGase that is activated. It is interesting that the changes in expression of TGase isoform mRNAs differs between

the various diseases, which indicates that there may be specificity regarding the involvement of these enzymes. The mRNAs for TGase 1 and 2 are increased in PSP, AD and HD. TGase 3 mRNA is also increased in HD. However, it is not yet known if the mRNAs are translated into active enzymes in all cases. It is vital to determine which isoforms are activated in each disease as substrates of these isoforms may include proteins and polyamines critical in the disease process. It is also important to determine whether the three novel TGases, TGase 5, 6, and 7 are expressed in the brains of diseased patients, as these novel TGases may also contribute to the increased TGase activity identified in diseased brain and may have a vital role in the progression of the disease. Determination of the role of each TGase isoform would provide a basis to form new strategies for treatment of these pathologies. Table 1.2 provides a summary of the regions of the brain affected by each neurodegenerative disease.

Table 1.2. Summary of the various parts of the brain affected by the different neurodegenerative diseases.

Neurodegenerative disease	Affected brain regions
Alzheimer's disease	Cerebral cortex Hippocampus Amygdala
Parkinson's disease	Basal ganglia Pons
Progressive supranuclear palsy	Globus pallidus Pons
Huntington's disease	Striatum Cerebral cortex Hippocampus Amygdala

1.3 Aim of thesis

The TGase 6 gene product was identified in a small lung carcinoma cell line, called H69. This cell line has neuronal cell characteristics such as the expression of neuron-specific enolase and brain isozyme of creatine kinase, which suggests that TGase 6 expression may be specific to neuronal cells.

The aims of this thesis were to:

Aim 1

To identify, clone and sequence human and mouse TGase 6 transcripts in order to determine whether the presumptive TGase gene identified was a functional gene.

Aim 2

To determine the expression pattern of TGase 6 using Northern blotting and *in situ* hybridisation. Based on its expression pattern we may then be able to develop a hypothesis as to its functional role.

Aim 3

To generate antibodies for the two isoforms of TGase 6, TGase 6_S (with exon XII alternatively spliced) and TGase 6_L. The antibodies are then to be used in immunohistochemistry to verify results obtained by *in situ* hybridisation, and for the determination of TGase 6 expression in various areas of mouse brain.

Aim 4

To generate the full-length TGase 6 cDNA from amplified PCR fragments in order to recombinantly express the enzyme and to perform experiments to determine if any hypotheses made are valid.

Aim 5

To determine if antibodies to TGase 6 are present in sera of patients with a neurological manifestation.

2.0 Characterisation of a novel transglutaminase, TGase 6

2.1 Introduction

The genes encoding TGase 2 and TGase 3 have been mapped to human chromosome 20q11. Previous work in the laboratory identified a third gene, TGM6, which is adjacent to the TGM3 gene on chromosome 20q11 (Grenard et al., 2001). The aim of this chapter was to identify, clone and sequence human and mouse TGase 6 transcripts in order to determine whether the presumptive TGase gene identified is a functional gene.

2.2 Materials and Methods

2.2.1 Cell culture

The H69 (European Collection of Cell Cultures, Salisbury, UK) cell line was established from the pleural fluid of a 55 year old Caucasian male with small cell carcinoma of the lung prior to treatment. When inoculated into athymic mice, these cells form tumours with a typical histology of the original biopsy specimen. Elevated levels of the biomarkers L-dopa decarboxylase, bombesin-like immunoreactivity, neuron-specific enolase and brain isozyme of creatine kinase were observed suggesting that they are derived from the neuronal lineage. The cells were cultured in RPMI 1640 medium (Life Technologies) supplemented with 10% (v/v) heat inactivated foetal bovine serum, 100 units/ml penicillin, 100µg/ml streptomycin and 2mM glutamine. Cultures were incubated at 37°C, 5% CO₂ and the spherical, floating aggregates were maintained in medium at 60,000 to 300,000 cells/ml. Cells were passaged by centrifugation of the cell suspension at 1,500xg for 5 min and the cell pellet was then resuspended in fresh media.

2.2.2 RNA extraction using TRIzol for Northern blotting

10⁶ H69 cells were harvested by centrifugation at 1,500xg, for 5 min, and the cell pellet washed by resuspension of the cell pellet in 3ml phosphate-buffered saline (PBS) warmed to room temperature (RT), by pipetting and centrifugation at 1,500xg for 5 min, repeated 3 times. The cell pellet was resuspended in 1.5ml ice-cold TRIzol Reagent (Invitrogen) and the solution left at RT for 5 min to allow complete lysis. The solution was split and 750µl was transferred to each of two eppendorf tubes. 175µl chloroform was added to each tube, the mixture shaken for 15 sec and incubated for 3 min at RT to precipitate any protein. Two heavy phase-lock gels (PLG) (Eppendorf) were pelleted by centrifugation at 12,000xg for 30 sec. Each RNA extraction mixture was transferred to a PLG to bind DNA and centrifuged at 13000xg for 2 min. The upper aqueous phase of each was transferred to a new 1.5ml eppendorf tube. An equal volume (~ 500µl) of isopropanol was added and each mixture incubated at -20°C for 20 min to allow precipitation of the RNA. Each mixture was centrifuged at 12,000xg for 10 min at 4°C. The supernatant was removed, each pellet washed in 0.5ml 75%

ethanol and centrifuged at 7,500xg for 5 min at 4°C. The supernatant was removed and each pellet air-dried briefly. The pellets were resuspended in RNase-free H₂O to a volume of 200µl. Each sample was heated to 65°C for 5 min to help resuspend the pellet and then incubated on ice for 1 hr. The samples were then combined and the total RNA was ethanol precipitated. Briefly, 1µl 0.5M EDTA and 20µl phenol/chloroform (Invitrogen) were added to the sample and centrifuged at 16,000xg for 3 min at 4°C. The top aqueous phase was transferred to a new 1.5ml eppendorf and 22µl 4M ammonium acetate (Invitrogen) and 88µl 100% ethanol was added. The sample was vortexed and frozen on dry ice. The sample was centrifuged at 16,000xg for 15 min at 4°C and the supernatant discarded. The pellet was resuspended in 10µl RNase-free H₂O and either stored at -80°C or converted to mRNA using the Fast-Track Kit (Invitrogen) as described in section 2.2.4.

2.2.3 RNA extraction using TRI reagent

10⁶ H69 cells were harvested by centrifugation at 1,500xg for 3 min. The cell pellet was resuspended in 1ml Tri reagent (a mixture of guanidine thiocyanate and phenol in a mono-phase solution) (Sigma) and the sample left at RT for 5 min to allow complete dissociation of nucleoprotein complexes. 200µl chloroform was added to precipitate proteins, and samples incubated for 10 min at RT. The sample was centrifuged at 12,000xg for 15 min at 4°C. Centrifugation separates the mixture into three phases: a red organic phase (containing protein), an interphase (containing DNA) and a colourless upper phase (containing RNA). The top aqueous phase was transferred to a new 1.5ml eppendorf tube and 0.5ml isopropanol was added to precipitate the RNA. The sample was incubated for 10 min at RT and then centrifuged at 12,000xg for 10 min at 4°C. The RNA pellet was washed with 1ml 75% ethanol, air-dried for 10 min and resuspended in 20µl RNase-free H₂O. The total RNA was either stored at -80°C or converted to mRNA.

2.2.4 Conversion of total RNA to mRNA

Total RNA was converted to mRNA using the MicroFast Track kit (Invitrogen). 500µg total RNA was added to 1ml MicroFast Track™ 2.0 Lysis buffer (1ml Stock

buffer (200mM NaCl, 200mM Tris, pH 7.5, 1.5mM MgCl₂, 2% SDS) and 20μl Protein/RNase Degradar (proprietary mixture of proteases)), heated to 65°C for 5 min and then immediately placed on ice for 1 min. To remove any insoluble material the sample was centrifuged at 4,000xg for 5 min at RT and the supernatant transferred to a new, sterile microcentrifuge tube. 5M NaCl solution (63μl) was added to adjust the NaCl concentration of the lysate to 0.5M final concentration. The solution was added to a vial containing oligo(dT) cellulose, and the cellulose was left to swell for 2 min. The tube was rocked gently for 20 min on a rocking platform to increase the efficiency of mRNA binding to oligo(dT) cellulose. The oligo(dT) cellulose was pelleted by centrifugation at 4,000xg for 5 min at RT. The oligo(dT) cellulose was washed three times by gentle resuspension in 1.3ml Binding buffer (10mM Tris-HCl, pH 7.5, 500mM NaCl) and centrifuged as above. The resin was resuspended in 0.3ml Binding Buffer and transferred to a spin-column (inside the spin-column/microcentrifuge tube set). The sample was centrifuged at 4,000xg for 10 sec at RT, and the resin washed three times by addition of 500μl Binding buffer to the resin bed and centrifugation at 4,000xg for 10 sec at RT. 200μl Low Salt Wash Buffer (10mM Tris-HCl, pH 7.5, 250mM NaCl) was added and the resin gently resuspended with a sterile pipette tip. The sample was centrifuged at 4,000xg for 10 sec at RT. The low salt wash removes SDS and non-polyadenylated RNAs such as rRNAs. This step was repeated and then the spin column placed into a new microcentrifuge tube. 100μl Elution Buffer (10mM Tris-HCl, pH 7.5) was mixed into the cellulose bed with a sterile pipette tip, and the tube centrifuged at 4,000xg for 10 sec at RT. This step was repeated resulting in 200μl of eluate, which contains the mRNA. The mRNA was precipitated with 10μl 2mg/ml glycogen carrier, 30μl 2M sodium acetate and 600μl 100% ethanol, and the solution placed on dry ice. The solution was centrifuged at 16,000xg for 15 min at 4°C to collect the mRNA and the pellet resuspended in 10μl Elution Buffer.

2.2.5 cDNA synthesis

cDNA was prepared from human cell lines or mouse brain tissue using the cDNA Cycle Kit (Invitrogen) or Superscript II RNase H⁻ Reverse Transcriptase (Life Technologies).

2.2.5.1 cDNA synthesis using the cDNA Cycle kit (Invitrogen)

0.5µg total RNA or 50ng mRNA was added to a nuclease-free microcentrifuge tube with 5µM Oligo(dT) primer and sterile H₂O to a total volume of 12.5µl. The mixture was heated to 65°C for 10 min to remove the secondary structure and then placed at RT for 2 min. 1 unit RNase inhibitor, 1x RT Buffer, 5mM dNTPs, 4mM Sodium Pyrophosphate and 5 units AMV Reverse Transcriptase was added to a total volume of 20µl, and the reaction incubated at 42°C for 60 min. The enzyme was heat inactivated at 95°C for 2 min, the reaction chilled on ice for 1 min, then 5 units AMV Reverse Transcriptase added for a second round of reverse transcription as above. After inactivation of the enzyme, the DNA was purified by phenol-chloroform extraction. Briefly, 1µl 0.5M EDTA, pH 8.0 and 20µl phenol-chloroform was added. The reaction was vortexed, centrifuged for 3 min and the top aqueous layer carefully removed and transferred to a new, sterile microcentrifuge tube. 22µl 4M ammonium acetate and 88µl 100% ethanol was added to precipitate the DNA and the tube placed on dry ice for a minimum of 10 min. DNA was collected by centrifugation at 16,000xg for 15 min at 4°C. The pellet resuspended in 20µl nuclease-free H₂O, and stored at -20°C.

2.2.5.2 cDNA synthesis using SuperScript II

5µg total RNA or mRNA was added to a nuclease-free microcentrifuge tube with 5µM oligo(dT) primer, 1mM dNTP Mix (1mM each dATP, dGTP, dCTP and dTTP) and sterile, distilled H₂O to a total volume of 13µl. The solution was heated to 65°C for 5 min and then chilled on ice for 1 min. 1x first-strand buffer (250mM Tris-HCl, pH 8.3, 375mM KCl, 15mM MgCl₂) and 10mM dithiothreitol (DTT) was added to a volume of 19µl, and the reaction mixture pre-warmed to 42°C for 2 min. 1µl (200 units) SuperScript II RNase H⁻ reverse transcriptase was added and the reaction carried out at 42°C for 50 min, before termination by enzyme inactivation at 70°C for 15 min.

2.2.6 Amplification of TGase 6 from various cell lines

cDNA had previously been prepared from the following human cell lines, osteosarcoma cells (MG-63), fibrosarcoma cells (HT 1080), erythroleukemia cells (HEL), mammary epithelium carcinoma cells (MCF7), HELA cells, embryonal kidney epithelial cells (293), platelets and primary keratinocytes. RT-PCR analysis was carried out with 2.5 units of *Taq* DNA Polymerase (Fisher Scientific) and 5 μ l cDNA from each cell line in 100 μ l of 1x Assay Buffer B (10mM Tris/HCl, pH 8.3, 50mM KCl containing 2.5mM MgCl₂), 0.2mM dNTPs and 5 μ M of the TGase 6 specific oligonucleotide primers corresponding to putative exon II and exon IX sequences:

5'-ATCAGAGTCACCAAGGTGGAC (fP1) and

5'-TCACAGTCCAGGGCTCTGCTCAG (rP1);

5'-CAGCCTCAGTCACCGCCATCCGC (fP2) and

5'-GATACTTGTAGGTCAGTGATG (rP2).

The PCR cycles were 45 sec at 94°C (denaturation), 2 min at 55°C (annealing) and 3 min at 72°C (elongation) and were carried out with a Robocycler (Stratagene). A total of 37 cycles were made, with the first cycle containing an extended denaturation period (6 min) during which the polymerase was added (hot start), and the last cycle containing an extended elongation period (10 min). The PCR products were analysed on 1% agarose gel.

2.2.7 Agarose gel electrophoresis

10 μ l of amplified products were mixed with 2 μ l 6x Sample buffer (3% Glycerol, 0.25% Bromophenol blue in distilled H₂O) analysed on 1% agarose gels (1% agarose dissolved in 1x TAE (4mM Tris acetate, 0.2mM Na₂EDTA, pH 8.5) by heating for 2 min in a microwave). The gels were electrophoresed in 1x TAE buffer containing 0.25mg/l ethidium bromide (Sigma) in a RunOne System.

2.2.8 Cloning of TGase 6 by anchored PCR

Poly(A⁺) RNA was prepared from about 10⁶ cells or 10µg total RNA by oligo(dT)-cellulose column chromatography using the Micro-Fast Track Kit as described above in section 2.2.4. The poly(A⁺) RNA was reverse transcribed into DNA into a total volume of 20µl using the cDNA Cycle Kit as described in section 2.2.5.1. Overlapping fragments of TGase 6 were amplified by PCR using oligonucleotides fP1 and 5'-AGAAACACATCGTCCTCTGCACACC (rP3), 5'-CAGGCTTTCCTCTCACCGCAAACAC (fP3) and 5'-CGTACTTGACTGGCTTGTACCTGCC (rP4), 5'- TCTACGTCACCAGGGTCATCAGTGC (fP4) and 5'- GCCTGTTACCGCCTTGCTGT (rP5), 5'- CATCACTGACCTCTACAAGTATCC (fP5) and 5'- ACGGCGTGGGATTCATGCAGG (rP6), 5'-CATCCTCTATACCCGCAAGCC (fP6) and 5'- AGGTTGAGGCAGGATTAAGTACTGAGGCCTC (rP7).

PCRs were carried out with 1.25 units of AmpliTaq Gold DNA polymerase (PE Biosystems) and 2µl cDNA in a total of 50µl of supplied reaction buffer (15mM Tris-HCl, pH 8.0, 50mM KCl) supplemented with 2.5mM MgCl₂, 0.2mM dNTPs and 5µM of the appropriate gene-specific primers. Forty PCR cycles were carried out in a Gene Amp 9600 thermal cycler (PE Biosystems). Each cycle consisting of denaturation at 95°C for 45 sec, an elevated annealing temperature of 65°C for 1 min for the initial 5 cycles and an annealing temperature of 60°C for the remaining cycles, and an extension at 72°C for 1 min, with the first cycle containing an extended denaturation period (10 min) for activation of the polymerase and the last cycle containing an extended elongation period (10 min). Amplified products were analysed on 1% agarose gels, extracted from the gel and sequenced.

2.2.9 QIAquick gel extraction

The DNA fragment was excised from the agarose gel with a clean, sharp scalpel. The gel slice was weighed in a microcentrifuge tube and 3 volumes of Buffer QG to 1 volume of gel were added. The mixture was incubated at 50°C for 10 min, with

vortexing of the tube every 2-3 min during the incubation to help dissolve the gel. After the gel slice has dissolved, the colour of the mixture was checked to ensure it was still yellow. If the colour of the solution had become orange or violet, 10µl 3M sodium acetate, pH 5.0 was added to return the pH to ≤ 7.5 , as the adsorption of DNA to the QIAquick membrane is efficient only at pH ≤ 7.5 . To bind the DNA, the sample was applied to a QIAquick spin column and centrifuged at 16,000xg for 1 min. The column was washed by adding 750µl Buffer PE and centrifuged at 16,000xg for 1 min. The flow-through was discarded and the column centrifuged for an additional 1 min. The column was placed in a clean 1.5ml microcentrifuge tube and 30µl nuclease-free H₂O was added to the centre of the QIAquick membrane and left to stand for 1 min. The DNA was eluted from the column by centrifugation at 16,000xg for 1 min.

2.2.10 Sequencing

2.2.10.1 Sequencing performed using the ABI PrismTM dRhodamine Terminator Cycle Sequencing Ready Reaction Kit (PE Biosystems) and an ABI Prism 310 Genetic Analyzer.

For each reaction, 4µl Terminator Ready Reaction Mix (A-Dye Terminator labelled with dichloro[R6G], C-Dye terminator labelled with dichloro[TAMRA], G-Dye terminator labelled with dichloro[R110], T-Dye Terminator labelled with dichloro[ROX], deoxynucleoside triphosphates (dATP, dCTP, dITP, dTTP), AmpliTaq DNA Polymerase, FS, with thermally stable pyrophosphatases, MgCl₂, Tris-HCl buffer, pH 9.0), 50ng PCR product or 500ng double-stranded DNA, 5µM primer and deionised H₂O to a final volume of 10µl were added to a 0.5ml microcentrifuge tube (Treff). The reagents were mixed by pipetting and centrifuged briefly. To sequence the DNA, the tubes were placed in a GeneAmp 9600 Thermal cycler. 25 cycles were performed consisting of denaturation at 96°C for 20 sec, an annealing temperature of 50°C for 10 sec and an extension period at 60°C for 4 min. The contents of the tubes were centrifuged briefly and then the reaction products purified. For each sequencing reaction, 10µl deionised water, 2µl 3M sodium acetate, pH 5.0, and 50µl 95% ethanol were added and incubated for 20 min at RT in the dark. The samples were centrifuged at 16,000xg for 20 min at RT, the pellet gently rinsed

with 250µl 70% ethanol, and dried at 90°C for 5 min. Each pellet was resuspended in 25µl Template Suppression reagent, vortexed and then heated at 95°C for 2 min to denature and placed on ice. Each sample was then loaded into the ABI Prism™ 310 Genetic Analyzer.

2.2.10.2 Sequencing performed using the Thermo Sequenase™ Cy5/Cy5.5 Dye Terminator Sequencing Kit (Amersham Pharmacia) and a Long-Read Tower (Visible Genetics)

2.2.10.2.1 Sequencing reactions using Cy5 dye-labelled ddNTP terminators

0.2µl Cy5-labelled ddNTP (8.8µM Cy5-labelled ddATP, 7.0µM Cy5-labelled ddCTP, 17.6µM Cy5-labelled ddGTP or 14.3µM Cy5-labelled ddTTP), 1.1mM dNTP mix (1.1mM each dATP, dCTP, dGTP and dTTP), and 1.6µl RNase-free water were added to an appropriately labelled tube (“A mix”, “C mix”, “G mix” or “T mix”). The reagents were mixed, centrifuged briefly and placed on ice in the dark. For each template being sequenced, four new tubes were labelled “A”, “C”, “G” and “T”, and placed on ice. 2µl of the “A mix” was dispensed into the tube labelled “A” and this process repeated for C, G, and T, then each tube was placed on ice in the dark. For each template, a master mix of 0.2-0.5µg plasmid DNA or 50-200fmol PCR product, 4µM primer, 3.5µM reaction buffer (15mM Tris-HCl, pH 9.5, 3.5mM MgCl₂), 1µl Thermo Sequenase DNA polymerase (10 units/µl with 0.0017 units/µl *Thermoplasma acidophilum* inorganic pyrophosphatase in 40mM Tris-HCl, pH 8.0, 1mM DTT, 0.1mM EDTA, 0.5% Tween-20, 0.5% Nonidet P-40, 50% glycerol) and RNase-free H₂O to a final volume of 27µl. The reagents were mixed thoroughly, centrifuged briefly and 6µl of master mix was aliquoted into each tube labelled A, C, G, and T. To sequence the DNA, the tubes were placed in a GeneAmp 9600 thermal cycler. 30 cycles were performed consisting of denaturation at 95°C for 30 sec, annealing at 55°C for 30 sec and an extension period of 72°C for 2 min. The contents of the tubes were centrifuged briefly and placed on ice.

2.2.10.2 Sequencing reactions using Cy5.5 dye-labelled ddNTP terminators

1µl of Cy5.5 ddNTP termination mix (ddA mix (150µM each dATP, dCTP, dGTP and dTTP, 3µM Cy5.5 ddATP); ddC mix (150µM each dATP, dCTP, dGTP and dTTP, 3µM Cy5.5 ddGTP); ddG mix (150µM each dATP, dCTP, dGTP and dTTP, 3µM Cy5.5 ddGTP); ddT mix (150µM each dATP, dCTP, dGTP and dTTP, 3µM Cy5.5 ddTTP)) was dispensed into appropriately labelled tube (“A mix”, “C mix”, “G mix” and “T mix”). The tubes were placed on ice in the dark. For each template being sequenced, a master mix of 0.2-0.5µg plasmid DNA or 50-200fmol PCR product, 4µM primer, 3.5µl reaction buffer, 2µl Thermo Sequenase DNA polymerase and RNase-free H₂O to a final volume of 31.5µl. The reagents were mixed, centrifuged briefly and 7µl of master mix was aliquoted into each tube labelled A, C, G and T and placed on ice in the dark. The DNA was sequenced on a thermal cycler as for Cy5.

2.2.10.3 Purification of sequencing reactions using ethanol precipitation

2µl 7.5M ammonium acetate, 2µl 10mg/ml glycogen solution and 30µl chilled 100% ethanol were added to each of the individual reactions, mixed thoroughly and placed on ice for 20 min. The reactions were centrifuged at 16,000xg for 20 min at 4°C and the pellet washed with 200µl chilled 70% ethanol. The reactions were centrifuged at 16,000xg for 5 min and 6µl Formamide loading dye (deionised formamide containing a proprietary dye) was added to each tube. The reactions were vortexed to ensure complete resuspension of the pellets. Just prior to loading the samples onto the gel, each sample was heated to 70°C for 3 min and quenched on ice immediately. 1.5µl of each sample was loaded into separate lanes of either a long or short polyacrylamide sequencing gel in a Long Read TowerTM automated DNA sequencer (Visible Genetics, Inc.). The gels were electrophoresed in 1x TBE buffer (8.9mM Tris, pH 8.0, 8.9mM Boric acid, 0.2mM EDTA), at 60°C, 1500V, 50% Laser power, 1 sec sampling for 180 min, for Long Read gels, and at 53°C, 1300V, 50% laser power, 0.5 sec sampling for 35 min for Short Read gels.

2.2.11 Rapid amplification of 5'- mRNA end

A modified RACE protocol was used to determine the transcription start site and obtain additional sequence information of exon I of TGase 6. Total RNA was extracted from H69 cells using Tri Reagent as described in section 2.2.3. The total RNA was purified using RNase free amplification grade DNase I (GIBCO) to eliminate any residual genomic DNA from the RNA sample. 10µg total RNA, 1x Reaction buffer (20mM Tris-HCl, pH 8.4, 50mM KCl, 2mM MgCl₂), 20 units Amplification grade DNase 1 and DEPC-treated H₂O to a final volume of 50µl were added to a 0.5ml microcentrifuge tube on ice. The sample was incubated for 15 min at RT. The reaction was terminated by the addition of 0.5mM EDTA. The sample was incubated at 65°C for 15 min to heat inactivate the DNase 1, then placed on ice for 1 min. The reaction was collected by brief centrifugation.

cDNA was synthesised from 5µg of purified total RNA by reverse transcription using SuperScript II RNaseH⁻ reverse transcriptase as described in section 2.2.5.2. The cDNA was purified from nucleotides using the Concert Gel Extraction System (GIBCO). 20µl cDNA was added to 25µl Gel Solubilisation Buffer (containing concentrated sodium perchlorate, sodium acetate and TBE-solubilizer) and the mixture pipetted into the cartridge of a spin column unit (cartridge inserted into 2ml collection tube). The spin column unit was centrifuged at 12,000xg for 1 min. 700µl wash buffer (containing Tris-HCl, NaCl, and EDTA) was pipetted onto the cartridge and incubated for 5 min at RT. The spin column unit was centrifuged at 12,000xg for 1 min and the cartridge placed in a new, sterile 1.5ml microcentrifuge tube. 40µl pre-warmed TE buffer (10mM Tris-HCl, pH 8.0, 0.1mM EDTA) was added directly to the centre of the cartridge and incubated for 1 min at RT. cDNA was eluted from the cartridge by centrifugation at 12,000xg for 2 min.

The purified cDNA was tailed in the presence of 200µM dCTP with 10 units of terminal deoxynucleotidyl transferase (Promega) for 30 min at 37°C to anchor the PCR at the 5' end. The PCR reaction was anchored by performing a total of 5 cycles of one-sided PCR at a lower annealing temperature (37°C) with the abridged anchor primer, 5'-GGCCACGCGTCGACTAGTACGGGIIGGGIIGGGIIG (Life

Technologies) only. Following transfer of 25% of this reaction at 94°C to a new tube containing abridged anchor primer and TGase 6-specific primer rP3 or 5'-GATGTCTGGAACACAGCTTTGG (rP8), the first round of amplification was carried out for a total of 40 cycles in a Gene Amp 9600 thermal cycler. Each cycle consisted of denaturation at 94°C for 45 sec, an annealing temperature of 55°C for 1 min, and extension at 72°C for 1.5 min, with the first cycle containing an extended denaturation period (10 min) and the last cycle containing an extended elongation period (7 min). Nested PCR was done with the universal amplification primer (UAP), 5'-CUACUACUACUAGGCCACGCGTCGACTAGTAC (Life Technologies) and TGase 6-specific primer rP1, rP3, 5'-GCGTGAGGCTGAACGACTGG (rP10) or 5'-CAGCTCAGGGCAGGGGTACTC (rP11) and 1µl from the first round PCR (PCR product amplified either with rP3 or rP8) as described above with the exception of annealing at 60°C.

2.2.12 PCR amplification of 5' end

Forward primers were designed based on the genomic sequence upstream of exon II, utilising any potential ATG start sites positioned appropriately upstream of a splice donor/acceptor site to maintain the correct reading frame:

5'-GGGAAACCAGTGCAGGATCCATG (fP5'1)

5'-GGAGTCCAGCTGGCCTTCACATG (fP5'2)

5'-GCTATTACACCACATATGGCAG (fP5'3)

5'-GAAGGACCCACTTGCCAGATG (fP5'4)

5'-CAAGTCTTCCAGGTCCAGAGG (fP5'5)

5'-CCACATGGAAAGGAACCAAG (fP5'6)

5'-CAGTTGATATTCTTGAAAGGGAAG (fP5'7)

5'-GATTGGACAAAGGAGAGATGAAC (fP5'8)

5'-TGGCTCAGCAGAAGGAAGGC (fP5'9)

PCR was carried out with 1.25 units of AmpliTaq Gold DNA polymerase (PE Biosystems) and 1µl H69 DNA in a total of 50µl of supplied reaction buffer supplemented with 2.5mM MgCl₂, 0.2mM dNTPs and 5µM of the appropriate forward primer and reverse primers located in exon II. Forty PCR cycles were carried out as described in section 2.2.8.

2.2.13 cDNA synthesis from mouse brain

Mouse brain (strain BalbC) was washed in PBS and frozen on dry ice. The frozen brain was homogenised with Tri Reagent using a Teflon pestle. The homogenate was centrifuged at 5000xg for 3 min. Total RNA was extracted using Tri Reagent as described in section 2.2.3 and cDNA was synthesised from 2µg total RNA by reverse transcription using SuperScript II as described in section 2.2.5.2.

2.2.14 PCR amplification from mouse brain cDNA/genomic DNA

PCR was carried out with 1.25 units of AmpliTaq Gold DNA polymerase (PE Biosystems) and either 1µl mouse brain cDNA (see above), or 100ng mouse genomic DNA in a total of 50µl of supplied reaction buffer (15mM Tris-HCl, pH 8.0, 50mM KCl) supplemented with 25mM MgCl₂, 0.2mM dNTPs and 5µM of the selected TGase 6-specific primers. Forty PCR cycles were carried out as described in section 2.2.8 and amplified products were analysed on 1% agarose gels, extracted and sequenced.

2.2.15 Northern blotting using ³²P

3µg of poly(A⁺) RNA from H69 cells was separated in a 1.2% agarose gel containing formaldehyde and a 0.24-9.5kb RNA ladder (Life Technologies) was used to calibrate the gel. For a medium gel apparatus (EasyCast) using a 12 well comb, 100ml 1.2% agarose gel solution was prepared. 1.2g agarose was dissolved in 84ml DEPC-treated distilled H₂O by heating for ~ 2 min in a microwave. The solution was mixed and cooled to ~60°C. 1% formaldehyde and 1x MOPS (20mM MOPS, pH 7.0, 5mM Na acetate, 2mM EDTA) was added to a total volume of 100ml. The gel solution was mixed and immediately poured into the gel apparatus in a fume hood. The gel was left to set for 45 min. 1x MOPS running buffer was added to the electrophoresis tank until the buffer was even with but not covering the surface of the gel. Samples, 3µg poly(A⁺) RNA and 3µg RNA ladder (mixed with 1.5µg ethidium bromide) were each mixed with loading buffer (50% formamide, 6% formaldehyde, 1x MOPS, 10% glycerol, 0.05% Bromophenol blue, and DEPC-treated H₂O) to a final volume of

40 μ l. The samples were heated at 60°C for 15 min, and then chilled on ice. The samples were loaded into the gel and the gel electrophoresed at a constant 70V, 40mAmp, for 10 min until the samples had entered the gel. The gel was then completely covered with 1x MOPS running buffer and run at 70V, 40mAmp for ~3 hrs. The RNA was transferred to a Nytran nylon transfer membrane (Schleicher & Schuell) by capillary transfer using 20x SSC (0.3M trisodium citrate, pH 7.0, 3M NaCl) as transfer buffer. The RNA was immobilised by UV crosslinking at 1200 μ Joules using a UV Stratalinker 1800linker (Stratagene).

2.2.15.1 Preparation of probe

A ~730bp DNA fragment of human TGase 6, encoding the 3' end of the gene was generated by PCR using oligonucleotides 5'-AGGGCTGGGGGTCGCTGTC (fP7) and 5'-AATCCACCCCTCTCAGTCCCT (rP11) (nucleotides 1427-2156 in Fig. 2.3). A ³²P-labelled probe corresponding to this fragment was prepared using the Multiprime DNA Labelling System (Amersham). In summary, 25ng of DNA was denatured by heating at 100°C for 2 min and then chilled on ice. The following reaction was then set up on ice to synthesise labelled DNA using random primers: 25ng DNA, 1x buffer (dATP, dGTP and dTTP in a concentrated buffer solution containing Tris-HCl, pH 7.8, MgCl₂ and 2-mercaptoethanol), 5 μ M random primer (random hexanucleotides in an aqueous solution containing nuclease-free BSA), 5 μ l [α -³²P] dCTP, 2 units DNA polymerase I 'Klenow' fragment (in 50 μ M potassium phosphate, pH 6.5, 10mM 2-mercaptoethanol and 50% glycerol) and H₂O to a final reaction volume of 50 μ l. The reaction solution was mixed gently, then incubated at 37°C for 1 hr. Unincorporated label was removed using a ProbeQuantTM G-50 Microcolumn (Amersham Pharmacia). To prepare the G-50 Microcolumn, the resin in the column was resuspended by vortexing. The cap was loosened one-fourth turn and the bottom closure snapped off. The column was placed in a 1.5ml microcentrifuge tube for support and centrifuged at 735xg for 1 min. The column was placed in a new 1.5ml tube and 50 μ l reaction solution was slowly applied to the top-centre of the resin, ensuring that the resin bed was not disturbed. The column was centrifuged at 735xg for 2 min and the purified sample collected in the microcentrifuge tube.

2.2.15.2 Determination of cpm/ μ l of labelled probe

1 μ l of purified sample and 99 μ l 150mM STE buffer, pH 8.0 (0.438g NaCl dissolved in 40ml TE buffer (10mM Tris-HCl, pH 8.0 and 1mM EDTA) with stirring. The solution was titrated to pH 8.0 using HCl and the volume adjusted to 50ml with TE buffer) were mixed in a microcentrifuge tube. 100 μ l STE buffer, and 100 μ l STE buffer and sample were each added to 10ml sample cocktail in a scintillation vial. The samples were counted using a standard ^{32}P counting programme. The total cpm incorporated into the purified sample was determined by multiplying cpm/ μ l by the volume of the sample recovered from the column.

2.2.15.3 Hybridisation and detection

The membrane was pre-hybridised with 10ml of hybridisation solution (50% deionised formamide, 0.25M NaCl, 1mM EDTA, 7% SDS, 0.25M Na phosphate pH 7.2, milliQ H₂O to a final volume of 20ml) in a shaking incubator. The membrane was then hybridised with the probe dissolved in 10ml hybridisation solution overnight at 42°C in a shaking incubator. The membrane was washed in washing buffer (0.1x SSC, 1% SDS) for 5 min at RT, then for 20 min at 65°C, three times. When repeating the experiment the membrane was washed in 2x SSC, 0.1% SDS for 5 min at RT, in 0.1x SSC, 0.1% SDS at 65°C for 15 min, in 0.2x SSC, 0.1% SDS at 68°C for 15 min three times. The membrane was then exposed to BioMax MR film (Kodak) using a Spectroline cassette (Spectronics Corporation) for 24 hrs for the first exposure and 10 days for the second exposure.

2.2.16 Northern blotting using DIG-labelling

5 μ g of poly(A⁺) RNA extracted from H69 cells was separated in a 1.2% agarose gel containing formaldehyde and transferred as above.

2.2.16.1 Preparation of probe

A ~730bp DNA fragment of human TGase 6, was generated by PCR as described in section 2.2.15.1 and then sub-cloned into the pCRII vector by TA-cloning producing plasmid, pCRII-northern6.

2.2.16.1.1 TA-cloning

TA-cloning was performed using the TA-cloning Kit Dual Promoter (Invitrogen). *Taq* polymerase has a non template-dependent activity that incorporates a single deoxyadenosine (A) to the 3' ends of PCR products. The pCRII vector has a single deoxythymidine (T) residue which allow PCR inserts to ligate efficiently into the vector.

2.2.16.1.2 Ligation of PCR product into vector

To ligate the PCR product into the vector a 10 μ l ligation reaction was set up as follows: 10ng PCR product, 1x Ligation buffer (6mM Tris-HCl, pH 7.5, 6mM MgCl₂, 5mM NaCl, 0.1mg/ml bovine serum albumin (BSA), 7mM β -mercaptoethanol, 0.1mM ATP, 2mM DTT, 1mM spermidine), 50ng pCRII vector, 4 units T4 DNA Ligase and sterile H₂O to a final volume of 10 μ l. The ligation reaction was incubated at 14.5°C overnight, then centrifuged briefly and placed on ice (or stored at -80°C).

2.2.16.1.3 Transformation

For transformation, a 50 μ l vial of frozen One Shot competent cells (TOF10F') (Invitrogen) were thawed on ice, 2 μ l of the ligation reaction was pipetted into the cells and a pipette tip used to stir the mixture gently. The vial was incubated on ice for 30 min. The cells were heat shocked for 30 sec in a 42°C waterbath and then placed on ice. 250 μ l SOC medium (2% Tryptone, 0.5% Yeast extract, 10mM NaCl, 2.5mM KCl, 10mM MgCl₂, 10mM MgSO₄, 20mM glucose) was added and the vial incubated for 1 hr at 37°C in a shaking incubator at 225rpm. The vial was then placed on ice. 100 μ l and 200 μ l of the transformation reaction were plated onto 2 Luria-Bertia (LB)

plates containing 30µg/ml kanamycin and 40mg X-Gal (Sigma) respectively, and 40µl 100mM IPTG. The plates were left for 5 min to ensure the liquid is absorbed, then inverted and incubated at 37°C overnight. The insertion of the DNA fragment into the plasmid disrupts the Lac Z gene allowing blue/white screening of the colonies.

2.2.16.1.4 Preparation of X-Gal/LB agar plates for transformation

1l of LB agar was prepared by dissolving 10g Tryptone, 10g NaCl and 15g Agar in 950ml deionised H₂O. The pH was adjusted to 7.0 with NaOH and the volume adjusted to 1l. The medium was autoclaved and when cooled to ~55°C, 30µg/ml kanamycin or 50µg/ml ampicillin was added. The medium was poured into 10cm plates and left to harden. 40mg X-Gal stock solution was pipetted onto the centre of the plate and spread evenly with a sterile spreader. The plates were incubated at 37°C for 30 min to allow the solution to diffuse. Plates used for transformation reactions with TOP10F' cells, 40µl 100mM IPTG also needs to be spread on to the plates. TOP10F' expresses the *lac* repressor (*lacI^q*) which represses transcription from the *lac* promoter, so to perform blue-white screening for inserts IPTG must be added to the plates to express LacZ α .

2.2.16.1.5 Isolation of plasmid DNA

Single, well-isolated white colonies were scraped off a LB agar plate containing kanamycin using sterile pipette tips. Each pipette tip was ejected into 3ml of LB medium (10M Tryptone, 5M Yeast extract, 20M NaCl, 0.1mM NaOH) containing 30µg/ml kanamycin, to inoculate the medium. The cultures were incubated overnight at 37°C in a shaking incubator at 225rpm. A Wizard Plus SV Minipreps DNA Purification System (Promega) was used to isolate the plasmid DNA. 3ml of bacterial culture was divided between two 1.5ml microcentrifuge tubes and the culture harvested by centrifugation at 10,000xg for 5 min at RT. The supernatant was removed from each tube, 125µl Cell Resuspension Solution (50mM Tris-HCl, pH 7.5, 10mM EDTA, 100µg/ml RNase A) was added to each tube, the cell pellets completely resuspended by vortexing and then the two suspensions were pooled.

250µl Cell Lysis Solution (0.2M NaOH, 1% SDS) was added, the cell suspension incubated for 5 min until the cell suspension has cleared. 10µl Alkaline Protease Solution was added, and the suspension incubated for 5 min at RT. Alkaline protease inactivates endonucleases and other proteins released during the lysis of the bacterial cells which can adversely affect the quality of the isolated DNA. 350µl Neutralization Solution (4.09M guanidine hydrochloride, 0.759M potassium acetate, 2.12M glacial acetic acid, pH 4.2) was added and the suspension mixed immediately by inversion of the tube 4 times. The bacterial lysate was centrifuged at 14,000xg for 10 min at RT. The cleared lysate was transferred to a prepared plasmid DNA purification unit (spin column inserted into a 2ml collection tube) by decanting, ensuring that none of the white precipitate was disturbed or transferred with the supernatant. The supernatant was centrifuged at 14,000xg for 1 min at RT and then 750µl Column Wash solution (60mM potassium acetate, 10mM Tris-HCl, pH 7.5, 60% ethanol) was added to the spin column, centrifuged at 14,000xg for 1 min at RT. The wash procedure was repeated using 250µl Column Wash solution and centrifugation for 2 min. The spin column was transferred to a new, sterile 1.5ml microcentrifuge tube and the plasmid DNA eluted by adding 50µl nuclease-free H₂O to the spin column and centrifugation at 14,000xg, 1 min at RT.

2.2.16.1.6 DIG labelling of probe

The cDNA containing vector (pCRII-northern6) was linearised by restriction with restriction endonuclease AflIII (Promega). Digoxigenin-UTP (DIG)-labelled single-strand antisense RNA probes were prepared using an RNA Labelling Kit (Roche) following the manufacturer's instructions. Briefly, 1µg of purified plasmid DNA in 13µl of RNase free, DEPC treated double distilled H₂O was added to 1x labelling mix (for efficient *in vitro* transcription of linearised template DNA), 1x transcription buffer (also for efficient transcription of linearised template DNA), and 40 units RNA polymerase T7 (to synthesise RNA from the DNA template) on ice. The reaction was incubated for 1 hr at 42°C. 20 units RNase-free DNase I was added and then incubated for 15 min at 37°C to degrade template DNA. The reaction was stopped by the addition of 0.2M EDTA, pH 8.0. To determine the yield of DIG-labelled RNA, a direct detection method was used. In summary, 1µl spots of a dilution series of the

generated DIG-labelled RNA and labelled control RNA (10ng/ μ l of known concentration) were applied to a strip of nylon membrane. The expected yield of a standard labelling reaction is 20 μ g of labelled RNA so a dilution series (0pg/ μ l, 0.01pg/ μ l, 0.03pg/ μ l, 0.1pg/ μ l, 0.3pg/ μ l, 1pg/ μ l, 3pg/ μ l, 10pg/ μ l, 1ng/ μ l, 10ng/ μ l) of the generated DIG-labelled RNA was made on this basis. The nucleic acid was fixed to the membrane by crosslinking with 1200 μ Joules using an UV-Stratalinker. Immobilised RNA was subsequently visualised using detection reagents from DIG Northern Starter Kit (Roche). For this purpose, the membrane was transferred to a small plastic container with 20ml Washing buffer (0.1M Maleic acid; pH 7.5, 0.15M NaCl, 0.3% (v/v) Tween 20) and incubated for 2 min at RT on a rocking platform. The membrane was then incubated in 10ml Blocking solution (1x working solution prepared by diluting 10x Blocking reagent 1:10 in Maleic acid buffer [0.1M Maleic acid, 0.15M NaCl, pH 7.5]) for 30 min, 10ml Antibody solution (Anti-digoxigenin-AP, Fab, diluted 1:10000 (75mU/ml) in Blocking solution) for 20 min which binds to the DIG-label, washed with 20ml Washing buffer twice for 15 min and then immersed in 10ml 0.1M Tris-HCl, 0.1M NaCl, pH 9.5 (Detection buffer) for 10 min. The membrane was placed RNA side up in a pouch made by taping together two transparency sheets and CDP-Star was applied until the membrane was evenly soaked. The membrane was incubated for 5 min at RT. Enzymatic dephosphorylation of CDP-Star by alkaline phosphatase leads to a light emission at a maximum wavelength of 465nm. The intensities of the dilution series of DIG-labelled RNA and control RNA were compared by exposure of the membrane to BioMax MR film (Kodak).

2.2.16.2 Hybridisation and detection

The nucleic acid was fixed to the membrane by UV-crosslinking and then the membrane was rinsed briefly in H₂O and allowed to air dry. The membrane was pre-hybridised in 15 ml DIG Easy Hyb solution (Roche) for 30 min at 42°C with gentle agitation. The DIG-labelled probe was denatured by boiling for 5 min and rapidly cooled on ice. The membrane was then hybridised overnight at 42°C in 3.5 ml of DIG Easy Hyb solution containing 100ng/ml DIG-labelled RNA probe. The membrane was washed in 2x SSC, 0.1% SDS for 5 min at RT twice, and then in 0.1x SSC, 0.1%

SDS for 15 min at 65°C twice. Detection was carried out using solutions provided in a DIG Northern Starter Kit as described in section 2.2.16.1.6. After high stringency washes, the membrane was rinsed briefly in Washing buffer. The membrane was then incubated for 30 min in 100ml Blocking solution and then incubated for 30 min in Antibody solution. The membrane was then washed for 15 min with 100ml Washing buffer twice, and immersed for 5 min in 20ml Detection buffer. The membrane was placed RNA side facing up in a pouch made by taping together two transparency sheets and CDP-Star was applied until the membrane was evenly soaked. The membrane was incubated for 5 min at RT and then exposed to BioMax MR film (Kodak) for 15 min and 25 min in a Spectroline cassette and developed.

2.2.17 Primer extension analysis

Oligonucleotide rP5'1 5'-ACCGTCAGGACAGTGTGGC containing a 5' Cy5 label was purchased from Sigma Genosys. Primer (5µM) was hybridised to 0.9µg poly(A⁺) RNA from H69 cells by incubation for 10 min at 65°C and chilled on ice for 2 min. Reverse transcription was performed with 200 units of Superscript II RNase H⁻ reverse transcriptase (Life Technologies) in a total of 20µl for 60 min at 42°C, according to the manufacturer's instructions. The enzyme was heat inactivated for 15 mins at 70°C and primer extension products precipitated with ethanol. The primer extension products were analysed separately or after mixing with 5fmol/well ALFexpressTM SizerTM 50-500 (Amersham Biosciences), which was used as a marker. The marker was also run independently. The reactions were separated on the Long-Read TowerTM automated DNA sequencer (Visible Genetics, Inc.) using a short 6% acrylamide gel. The gel was run at 53°C, at 1300V, at 50% Laser power, 0.5 sec sampling for 35 min. The reactions were analysed using a fragment analysis software package (GeneObjectsTM).

2.3 Results

2.3.1 Detection of TGase 6 transcripts in various cell lines

RT-PCR analysis of different human cell lines using TGase 6-specific primers for exons II and IX proved unsuccessful as PCR products for TGase 6 could not be amplified from any of the cell lines tested including osteosarcoma cells (MG-63), fibrosarcoma cells (HT 1080), erythroleukemia cells (HEL), mammary epithelium carcinoma cells (MCF7), HELA cells, embryonal kidney epithelial cells (293), platelets and primary keratinocytes (undifferentiated). However, a match for TGase 6 was found with an EST sequence (IMAGE clone AW245863, GenBank) in the database (see Appendix 1). Unfortunately, the exact lung small cell carcinoma line was unobtainable, so different lung carcinoma cell lines were analysed. A TGase 6 product was amplified from RNA isolated from the small lung carcinoma cell line, H69.

2.3.2 Determination of the cDNA sequence of TGase 6

Based on sequence homology to known TGases and genomic information available, exon II to exon XIII of TGase 6 was predicted. This sequence information enabled primers to be derived so that the TGase 6 sequence could be amplified by PCR. Overlapping fragments of TGase 6 were amplified by PCR using the strategy outlined in Fig. 2.1.

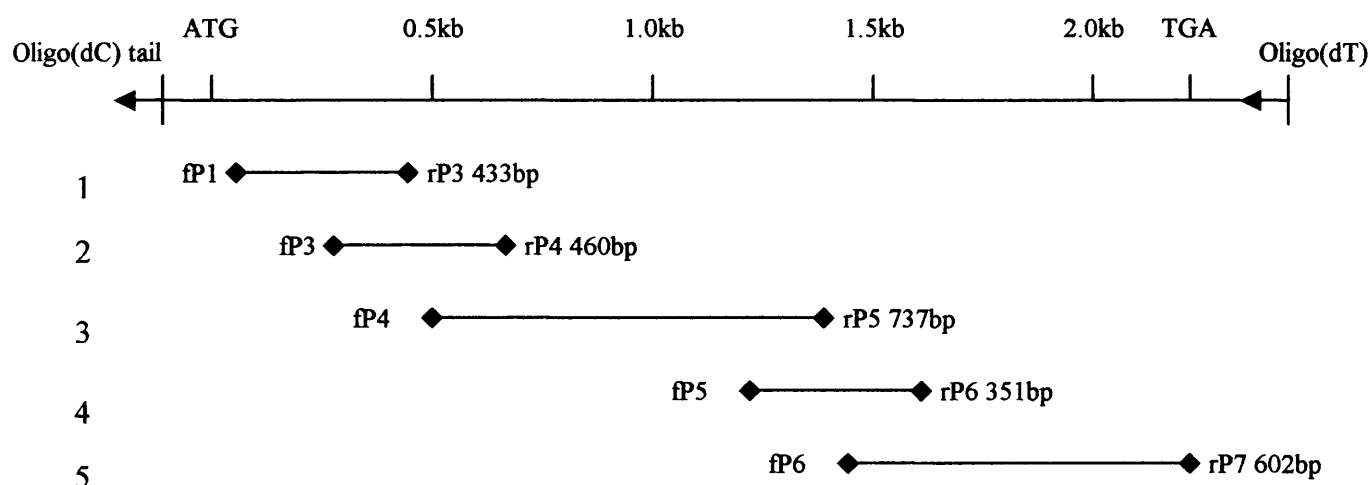


Fig. 2.1 PCR strategy for amplification of cDNA sequences of TGase 6. The top line represents the cDNA for TGase 6 with the start and stop codons indicated. Below is an outline of the PCR strategy, showing the consecutive PCR reactions (labelled 1-5) performed with oligonucleotide primers to obtain PCR products visible in ethidium bromide-stained agarose gels. Sequences of primers fP1 to rP7 are given in Appendix 2, Table 1. The length of the final PCR products is given on the right. The oligo(dT) primer (Invitrogen) was used as the 3' anchoring primer for reverse transcription.

PCR with fP1 and rP3 resulted in weak amplification of a PCR product. The PCR product was extracted from the agarose gel and used in a subsequent PCR reaction to reamplify the product with the same primers (Fig. 2.2, lane 1). The process was also repeated for primer sets fP4 and rP5 and fP5 and rP6 (Fig. 2.2, lanes 3 and 4). Only PCR amplification with primers fP3 and rP4 and fP6 and rP7 resulted in significant amplification of the product in a single round of PCR for it to be analysed directly (Fig. 2.2, lanes 2 and 5). Each PCR product was gel purified from at least 2 independent PCR reactions and sequenced in both directions to prevent any sequence errors introduced by PCR. The TGase 6 sequence is given in Fig. 2.3, with PCR products 1 to 5 corresponding to nucleotides 10-444, 348-814, 644-1386, 1313-1663, and 1610-2236, respectively.

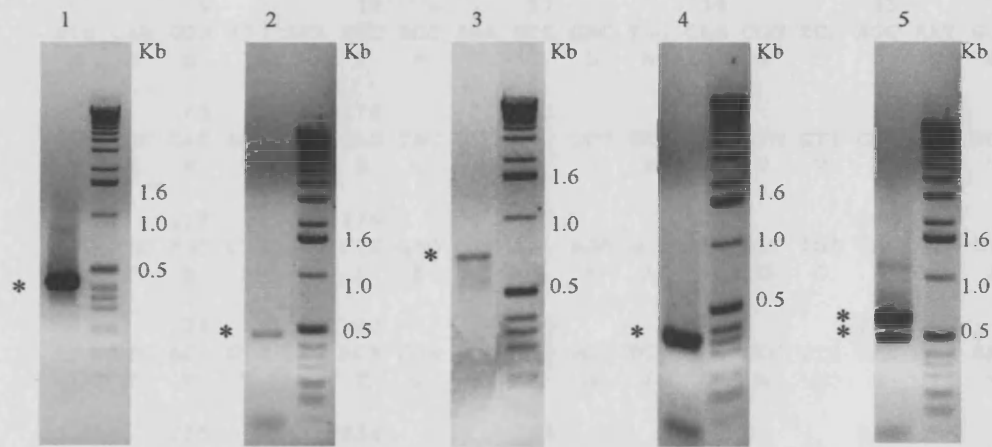


Fig.2.2 Amplification of TGase 6 from cDNA of H69 cells. 1-5 represent PCR products obtained using reactions designated 1-5 in the PCR strategy (Fig. 2.1). The PCR products were separated on 1% agarose gels calibrated with the 1kb ladder. Sizes of the DNA ladder standards are given on the right in bp.

All the primer combinations resulted in the amplification of one PCR product except for the combination of fp6 and rP7, which resulted in the amplification of two products (Fig. 2.2, lane 5). Sequencing of both products and comparison to the genomic sequence (AL049650, GenBank) revealed that there are two splice variants of human TGase 6, with exon XII alternatively spliced. The absence of exon XII results in a frameshift causing premature termination of translation within exon XIII (Fig. 2.4).

9 18 27 36 45 54
ATG CAG GGG ATC AGA GTC ACC AAG GTG GAC TGG CAG CGG TCG AGG AAT GGC GCT
 M A G I R V T K V D W Q R S R N G A

63 72 81 90 99 108
 GCC CAC CAC ACC CAG GAG TAC CCC TGC CCT GAG CTG GTG GTT CGC AGG GGC CAG
 A H H T Q E Y P C P E L V V R R G Q

117 126 135 144 153 162
 TCG TTC AGC CTC ACG CTG GAG CTG AGC AGA GCC CTG GAC TGT GAG GAG ATC CTC
 S F S L T L E L S R A L D C E E I L

171 180 189 198 207 216
 ATC TTC ACG GTG GAG ACA GGA CCC CGG GCT TCT GAG GCC CTC CAC ACC AAA GCT
 I F T V E T G P R A S E A L H T K A

225 234 243 252 261 270
 GTG TTC CAG ACA TCG GAG CTG GAG CGG GGT GAG GGC TGG ACA GCA GCA AGG GAG
 V F Q T S E L E R G E G W T A A R E

279 288 297 306 315 324
 GCT CAG ATG GAG AAA ACT CTG ACC GTC AGT CTC GCC AGC CCT CCC AGT GCT GTC
 A Q M E K T L T V S L A S P P S A V

335 342 351 360 369 378
 ATT GGC CGC TAC CTG CTG AGC ATC AGG CTT TCC TCT CAC CGC AAA CAC AGC AAC
 I G R Y L L S I R L S S H R K H S N

387 396 405 414 423 432
 CGG AGG CTG GGC GAG TTT GTT CTC CTT TTC AAC CCA TGG TGT GCA GAG GAC GAT
 R R L G E F V L L F N P W C A E D D

445 450 459 468 477 486
 GTG TTT CTG GCC TCA GAG GAG GAG AGA CAG GAG TAC GTG CTC AGC GAC AGC GGC
 V F L A S E E E R Q E Y V L S D S G

495 504 513 522 531 540
 ATC ATC TTC CGA GGC GTG GAG AAG CAC ATA CGA GCC CAG GGC TGG AAC TAC GGG
 I I F R G V E K H I R A Q G W N Y G

549 558 567 577 585 594
 CAG TTT GAG GAG GAC ATC CTG AAC ATC TGC CTC TCC ATC CTG GAT CGA AGC CCC
 Q F E E D I L N I C L S I L D R S P

603 612 621 630 639 648
 GGT CAC CAA AAC AAC CCA GCC ACC GAC GTG TCC TGC CGC CAC AAC CCC ATC TAC
 G H Q N N P A T D V S C R H N P I Y

657 666 675 684 693 702
 GTC ACC AGG GTC ATC AGT GCC ATG GTG AAC AGC AAC AAC GAC CGA GGT GTG GTG
 V T R V I S A M V N S N N D R G V V

711 720 729 738 747 756
 CAA GGA CAG TGG CAG GGC AAG TAC GGC GGC GGC ACC AGC CCG CTG CAC TGG CGC
 Q G Q W Q G K Y G G G T S P L H W R

765 774 783 792 801 810
 GGC AGC GTG GCC ATT CTG CAG AAG TGG CTC AAG GGC AGG TAC AAG CCA GTC AAG
 G S V A I L Q K W L K G R Y K P V K

819 828 837 846 855 864
 TAC GGC CAG TGC TGG GTC TTC GCC GGA GTC CTG TGC ACA GTC CTC AGG TGC TTG
 Y G Q C W V F A G V L C T V L R C L

873 882 891 900 909 918
 GGG ATA GCC ACA CGG GTC GTG TCC AAC TTC AAC TCA GCC CAC GAC ACA GAC CAG
 G I A T R V V S N F N S A H D T D Q

927 936 945 954 963 972
 AAC CTG AGT GTG GAC AAA TAC GTG GAC TCC TTC GGG CGG ACC CTG GAG GAC CTG
 N L S V D K Y V D S F G R T L E D L

981 990 999 1008 1017 1026
 ACA GAA GAC AGC ATG TGG AAT TTC CAT GTC TGG AAT GAG AGC TGG TTT GCC CGG
 T E D S M W N F H V W N E S W F A R

1035 1044 1053 1062 1071 1080
 CAG GAC CTA GGC CCC TCT TAC AAT GGC TGG CAG GTT CTG GAT GCC ACC CCC CAG
 Q D L G P S Y N G W Q V L D A T P Q

1089 1098 1107 1116 1125 1134
 GAG GAG AGT GAA GGT GTG TTC CGG TGC GGC CCA GCC TCA GTC ACC GCC ATC CGC
 E E S E G V F R C G P A S V T A I R

1143 1152 1161 1170 1179 1188
 GAG GGT GAT GTG CAC CTG GCT CAC GAT GGC CCC TTC GTG TTT GCG GAG GTC AAC
 E G D V H L A H D G P F V F A E V N

1197 1206 1215 1224 1233 1242
 GCC GAC TAC ATC ACC TGG CTG TGG CAC GAG GAT GAG AGC CGG GAG CGT GTA TAC
 A D Y I T W L W H E D E S R E R V Y

1251 1260 1269 1278 1287 1296
 TCA AAC ACG AAG AAG ATT GGG AGA TGC ATC AGC ACC AAG GCG GTG GGC AGT GAC
 S N T K K I G R C I S T K A V G S D

1310 1319 1328 1337 1346 1355
 TCC CGC GTG GAC ATC ACT GAC CTC TAC AAG TAT CCG GAA GGG TCC CGG AAA GAG
 S R V D I T D L Y K Y P E G S R K E

1364 1373 1382 1391 1400 1409
 AGG CAG GTG TAC AGC AAG GCG GTG AAC AGG CTG TTC GGC GTG GAA GCC TCT GGA
 R Q V Y S K A V N R L F G V E A S G

1418 1427 1436 1445 1454 1463
 AGG AGA ATC TGG ATC CGC AGG GCT GGG GGT CGC TGT CTC TGG CGT GAC GAC CTC
 R R I W I R R A G G R C L W R D D L

1472 1481 1490 1499 1508 1517
 CTG GAG CCT GCC ACC AAG CCC AGC ATC GCT GGC AAG TTC AAG GTG CTA GAG CCT
 L E P A T K P S I A G K F K V L E P

1526 1535 1544 1553 1562 1571
 OCC ATG CTG GGC CAC GAC CTG AGA CTG GCC CTG TGC TTG GCC AAC CTC ACC TCC
 P M L G H D L R L A L C L A N L T S

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

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

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

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

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

```

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

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

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

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

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

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

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

2228      2237
TAA TCC TGC CTC AAC CT

```

Fig. 2.3 Nucleotide and deduced amino acid sequence for the long form of TGase 6. The initiation and stop codon, respectively, are underlined. The predicted active site Cys residue is shown in red, and the His and Asp residues of the catalytic triad are in pink.

9 18 27 36 45 54
 ATG CAG GGG ATC AGA GTC ACC AAG GTG GAC TGG CAG CGG TCG AGG AAT GGC GCT
 M A G I R V T K V D W Q R S R N G A
 63 72 81 90 99 108
 GCC CAC CAC ACC CAG GAG TAC CCC TGC CCT GAG CTG GTG GTT CGC AGG GGC CAG
 A H H T Q E Y P C P E L V V R R G Q
 117 126 135 144 153 162
 TCG TTC AGC CTC ACG CTG GAG CTG AGC AGA GCC CTG GAC TGT GAG GAG ATC CTC
 S F S L T L E L S R A L D C E E I L
 171 180 189 198 207 216
 ATC TTC ACG GTG GAG ACA GGA CCC CGG GCT TCT GAG GCC CTC CAC ACC AAA GCT
 I F T V E T G P R A S E A L H T K A
 225 234 243 252 261 270
 GTG TTC CAG ACA TCG GAG CTG GAG CGG GGT GAG GGC TGG ACA GCA GCA AGG GAG
 V F Q T S E L E R G E G W T A A R E
 279 288 297 306 315 324
 GCT CAG ATG GAG AAA ACT CTG ACC GTC AGT CTC GCC AGC CCT CCC AGT GCT GTC
 A Q M E K T L T V S L A S P P S A V
 333 342 351 360 369 378
 ATT GGC CGC TAC CTG CTG AGC ATC AGG CTT TCC TCT CAC CGC AAA CAC AGC AAC
 I G R Y L L S I R L S S H R K H S N
 387 396 405 414 423 432
 CGG AGG CTG GGC GAG TTT GTT CTC CTT TTC AAC CCA TGG TGT GCA GAG GAC GAT
 R R L G E F V L L F N P W C A E D D
 441 450 459 468 477 486
 GTG TTT CTG GCC TCA GAG GAG GAG AGA CAG GAG TAC GTG CTC AGC GAC AGC GGC
 V F L A S E E E R Q E Y V L S D S G
 495 504 513 522 531 540
 ATC ATC TTC CGA GGC GTG GAG AAG CAC ATA CGA GCC CAG GGC TGG AAC TAC GGG
 I I F R G V E K H I R A Q G W N Y G
 549 558 567 576 585 594
 CAG TTT GAG GAG GAC ATC CTG AAC ATC TGC CTC TCC ATC CTG GAT CGA AGC CCC
 Q F E E D I L N I C L S I L D R S P
 603 612 621 630 639 648
 GGT CAC CAA AAC AAC CCA GCC ACC GAC GTG TCC TGC CGC CAC AAC CCC ATC TAC
 G H Q N N P A T D V S C R H N P I Y
 657 666 675 684 693 702
 GTC ACC AGG GTC ATC AGT GCC ATG GTG AAC AGC AAC AAC GAC CGA GGT GTG GTG
 V T R V I S A M V N S N N D R G V V
 711 720 729 738 747 756
 CAA GGA CAG TGG CAG GGC AAG TAC GGC GGC GGC ACC AGC CCG CTG CAC TGG CGC
 Q G Q W Q G K Y G G G T S P L H W R
 765 774 783 792 801 810
 GGC AGC GTG GCC ATT CTG CAG AAG TGG CTC AAG GGC AGG TAC AAG CCA GTC AAG
 G S V A I L Q K W L K G R Y K P V K
 819 828 837 846 855 864
 TAC GGC CAG TGC TGG GTC TTC GCC GGA GTC CTG TGC ACA GTC CTC AGG TGC TTG
 Y G Q C W V F A G V L C T V L R C L
 873 882 891 900 909 918
 GGG ATA GCC ACA CGG GTC GTG TCC AAC TTC AAC TCA GCC CAC GAC ACA GAC CAG
 G I A T R V V S N F N S A H D T D Q

927 936 945 955 963 972
 AAC CTG AGT GTG GAC AAA TAC GTG GAC TCC TTC GGG CGG ACC CTG GAG GAC CTG
 N L S V D K Y V D S F G R T L E D L

981 990 999 1008 1017 1026
 ACA GAA GAC AGC ATG TGG AAT TTC CAT GTC TGG AAT GAG AGC TGG TTT GCC CGG
 T E D S M W N F H V W N E S W F A R

1035 1044 1053 1062 1071 1080
 CAG GAC CTA GGC CCC TCT TAC AAT GGC TGG CAG GTT CTG GAT GCC ACC CCC CAG
 Q D L G P S Y N G W Q V L D A T P Q

1089 1098 1107 1116 1125 1134
 GAG GAG AGT GAA GGT GTG TTC CGG TGC GGC CCA GCC TCA GTC ACC GCC ATC CGC
 E E S E G V F R C G P A S V T A I R

1143 1152 1161 1170 1179 1188
 GAG GGT GAT GTG CAC CTG GCT CAC GAT GGC CCC TTC GTG TTT GCG GAG GTC AAC
 E G D V H L A H D G P F V F A E V N

1197 1206 1215 1224 1233 1242
 GCC GAC TAC ATC ACC TGG CTG TGG CAC GAG GAT GAG AGC CGG GAG CGT GTA TAC
 A D Y I T W L W H E D E S R E R V Y

1251 1260 1269 1278 1287 1296
 TCA AAC ACG AAG AAG ATT GGG AGA TGC ATC AGC ACC AAG GCG GTG GGC AGT GAC
 S N T K K I G R C I S T K A V G S D

1305 1314 1323 1332 1341 1350
 TCC CGC GTG GAC ATC ACT GAC CTC TAC AAG TAT CCG GAA GGG TCC CGG AAA GAG
 S R V D I T D L Y K Y P E G S R K E

1359 1368 1377 1386 1395 1404
 AGG CAG GTG TAC AGC AAG GCG GTG AAC AGG CTG TTC GGC GTG GAA GCC TCT GGA
 R Q V Y S K A V N R L F G V E A S G

1413 1422 1431 1440 1449 1458
 AGG AGA ATC TGG ATC CGC AGG GCT GGG GGT CGC TGT CTC TGG CGT GAC GAC CTC
 R R I W I R R A G G R C L W R D D L

1467 1476 1485 1494 1503 1512
 CTG GAG CCT GCC ACC AAG CCC AGC ATC GCT GGC AAG TTC AAG GTG CTA GAG CCT
 L E P A T K P S I A G K F K V L E P

1521 1530 1539 1548 1557 1566
 CCC ATG CTG GGC CAC GAC CTG AGA CTG GCC CTG TGC TTG GCC AAC CTC ACC TCC
 P M L G H D L R L A L C L A N L T S

1575 1584 1593 1602 1611 1620
 CGG GCC CAG CGG GTG AGG GTC AAC CTG AGC GGT GCC ACC ATC CTC TAT ACC CGC
 R A Q R V R V N L S G A T I L Y T R

1629 1638 1647 1656 1665 1674
 AAG CCA GTG GCA GAG ATC CTG CAT GAA TCC CAC GCC GTG AGG CTG GGG CCG CAA
 K P V A E I L H E S H A V R L G P Q

1683 1692 1701 1710 1719 1728
 GAA GAG AAG AGA ATC CCA ATT ACA ATA TCT TAC TCT AAG TAT AAA GAA GAC CTG
 E E K R I P I T I S Y S K Y K E D L

1737 1746 1755 1764 1773 1782
 ACA GAG GAC AAG AAG ATC CTG TTG GCT GCC ATG TGC CTT GTC ACC AAA GGA GAG
 T E D K K I L L A A M C L V T K G E

1791 1800 1809 1818 1827 1836
 AAG CTT CTG GTG GAG AAG GAC ATT ACT CTA GAG GAC TTC ATC ACC ATC AAG CGT
 K L L V E K D I T L E D F I T I K R

1845	1854	1863	1872	1881	1890
GCC TAC CCT GGA GCC TCA GGA GAG GGC CTC AGT CCA GTT <u>TGA</u> CAT CAC CCC CTC					
A Y P G A S G E G L S P V *					
1899	1908	1917	1926	1935	1944
CAA AAG TGG CCC AAG GCA GCT GCA GGT GGA CCT TGT AAG CCC TCA CTT CCC GGA					
1953	1962	1971	1980	1989	1998
CAT CAA GGG CTT TGT GAT CGT CCA TGT GGC CAC TGC CAA GTG ATG GAT CAT GAG					
2007	2016	2025	2034	2043	2052
GGA CTG AGA GGG GTG GAT TTG GCC CCT GTC CTC CTC CTG CCC ATT CTT TGT CTC					
2061	2070	2079	2088	2097	
TTC CAC ATG GGA GCC AGG AGG CCT CAG TTA ATC CTG CCT CAA CCT					

Fig. 2.4 Nucleotide and deduced amino acid for the short form of TGase 6. The initiation and stop codon, respectively, are underlined. The predicted active site Cys residue is shown in red, and the His and Asp residues of the catalytic triad are in pink. The sequence difference between the two alternatively spliced variants as a result of the reading frame shift is shown in bold.

Since exon I contains largely non-coding sequence and is very diverse between different TGase genes it cannot be identified based on sequence homology. To obtain more information on the 5'-end of the TGase 6 cDNA, a 5'-rapid amplification of cDNA ends approach was used. A poly(dC) tail was added to the cDNA using terminal deoxynucleotidyl transferase to anchor the PCR reaction with an oligo(dG) primer (abridged anchor primer (AAP)). The reaction was anchored with the abridged anchor at low annealing temperature, and a first round of amplification was performed with abridged anchor primer and TGase 6-specific primers rP1, rP3 or rP8 (Fig. 2.5). Subsequent reactions were nested with universal amplification primer (UAP) and TGase 6-specific primers.

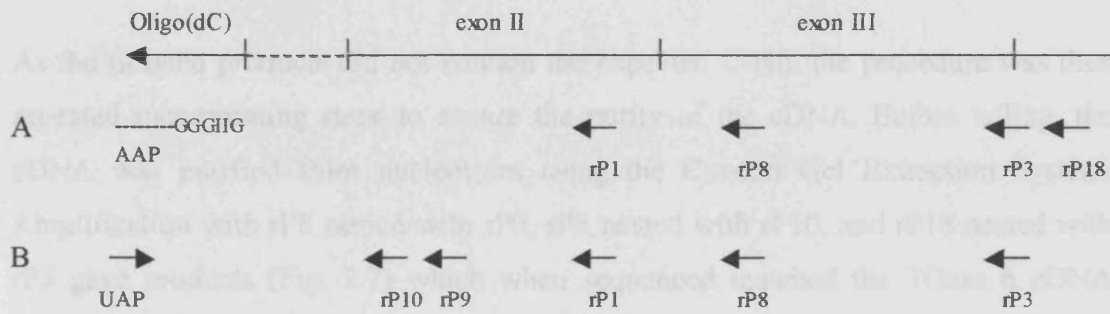


Fig. 2.5 PCR strategy for amplification of cDNA sequences of TGase 6 at the 5' end. The top line represents the cDNA for TGase 6 with exon I indicated by a dotted line. Below is an outline of the PCR strategy used: A outlines the position of the primers used in the first cycle of PCR reactions. B outlines the position of the primers used in the subsequent nested PCR reactions.

Reactions with rP3 nested with rP8, rP8 nested with rP1, and rP1 nested with rP10 yielded PCR products (Fig. 2.6A), which by sequencing were found to be unrelated to TGase 6. The procedure was repeated several times with different primer combinations. Products were obtained for a combination of rP3 nested with rP8 and rP3 nested with rP1 (Fig. 2.6B) which when sequenced were also found to be unrelated to TGase 6.

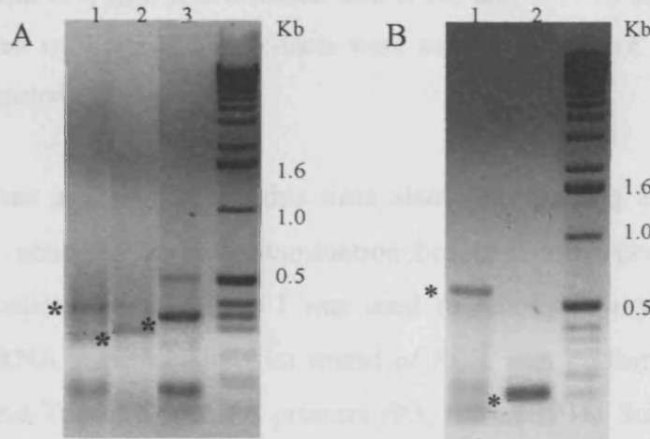


Fig. 2.6 Amplification of the 5'-end of TGase 6 cDNA. A. Lane 1 represents the products of PCR reactions with primers rP3 nested with rP8; lane 2, products of reactions with rP8 nested with rP1; and lane 3, products amplified with rP1 nested with rP10. B. Lane 1 represents PCR products of reactions with rP3 nested with rP8, and lane 2, PCR products of reactions with rP3 nested with rP1. PCR products were analysed by electrophoresis in 1% agarose gels calibrated with the 1kb ladder. * indicates the PCR product which was gel extracted and sequenced.

As the isolated products did not contain the expected C-tail, the procedure was then repeated incorporating steps to ensure the purity of the cDNA. Before tailing, the cDNA was purified from nucleotides using the Concert Gel Extraction System. Amplification with rP8 nested with rP9, rP8 nested with rP10, and rP18 nested with rP3 gave products (Fig. 2.7) which when sequenced matched the TGase 6 cDNA sequence up to exon II, followed by a 97bp sequence which appeared to be the adjacent genomic sequence of intron 1, which was then C-tailed.

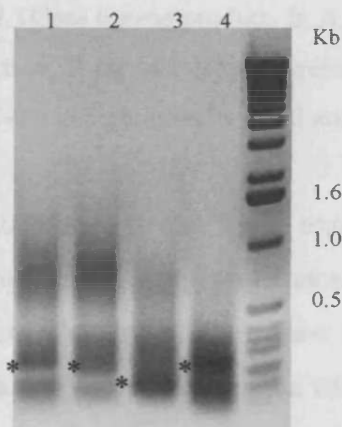


Fig. 2.7 Amplification of the 5'-end of TGase 6. Lane 1 represents the product obtained with rP8 and nesting with rP9, lane 2, rP8 nested with rP10, lane 3, rP18 nested with rP10, and lane 4, rP18 nested with rP3. PCR products were analysed as above. * indicates product which was gel extracted and sequenced.

The procedure was again repeated, this time also incorporating a step to purify the total RNA from genomic DNA contamination before it was reverse transcribed to cDNA. Amplification grade DNase I was used to eliminate any residual genomic DNA from the RNA sample. The first round of PCR was performed with abridged anchor primer and TGase 6-specific primers rP3, rP8 or rP18. Subsequent reactions were nested with UAP and appropriate reverse primers and products were obtained for a combination of rP3 nested with rP9, rP18 nested with rP9 and rP8 nested with rP1 (Fig 2.9A). When sequenced the product obtained from rP18 nested with rP9 was found to be derived from cDNA but could not yield a functional protein as no open reading frame was apparent. Sequence analysis revealed that the sequence matched the TGase 6 cDNA up to the start of exon II, followed by a section of genomic DNA from intron I, and then ending in the C-tail. A schematic is shown in Fig. 2.8.

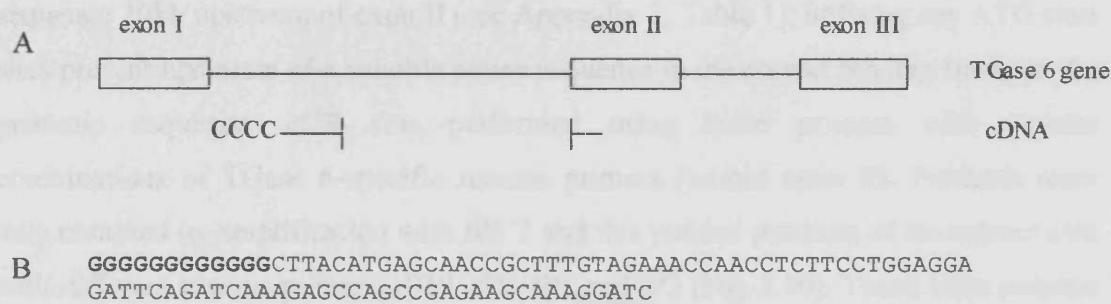


Fig. 2.8 Schematic of identified TGase 6 gene product. In A, the regions of the cDNA product are compared to the gene structure. B represents the sequence of the amplified product. The sequence derived from added C-tail is highlighted in bold and the start of exon II underlined.

This product was unlikely to represent the “true” exon I sequence as no ATG was present in the correct reading frame. The significance of this finding is unclear but could suggest that a promoter is located within intron I. The procedure was repeated again and products were obtained for a combination of rP10 nested with rP9, and rP8 nested with rP9 (Fig. 2.9B). Again products sequenced were derived from cDNA, but not exon I of TGase 6. Therefore, despite the use of a variety of approaches the exact sequence position of exon I remained elusive.

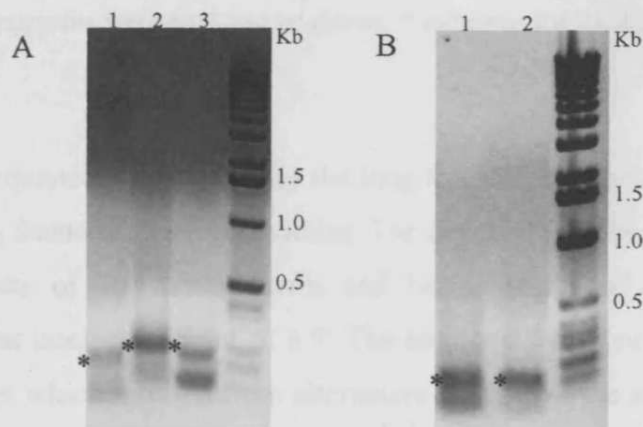


Fig. 2.9 Amplification of the 5'-end of TGase 6 cDNA. A. Lane 1 represents the PCR products obtained with rP3 and nesting with rP9, lane 2, rP18 nested with rP9, and lane 3, rP8 nested with rP1. B. Lane 1 represents the PCR products obtained with rP10 and nesting with rP9, and lane 2, rP8 nested with rP9. PCR products were analysed as above. * indicates the PCR products gel extracted and sequenced.

In an alternative approach, nine forward primers were designed based on the genomic sequence 20kb upstream of exon II (see Appendix 2, Table 1), utilising any ATG start sites present upstream of a suitable splice sequence in the correct reading frame in the genomic sequence. PCR was performed using these primers with various combinations of TGase 6-specific reverse primers (within exon II). Products were only obtained by amplification with fP5'2 and this yielded products of the correct size with different reverse primers, rP19, rP9, rP1 and rP3 (Fig. 2.10). These were isolated and sequenced, and thereby identified the exon I/exon II junction.

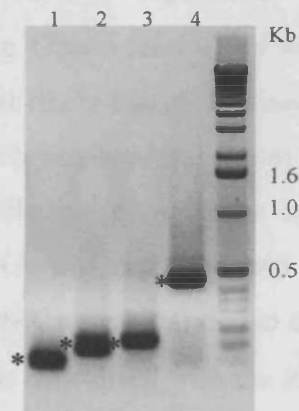


Fig. 2.10 Amplification of the 5'-end of TGase 6 cDNA. Lane 1 represents the PCR product amplified with fP5'2 and rP19, lane 2, fP5'2 and rP9, lane 3, fP5'2 and rP1 and lane 4, fP5'2 and rP3. DNA fragments were analysed as above. * indicates the PCR products gel extracted and sequenced.

The obtained sequence information for the long form of TGase 6 (Fig. 2.3) contained an open reading frame of 2109 nucleotides. The deduced protein for the long form of TGase 6 consists of 708 amino acids and has a calculated molecular mass of 79,466Da and an isoelectric point of 6.9. The obtained sequence information for the shorter transcript which resulted from alternative splicing of the sequence encoded by exon XII, (Fig. 2.4) contained an open reading frame of 1878 nucleotides. The deduced protein for the short form of TGase 6 consists of 626 amino acids and has a calculated molecular mass of 70,671Da and an isoelectric point of 7.6. The sequences for human TGase 6 have been submitted to Genbank: AF540969 (long transcript) and AF540970 (spliced transcript) (see Appendix 3 and 4).

2.3.3 Determination of transcript size

A DNA probe spanning the sequence that encodes the two C-terminal β -barrel domains of TGase 6 labelled with ^{32}P was used to detect the novel TGase 6 on a Northern blot of H69 cell mRNA. Total RNA was extracted from H69 cells using Tri Reagent and poly(A⁺) RNA isolated using oligod(T) cellulose as described in Materials and Methods. 3 μg mRNA was separated in a formaldehyde gel, transferred onto a nylon membrane and hybridised with the ^{32}P -labelled DNA probe. As the mRNA could not be detected the experiment was repeated with increasing concentration of mRNA to 5 μg . Only a faint signal was obtained which did not give a conclusive result. This is most likely due to the low expression of TGase 6. A new approach was then taken which involved the use of a DIG-labelled RNA probe and indirect detection using an alkaline phosphatase labelled secondary antibody and CDP-Star substrate. Total RNA was extracted and converted to mRNA as before. 100ng mRNA and 1 μg total RNA were separated in a gel and transferred onto a nylon membrane. The membrane was hybridised with the RNA probe overnight at 68°C. As no signal could be detected the procedure was then repeated but 1.25 μg mRNA was used, hybridisation carried out at 42°C and the stringency of the wash reduced. Hybridisation and antibody detection revealed faint bands. So, the experiment was repeated but the concentration of mRNA was increased to 5 μg . Hybridisation revealed the presence of three bands corresponding to 6.3kb, 2.0kb and 1.1kb (Fig. 2.11). The 6.3kb transcript may represent the full-length TGase 6 mRNA while the predominant smaller species of 2.0kb may represent an alternatively processed form. Alternatively, it is possible that the mRNA has partially degraded or that the probe used is cross-hybridising with other RNAs.

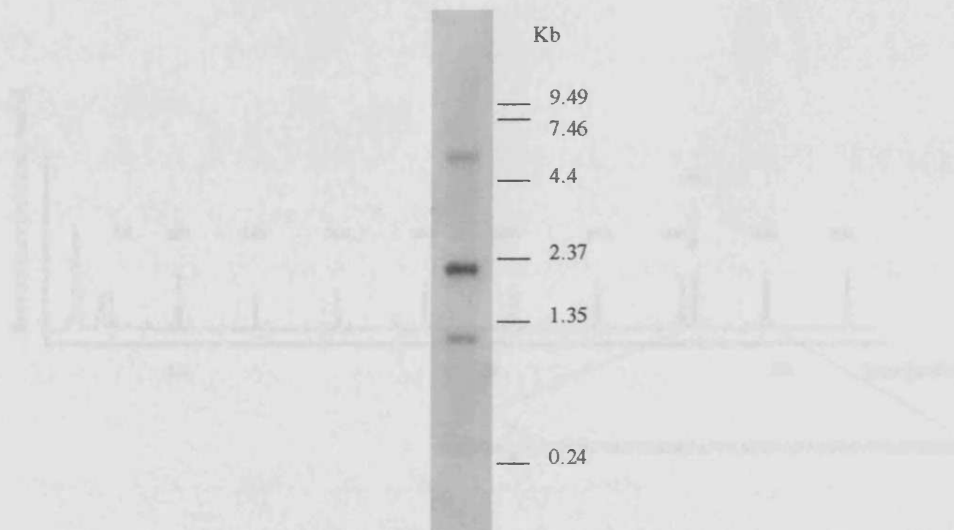


Fig. 2.11 Size of transcripts of TGase 6 in H69 cells. Northern blot containing 5 μ g of poly(A⁺) RNA extracted from H69 cells probed with a ~730bp fragment corresponding to the two C-terminal β -barrel domains of TGase 6. The migration position of RNA size markers is indicated on the right.

2.3.4 Primer extension analysis

A primer extension experiment was carried out in order to determine the major transcription initiation site for the TGM6 gene. For this purpose, poly(A⁺) RNA from H69 cells was hybridised to reverse primer rP5'1 containing a 5' Cy5 label. Reverse transcription was then performed and the primer extension products separated on a DNA sequencer using an acrylamide gel. A fragment size marker was not only run in parallel on the same gel but also mixed with the primer extension product to exclude loading induced differences in migration. The products were analysed in comparison to the fragment size marker using the fragment analysis software package (GeneObjectsTM). The primer extension experiment located the major transcription initiation site used in H69 cells, 408 nucleotides upstream of the reverse primer (Fig. 2.12) and 455 nucleotides upstream of the translation start codon.

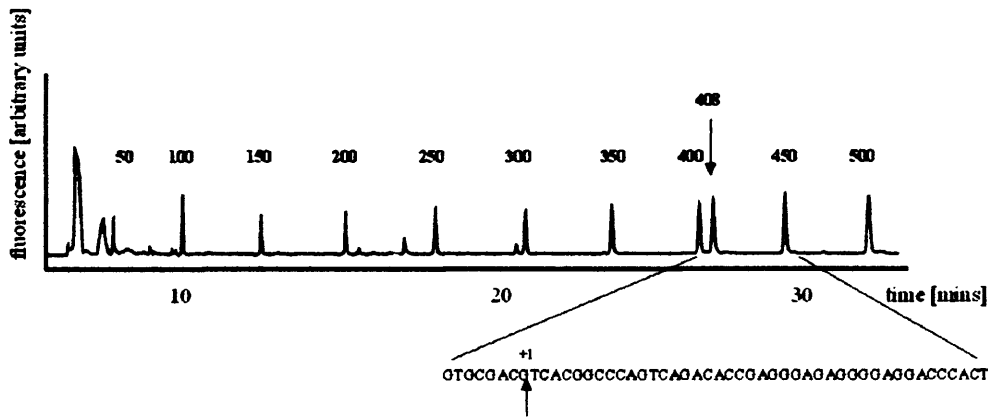


Fig. 2.12 Primer extension analysis of poly(A⁺) RNA isolated from H69 cells. Extension products were analysed on a Long-Read TowerTM automated DNA sequencer using a Fragment Analysis computer package. ALFexpressTM SizerTM 50-500 was added as a marker and each marker peak is labelled. The peak representing the major transcription start site is indicated by an arrow. The sequence between 400 and 450bp spanning the initiation site is shown underneath the fragment analysis.

```

human                +1GCAGCGTGCAACCATGTACTTCGGGGTCCTGGCGCTGAGCCTGTT 45
human TTTTCTCATTGAGGGTTGTTGTGAGGGCTAAGAATGGCCAACCATAAATCCGAGTTCCCTCAGTCTTCTCTCTTTCCCGT 127
human TCTTTTCTTTGCCAGGCTTGTGACCTTTCTGAGCCTCAGTTTCTCCACAGTCCCATTCGCAACTCCGTGTTTCCAAGGTCTC 209
human TTGGCAGGAGCCACTCCACAGTGAACCCCTCTAGGTGCAATTACAGGATGATTTTGTGTCTCGTGGGTGTCGCTGCTAGGG 291
human TGGAGGGCCACGCTCCTCTTTTGAATTCTGACTTTCTGCCTAGGAGTGTCCTCAAGGTTTTCTTGTAAATTGGTCTGG 373
human CTGGCTGCTGTGACACGCCACACTGTCCTGACGGTGCACACACTGCTGTGTGGAGGAACAGAGGAGTCCAGCTGGCC TTCAC 455
human ATGGCAGGG 464

```

Fig. 2.13 Human TGase 6 sequence showing the 5' end of the cDNA starting at the major transcription start site identified to the end of exon I. The start codon for translation is underlined.

To confirm that the major transcription initiation site is located in the proposed region, PCR amplification was carried out. Forward primers were designed based on the genomic sequence close to the proposed transcription initiation site. PCR was performed using these primers with various combinations of reverse primers within exon II. Products were not obtained in the first round of amplification with primers fp5'11 and rP1 (Fig. 2.14, A, I1), however, upon a second round of amplification with nested primers, products of the expected size, 529bp, and 576bp, were obtained (from

a combination of primers fP5'10 with rP19 and rP9, respectively (Fig. 2.14 B, lane 1 and 2). PCR amplification using fP5'2 and rP9 served as a positive control and yielded a product of the expected size, 146bp (Fig. 2.14 B, lane 4). PCR amplification using fP5'10 and rP9 from a first round reaction (Fig 2.4 I2) without suitable template served as a negative control, i.e, the reverse primer in the nested reaction is downstream from the reverse primer used in the first round reaction, and therefore should not yield a product, as the agarose gel indicates (Fig. 2.14 B, lane3).

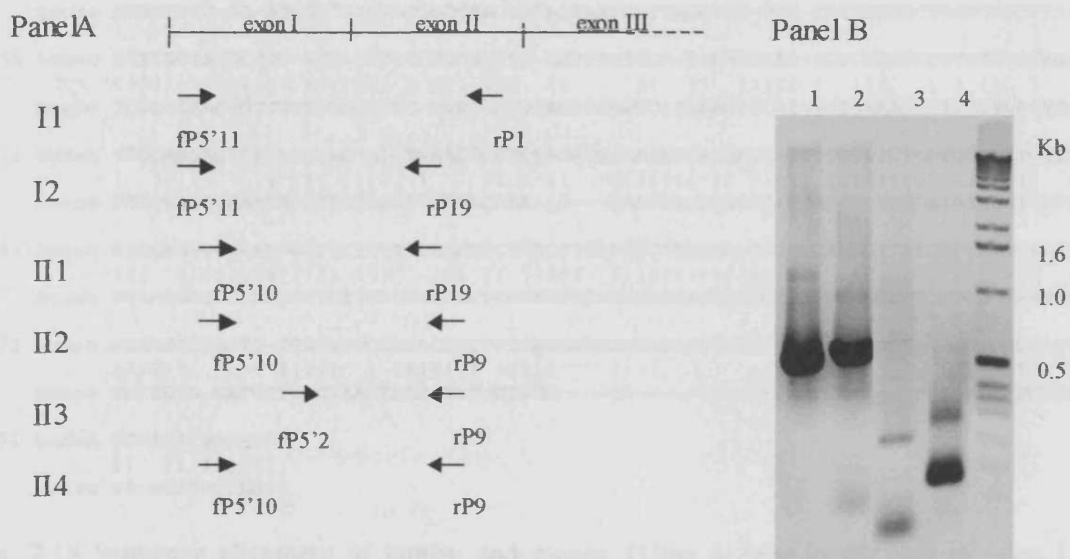


Fig. 2.14 Amplification of the 5' end of human TGase 6 cDNA. Panel A. PCR strategy for amplification of cDNA sequences of TGase 6 at the 5' end. The top line represents the cDNA for TGase 6. Below is an outline of the PCR strategy used whereby I outlines the position of the primers used in the first cycle of PCR reactions and II outlines the position of the primers used in the subsequent nested PCR reactions. Panel B. Lane 1 represents the PCR product amplified with fP5'10 and rP19, lane 2, fP5'10 and rP9, lane 3, fP5'10 and rP9, and lane 4, fP5'2 and rP9. DNA fragments were analysed as before.

These results demonstrate that the region adjacent to the proposed transcription initiation site is transcribed. Matches for the human TGase 6 sequence at the 5' end with mouse EST sequences were found in the database. Clones BY357397, BB612839, BB630356 and AK037229 (see Appendix 5, 6, 7 and 8, respectively) were identified and their respective 5' ends are given in Fig. 2.15.

```

-240 human CCTCT-CAAGAGTCTAGTG-TCCACCAAGGAAGTCTAGTGGCCCTGAAGGAGGATCATGGACC--TCTGCCTCTTTAG-ATG
      ||      ||      ||      ||      ||      ||      ||      ||      ||      ||      ||      ||      ||      ||
mouse  GTGCTGGGTATGTGCAGGGAT-TGCAGAGGA----TGT-G-AAGTG--ATAGG-TGCCGTCCCTGTAAG-ATCA TAGGA--

-163 human CCTCTCAAGAGTCTAGTGTCCACCAAGGAAGTCTAGTGGCCCTGAAG-GAG-GATCATGGACCTCTGCCTCTTTAGATG-CC
      ||      ||      ||      ||      ||      ||      ||      ||      ||      ||      ||      ||      ||      ||
mouse  CCTGTGCTGGG--TA-TGTGC----AGGGATTCAGAGGATGTGAAGTGATAGGTGCCGTCCCTGTAAGATCAT--AGGACC

-84  human T-GCAGAACAGA-GTGG--AGCTG-----TGGGCC--AGGAGTCTTCTGCTCTCACCCAGGAGG-GG-AGAGGGA
      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
mouse  TAG-AGGACAGCTGTGGTCATGGACTCTTCTCATCTTGCCTGAAGGAGAAGGGA-----CACCTGGAGTAGGTATCATGA

-21  human GCCACAGACT--GACCCGG-CACTGCAGCGTG-CAACCATGTACTTCGGGGTCCCTGGCGCTGAGCCTGTTTTTCTCATTG
      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
mouse  CTCTTGGACTCTGGTCTGGACACTGCAGCTTTACTATCATGTGCCTCCAAGT---GGCTCTGAGC---TCTCATGT-ACTTG

+58  human AGGGTTGTGTGAGGGCTAAGAATGGCCAACCATAA-ATCCG--AGTCCCTCAGTCTTCTCTCTT-TCCCCGTT-CTTTT-
      ||      ||      ||      ||      ||      ||      ||      ||      ||      ||      ||      ||      ||      ||
mouse  AGGGATGC-GA-ATGG-TAAGAACGGCTGACCATAGTAACCATTAGCTCC-TCC-TATTCTCCCTGGTCCGCTTAGCATTAG

+134 human TCTTGCCAGGCTTG-TGA-CCTTTCTGAGCCTC-AGTTTCCCA-CAGTCCCAT-TCGCAACTCCGTTTCCAAGGTCTCT
      ||      ||      ||      ||      ||      ||      ||      ||      ||      ||      ||      ||      ||      ||
mouse  TCTTGTGAC-CTTCTGGGCTGT-TCCTCTGTAGCGCAGCCCTCAACCCCATGTTTCAAAGGCATCTTGACTGGGTCCAC

+211 human TGGCAGGAGCCACTCCAC-AG-TGAACCCTCTAGGTGCAATTACAGGATGATTTTGTGTC TCGTGGGTGTCTGCTGCTAGG
      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
mouse  TCCCACGCCGAACACCTTCTAGGTGCAAACCAA-AG---CAATTACTGGGTGATTATGTGTCATGGATGTCATTGCTGGG

+291 human GTGGAG-GGCCACGCCTCTCTTTTGAATCTGACTTTCCCTGCCTAGGAGTGTCCCAAGGTTTTCTTGTAAAATTGGTCT
      ||      |      ||      ||      ||      ||      ||      ||      ||      ||      ||      ||      ||      ||
mouse  GTGAGGAGGCCACGCCTCTTCTCCGGAAC TCCGACTTCACTGCCTAGGAGTGTCTCTGATGTCTT ATTAATGTGGATCT

+372 human GGCTGGCTGCTG-TGACGCGCCACACTGTCTGACGGTGCACACACTGCTGTGTGG-AGGAACAGAGGAGTCCAGCTGGCC-
      ||      ||      |      ||      |      ||      |      ||      |      ||      |      ||      |      ||      |
mouse  GGCTGGA-GATCCTGACGAAC TACTCTGCTGCC----CA-----GGGAGC-----AGGCGTCCATCTGGCTG

+451 human TTCACATGGCAGGG
      ||      ||      ||      ||      ||
mouse  TT-GCATGGCAGGG

```

Fig. 2.15 Sequence alignment of human and mouse TGase 6 gene in the area of exon I. Highlighted green is the major transcription initiation site identified in this work, blue is the start of the EST image clone BY357397, red, clone BB612839, and yellow, clone BB630356 and clone AK037229.

2.3.5 3'-untranslated region

The likely 3' end of the TGase 6 cDNA was defined through an EST sequence available in the GenBank™ database. A blast search of the 3' end revealed an exact sequence match with only one EST clone (AA961594, GenBank, see Appendix 9), which is located in between extensive repeat regions either side. The obtained sequence information revealed that the last exon, exon XIII contains a possible consensus polyadenylation signal ATAAA, 3893bp downstream of the termination codon (Fig. 2.16). This would suggest that the size of the mRNA encoding full-length TGase 6 is 6476 nucleotides. This is consistent with the result obtained by Northern blotting. CAYTG signals which bind to U4 small nuclear RNA (Berget, 1984) are found 27bp, 33bp, 41bp and 48bp downstream of the polyadenylation signal. Another

consensus sequence, YGYGTTY, which is present in many eukaryotic transcripts and contributes a signal for efficient 3' processing (McLauchlan et al., 1985), is found 111bp, and 123bp downstream of the polyadenylation signal. To further confirm that the clone identified is part of the human TGase 6 transcript, PCR amplification could be performed using forward primers positioned within the 3' end of TGase 6 and a reverse primer within the region spanned by the clone. Amplification of an appropriately sized product would indicate that the clone is part of the TGase 6 transcript and the polyadenylation signal found in the clone is the polyadenylation signal for TGase 6. The 3' end sequence of mouse TGase 6 is not matched by any EST sequences in the GenBankTM database.

```

3841 gtaagcatcttccctggaacaggaggccatctgccactccacgtataaagatgctccttcccgcagggag
3911 gctccacgtgtcccagatccatttccccttgatgatggctccagacacatccttccagcaaatgccagcta
3981 tctgtatgcctcgagctcggctgttgaggctgagttcctagaaagctgattctgagatggagtttagga
4051 tgcaggatgttgatctaacagtgatctagggttcaacacccatggaagggagtgggggagcatgagtggtg

```

Fig. 2.16 Structure of the 3'-untranslated region of the human TGase 6 gene. The 3' flanking sequence is shown with the sequences homologous to known consensus sequences for 3' processing of transcripts (ATAAA, CAYTG, and YGTGTTY) underlined.

2.3.6 Sequencing of mouse TGase 6

Having cloned and sequenced human TGase 6, the next objective was to isolate mouse TGase 6 cDNA so that the expression of the gene could be investigated in a more accessible model.

PCR amplification was used to amplify mouse TGase 6 products using mouse brain cDNA with various combinations of human TGase 6-specific primers (Fig. 2.17A). Initially PCR products were obtained for 3 primer combinations as indicated in Fig. 2.18.

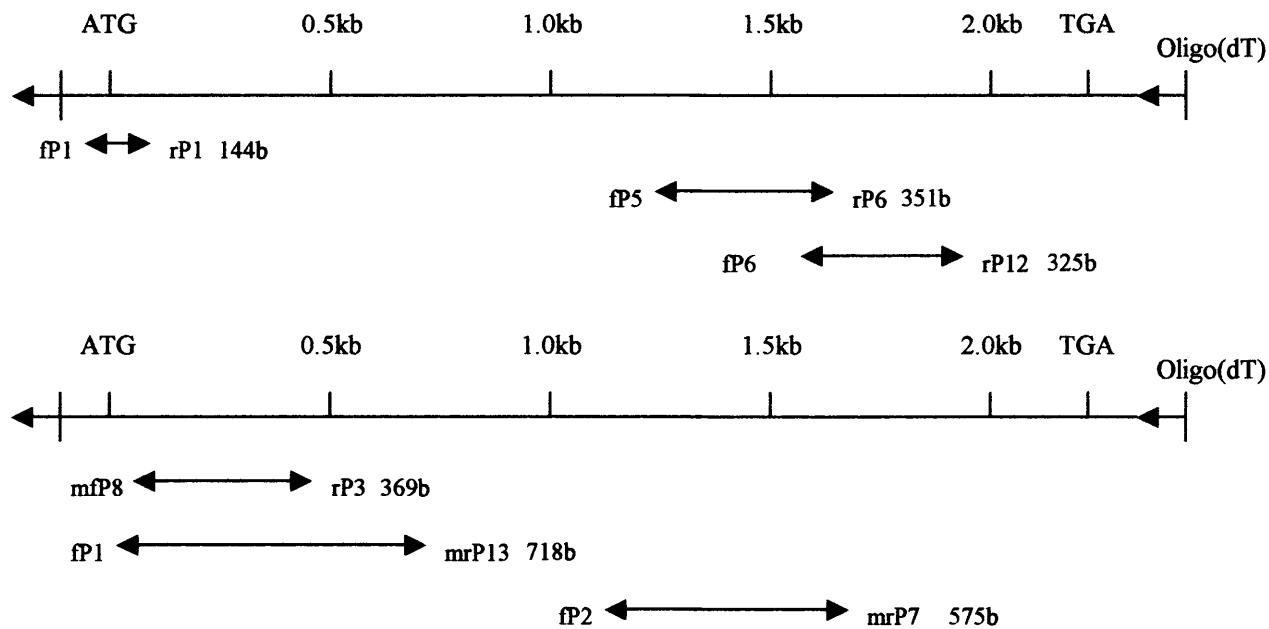


Fig. 2.17 PCR strategy for amplification of cDNA sequences of mouse TGase 6. The top line in A and B represents the cDNA for mouse TGase 6 with the start and stop codons indicated. A shows an outline of the PCR strategy, showing the PCR reactions performed with human TGase 6 oligonucleotide primers which were able to amplify PCR products corresponding to the mouse sequence. B shows an outline of the PCR strategy, showing the PCR reactions performed using a combination of human TGase 6 oligonucleotide primers and mouse TGase 6 oligonucleotide primers designed from the mouse sequence obtained from the products amplified in strategy A. Sequence of primers is given in Appendix 2, Table 1 and 2. The length of the final PCR product is given on the right. The oligo(dT) primer was used as the 3' anchoring primer.

The PCR products obtained were sequenced and gave a similar sequence to human TGase 6, but there were several nucleotide differences, predominantly in the third nucleotide of a codon, confirming that the sequence is infact derived from the mouse orthologene (Fig. 2.20).

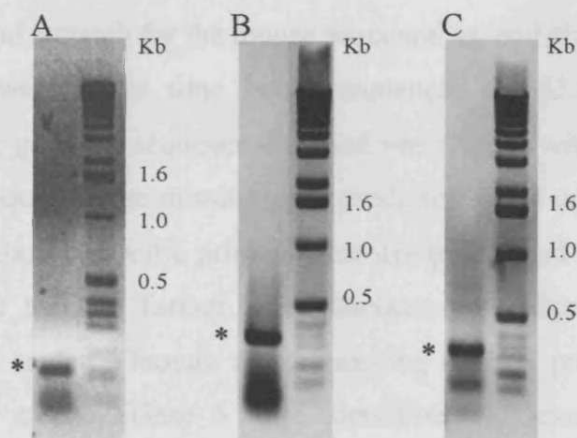


Fig. 2.18 Amplification of TGase 6 from mouse brain using human TGase 6-specific primers. A represents the PCR product amplified with fP1 and rP1, B, fP5 and rP6, and C, fP6 and rP12. The PCR products were separated on 1% agarose gels calibrated with the 1kb ladder. * indicates the PCR products gel extracted.

The mouse sequence obtained was then used to design mouse specific TGase 6 primers. PCR amplification was repeated using the new mouse TGase 6 specific primers (see Appendix 2, Table 1 and 2) in various combinations with human TGase 6-specific primers (Fig. 2.17B). PCR products were obtained with a combination of mfP8 and rP3, fP1 and mrP13, and fP2 and mrP7 (Fig. 2.19).

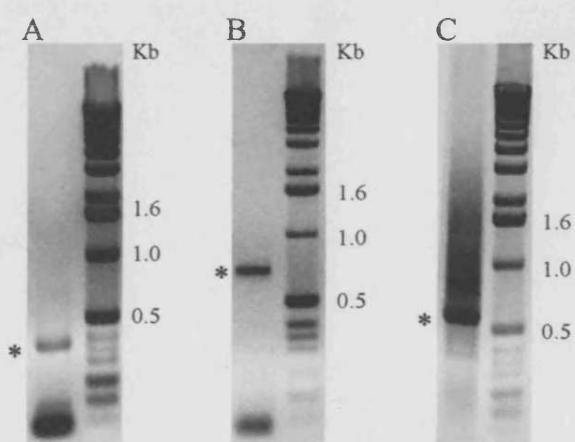


Fig. 2.19 Amplification of TGase 6 from mouse brain. A represents the PCR product amplified with mfP8 and rP3, B, fP1 and mrP13, and C, fP2 and mrP7. The PCR products were separated on 1% agarose gels calibrated with a 1kb ladder * indicates the PCR products gel extracted.

A blast search found a match for the mouse sequence we had obtained with a genomic sequence which was at the time being sequenced (AL833804, GenBank) (see Appendix 10). The genomic sequence obtained was aligned with the available mouse TGase 6 cDNA sequence, the missing parts predicted based on the human sequence and new mouse TGase 6-specific primers were designed based on the alignment. The new primers were used in further PCR reactions to bridge the gaps within the sequence obtained so far. Through the sequencing of PCR products two additional splice variants of mouse TGase 6 were identified with exon III, or exon III/IV alternatively spliced. A match for mouse TGase 6 was found with EST IMAGE clones BB612839 and BB630356 (see Appendix 6 and 7, respectively). A comparison of these sequences to the genomic sequence found considerable homology at the 5' end including exon I. However, one clone (BB630356) corresponds to the alternatively spliced RNA lacking exons III/IV. The EST clone (AL833804) matched the upstream sequence of exon I, close to the transcription initiation site determined for human TGase 6 (Fig. 2.15). The complete mouse TGase 6 sequence was determined and shows 85% identity with the human sequence on the amino acid level (Fig. 2.20). The sequences for mouse TGase 6 have been submitted to GenBank: AY159126 (long transcript), AY177607 (exon III spliced transcript) and AY177606 (exon III/IV spliced transcript) (see Appendix 11, 12 and 13, respectively).

S A H D T D Q N L S V D K Y V D S F G R T L E D L T E 326
 human CTGAGCCACGACAGACCAGAACCTGAGTGTGGACAAATACGTGGACTCCTTCGGGCGGACCCTGGAGGACCTGACAGAA 1433
 mouse *****C**TGGC**T*****T*****T**AT**C**A*****
 * * * * * G * * * * * * * * * * * Y * * * * *

D S M W▼N F H V W N E S W F A R Q D L G P S Y N G W Q 353
 human GACAGCATGTG•GAATTTCCATGTCTGGAATGAGAGCTGGTTGCCCAGGACCTAGGCCCTCTTACAATGGCTGGCAG 1514
 mouse *****.***C*****C*****A**A*****G*****A**C**TG*****
 * D * * *

V L D A T P Q E E S E G▼V F R C G P A S V T A I R E G 380
 human GTTCTGGATGCCACCCCGAGGAGAGTGAAG•GTGTGTTCGGTGCAGCCAGCCTCAGTCACCGCCATCCGCGAGGGT 1595
 mouse *****T*****A*****C*****A*****T**T**T*****G*****T*****T*****
 * * * * * * * * * * * * * * * * * M * * * * * * * * * * *

D V H L A H D G P F V F A E V N A D Y I T W L W H E D E 408
 human GATGTGCACCTGGCTCAGATGGCCCTTCGTGTTGCGGAGGTCAACGCCGACTACATCACTGGCTGTGGCAGGAGATG 1676
 mouse **C*****T**A**C**T*****C*****T**A*****A*****CA
 * K

S R E R V Y S N T K K I G R C I S T K A V G S D S R V 435
 human AGAGCCGGGAGCGTGTACTCAAACACGAAGAAGATTGGGAGATGCATCAGCACCAAGGCGGTGGGCACTGCCCGC 1758
 mouse *****G*****G*****GG*****C*****C**C**G*****T*****A*****T*GT*
 R * * * * * * * * * * * D *

D I T D L Y K Y P E G▼S R K E R Q V Y S K A V N R L 461
 human TGGACATCACTGACCTCTACAAGTATCCGGAAG•GGTCCCGAAAGAGAGGCGGTGTACAGCAAGGCGGTGAACAGGCTG 1838
 mouse *****T**A**C**T*****A*****G*****G*****A*****A**T*****G**AA**
 * * * * * G * K K *

F G V E A S G R R I W I R R A G G R C L W R D D L L E 488
 human TTCGGCGTGAAGCCTCTGGAAGGAGAATCTGGATCCGAGGGCTGGGGTGCCTGCTCTGGCGTGACGACCTCCTGGAG 1919
 mouse C*AA**T*****GG**G*****G*GG*****CA**C**TG**G**G*****C*****A
 L S * * * W * * * R R * * * * S V * G V * * * * *

P A T K P S I A G K F K V L E P P M L G H D L R L A L 515
 human CCTGCCACCAAGCCAGCATCGTGGCAAGTTCAGGTGCTAGAGCCTCCCATGCTGGGCCAGGACCTGAGACTGGCCCTG 2000
 mouse **G**T*****A*****A*****A*****G*****A**GG**C*****G*****AG*****T**
 * V * * * * * T * * * * * * * * * * * * * * * V * * * Q * * * K * * *

C L A N L T S R A Q R V R V N L S G A T I L Y T R K P 542
 human TGCTTGCCAACTCACCTCCCGGCCAGCGGGTGAAGGTCAACCTGAGCGGTGCCACCATCCTCTATACCCGCAAGCCA 2081
 mouse *****A*****G**A*****CC**A*****TG*****C*****T
 * * T * * * * A * * * * * * * * * * * V * * * * * * * * *

V A E I L H E S H A V R L G P Q E E▼K R I P I T I S Y 569
 human GTGGCAGAGATCCTGCATGAATCCACGCGGTGAGGCTGGGGCCGCAAGAAG•AGAAGAGAATCCCAATTACAATATCTTA 2162
 mouse *****G*****A*****A*****A**T*****.A*****AG*****G**C*****
 * * * * * R * * * * * T * * * * * L * * * * * K * * * V * * *

S K Y K E D L T E D K K I L L A A M C L V T K G E K L 596
 human CTCTAAGTATAAAGAAGACCTGACAGAGGACAAGAAGATCCTGTTGGCTGCCATGTGCCTTGTACCAAGGAGAGAGCT 2243
 mouse ****C***C***G*****A*****GT**G**C*****
 * Q * * G * * * * * * * * * * * * * * * S * * * *

L V E K D I T L E D F I T I K ▼V L G P A M V G V A V T 623
 human TCTGGTGGAGAAGGACATTACTCTAGAGGACTTCATCACCATCAAG•GTTCTGGGCCAGCCATGGTGGGAGTGGCAGTTA 2322
 mouse C*****C**G*****.***G*****G*****AG*****A*****A**C**
 * V * * * T * *

V E V T V V N P L I E R V K D C A L M V E G S G L L Q 650
 human CAGTGAAGTGACAGTAGTCAACCCCTCATAGAGAGTGAAGGACTGTGCGCTGATGGTGGAGGGCAGCGCCCTTCTCC 2405
 mouse *T*****CTG**G*****TCG*****C**A*****T*****A*****T*****
 * * * * * L * * * * * S * * * * * V * * * * * * * * *

E Q L S I D▼V P T L E P Q E R A S V Q F D I T P S K 676
 human AGGAACAGCTCAGCATGA•CGTGCCTACCCTGGAGCCTCAGGAGAGGGCCTCAGTCCAGTTGACATCACCCCTCCAAA 2483
 mouse *AGG*****A**T**TG*•GG**C**G**T**GC*****A*****TGA**T**A**CA**A*****T*****G
 G * * * * * E * * * * * S * * * * * K * * * * * L * * * * * N * * * *

S G P R Q L Q V D L V S P H F P D I K G F V I V H V A 703
 human AGTGGCCCAAGCAGCTGCAGTGGACCTGTAGCCCTCACTTCCCGGACATCAAGGGCTTGTGATCGTCCATGTGGCC 2564
 mouse *****C*****C**T**A**G*****A*****A*****A**C**T
 * * * * * * * * * * * * * * * * * S Q * * * * * * * * * * * I * * *

```

          T A K
human ACTGCCAAGTGATGGATCATGAGGGACTGAGAGGGGTGGATTGGCCCTGTCTCCTCCTGCCATTCTTTGTCTCTTCC 706
mouse **A*****GC**C***A***G*GA**ACTTTG*CCAT*TG*C*GT*T***CATAT*AGAGACAGGGG*CT* 2645
      * * *

human ACATGGGAGCCAGGAGGCCTCAGTTAATCCTGCCCAACCT 2689
mouse *GG*T**TC*TGCTTCTGTCTTTGGCC**ATGAG*TT*GG

```

Fig. 2.20 Sequence alignment for human and mouse TGase 6. * indicate sequence identity, gaps are indicated by a dash (-). Exon boundaries are indicated by ▼.

2.3.7 Proximal promoter sequence

The transcription start site for the mouse TGase 6 has not been determined experimentally but based on the EST identified in the GenBank database it is likely to be close to the site identified in the human gene. The proximal promoter regions for human and mouse TGase 6 were analysed for potential binding sites of transcription factors using MatInspector (Genomatix, Munich, Germany) software package. No classical TATA-box sequence was identified in either sequence indicating that the human and mouse TGase 6 promoter may be a TATA-less promoter. However, other potential transcription factor binding sites were identified and a comparison of the promoter regions of human and mouse was made. Binding sites for hepatic nuclear factor 4 (NF4) was identified in both mouse and human (Fig. 2.21). Binding sites for AP1 and Ets-1 were identified in human TGase 6 and a binding site for NFκB was identified in mouse TGase 6.

A

AP1

-210 cctagcttctgagtcagaaccagctctccccttttcttatctacctcctctcaagagtctagtgtccac
 Ets-1

-140 caaggaagtctagtgccctgaaggaggatcatggacctctgcctctttagatgctgcagaacagagtg
 NF4

-70 gagctgtgggccaggagtcttctgctctcaccaggagggagagggagccacagactgaccggcact

+1 **gcagcgtgcaaccatgtacttcggggctctggcgctgagcctgttttttctcatttgagggttgtgtga**

+71 gggctaagaatggccaaccataaatccgagttccctcagtcttctctctttccccgttctttttcttgcc

B

NF4

-280 cagtaataagaggagtgtgtattccttttgctaaagaattcttcaaagatggagaggaccaaagcttcag

-210 acgaggctgtgtgggttcaagctgcccttctagagctcagggctcctcagcagaaacaagacctgtgctg

-140 ggtatgtgcagggattgcagaggatgtgaagtgataggtgccgtccctgtaagatcataggacctagagg

NFκB

-70 acagctgtggtcatgggactcttctcatcttgccctgaaggagaaggacaccctggagtaggtatcatga

+1 **ctcctggactctggtctggacactgcagctttactatcatgtgcctccaagtggctctgagctctcatgt**

+71 acttgagggatgccaatggtaagaacggctgaccatagtaaccattagctcctcctattctccctgggtcc

Fig. 2.21 The nucleotide sequence of the proximal 5' region of the human and mouse TGase 6 genes. A represents the human sequence and B, the mouse sequence. Consensus sequences for potential regulatory elements are underlined. The major transcription start sites are indicated in bold. The numbering is based on the transcription start site designated as +1.

2.4 Discussion

In this study, a cDNA encoding a novel member of the TGase gene family, TGase 6, was isolated from the human small cell carcinoma cell line H69. This confirmed that the TGM6 gene is a transcribed gene and not a pseudogene. Two related transcripts were obtained which encoded proteins of 708 and 629 amino acids with a molecular mass of 79,466Da and 70,671Da, respectively. The structure of the gene encoding TGase 6 (TGM6) can be inferred from comparison of the cDNA to the available genomic sequence. TGase 6 consists of 13 exons and 12 introns and the size of the exons is comparable to the other members of the TGase family (Table 2.2). A comparison of the determined cDNA sequence to genomic sequences in the NCBI database revealed the organisation of the TGase 6 gene. All intron/exon splice sites are consistent with the known GT/AG donor/acceptor site rule common to eukaryotic genes (Mount, 1982). The size of the introns varied considerably from 80bp to 13488bp.

Table. 2.1 Nucleotide sequence of the exon/intron/exon junctions for human TGase 6. The nucleotide sequence of each 5'- and 3'-splice junction is shown. Capital letters represent exon sequences and small letters are intron sequences. Residues consistent with the splice site consensus sequence (MAG/GTRAG and YAG/G) are underlined.

| Donor sequence | Intron (size in bp) | Acceptor sequence |
|--|---------------------|---|
| TCAC ATG <u>GCA</u> Ggtaagtgggca
M A | I (13483) | tccccacc <u>cag</u> GG ATC AGA GT
G I R |
| ATG GAG <u>ACA</u> Ggtaactgggct
V E T | II (567) | ctctcct <u>cag</u> GA CCC CGG GC
G P R |
| TGG TGT <u>GCA</u> Ggtaggagtggc
W C A | III (1069) | ctgctttt <u>cag</u> AG GAC GAT GT
E D D |
| C TAC GGG <u>CAG</u> gtctccagggg
Y G Q | IV (1293) | ctggggag <u>cag</u> TTT GAG GAG G
F E E |
| C AGT GCC <u>ATG</u> gtgagaagccc
S A M | V (1514) | cctctgccc <u>cag</u> GTG AAC AGC A
V N S |
| CTG TGC <u>ACA</u> Ggtaccctggga
L C T | VI (567) | ctgctctg <u>cag</u> TC CTC AGG TG
V L R |
| AC AGC ATG <u>TG</u> gtgggtcctgc
S M W | VII (2952) | ctctgtgg <u>cag</u> G AAT TTC CAT
N F H |
| GAG AGT <u>GAA</u> Ggtacgctcaat
E S E | VIII (80) | ctcctgccc <u>cag</u> GT GTG TTC CG
G V F |
| TAT CCG <u>GAA</u> Ggtaagggccac
Y P E | IX (13448) | atgccctg <u>cag</u> GG TCC CGG AA
G S R |
| CCG CAA <u>GAA</u> Ggtaagtgtacg
P Q E | X (12872) | ttccctt <u>cag</u> AG AAG AGA AT
E K R |
| C ACC ATC <u>AAG</u> gtgacctcagc
T I K | XI (293) | gtcccctt <u>cag</u> GTT CTG GGC C
V L G |
| TC AGC ATC <u>GAG</u> taagtgccagc
S I D | XII (1461) | ccttctc <u>cag</u> C GTG CCT ACC
V P T |

Table. 2.2 DNA structure of human TGases. Numbers represent sizes in bp.

| Exon number | FXIII | TGase 1 | TGase 2 | TGase 3 | TGase 6 | TGase 5 | Band 4.2 | TGase 4 |
|-------------|-------|---------|---------|---------|---------|---------|----------|---------|
| I | 63 | 89 | | | | | | |
| II | 148 | 319 | 10 | 7 | 7 | 10 | 10 | 63 |
| III | 189 | 189 | 180 | 174 | 174 | 180 | 186 | 174 |
| IV | 252 | 249 | 243 | 240 | 243 | 246 | 234 | 237 |
| V | 119 | 119 | 119 | 119 | 119 | 119 | 119 | 119 |
| VI | 108 | 108 | 129 | 129 | 129 | 129 | 104 | 108 |
| VII | 175 | 174 | 178 | 178 | 178 | 178 | 178 | 175 |
| VIII | 139 | 140 | 136 | 136 | 139 | 139 | 139 | 139 |
| IX | 104 | 104 | 104 | 104 | 104 | 104 | 104 | 104 |
| X | 89 | 89 | 243 | 246 | 243 | 240 | 243 | 252 |
| XI | 154 | 154 | - | - | - | - | - | - |
| XII | 288 | 283 | 273 | 309 | 342 | 369 | 300 | 279 |
| XIII | 161 | 161 | 161 | 158 | 155 | 161 | 161 | 170 |
| XIV | 137 | 137 | 134 | 134 | 134 | 134 | 134 | 137 |
| XV | 151 | 226 | 145 | 145 | 154 | 151 | 160 | 1026 |

Primer extension analysis located the transcription initiation site 455 nucleotides upstream of the translation start codon, and it is likely that the polyadenylation site is located 3893 nucleotides downstream of the termination codon. However, 3' extension analysis needs to be carried out to further confirm this experimentally. The two isolated gene products result from alternative splicing of exon XII. The sequence alterations due to the splicing result in a short protein which terminates just after the first C-terminal β -barrel domain. The β -barrel domains have been implicated in the regulation of enzyme-substrate interaction, and the lack of the second C-terminal β -barrel (Fig. 2.4) is therefore likely to be of biological significance.

The mouse cDNA sequence for TGase 6 has been isolated and the full-length amino acid sequence of the mouse TGase gene product was inferred from the cDNA sequence. The human and mouse gene products show a high degree of similarity, 85% identity, on the amino acid level, and no major differences were found. It was also determined that there are three splice variants of mouse TGase 6, with exon III, exon III/exon IV or exon XII alternatively spliced. The absence of exon XII, like in the human TGase 6 transcript, results in a frameshift. The gene structure of the mouse *tgm6* gene is identical to that of the human TGM6 gene. All intron/exon splice sites are consistent with the known GT/AG donor/acceptor site rule common to eukaryotic genes and are comparable to the splice sites found in the human sequence.

Table.2.3 Nucleotide sequence of the exon/intron/exon junctions for mouse TGase 6. The nucleotide sequence of each 5'- and 3'-splice junction is shown Capital letters represent exon sequences and small letters are intron sequences. Residues consistent with the splice site consensus sequence (MAG/GTRAG and YAG/G) are underlined.

| Donor sequence | Intron (size in bp) | Acceptor sequence |
|---|---------------------|---|
| TGC ATG <u>GCA</u> Ggtgagtgggca
M A | I (12381) | ttcctgctcag <u>GG</u> GTC AGA AT
G V R |
| ATG GAG <u>ACA</u> Ggtgagtggcct
V E T | II (504) | cttctgccag <u>GA</u> CCT CAC GC
G P H |
| TGG TGC <u>CCA</u> Ggtaggaaccgc
W C P | III (703) | ctgcttttcag <u>AG</u> GAC GAT GT
E D D |
| C TAC GGG <u>CAG</u> gtcttcagggg
Y G Q | IV (1332) | tctgggagcag <u>TTT</u> GAA GAG G
F E E |
| C AGT GCC <u>ATG</u> gtaaggaggcc
S A M | V (2207) | tctctgctcag <u>GTG</u> AAC AGC A
V N S |
| ATG TGC <u>ACA</u> Ggtatggaagga
M C T | VI (429) | ccactccacag <u>TC</u> CTC AGG TG
V L R |
| AC AGC <u>ATG</u> TGgttaggtcccag
S M W | VII (1288) | gtctctggcag <u>G</u> AAC TTC CAC
N F H |
| GAG AGC <u>GAA</u> Ggtagggatcat
E S E | VIII (66) | ctcccaccag <u>GT</u> ATG TTC CG
G M F |
| TAT CCA <u>GAA</u> Ggtgagggcctt
Y P E | IX (1334) | gcgccctgcag <u>GG</u> TCC CGA AG
G S R |
| CCA CTA <u>GAA</u> Ggtaagatcaga
P L E | X (5601) | ctacctttcag <u>AA</u> AAG AAG AT
E K K |
| C ACC ATC <u>AAG</u> gtaactcctgt
T I K | XI (245) | tgcttctccag <u>GTG</u> CTG GGG C
V L G |
| TA AGT ATT <u>GAG</u> taagtgatgg
S I E | XII (1149) | tttgtcttcag <u>G</u> GTG CCC AGC
V P S |

A comparison of TGase 6 to the previously characterised human TGases reveals that the structural requirements for TGase activity and Ca²⁺ binding are conserved (Fig. 2.22). The structure of four TGases has now been solved; human FXIIIa (Yee et al., 1994), TGase 2 (Liu et al., 2002), TGase 3 enzymes (Ahvazi et al., 2002, 2003) and a fish enzyme (fTGase, equivalent to mammalian TGase 2, Nogouchi et al., 2001), and all show a high degree of similarity. The

reaction centre is formed by the core domain and involves hydrogen bonding of the active site Cys to a His and Asp residue to form a catalytic triad similar to the Cys-His-Asn triad found in the papain family of cysteine proteases (Yee et al., 1996). The residues comprising the catalytic triad are conserved in TGase 6 (Cys274, His333, Asp356) and the core domain shows a high level of conservation as suggested by a sequence similarity of 59% when compared to the other members of the TGase family. A Tyr residue in barrel 1 domain of the a subunit of factor XIII is hydrogen bonded to the active site Cys residue and it has been suggested that the glutamine substrate attacks from the direction of this bond to initiate the reaction based on analogy to the cysteine proteases. In TGase 6, the Tyr residue is conserved (Tyr538). The Trp residue (Trp279 in FXIIIa) which is proposed to have a role in stabilisation of the oxyanion intermediate produced in the reaction mechanism (Pedersen et al., 1994) is conserved in TGase 6 (Trp238). All these residues thought to be involved in the catalytic reaction mechanism are conserved in the various TGase gene products, apart from band 4.2 protein, which is the only member of the TGase gene family which lacks catalytic activity. Due to the conservation of critical residues for enzyme function and domain folding and the substantial overall similarity of TGase 6 to the other members of the TGase family with catalytic activity, it is probable that the characterised cDNA of TGase 6 encodes an active TGase. A detailed discussion of structural similarities of TGase 6 to other TGases is given in Chapter 7. Analysis of the proximal promoter region of human TGase 6 identified binding site for Ap1 and Ets. This finding is consistent with other TGases, such as TGase 1, TGase 3 and TGase 5. As transcription factors may play a role in the cell type-specific expression, TGase 6 expression may be similar to TGase 1, TGase 3 and TGase 5.

Comparison of mouse TGase 6 to the already solved TGases using CLUSTAL (Higgins et al., 1994) revealed that mouse TGase 6 shows the highest similarity to the human TGase 3 (PDB accession no. 1L9M). Based on the sequence comparison and identification of conserved residues between the two enzymes, it can be predicted that TGase 6 may also have three Ca²⁺ ion binding sites, contain three *cis* peptide bonds similar to the other solved TGases, and the transamidation reaction mechanism may occur by a similar process to TGase 3. However, from a sequence and structural comparison of TGase 3 and TGase 6, it is not clear whether TGase 6 is able to bind and hydrolyse GTP like TGase 2 and TGase 3. The comparison of TGase 6 and TGase 3 and the predicted structure and function of TGase 6 is discussed fully in Chapter 7.

Fig. 2.22 Comparison of the amino acid sequence of human TGase 6 with the sequences of the other members of the TGase family. The sequences are arranged in order to show the TGase domain structure based on the crystal structure of factor XIII a-subunit: N-terminal propeptide domain (d1), β -sandwich domain (d2), catalytic core domain (d3), and β -barrel domains 1 (d4) and 2 (d5). Dashes indicate gaps inserted for optimal sequence alignment and underlined residues represent amino acids conserved in at least four gene products.

FXIII 100 MDGSRDVGWGNPLQRPPTTSPREDFDGSRRGGGNSI FVARCCGCSGNNADDDMKPPEPSDSRRGSSGSTRRRECSGSDRRRVRYSRGGGNVAAAGG 100
101

102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000

3.0 RNA expression of TGase 6

3.1 Introduction

Having identified and sequenced human and mouse TGase 6, it became clear that the residues known to be critical for TGase activity are conserved within TGase 6, and therefore TGase 6 may be an active enzyme with a crosslinking function. To deduce the biological role of TGase 6, its expression pattern was determined as a first step. The aim of this chapter was to determine the expression pattern of TGase 6 using Northern blotting and *in situ* hybridisation. Based on its expression pattern we may then be able to develop a hypothesis as to its functional role.

3.2 Materials and Methods

3.2.1 Northern Hybridisation

A human RNA Master Blot containing poly(A⁺) RNA of 50 different tissues was obtained from Clontech Laboratories, Inc. (Palo Alto, CA). A ~210bp fragment of human TGase 6 was generated by PCR using oligonucleotides fp2 and rP2 (nucleotides 1127-1335 in Fig. 2.3). A ³²P-labelled probe corresponding to this fragment was prepared using the Multiprime DNA Labelling System as described in section 2.2.15.1. Unlabelled dNTPs were removed using G-50 ProbeQuant spin columns. Hybridisation was performed under the conditions recommended by the manufacturer. Briefly, the membrane was prehybridised in 15ml ExpressHybTM Hybridization Solution containing 50µg/ml denatured herring sperm DNA (Sigma), for 30 min at 65°C. 100µl of ³²P labelled probe was denatured for 5 min at 100°C, and then incubated for 30 min at 68°C. The probe was then added to 10ml ExpressHyb solution and the membrane hybridised in this solution overnight at 65°C. The membrane was washed with solution 1 (2x SSC, 1% SDS) for 5 min at RT, three times with solution 1 for 30 min at 65°C and finally four times with solution 2 (0.1x SSC, 0.5% SDS) for 30 min at 55°C. The labelled membrane was exposed to BioMax MR film (Kodak, Rochester, NY) and the film developed after a 5 day exposure.

3.2.2 In situ hybridisation

3.2.2.1 Preparation of mouse TGase 6 RNA probes

A 325bp fragment of the mouse TGase 6 cDNA, corresponding to the 3' end of the gene product was generated by PCR using oligonucleotides fp6 and rP12 (nucleotides 1610-1935 in Fig. 2.3). PCR was carried out as described in section 2.2.8. The fragment was cloned into the pCRII vector (see Appendix 14) using TA cloning and the plasmid, pCRII-in situ6, was generated. For *in vitro* transcription, the cDNA was excised by restriction with PvuII and AflIII, and the 786bp fragment gel purified as described in section 2.2.9.

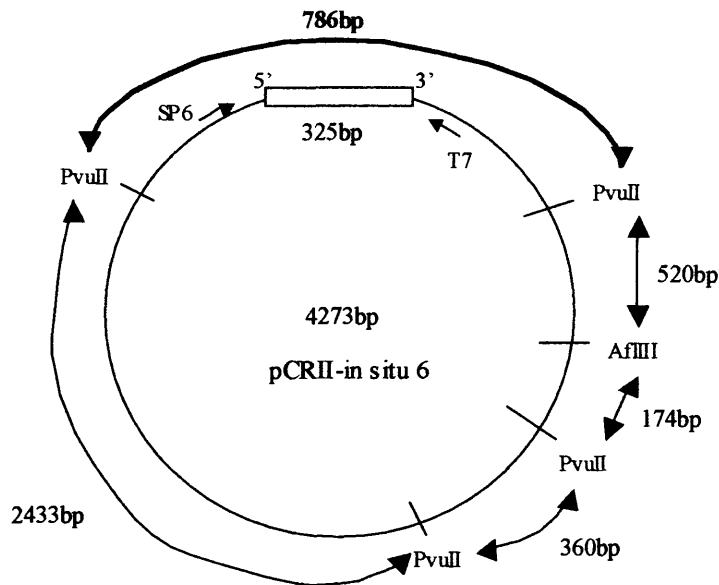


Fig. 3.1 Schematic representation of plasmid pCRII-in situ6, indicating the restriction endonuclease sites. The size of the fragments obtained are indicated and the fragment required is indicated in bold.

Digoxigenin-UTP (DIG)-labelled, single-strand antisense and sense RNA probes were prepared using the RNA Labelling Kit (Roche). 1µg purified DNA fragment in 13µl RNase-free, DEPC treated, double distilled H₂O was added to 2µl NTP Labelling mixture (0.1mM ATP, 0.1mM CTP, 0.1mM GTP, 0.065mM UTP, 0.035mM DIG-11-UTP, pH 7.5), 1x Transcription buffer, 20 units RNase inhibitor (to prevent the degradation of RNA during the labelling reaction) and 40 units of either RNA Polymerase SP6 or RNA Polymerase T7 (to synthesise RNA from the DNA template). The reactions were mixed gently and incubated at 37°C for 2 hrs. 20 units RNase-free DNase I was added and reactions were incubated at 37°C for 15 min to degrade template DNA. The reaction was terminated by the addition of 0.2M EDTA (pH 8.0) to a final concentration of 0.02M. To precipitate the RNA, 0.4M LiCl and 75µl chilled 100% ethanol was added and the reaction was left to precipitate at -80°C for 30 min. RNA was collected by centrifugation at 13,000xg for 15 min at 4°C. The RNA pellet was washed with 50µl chilled 70% ethanol, air-dried for 10 min, and then the pellet resuspended in 20µl sterile RNase-free DEPC treated H₂O.

To determine the yield of the DIG-labelled RNA, the RNA was compared to a dilution series of a DIG-labelled RNA standard. 1µl spots of a dilution series of the

generated DIG-labelled RNA and labelled control RNA of known concentration were applied to a strip of nylon membrane (as described in section 2.2.16.1.6). The nucleic acid was fixed to the membrane by crosslinking with 1200 μ Joules using an UV-Stratalinker. Immobilised RNA was subsequently visualised using detection reagents from the DIG nucleic acid detection kit (Roche). For this purpose, the membrane was rehydrated in 20ml Washing buffer (0.1M Maleic acid; pH 7.5, 0.15M NaCl, 0.3% (v/v) Tween 20) by incubation for 2 min at RT on a rocking platform. The membrane was then incubated in 10ml Blocking solution (1x working solution prepared by diluting 10x Blocking reagent 1:10 in Maleic acid buffer [0.1M maleic acid, 0.15M NaCl, pH 7.5]) for 30 min to block unspecific binding sites on the membrane, 10ml Antibody solution (Anti-digoxigenin-AP, diluted 1:5000 in Blocking solution) for 20 min, washed with 20ml Washing buffer twice for 15 min and then immersed in 10ml 0.1M Tris-HCl, 0.1M NaCl, pH 9.5 (Detection buffer) for 2 min. 200 μ l nitroblue tetrazolium salt (NBT)/ 5-bromo-4-chloro-3-indolyl phosphate (BCIP) solution (18.75mg/ml NBT, 9.4mg/ml BCIP in 67% (v/v) dimethyl formamide) was added to 10ml Detection buffer (substrate solution), and the membrane was incubated in this alkaline phosphatase substrate solution for 1 hr in the dark to develop the reaction. The enzyme catalysed reaction with NBT/BCIP produces an insoluble blue precipitate which allows visualisation of the DIG-labelled RNA molecules. The reaction was stopped by incubation of the membrane in DEPC-treated H₂O for 5 min. Fig. 3.2 shows the dilution series of the generated probes in relation to the control RNA probe. The intensities of the dye precipitate in the dilution series of DIG-labelled RNA and control RNA were compared to estimate the concentration of the generated RNA probes.

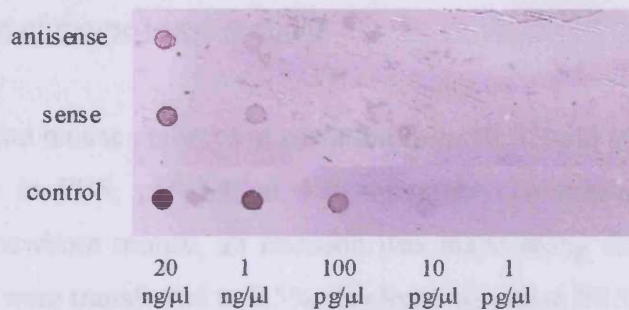


Fig. 3.2 Determination of generated DIG-labelled RNA probe. The membrane shows the intensities of the generated antisense and sense TGase 6 DIG-labelled RNA probes in comparison to the control RNA probe. The sense and antisense RNA probes concentration was estimated to be approximately 25ng/ μ l and 5ng/ μ l, respectively.

3.2.2.2 Preparation of mouse TGase 5 RNA probes

A 387bp fragment of mouse TGase 5, corresponding to the 3' end of the gene (nucleotides 1419-1806) was generated by PCR using oligonucleotides TGase 5 mfP1 and mrP1 (see Appendix 2 Table 3). PCR was performed as described in section 2.2.8. DIG-labelled, single-strand antisense and sense RNA probes were prepared as described for TGase 6 in section 3.2.2.1.

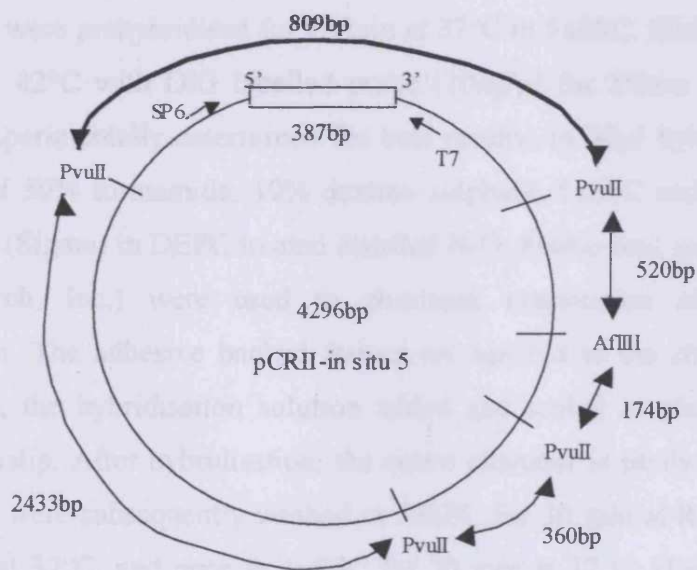


Fig. 3.3 Schematic representation of the restriction endonuclease sites within the pCRII vector containing the TGase 5 cDNA. The size of the fragments obtained are indicated and the fragment required is indicated in bold.

3.2.2.3 Preparation of mouse tissue sections

Newborn mouse and mouse embryos at gestation days 11, 13 and 16 were fixed in 4% paraformaldehyde in PBS, pH 7.4, at 4°C overnight. To improve penetration of fixative into the newborn mouse, an excision was made along the abdomen with a scalpel. The mice were transferred to 0.5% paraformaldehyde in PBS containing 0.42M EDTA for 4 days for demineralisation of skeletal tissues, washed in PBS, then processed through a graded ethanol series, and paraffin embedded by standard methods. Sagittal sections of 5µm thickness were cut, transferred onto gelatin coated glass slides and dried overnight at 50°C.

3.2.2.4 Hybridisation

Tissue sections were deparaffinised in xylene and rehydrated in a graded ethanol series of 100%, 95% and 70% respectively. The sections were washed with DEPC treated distilled H₂O, treated with proteinase K (Sigma) (4µg/ml in 100mM Tris-HCl, pH8, 50mM EDTA) at 4°C for 15 min before being acetylated with 0.25%(v/v) acetic anhydride (Sigma) in 0.1M triethanolamine, pH 8.0 (TEA) (Sigma) for 10 min.

All sections were prehybridised for 15 min at 37°C in 5xSSC. Slides were hybridised overnight at 42°C with DIG labelled probe (20ng/µl for TGase 6 and 25ng/µl for TGase 5, experimentally determined for best results) in 30µl hybridisation solution consisting of 50% formamide, 10% dextran sulphate, 5xSSC and 300µg/ml herring sperm DNA (Sigma) in DEPC treated distilled H₂O. Frame-Seal incubation chambers (MJ Research, Inc.) were used to eliminate evaporation of reagents during hybridisation. The adhesive backed frames are applied to the slide over the tissue section area, the hybridisation solution added and sealed in place with a flexible plastic coverslip. After hybridisation, the entire chamber is easily removed from the slide. Slides were subsequently washed in 2xSSC for 30 min at RT, twice in 2xSSC for 20 min at 37°C, and once in 1xSSC for 20 min at 37°C. Hybridised probe was subsequently visualised according to the manufacturer's instructions (Roche) and also as described in section 3.2.2.1. Briefly, slides were rinsed in buffer 1 (100mM Tris-HCl, pH 7.5, 150mM NaCl), then pre-incubated in buffer 1 containing blocking solution for 30 min. Subsequently, they were incubated with alkaline phosphatase-

labelled anti-DIG antibody (diluted 1:500) in the same buffer for 2 hrs, washed twice with buffer 1, then rinsed with buffer 3 (100mM Tris-HCl, pH 9.5, 150mM NaCl, 50mM MgCl₂). Colour development was conducted with a solution of NBT/BCIP in buffer 3 for up to 16 hours in the dark. The reaction was stopped by the addition of 100mM Tris-HCl and 50mM EDTA (pH 8.0). We found that development of the slides for 12 hours gave good results. Slides were then washed and mounted in a solution of 90% glycerol, 10% Na-PO₄.

3.2.3 Luxol Fast Blue/Cresyl violet stain

The Luxol Fast Blue/Cresyl violet stain is designed to show the myelin and nuclei of nerve cells. Tissue sections were deparaffinised in xylene and rehydrated in a graded ethanol series of 100%, 95% and 70% respectively. The sections were stained in 1% Luxol fast blue solution (Sigma) overnight at 37°C. The sections were washed in 95% ethanol, then in distilled H₂O. Differentiation was performed by immersing the sections in saturated lithium carbonate solution for 15 sec, then in 70% ethanol, and finally rinsing in H₂O. Differentiation is carried out to remove “background” colouring. This process was repeated until the grey and white matter were clearly distinguished and staining was only present in the myelin sheath. The tissue sections were subsequently stained in 1% cresyl violet solution for 10 min at RT, then washed in H₂O. The sections were differentiated in cresyl violet differentiator (70% ethanol, 10% acetic acid), then rinsed in 100% ethanol and cleared in xylene. Cresyl violet binds not only to the nuclei and Nissl substance but combines with the luxol fast blue anions present in the myelin. The sections were mounted in a solution of 90% glycerol, 10% Na-PO₄.

3.2.4 RNA samples

RNA samples from brain of normal and transgenic mice were obtained from P. Mastroberardino. RNA from wild-type mouse brain was used for the experiments in this chapter.

3.2.4.1 cDNA synthesis from RNA

cDNA was prepared from the RNA samples provided using SuperScript II RNase H-reverse transcriptase as described in section 2.2.5.2.

3.2.4.2. TGase screening

cDNA prepared from wild-type brain was used in PCR analysis. PCR reactions were prepared on ice as follows: 1µl cDNA, 1x PCR buffer, 2.5mM MgCl₂, 0.2mM dNTPs, 5µM of each primer, 2.5 units DNA *Taq* Polymerase (Promega) and nuclease-free H₂O to a total volume of 25µl. Primers for each of the 9 different TGases were used in the reactions:

Forty PCR cycles were carried out in a Gene Amp 9600 thermal cycler as described in section 2.2.8. Amplified products were analysed in 1% agarose gels calibrated with the 1kb ladder.

3.3 Results

3.3.1 Northern blotting

To obtain a first indication on the expression pattern of TGase 6 a dot blot Northern blot analysis of more than 50 adult and foetal human tissues was performed. TGase 6 showed widespread expression at a low level (detection required prolonged exposure times). TGase 6 mRNA expression was detected in the central nervous system (Fig. 3.4) which is in contrast to the other TGase gene products, including TGase 5 and TGase 7 (Grenard et al., 2001).

3.3.2 In situ hybridisation

3.3.2.1 TGase 6 expression in newborn mouse

To obtain a clearer understanding of TGase 6 expression on a cellular level, *in situ* hybridisation was performed on sagittal newborn mouse sections. A 325bp fragment corresponding to the 3' end of TGase 6 was used to generate digoxigenin-UTP labelled single stranded sense and antisense RNA probes. The 3' end was used for the probe as this area has the least homology between the different TGases and a similar human probe in Northern blotting gave no cross-hybridisation with other TGase gene products (Fig.2.11). The reporter molecule on the probe was detected using an AP-anti DIG antibody and detection visualised with NBT and BCIP. Fig. 3.5 gives a general overview showing that a specific reaction with the antisense probe resulted in a blue/brown precipitate.

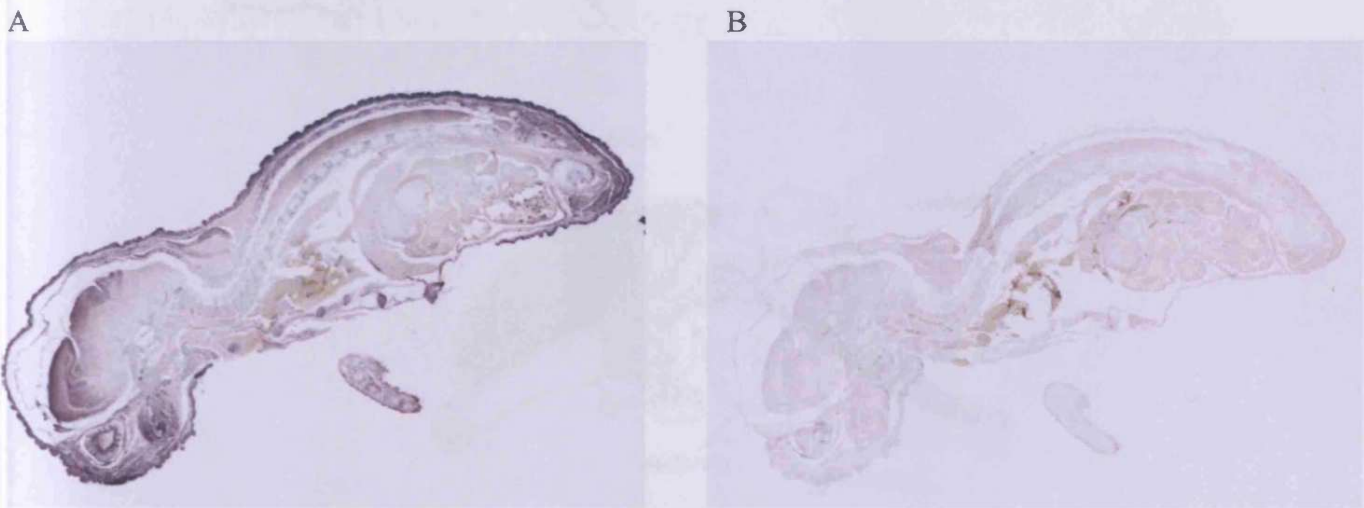
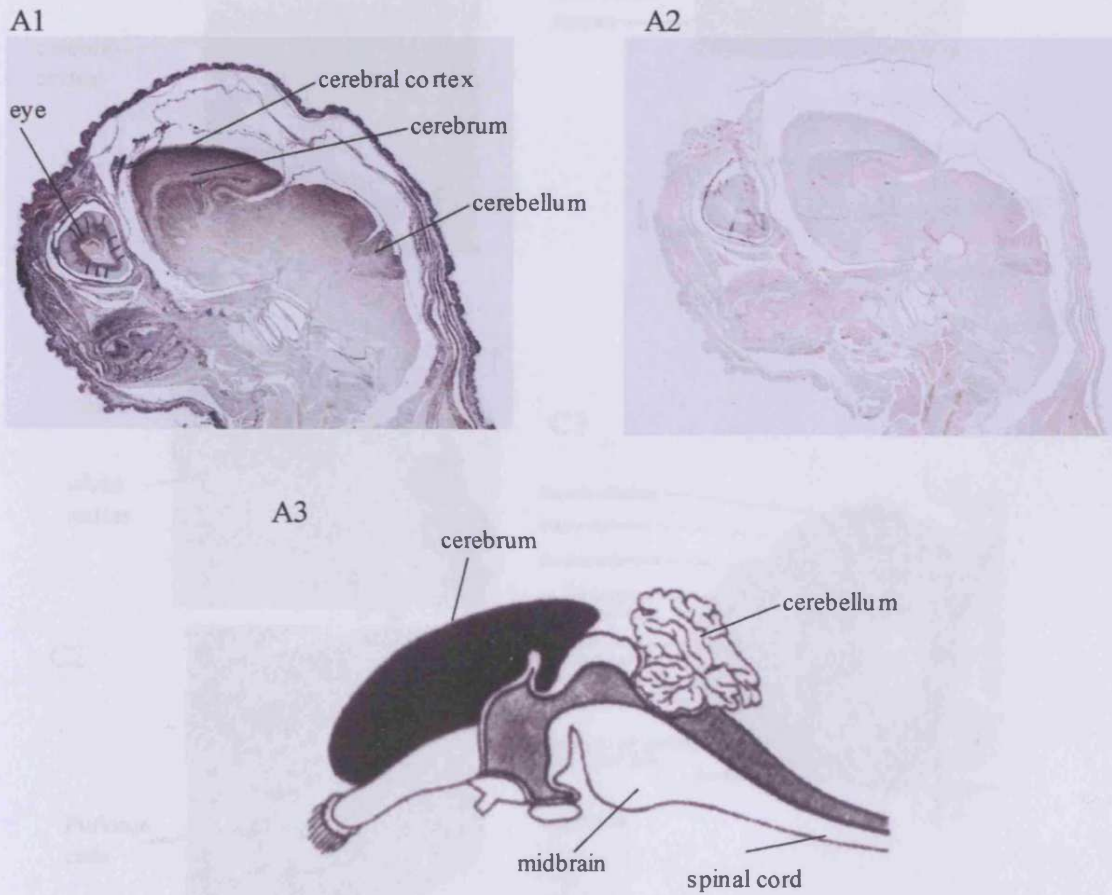
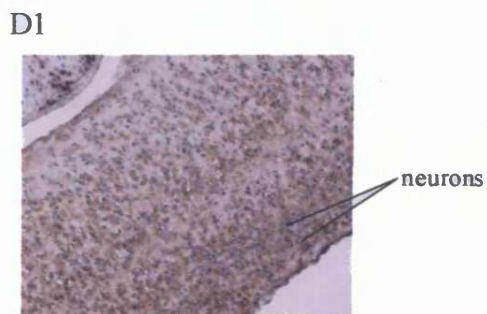
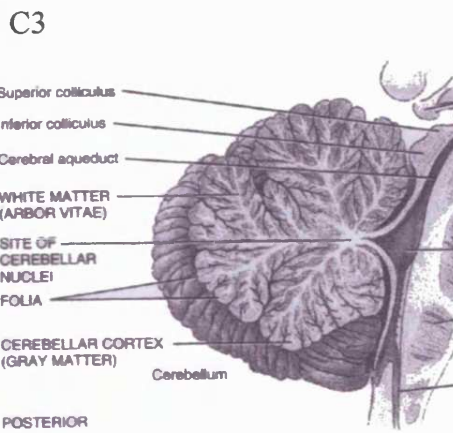
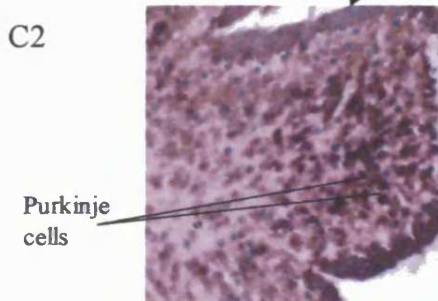
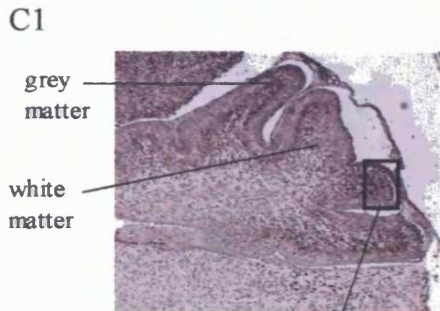
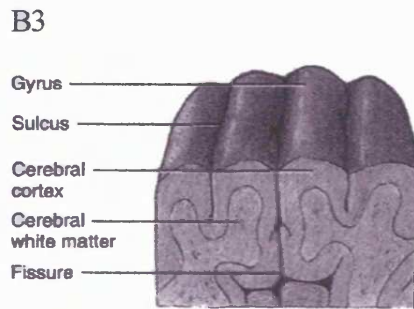
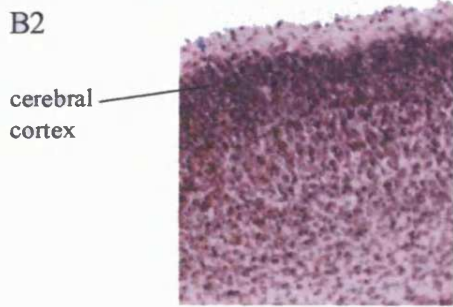
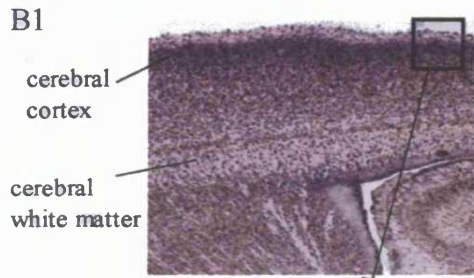


Fig. 3.5 *In situ* hybridisation analysis for TGase 6 on sagittal newborn mouse sections. An overview of the whole newborn mouse probed with the antisense and sense TGase 6 probes are given in A and B (x10), respectively.

The overview showed that TGase 6 is predominantly expressed in the following tissues: TGase 6 appears to be expressed in the neurons of the cerebral cortex (Fig. 3.6 A1 and B1). The cerebral cortex is the superficial grey matter layer of the cerebrum which contains billions of neurons. TGase 6 expression can also be detected in the Purkinje cells of the cerebellum (Fig. 3.6 C1). Purkinje cells are found in a layer between the cerebellar cortex which is the molecular layer and relatively free of

neuronal cells, and the cerebellar medulla. TGase 6 was also expressed in the neurons of the spinal cord (Fig. 3.6 D1) and the retinal cells of the eye (Fig. 3.6 E1). In skin, TGase 6 was specifically expressed in the epidermis and corresponding keratinising cells in hair follicles (Fig. 3.6 F1). TGase 6 was also expressed in developing vertebrae and other long bones where vasculature invasion into the cartilage and bone formation occurs (Fig. 3.6 G1).



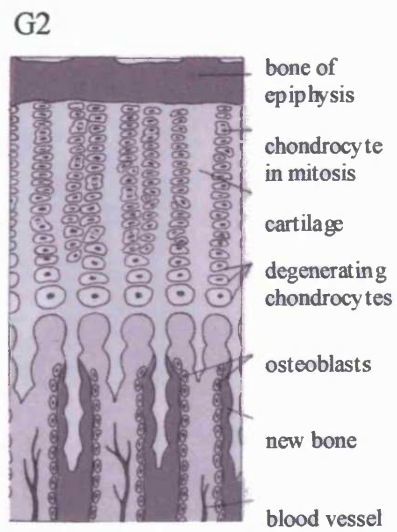
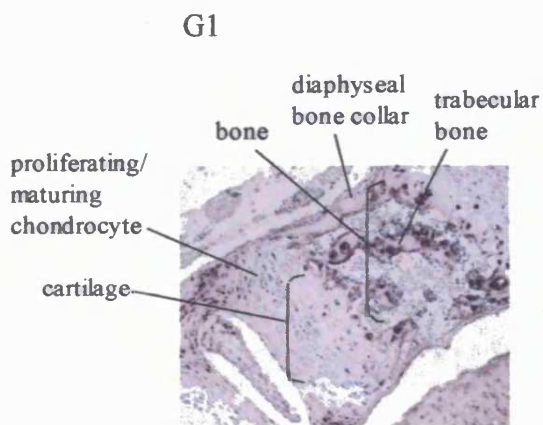
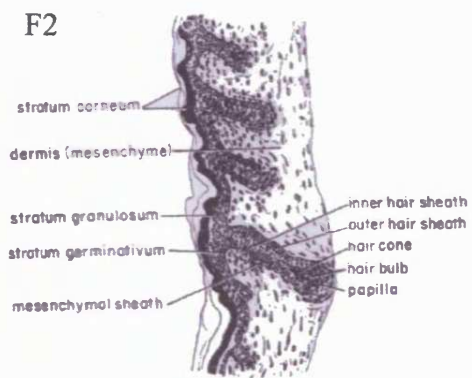
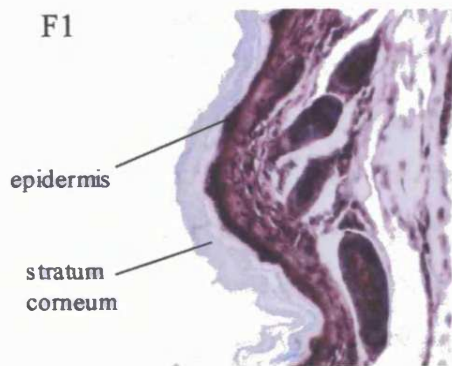
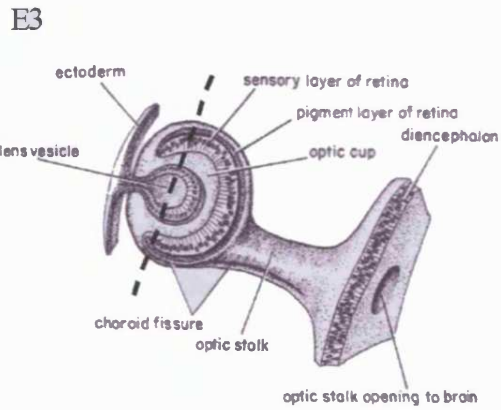
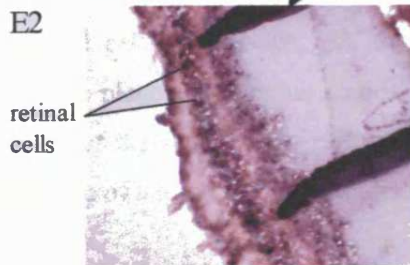
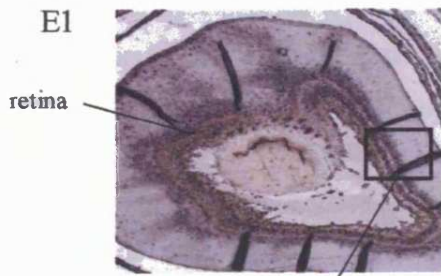




1871

1872





110



111



112



Fig 3.6. *In situ* hybridisation analysis for TGase 6 on sagittal newborn mouse sections. A1 shows an overview of the TGase 6 antisense probe staining in the head region (x10), A2, the TGase 6 sense probe staining (x10) and A3 shows a schematic overview of the rodent brain (Lippert, 1989). B1 (x100) and B2 (x200) show the staining in the cerebral cortex, B3, a schematic of the cerebrum (Tortora and Grabowski, 2000), C1 (x100) and C2 (x400), staining in the purkinje cells in the cerebellum, C3, a schematic of the cerebellum (Tortora and Grabowski, 2000), D1, staining in the spinal cord (x100), E1 (x100) and E2 (x400), staining in the retinal cells of the eye, E3 a schematic of the eye (Rugh, 1968), the dotted line represents the plane of the tissue section, F1, staining in the epidermis of the skin (x400), F2, a schematic of the skin (Rugh, 1968), , G1, staining in the developing vertebrae (x200) and G2, a diagram of a long bone growth plate (Martini, 1989).

To confirm that the staining detected with the TGase 6 probe was neuronal, a luxol fast blue/cresyl violet stain was performed on sections that had previously been used for *in situ* hybridisation with the TGase 6 probe and the staining compared. The stain is designed to show the myelin and nuclei of nerve cells (cresyl violet binds to the nuclei and nissl substance and combines with the luxol fast blue present in the myelin) In the cerebellum (Fig. 3.7A) the arrows on the figure point to the Purkinje cells stained with the TGase 6 probe after *in situ* hybridisation. In Fig. 3.7B, the arrows point to the same cells which luxol fast blue/cresyl violet was able to stain. This confirms that TGase 6 is expressed in neuronal cells. Experiments to identify in more detail cell populations in the nervous system which express TGase 6, by using specific antibodies to cell lineage markers is currently ongoing in the lab.

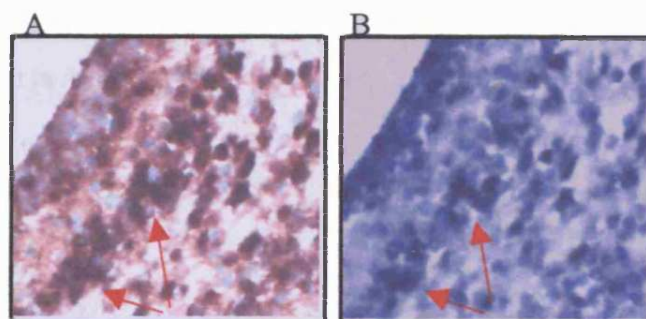


Fig. 3.7 Comparison of *in situ* hybridisation analysis with a luxol fast blue/cresyl violet stain. A. The arrows indicate TGase 6 probe binding in the Purkinje cells of the cerebellum on a newborn mouse tissue section (x400). B. The arrows indicate the same cells on the same tissue section stained with luxol fast blue/ cresyl violet (x400).

— — — — —

... ..

— — — — —

... ..

— — — — —

... ..



3.3.2.2 TGase 5 expression in newborn mouse

We compared TGase 6 expression with that of another novel TGase gene that had been proposed to be expressed in the central nervous system. For this purpose, probes were generated from the 3' end of TGase 5 and used in further *in situ* hybridisation experiments. TGase 5 showed weak staining in the cerebral cortex and cerebellum of the brain, but its expression does not appear to be neuron specific (Fig. 3.9 A, B and C). TGase 5 showed little staining in the eye (Fig. 3.9 D) and again TGase 5 appeared to be expressed in the spinal cord but not in a neuron specific fashion (Fig. 3.9 E). In skin, TGase 5 is abundantly expressed as expected (Aeschlimann et al., 1998) and is present only in all suprabasal cell layers (Fig. 3.9 F). TGase 5 expression was also detected in chondrocytes in the developing vertebrae (Fig. 3.9 G).

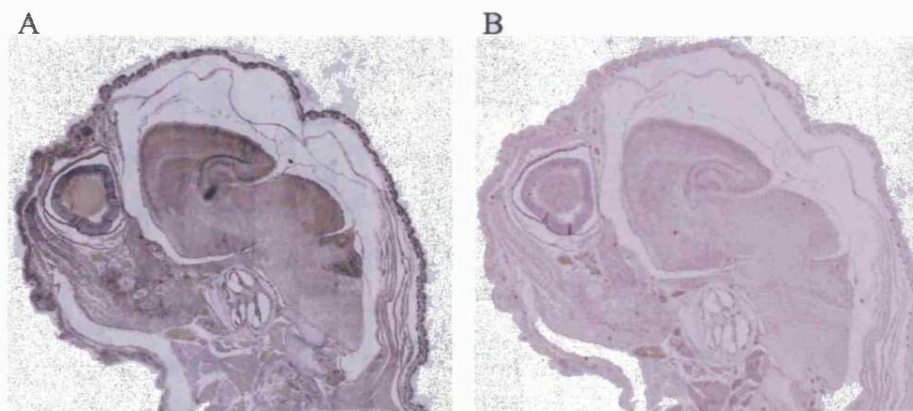


Fig. 3.8 *In situ* hybridisation analysis for TGase 5 on sagittal newborn mouse sections. An overview of the head region of newborn mouse probed with the antisense and sense TGase 5 probes are given in A and B (x10), respectively.

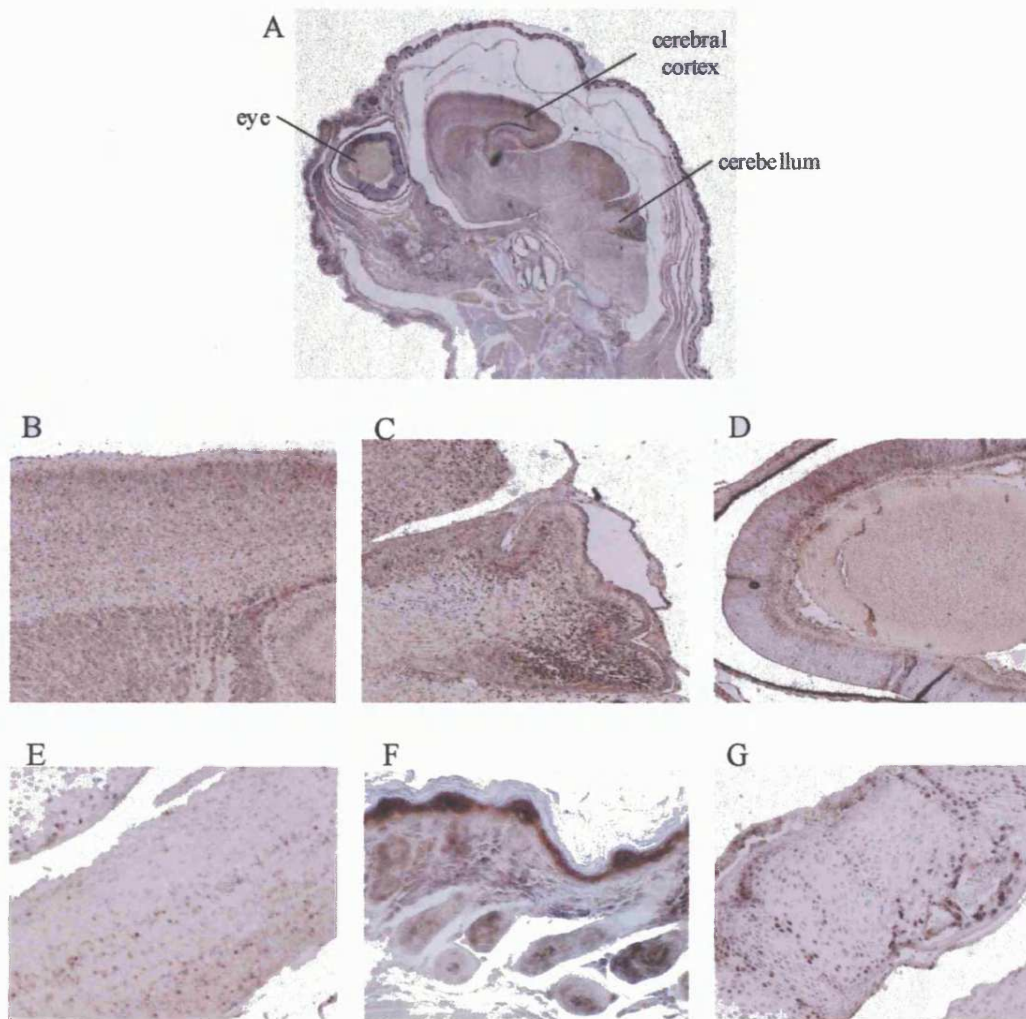


Fig. 3.9 *In situ* hybridisation analysis for TGase 5 on sagittal newborn mouse sections. A shows an overview of the TGase 5 probe binding in the head region (x10), B, weak staining in the cerebral cortex (x100), C, weak staining in the cerebellum (x100), D non-specific staining in the retinal cells of the eye (x100), E, few cells in the spinal cord are stained (x100), F, high levels of staining in the epidermis of the skin, particularly the suprabasal cell layers (x400) and G, staining in a subset of chondrocytes in the vertebrae (x200).

3.3.2.3 TGase 6 expression in mouse embryo development

After identifying that TGase 6 is expressed predominantly in the central nervous system in what appears to be the neuronal cell population, we were interested to see whether changes in TGase 6 expression are correlated with specific events in development. *In situ* hybridisation was therefore performed at different stages of mouse development. Mouse embryos at 11, 13 and 16 days of gestation were



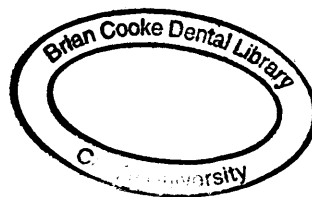
The following text is extremely blurry and illegible. It appears to be a list or a series of short paragraphs, possibly describing the items shown in the images above. The text is arranged in several lines across the width of the page.

The following text is also extremely blurry and illegible. It appears to be a continuation of the text from the previous block, possibly providing more details or a conclusion. The text is arranged in several lines across the width of the page.

obtained. The central nervous system is the first to develop and to differentiate, and one of the last to be completed (Rugh, 1968). For example, the cerebellum is not completely differentiated until some days after birth. Simplistically however, the primary parts of the brain can be identified soon after the neural groove, neural plate and head process stage at 7.5 days, and by 14 days the brain is typically that of a mammal.

3.3.2.3.1 11 day mouse embryo

The 11 day mouse embryo is comparable to the 30 day human embryo. Brain differentiation is continuing with the walls of the entire central nervous system actively proliferating neuroblastic cells, which begin to occlude some of the neural cavities (Rugh, 1968). Up to day 11, the major neuroblastic activity is occurring behind the hindbrain where cranial ganglia V to IX suddenly develop. The olfactory nerve (I) fibres also reach to the telencephalon at this stage (Fig. 3.10). Overall, little TGase 6 expression can be detected in brain at this stage, however, weak staining can be detected in the telencephalon (forebrain), and metencephalon (hindbrain), which may indicate TGase 6 is expressed in the cranial ganglia and olfactory nerve fibres. Extensive labelling was seen in the spinal cord.



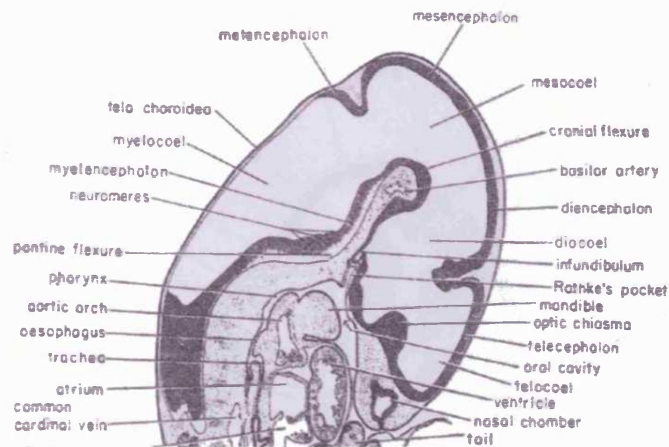
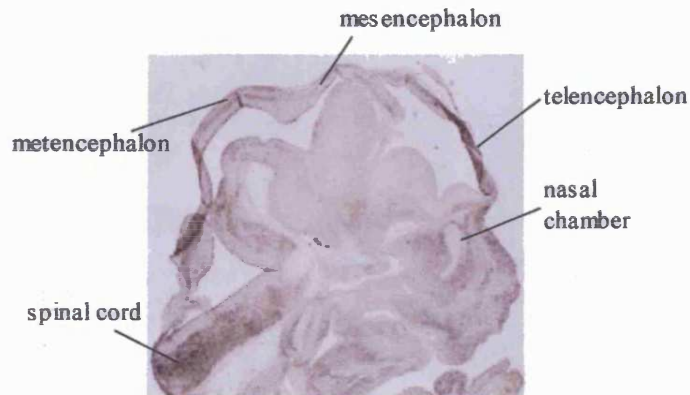


Fig. 3.10 *In situ* hybridisation analysis for TGase 6 on sagittal 11 day old mouse embryo. A shows that little staining for TGase 6 can be detected in the 11 day embryo (x50). B shows a schematic of a mid-sagittal section of a 11 day embryo (Rugh, 1968).

3.3.2.3.2 13 day mouse embryo

The 13 day mouse embryo is comparable to the 38 day human embryo. Brain differentiation is continuing with further thickening of the brain walls particularly in the roof of the mesencephalon, the floor of the myelencephalon and the metencephalon (cerebellum) (Rugh, 1968). By day 13, the expression of TGase 6 has increased in the brain with strong staining now apparent in the mesencephalon (Fig.3.11). It is known that by day 13, neurons are now present indicating that neurodifferentiation has started in the brain. However, staining in the telencephalon (cerebral cortex) remains comparably weak. This is likely to be due to the fact that little neuroblastic activity has occurred in this region at this stage.

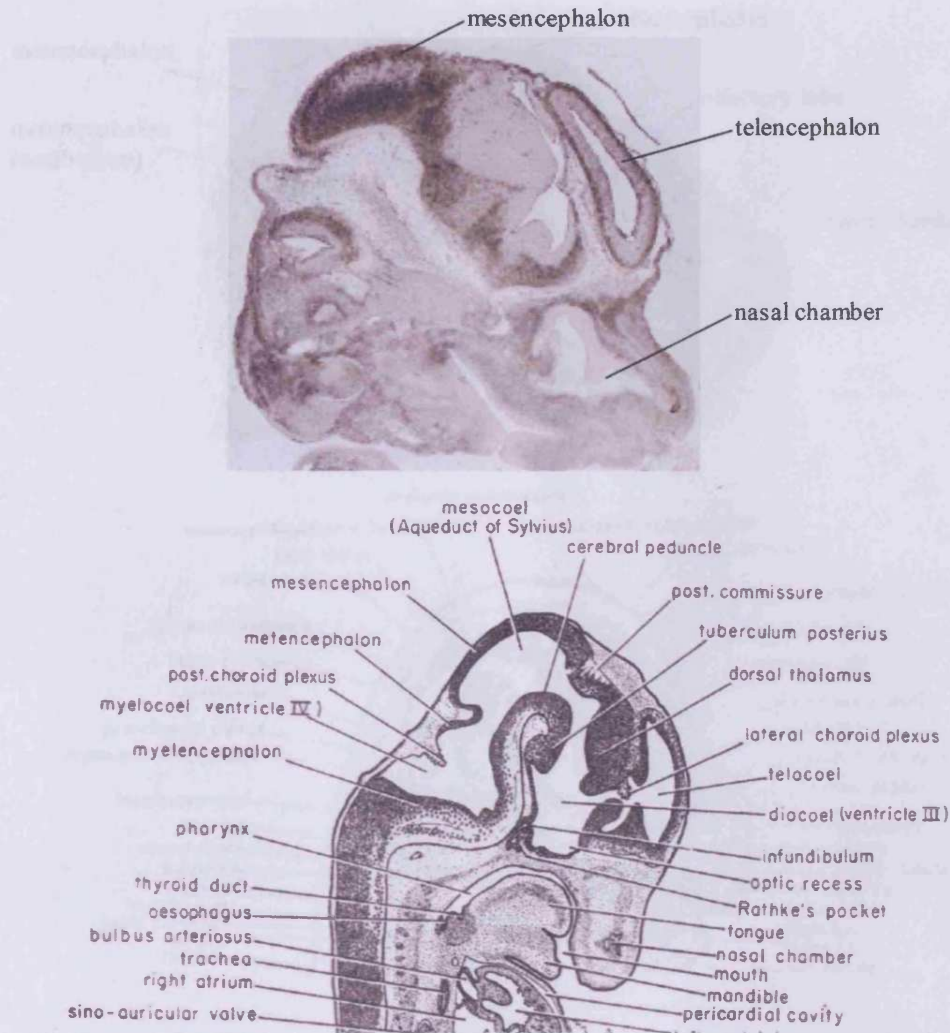


Fig. 3.11 *In situ* hybridisation analysis for TGase 6 on sagittal 13 day old mouse embryo. A shows that TGase 6 staining has increased in the mesencephalon (x50). B shows a schematic of a mid-sagittal section of a 13 day embryo (Rugh, 1968).

3.3.2.3.3 16 day mouse embryo

The 16 day mouse embryo is comparable to the 10.4-week human embryo and the brain, except for the cerebellum, is completing its development. From days 14-16, the major neuroblastic activity occurs in the cerebral cortex (telencephalon) and the vibrissae develop (Rugh, 1968). By day 16, TGase 6 is highly expressed in the cerebral cortex (Fig. 3.12) and the expression pattern is comparable to the expression in the fully developed brain (Fig. 3.6).

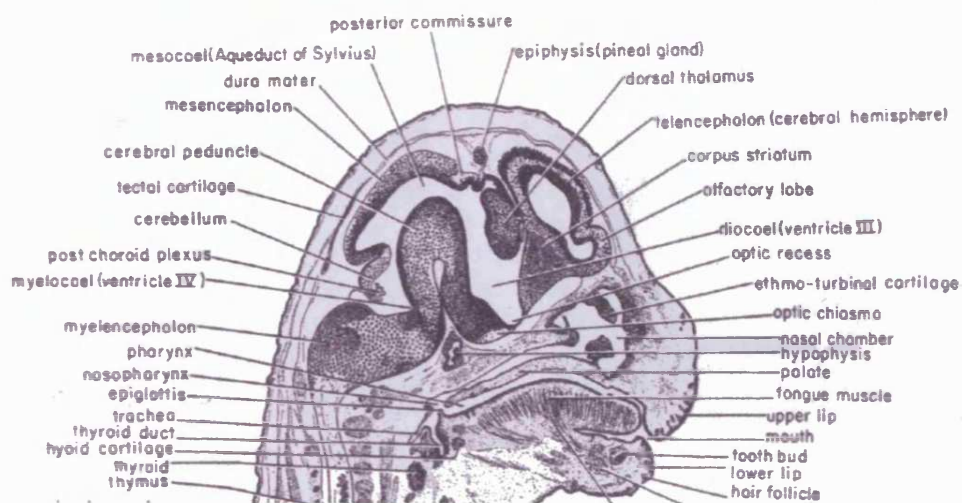
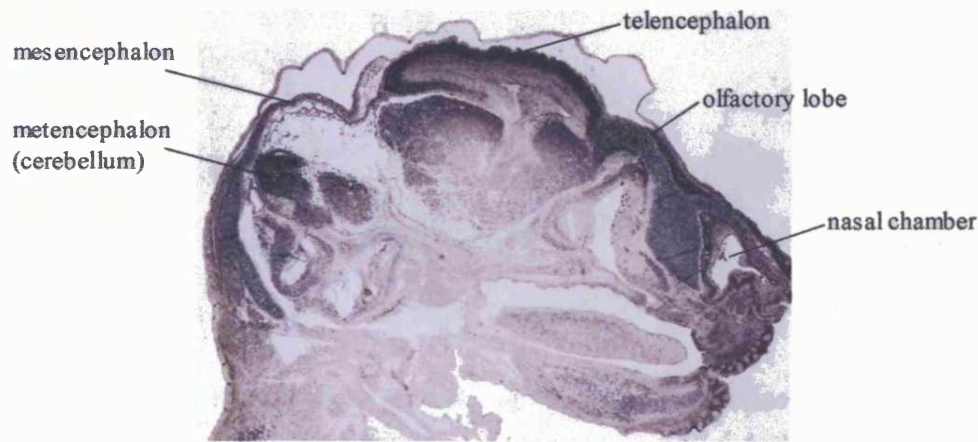


Fig 3.12 In situ hybridisation analysis for TGase 6 on sagittal 16 day mouse embryo. A shows that the embryo is expressing TGase 6 at a high level in the cerebral cortex, mesencephalon and the cerebellum (x50). B shows a schematic of a mid-sagittal section of a 16 day embryo (Rugh, 1968).

Therefore, it can be summarised that the expression of TGase 6 is developmentally regulated and occurs in conjunction with neuronal development. Fig. 3.13 gives an overview of how the expression increases with development of the brain. Fig. 3.14 focuses on the cerebrum and shows how the expression of TGase 6 increases in the cerebral cortex (a region rich in neurons) as the mouse embryo develops.



The first of these is the *Trilobites*, which were the most numerous and diverse of the early life forms. They were small, three-lobed creatures that lived in the oceans. The second is the *Graptolites*, which were small, plant-like organisms that lived in the oceans. The third is the *Crinoids*, which were large, stalked creatures that lived in the oceans.

The *Trilobites* were the most numerous and diverse of the early life forms. They were small, three-lobed creatures that lived in the oceans. The *Graptolites* were small, plant-like organisms that lived in the oceans. The *Crinoids* were large, stalked creatures that lived in the oceans.

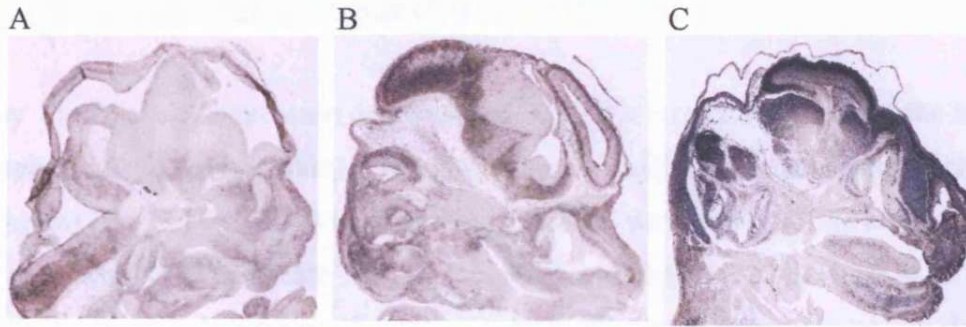


Fig. 3.13 *In situ* hybridisation analysis of TGase 6 expression during mouse brain development. Comparative analysis of expression at days 11 (A), 13 (B) and 16 (C) (x50).

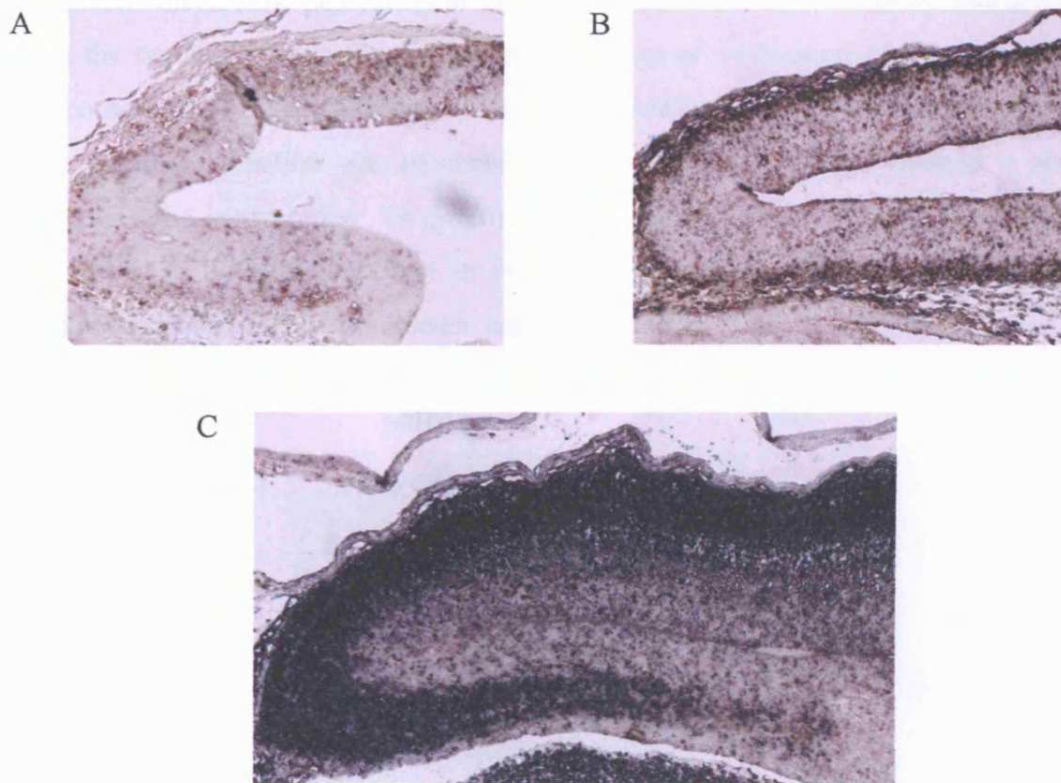


Fig. 3.14 Comparative *in situ* hybridisation analysis on the cerebrum of sagittal mouse embryo sections. A shows TGase 6 expression in the day 11 mouse embryo (x100), B, TGase 6 expression in the day 13 mouse embryo (x100), and C, TGase 6 expression in the 16 day mouse embryo (x100).

3.3.2.4 TGase 5 expression in mouse embryo development

At day 11, TGase 5 expression can be detected in the three main areas of the brain, the telencephalon (forebrain), mesencephalon (midbrain) and metencephalon (hindbrain) (Fig. 3.15). This is in contrast to TGase 6, which shows little expression at this stage. This could indicate that the TGase 5 expression is non-neuronal as little neurodifferentiation has taken place at this stage of embryo development. By day 13 and day 16, the expression of TGase 5 has increased in the brain with staining again visible in most areas of the brain (results not shown). While the expression of TGase 5 and TGase 6 is overlapping, the expression of TGase 5 in the cerebrum appears more widespread whereas the expression of TGase 6 appears to be specific to distinct layers within the cerebral cortex. The apparent high level of expression of TGase 5 (Fig. 3.15) could be an experimental factor where the probe concentration was too high, or the development reaction was incubated for too long as some background is also visible with the sense probe. To determine if the TGase 5 is expressed at a higher level earlier in development than in the mature organ, quantitative PCR analysis would need to be performed on mouse brain tissues.

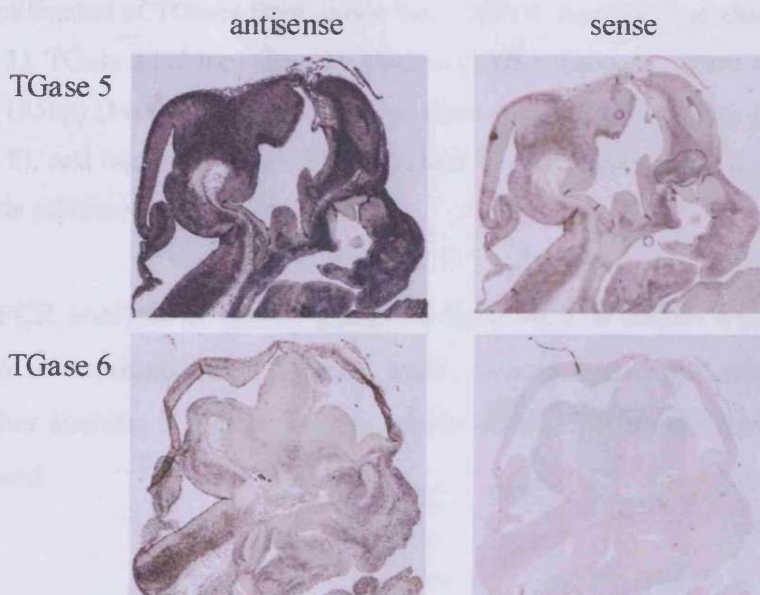


Fig. 3.15 Comparison of *in situ* hybridisation analysis for TGase 5 and TGase 6 on sagittal mouse embryo sections at day 11 (x50).

3.3.3 Amplification of TGases from mouse brain cDNA

To confirm that TGase 5 and TGase 6 are expressed in the brain and to determine which other members of the TGase family are expressed in brain, PCR was performed from mouse brain cDNA using primer sets for each of the 9 different TGases (see Appendix 2, Table 4). These primer sets have been previously verified (Stephens et al., 2004). PCR products of the expected size were amplified for TGase 1, TGase 2, TGase 3, TGase 5, TGase 6 and TGase 7 (Fig. 3.16). No products could be amplified for fXIIIa or band 4.2 protein. This confirms the *in situ* hybridisation analysis with regards to expression of TGase 5 and TGase 6 in the brain. It also confirms previous work by other groups which showed that TGase 1, TGase 2 and TGase 3 are also expressed in the brain.



Fig. 3.16 Amplification of TGases from mouse brain cDNA. Amplified products for TGase 1 (182bp) (lane 1), TGase 2 (177bp) (lane 2), TGase3 (157bp) (lane 3), TGase 4 (252bp) (lane 4), TGase 5 (135bp) (lane 5), TGase 6 (171bp) (lane 6), TGase 7 (126bp) (lane 7), fXIIIa (169bp) (lane 8), and band 4.2 protein (120bp) (lane 9) were analysed by electrophoresis in 1% agarose gels calibrated with the 1kb ladder.

Preliminary PCR analysis of cDNA extracted from various mouse tissues indicated that TGase 6 is expressed in the brain, heart, pancreas and placenta (results not shown). Further analysis is required as the results obtained were not consistent for all primer sets used.

3.4 Discussion

The initial expression pattern of TGase 6 obtained from Northern dot blot analysis indicated that TGase 6 is present in the central nervous system. This is consistent with the fact that TGase 6 was isolated from small cell carcinoma cell line, H69, which has neuronal cell characteristics, such as expression of neuron-specific enolase and brain isozyme of creatine kinase. *In situ* hybridisation revealed that TGase 6 is expressed in the brain, particularly within the cell layers rich in neurons of the cerebral cortex of the cerebrum, and the cerebellum, and that during development the expression coincides with neurogenesis. TGase 6 expression was also detected in the retinal cells of the eye, the epidermis of skin and in the developing long bones. This demonstrates that TGase 6 expression is not exclusively present in neuronal cells. *In situ* hybridisation also revealed that TGase 5 is expressed in the brain, but the staining did not appear to be neuron specific or coincide with neurogenesis during development. Analysis of the TGase expression profile in wild type mouse brain using PCR amplification confirmed the expression of the TGase 6 transcript in the brain. The results also corroborated the data in the literature which suggests that TGase 1, TGase 2, and TGase 3 are expressed in the brain (Kim et al., 1999). Studies determined that the mRNA and protein for TGase 2 is abundantly expressed in the four regions of human brain studied, amygdala (which is part of the limbic system that encircles the upper part of the brain stem and the corpus callosum), cerebellum, corpus callosum (found below the cerebrum), and cerebral cortex. The TGase 3 enzyme was also observed to a minor degree in all the tissues tested, but was found more significantly in the amygdala and a significant amount of TGase 1 mRNA expression was observed in the cerebellum and corpus callosum. Many studies have suggested that one of the processes responsible for the generation of abnormal inclusions in neurodegenerative diseases is TGase crosslinking. TGase catalysed isodipeptide bonds have been identified in abnormal inclusions in PD, AD, PSP and HD. Due to the finding that TGase 6 is expressed in the neurons of the brain, it may be possible that TGase 6 contributes to the generation of abnormal inclusions. It is also interesting that TGase 6 is predominantly expressed in the cerebral cortex and cerebellum, two areas of the brain greatly affected in AD and HD. Studies on AD and HD have mainly focused on TGase 2 enzyme which is abundant in many tissues and is also expressed in neurons of the brain. The expression of TGase 2 is increased significantly in diseased brain.

As TGase 6 is expressed more selectively in the brain, it would be interesting to determine if its expression is specifically increased in association with distinct degenerative conditions in the brain, particularly if the expression increases in specific areas of the brain affected by neurodegenerative diseases.

4.0 Antibody generation and characterisation of protein expression of TGase 6

4.1 Introduction

To extend the mRNA expression data to the protein level we have raised antibodies to the enzyme. Detection of the protein can provide additional information with regard to subcellular localisation and also protein levels present. Furthermore, antibodies can be used for detection of isoform specific distribution in tissue sections, detection of proteins in immunoblots and for purification of isoforms by immunoprecipitation. The aim of this chapter was to generate antibodies for the two isoforms of TGase 6, TGase 6_S (with exon XII alternatively spliced) and TGase 6_L. The antibodies were then used in immunohistochemistry to verify results obtained by *in situ* hybridisation, and for the determination of TGase 6 expression in various areas of mouse brain.

4.2 Materials and Methods

4.2.1 Expression of the β barrel 1 and 2 domain of human TGase 6

4.2.1.1 Isolation of DNA fragment encoding C-terminal β barrel 1 and 2 domain of human TGase 6

cDNA was prepared from poly(A⁺) RNA of H69 cells with the cDNA Cycle Kit as described in section 2.2.5.1. The C-terminal β barrel 1 and 2 domain was amplified by PCR using TGase 6-specific oligonucleotides

5'-GGATCCCCGAGGGCTGGGGGTCGCTGTC (fP8) and

5'-GAATTCAATCCACCCCTCTCAGTCCCT (rP19). fP8 has a BamHI site added and rP19 has an EcoRI site (underlined). PCR was carried out as described in section 2.2.8. Amplified products were analysed on 1% agarose gels, extracted using the QIAquick Gel Extraction Kit (Qiagen) and cloned into the pCRII vector by taking advantage of the 3' A-overhangs generated by the *Taq* DNA polymerase using the Original TA-Cloning Kit (Invitrogen) as described in section 2.2.16.1. Plasmid DNA was prepared with the Wizard Miniprep DNA Purification System (Promega). The DNA was finally eluted from the column matrix using 50 μ l of nuclease-free H₂O. The nature of the product was confirmed by sequencing as described in section 2.2.10.1.

4.2.1.2 Construction of expression vector

An expression vector containing the human C-terminal β barrel domains cDNA of TGase 6 was constructed by inserting the cloned TGase 6 fragment into the vector pGEX2T (Amersham Pharmacia) (see Appendix 15), at the 3' end of the GST sequence (Fig. 4.1) using the engineered EcoRI and BamHI site. To avoid re-ligation of the vector, the linearised vector was dephosphorylated with calf intestinal alkaline phosphatase (CIP) (10 units) (NE Biolabs) for 30 min at 37°C. The insert was ligated into the EcoRI/BamHI site of the vector using 1 unit T4 DNA ligase and the plasmid was transformed into *E. coli* BL21 (Life Technologies) as described in section 2.2.16.1.3. Plasmid DNA was prepared and sequenced as before using primers T7 and M13 (Promega).

PLSDLVPRGSTSNGRQCACIRRWPRAGGRCLWRDDLLEPATKPSIAGKFKVLEPPMLGHDLRLALCLAN
LTSRAQRVRVNLSGATILYTRKPVAEILHESHAVRLG PQEEKRIPITISYSKYKEDLTEDKKILLAAMC
LVTKGEKLLVEKDITLEDFITIKVLGPAMVGVAVTVEVTVVNPLIERVKDCALMVEGSGLLQEQLSIDV
PTLEPQERASVQFDITPSKSGPRQLQVDLVSPHFDPDIKGFVIVHVATAK*

Fig. 4.1 Sequence showing the amino acid sequence of the TGase 6 β -barrel domains ligated at the 3' end of the GST sequence. The GST sequence is highlighted red, and the fragment of the pCRII vector attached to the 5' end of the insert is underlined.

4.2.1.3 Expression of the GST-fusion protein

E. coli BL21 transformed with the pGEX2T expression vector containing the β barrel 1 and 2 domain of human TGase 6 was inoculated into 50ml of Luria-Bertani (LB) broth containing 50 μ g/ml Ampicillin, and grown overnight at 37°C in a shaking incubator at 225rpm. The overnight culture was then expanded in 450ml of 2xYT medium (16M Tryptone, 10M Yeast extract, 5M NaCl, pH 7.0) containing 50 μ g/ml Ampicillin, and allowed to grow at 37°C in a shaking incubator until an absorbance of 0.5 at 600nm was reached (approximately 3 hrs). Expression was then induced by the addition of isopropyl β D-thiogalactoside (IPTG) to a final concentration of 0.1mM. After incubation overnight at 28°C in a shaking incubator, the suspension was centrifuged for 30 min at 3000xg at 4°C to collect the bacteria. To extract the expressed protein, the bacterial pellet was resuspended in 20ml of osmotic shock buffer (20mM Tris/HCl, pH 8.0, 2.5mM EDTA, 5mM DTT, 100mM NEM and 100mM PMSF). 0.1% Sodium deoxycholate was added, the suspension vortexed and incubated for 5 min at RT. Lysozyme was then added to a concentration of 1.5mg/ml to facilitate the breakdown of the bacterial outer coat and the cell suspension was incubated at RT for 5 min to allow complete lysis of the bacteria. 1mg/ml DNase 1 (Amersham Pharmacia) and 15mM MgCl₂ were added and the suspension was incubated for 15 min at 37°C to degrade the genomic DNA. 25% sucrose was added to stabilise the protein, and the suspension was then centrifuged at 10,000xg, for 10 min, at 4°C and the supernatant and pellet stored at -20°C for further analysis.

4.2.1.4 Purification of GST-fusion protein by affinity chromatography

The extract (containing the GST-fusion protein) was applied to glutathione-Sepharose 4B (Amersham Pharmacia), and incubated for 30 min at RT with gentle shaking. The Sepharose beads were collected by centrifugation for 5 min at 500xg. After washing the beads 3 times with osmotic shock buffer by repeated centrifugation and resuspension, the GST-fusion protein was eluted with elution buffer containing an excess of glutathione (100mM Tris/HCl, pH 8.0, 20mM glutathione, 50mM NaCl, 5mM DTT and 0.1% sodium deoxycholate). The eluate was analysed by SDS-PAGE and Western blotting.

4.2.2 SDS-PAGE

Samples for SDS-PAGE analysis were mixed with an equal volume of 2x sample buffer (200mM Tris/HCl, pH 6.8, 4% SDS, 10mM EDTA, 30% glycerol, 0.3% bromophenol blue and 2% mercaptoethanol) and heated to 100°C for 5 min before loading. SDS-PAGE was carried out in 4-20% polyacrylamide gradient gels (Invitrogen) electrophoresed in running buffer (25mM Tris-HCl, pH 8.3, 192mM Glycine, 0.1% SDS) at 125V, 40mAmp for 2 hrs. The gels were stained with 0.1% (w/v) Coomassie brilliant blue solution (7% acetic acid, 50% methanol, 0.1% Coomassie blue R-250). The gels were destained in destain solution I (10% acetic acid, 25% isopropanol), followed by destain solution II (10% acetic acid, 10% isopropanol) and destain solution III (7% acetic acid).

4.2.3 Western blotting

Proteins were separated on SDS-polyacrylamide gels and transferred onto nitrocellulose membranes (Schleicher and Schuell) in a blotting apparatus (Novex) with transfer buffer (25mM Tris, 192mM glycine, 20% methanol), at 125V, 100mAmp for 2 hrs. Membranes were blocked with 5% non-fat dry milk in TBS (50mM Tris-HCl, pH 7.4, 150mM NaCl) to saturate unspecific binding sites on the membrane. Membranes were then incubated with the primary antibody (diluted in 5% milk in TBS) at RT for 1 hr. After washing the membranes 3 times with TBS containing 0.05% Tween-20, the membranes were incubated with the peroxidase-

conjugated secondary antibody (DAKO) (diluted in 5% milk in TBS) for 1 hr at RT. After the membranes had been washed twice with TBS containing 0.05% Tween-20 and twice with distilled H₂O, the immunological reaction was revealed using the enhanced chemiluminescence (ECL) kit (Amersham) according to the manufacturer's protocol.

4.2.4 Membrane stripping

To strip previously bound antibodies from the membrane, the membranes were incubated in 20ml stripping buffer (0.06M Tris/HCl, pH 6.8, 2% SDS, 0.007% 2-mercaptoethanol, and H₂O to a final volume of 20ml) for 30 min at 50°C. The membranes were then rinsed in TBS, and either used immediately, or wrapped in saran wrap and stored at 4°C.

4.2.5 Peptides

Peptides were synthesised, purified by HPLC and characterised by mass spectrometry by Sigma Genosys (see Appendix 16).

4.2.6 Peptide coupling for TGase 6_s

Keyhole Limpet Hemocyanin (Sigma) was reconstituted to 5mg/ml and dialysed into distilled H₂O to remove traces of contaminants. The pH was then adjusted to 7.5 with HCl. Peptides were dissolved in distilled H₂O to a concentration of 12mg/ml. 1ml (5mg) of hemocyanin (carrier protein) solution was added to 3.5mg sulfo-NHS (Pierce) and the solution was mixed. The solution was transferred to a 1.5ml eppendorf tube containing 3.1mg 1-ethyl-3-(3-dimethylaminopropylcarbodiimide) (EDC) (Pierce), mixed and incubated for 5 min at RT. This allows the EDC to generate an active ester at the carboxylate groups on hemocyanin. Either 500µl of 12mg/ml peptide (4µmol) was added or, as a control reaction, 500µl of 2.7mg/ml monodansylcadaverine (4µmol) (Sigma). The reactions were incubated for 4 hrs at RT, whilst stirring and maintaining the pH at 7.5 by addition of HCl. The reactions were then incubated overnight at RT whilst stirring. The conjugate was separated from low molecular weight contaminants on PD-10 columns (containing Sephadex G-

25M) (Amersham Pharmacia) equilibrated in PBS. 0.5ml fractions were collected and analysed for protein by measuring OD at 280nm and OD at 335nm for control-conjugate. The fractions containing the peptide-hemocyanin conjugate were pooled.

4.2.6.1 Peptide coupling for TGase 6_L

Hemocyanin was reconstituted to 5mg/ml in 20mM Hepes, pH 7.4 and dialysed into 20mM Hepes, pH 7.4, to remove traces of contaminants. The peptide was dissolved in distilled H₂O to a concentration of 14mg/ml. 1ml (5mg) of hemocyanin solution was added to 6.1mg sulfo-GMBS (heterobifunctional crosslinking reagent) (Pierce) and incubated for 45 min at 4°C, whilst stirring. This allows the sulfo-GMBS to react with lysine side chains of hemocyanin and thereby add maleimido groups. The conjugate was separated from low molecular weight contaminants on a PD-10 column equilibrated in Hepes. 0.5ml fractions were collected, the protein fraction identified by measuring OD at 280nm and the fractions containing the activated protein were pooled. The pH of the pooled conjugate was adjusted to 6.8 by addition of a defined volume of dilute HCl (determined experimentally). Either 500µl of 14mg/ml peptide (4µmol) was added, or as a control, 500µl of 1.5mg/ml 7-mercapto-4-methyl-coumarin (4µmol) (Pierce). The reactions were incubated overnight at 4°C whilst stirring. The conjugate was separated from low molecular weight contaminants on a PD10 column equilibrated in PBS. 0.5ml fractions were collected and analysed for protein by measuring OD at 280nm for peptide-conjugate and in case of control-conjugate, measuring OD at 358nm. The fractions containing the peptide-hemocyanin conjugate were pooled.

4.2.6.2 Preparation of control conjugate

As a control, the peptide coupling experiment for TGase 6_S and TGase 6_L was repeated as above, however, the peptides were coupled to 5mg/ml BSA (Promega) instead of hemocyanin.

4.2.6.3 Protein concentration determination

Protein concentrations were determined using the BCA Protein Assay Reagent Kit (Pierce). Briefly, 10µl of each standard (0-2,000µg/ml BSA in TBS) or unknown sample were pipetted in replicate into a microtitre plate. 200µl BCA Working Reagent (50 parts BCA Reagent A (sodium carbonate, sodium bicarbonate, bicinchoninic acid and sodium tartrate in 0.1M NaOH) mixed with 1 part BCA Reagent B (4% cupric sulphate)) was added to each well and the plate mixed thoroughly on a plate shaker for 30 sec. The plate was covered and incubated for 30 min at 37°C. The plate was equilibrated to RT and the absorbance measured at 562nm on a plate reader.

4.2.7 Production of antisera

Hemocyanin-peptide conjugate was sent to MicroPharm Ltd for raising of antisera in goats. 1mg peptide conjugate mixed with complete Freund's adjuvant was used for primary immunisation and 0.5mg peptide conjugate mixed with incomplete Freund's adjuvant for re-immunisation. The goats were re-immunised every 4 weeks. A 5ml blood sample was taken pre-immunisation and 6 weeks post primary immunisation. The goats were bled after 10,14 and 15.5 weeks. Each bleed yielded ~300ml of serum. All batches of serum were tested by Western blotting and found to give comparable results.

Antisera for TGase 5 was provided by D. Aeschlimann.

4.2.8 ELISA (Enzyme Linked ImmunoSorbent Assay)

Each well in a 96 well microtitre plate (NUNC, Dynatech) was coated with 100µl 5µg/ml antigen (Hemocyanin-peptide conjugate, Hemocyanin, BSA-peptide conjugate, and BSA) dissolved in TBS (50mM Tris, pH 7.4, 0.15M NaCl) overnight at 4°C. The plate was washed 5 times with TBS/Tween (TBS containing 0.01% Tween 20) and each well was incubated with 200µl 1% BSA/TBS for 2 hrs at RT to block residual protein binding sites on the plastic. A factor 2 dilution series ranging from 1:50 to 1:12800 of antiserum in 1% BSA/TBS was prepared and 100µl of each

was pipetted into duplicate wells and incubated for 1 hr at RT. The plate was washed as above and wells incubated with 100 μ l of peroxidase conjugate anti-goat-IgG (DAKO, diluted 1:10,000 in 1% BSA/TBS) for 1 hr at RT. After washing as above, 100 μ l of substrate solution (20mg 5-amino-2-hydroxy benzoic acid was dissolved in 25ml H₂O whilst heating slightly, the pH was adjusted to 6.0 by addition of 1M NaOH and 9ml of this solution was mixed with 1ml 0.05% H₂O₂) was added to each well and the plate was incubated for 30 min at RT. The reaction was stopped by the addition of 100 μ l 1M NaOH to each well. After 15 min the absorbance was read at 490nm in an ELISA plate reader.

4.2.9 Analysis of mouse brain tissue

Mouse brain (BalbC strain) was dissected and transported on wet ice. Cerebellum, medulla, cerebrum and spinal cord tissue were separated. Each tissue sample was homogenised in 1ml of a solution consisting 500 μ l 8M Urea and 500 μ l 2x sample buffer, using an eppendorf homogeniser and stored at -20°C. 10 μ g (wet tissue weight) of each brain sample was separated by SDS-PAGE for Western blotting. Western blotting was performed as described in section 4.2.3, however the membranes were incubated with TGase 6_S and TGase 6_L sera of a dilution of 1:500 in 5% milk/TBS.

4.2.10 Affinity purification of specific antibodies from TGase 6_S antiserum

ECH Sepharose 4B (Amersham Pharmacia) gel was washed with several volumes of H₂O. 6 μ M EDC and 6 μ M N-hydroxysulfosuccinimide (sulfo-NHS) (Pierce) in 1ml H₂O (freshly prepared) was immediately after dissolution, added to the gel matrix. The reaction was incubated for 5 min at RT and then the gel matrix collected by centrifugation and washed with 1 volume of H₂O by repeated centrifugation and resuspension. 2.06mg TGase 6_S peptide in 1.3ml H₂O, was added to the activated gel matrix and the pH maintained between 4.5 and 6.0 for the first hour. The reaction was then incubated overnight at RT. The gel matrix was washed five times with H₂O as described above and then the remaining active groups were blocked by incubation in an excess of 0.1M ethanolamine for 20 min at RT. The matrix was washed alternatively with low and high pH buffer three times to remove excess ligand: 1

volume 0.1M Na-acetate, pH 4.0; 1 volume 0.1M Tris/HCl, pH 8.3. The gel matrix was finally washed with TBS and the column stored at 4°C. For antibody isolation, 1ml of serum (TGase 6_S) was added and run into to the column matrix and incubated for 30 min. The column was washed three times with 1 volume of TBS. The specific antibodies were eluted in 2 volumes of 3M KSCN (potassium thiocyanate) and the fractions collected. Each fraction was dialysed into TBS immediately. The fractions were analysed for protein by measuring OD 280nm and antibody containing fractions pooled.

4.2.10.1 Affinity purification of specific antibodies from TGase 6_L antiserum

~0.4g thiol Sepharose 4B (Amersham Biosciences) was rehydrated in 25ml H₂O. The gel matrix was collected by centrifugation at 5000xg for 5 min. The gel was resuspended in 1ml H₂O and transferred to a column by pouring in one continuous motion. The remainder of the column was immediately filled with binding buffer (20mM Tris/HCl, pH 7.4, 150mM NaCl, 1mM EDTA). The gel was washed with 60ml H₂O and then finally with 5ml binding buffer. 2mg of TGase 6_L peptide dissolved in binding buffer was added to the gel matrix and incubated with the matrix for 1 hr at RT. The gel matrix was then washed with 10ml TBS, the fractions collected and the contribution of 2-thiopyridine measured at OD343nm. 2-thiopyridine is released when the gel matrix reacts with the peptide and therefore analysis at OD343nm allows confirmation that the peptide is bound to the gel matrix. For antibody isolation, 1ml TGase 6_L serum was loaded onto the column and incubated in the column matrix for 1 hr at 4°C. The gel matrix was washed with 5ml of TBS. The specific antibodies were eluted with 2ml of 3M KSCN and 0.5ml fractions collected. The fractions were analysed for protein by measuring OD280nm and antibody containing fractions pooled.

4.2.11 Immunohistochemistry

5µm sagittal tissue sections of newborn mouse and mouse embryos at gestation days 11, 13 and 16 fixed in paraformaldehyde and embedded in paraffin (see section 3.2.2.3) were prepared. Tissue sections were dewaxed in xylene, and rehydrated in a graded ethanol series of 90%, 70%, 60% and 40%. The sections were then rinsed in TBS

twice. The sections were blocked for endogenous peroxidase in methanol containing 1% H₂O₂ for 20 min twice and then rinsed in TBS. The sections were blocked for 15 min in 1% BSA/TBS and incubated for 1 hr with affinity purified primary antibodies diluted in 1% BSA/TBS (protein precipitates were removed by centrifugation at 10,000xg). The sections were washed in TBS three times followed by incubation for 1 hr with peroxidase conjugated anti-goat-IgG (DAKO) diluted in 1% BSA/TBS. The sections were rinsed in TBS and then incubated in development solution (20mg 3-amino-9-ethyl-carbazole (AEC) tablet (Sigma) was dissolved in 10ml dimethylformamide, mixed with 1x imidazole buffer, pH 7.0 (6mM NaCl, 4mM citric acid, 7mM imidazole), 20µl H₂O₂ and H₂O to a final volume of 100ml) for 30 min. The sections were rinsed in TBS and mounted in a solution of 90% glycerol, 10% Na-PO₄.

4.3 Results

4.3.1 Expression of the β barrel 1 and 2 domain of human TGase 6

We decided to generate a fragment consisting of the last two domains of the protein for use as an antigen to raise polyclonal antibodies. This portion of the protein was chosen as these domains contain the lowest proportion of sequence similarity among TGase gene family members and has no catalytic activity. The protein fragment was expressed as a GST-fusion protein to facilitate purification.

4.3.1.1 Generation of the expression construct for the β barrel 1 and 2 domain of TGase 6

To clone the β barrel domain of TGase 6, PCR was performed to amplify the region required. BamHI and EcoRI restriction sites were added to the 5' end of the TGase 6-specific primers so the domain could be ligated in frame into the desired expression vector. The PCR product was initially ligated into the pCRII vector using TA-Cloning. The insert was sub-cloned from the pCRII vector into the linearised pGEX2T expression vector using EcoRI and BamHI, and transformed into BL21, which is a protease negative strain of *E. coli*. Plasmid DNA was prepared and sequenced. Sequence analysis revealed a mutation within the BamHI in fP8, and as a consequence a 42bp fragment of the pCRII vector was attached to the 5' end of the insert as the next following BamHI site in the multiple cloning site of the vector was used (Fig. 4.1). However, this did not result in a frame-shift and based on its amino acid composition should have no major effect on the resultant protein but constitute an extended linker region between the GST and TGase 6 domains.

4.3.1.2 Expression and purification of the GST-fusion protein

The C-terminal domain of human TGase 6 was expressed as a fusion protein with glutathione S transferase (GST-hTGase 6) in *E. coli* BL21 (Fig.4.2). GST-fusion protein expression was induced by the addition of IPTG (Fig. 4.2, lane 2). The expressed protein was extracted by subjecting the bacteria to osmotic shock, followed

by separation of the soluble and insoluble fractions by centrifugation. The pellet was resuspended in 8M urea for SDS-PAGE analysis. The most prominent band migrated approximately as a 55kDa protein consistent with the expected size of the fusion protein. A significant proportion of the fusion protein was found to be present in a soluble form (Fig. 4.2, lane 3) and thus suitable for purification by affinity chromatography.

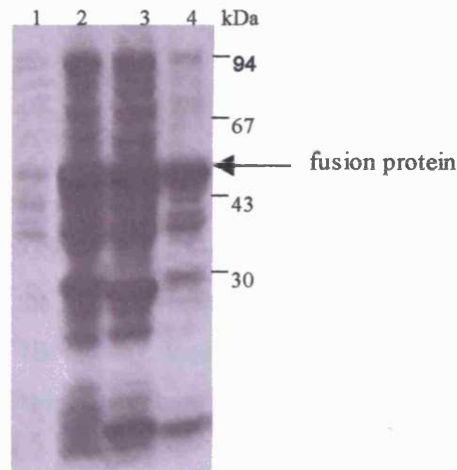


Fig. 4.2 Expression of C-terminal β barrel 1 and 2 domain of hTGase 6 in *E. coli* BL21. SDS-PAGE analysis and Coomassie-blue staining, calibrated with low molecular weight standards (indicated on the right): BL21 of *E. coli* transformed with pGEX2T- β barrel 1 and 2 domain of TGase 6. Lane 1 represents the pellet extracted in 8M urea prior to the addition of IPTG, and lane 2 represents the pellet extracted in 8M urea after incubation with IPTG overnight. The fusion protein was extracted by osmotic shock and was found in the soluble extract (lane 3) and the pellet extract (lane 4).

For purification of the fusion protein, the extract was applied to the glutathione-Sepharose 4B matrix and the GST-fusion protein was eluted from the glutathione-Sepharose 4B column with an elution buffer containing a large excess of free glutathione (20mM). The flow-through and eluate fractions were collected and analysed by SDS-PAGE (Fig. 4.3). The eluate does contain a band of the size predicted to be the molecular mass of the GST-fusion protein (55 kDa), however, the amount eluted was less than expected. A much more prominent band was detected in a sample of the Sepharose gel after elution indicating that a small proportion of the GST-fusion protein had been eluted from the gel matrix.

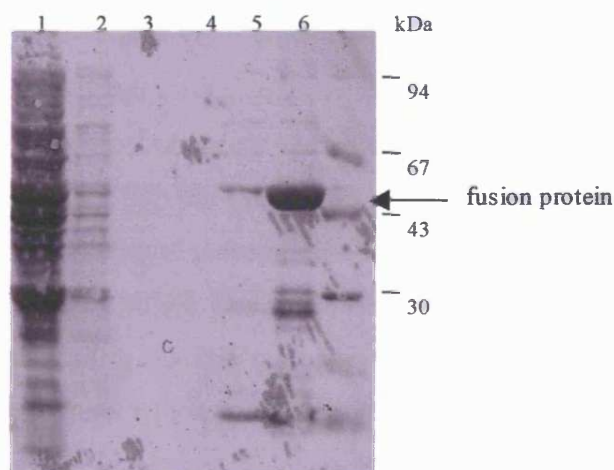


Fig. 4.3 Affinity purification of the GST-fusion protein on glutathione-Sepharose 4B. Fractions were analysed by SDS-PAGE and Coomassie blue staining. Lane 1 represents the soluble extract containing the fusion protein; lanes 2, 3 and 4 represent the flow-through of the glutathione-Sepharose 4B washes; lane 5, the eluted GST-fusion protein; and lane 6, a sample of the glutathione-Sepharose 4B gel boiled in SDS containing sample cocktail.

To improve elution efficiency tests were carried out, varying the concentrations of glutathione (from 20mM to 40mM) and NaCl (from 50mM to 400mM) in the elution buffer. However, the elution of the GST-fusion protein from the gel matrix could only be marginally improved. As it was critical to move this part of the project forward in a timely manner it was decided to pursue this further using a different approach.

4.3.2 Peptide design and synthesis

We decided therefore to synthesise peptides to raise antibodies. For this purpose, antigenic peptides for the two splice variants of TGase 6; TGase 6_s (exon XII spliced) and TGase 6_L (full-length) were designed with the assistance of Dr. K. Beck. The advantage of peptide synthesis is that it avoids folding problems of domain overexpression.

Several factors need to be considered when selecting a peptide for this purpose (Mimotopes, 2001). These include: 1) synthesis efficiency decreases with length; 2) the peptide must be soluble and therefore not too hydrophobic; 3) residues such as Cys, Met and Trp may cause oxidation problems (however, a terminal Cys is good for

coupling to a carrier), and a His residue may cause problems with acylation or enantiomerisation; and 4) clusters of bulky amino acids such as Val, Ile, Leu, Phe and Arg should be avoided. Considerations also need to be made with regards to epitope selection: 1) antibodies require access to the antigenic site and therefore the peptide should be selected from a surface accessible region; 2) an accessible region will be exposed to the environment and therefore the peptide region should be hydrophilic; 3) the peptide should adopt a native-like conformation and as a result the peptide region should be in a β -turn/coil or flexible region as it is difficult to imitate a native-like β -sheet or α -helical structure; 4) the peptide must be immunogenic and a peptide of 10-15 residues in length is ideal; and 5) the epitope should not be post-translationally modified.

To identify a suitable peptide region, the amino acid sequences of TGase 6_S and TGase 6_L were analysed using the ProtScale computer programme which can predict, on a probability scale, the structure of a selected region with regards to flexibility, surface accessibility, hydrophilicity, turns, β -strands and α -helices. A possible peptide region without clear secondary structure is identified when all of these parameters have a probability of greater than 0.5.

To confirm that the regions identified through the use of ProtScale are suitable peptide regions, the amino acid sequence of mouse TGase 6_S and TGase 6_L were aligned using the CLUSTALW computer programme (Higgins et al., 1994) in comparison to the sequences of those TGases for which the structure has been solved. Mouse TGase 6 showed the highest similarity to the activated form of TGase 3. Based on the conservation of residues between TGase 6 and TGase 3, the three-dimensional structures of TGase 6_S and TGase 6_L were predicted. This enabled the peptide regions proposed by ProtScale to be identified in the likely three-dimensional configuration. Based on this, it was possible to select peptides in a surface accessible region. The peptides chosen were located in the flexible hinge region between the catalytic domain and the β -barrel 1 domain. The peptides selected were TGase 6_S: TIRAYPGASGEGLS (Fig. 4.5); and TGase 6_L: CGWRDDLLEPVTKPS (Fig. 4.4).

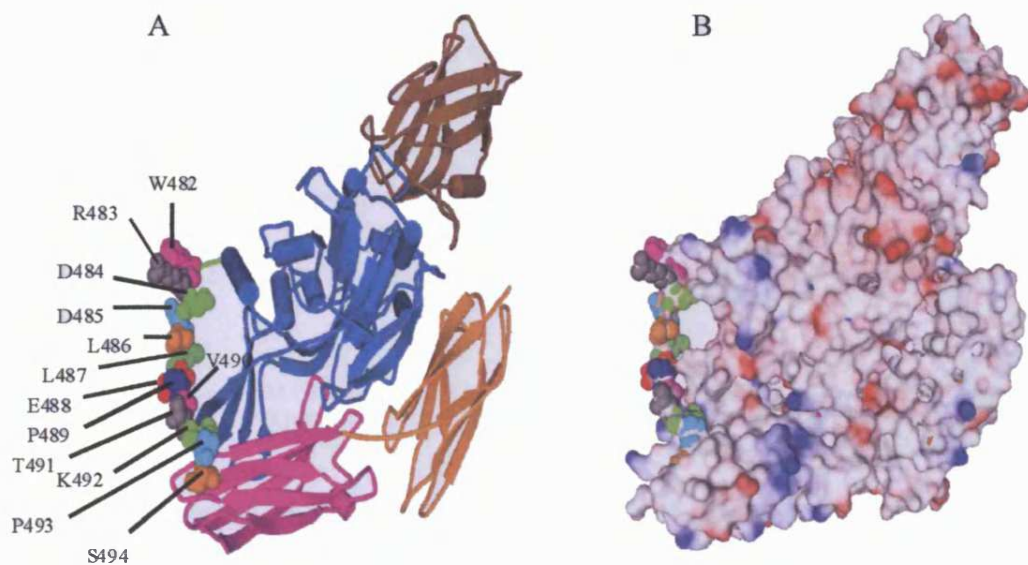


Fig. 4.4 Ribbon image of the predicted structure of TGase 6. The four domains are the N-terminal β -sandwich (brown), the catalytic core (blue), the β -barrel 1 (purple) and β -barrel 2 (red). Residues in the TGase 6_L peptide are located in connecting loop between the catalytic core and β -barrel 1 and are each coloured and shown in CPK style. B. The Connolly surface is shown for the TGase 6 structure, colour coded corresponding to the electrostatic potential (blue: positive charges, red: negative charges). This shows that the connecting loop incorporating the residues used for peptide synthesis are surface accessible.

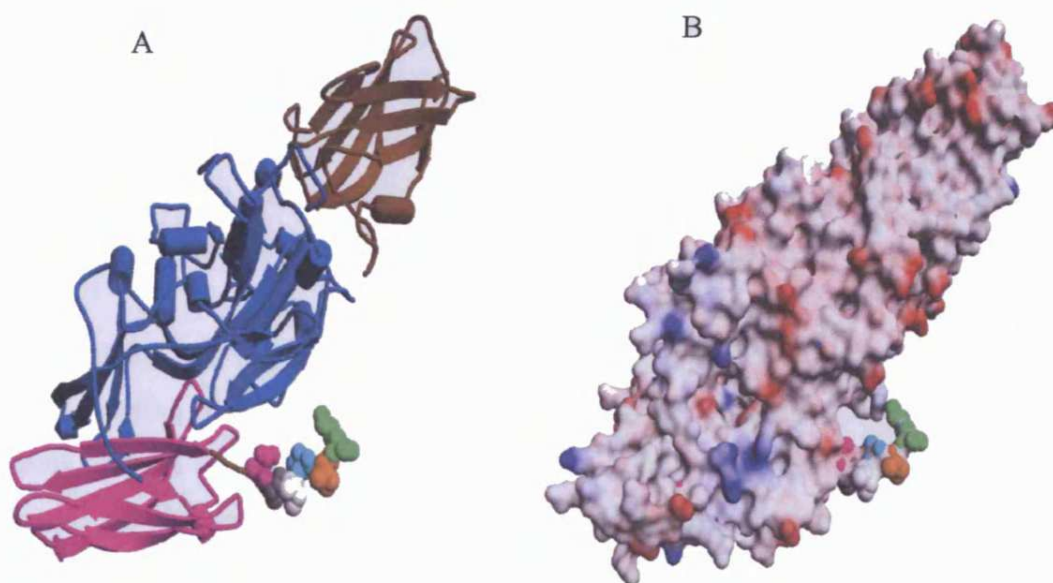


Fig. 4.5 Ribbon image of the predicted structure of TGase 6_S. The three domains are the N-terminal β -sandwich (brown), the catalytic core (blue), and the β -barrel 1 (purple). Residues



(18) (19) (20) (21) (22) (23) (24) (25) (26) (27) (28) (29) (30) (31) (32) (33) (34) (35) (36) (37) (38) (39) (40) (41) (42) (43) (44) (45) (46) (47) (48) (49) (50) (51) (52) (53) (54) (55) (56) (57) (58) (59) (60) (61) (62) (63) (64) (65) (66) (67) (68) (69) (70) (71) (72) (73) (74) (75) (76) (77) (78) (79) (80) (81) (82) (83) (84) (85) (86) (87) (88) (89) (90) (91) (92) (93) (94) (95) (96) (97) (98) (99) (100)



(101) (102) (103) (104) (105) (106) (107) (108) (109) (110) (111) (112) (113) (114) (115) (116) (117) (118) (119) (120) (121) (122) (123) (124) (125) (126) (127) (128) (129) (130) (131) (132) (133) (134) (135) (136) (137) (138) (139) (140) (141) (142) (143) (144) (145) (146) (147) (148) (149) (150) (151) (152) (153) (154) (155) (156) (157) (158) (159) (160) (161) (162) (163) (164) (165) (166) (167) (168) (169) (170) (171) (172) (173) (174) (175) (176) (177) (178) (179) (180) (181) (182) (183) (184) (185) (186) (187) (188) (189) (190) (191) (192) (193) (194) (195) (196) (197) (198) (199) (200)

in the TGase 6_s peptide are located at the C-terminal which differs from the β -barrel 2 sequence of full-length TGase 6 and are each coloured and shown in CPK style. B. The Connolly surface is shown for the TGase 6_s structure, colour coded corresponding to the electrostatic potential (blue: positive charges, red: negative charges). This shows that the residues used for peptide synthesis are surface accessible.

| | | |
|------------------------|-----|--|
| human TG6 _L | 462 | FGVEASGRRIWIRRAGGRCLWRDLDLEPATKPSIAGKFKVLEPPMLGHDLRLALCLA |
| mouse TG6 _L | 462 | LSVEAWGRRRRIRRASVGRVWRDLDLEPVTKPSITGKFKVLEPPVLGQDLKLALCLT |
| | | CGWRDLDLEPVTKPS |
| | | |
| human TG6 _s | 566 | YSKYKEDLTEDKKILLAAMCLVTKGEKLLVEKDITLEDFFITIKRAYPGASGEGLSPV |
| mouse TG6 _s | 566 | YSQYKGDLTEDKKILLAAMCLVSKGEKLLVEKDITLEDFFITIKRAYPGASGEGLSPV |
| | | TIRAYPGASGEGLSP |

Fig. 4.6 Comparison of peptides (shown in blue) with native human and mouse TGase 6 sequences. The C-terminal residue in the peptides was the carboxamide. The cysteine residue added to the TGase 6_L peptide is highlighted in red.

4.3.3 Peptide coupling

To ensure the immunogenicity of peptides, it is required for peptides to be coupled to a carrier protein. The main requirement for the carrier protein is that it must be antigenic in the species being used to raise antibodies (Mimotopes, 2001). Keyhole limpet hemocyanin is a suitable carrier protein as it is derived from a species phylogenetically distant from those used to raise antibodies. BSA is also a commonly used carrier protein. However, it must be taken into consideration that as well as raising antibodies to the peptides, antibodies will also be produced to the carrier protein. This, in general, should not cause a problem as these anti-carrier antibodies should not have any specificity for the 'target' proteins. Short peptides are typically coupled to a carrier protein to obtain the greatest antibody response.

A protocol for two-step coupling using EDC and sulfo-NHS was followed for the peptide coupling of TGase 6_s. This procedure allows the sequential coupling of two proteins without exposing the second protein to EDC and consequently modifying carboxyls on the second protein. EDC is able to generate an active ester at the carboxylate of a protein in the presence of sulfo-NHS forming an activated protein intermediate. The activated protein then reacts with the amine of the peptide to form the peptide-carrier conjugate.

Using this method, the TGase 6_s peptide was coupled to keyhole limpet hemocyanin. The peptide conjugates were separated from low molecular weight contaminants by molecular sieve chromatography using PD-10 columns. The fractions were then analysed by measuring OD 280nm (Fig. 4.7).

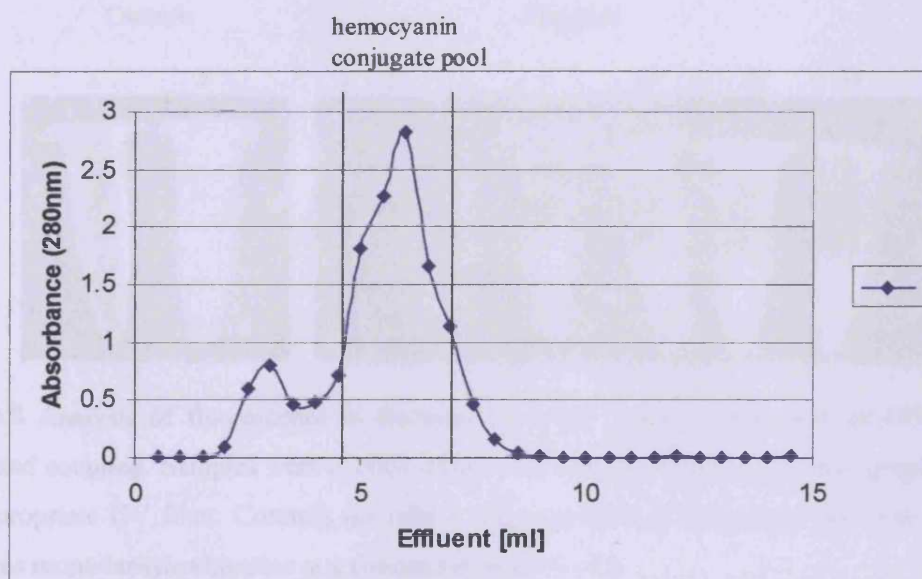


Fig. 4.7 Analysis of PD-10 fractions of TGase 6_s peptide-hemocyanin conjugate reaction. The reaction was applied to the column and eluted with PBS. 0.5ml fractions were collected and analysed for OD at 280nm. The vertical black lines indicate the fractions which contain the peptide-conjugate pool.

For the TGase 6_s coupling experiment, a conjugate was made using monodansylcadaverine instead of the peptide. This was a control experiment used to determine whether the coupling reaction was successful. The amine of monodansylcadaverine reacts with the active ester of the sulfo-NHS-hemocyanin complex forming a control conjugate. As this amine contains a fluorescent tag, it can be detected when coupled to the carrier protein.

Unfortunately, the dansyl signal at 335nm (excitation maxima) was too weak to be separated conclusively from nearby absorbance peaks. Therefore the fractions were analysed qualitatively on a UV transilluminator as no fluorimeter was available. From Fig. 4.8 it is possible to see that the fraction 1 does not fluoresce under UV light whereas by fraction 9, fluorescence is clearly visible indicating that fractions

constituting the protein peak contain the label and hence the coupling experiment has worked. No label is associated with the protein peak in the absence of conjugation reagents. Based on this analysis, fractions 9, 10, 11 and 12 were expected to contain the peptide-conjugate and were pooled for further use.

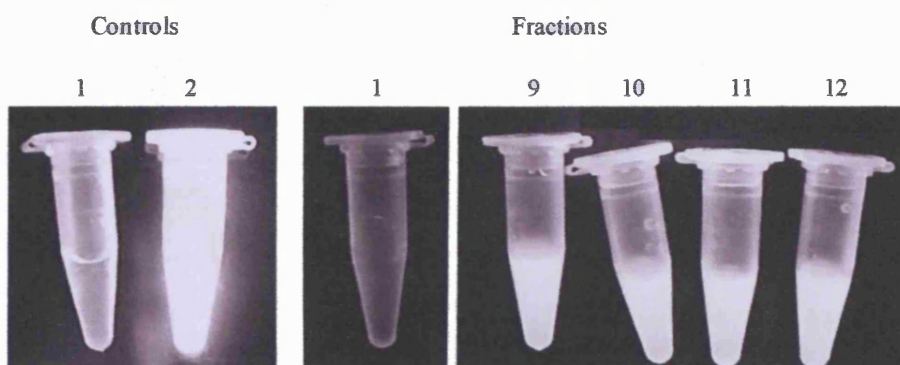


Fig. 4.8 Analysis of fluorescence in fractions from the control experiment for EDC/NHS-mediated coupling. Samples were excited using a UV transilluminator and photographed with an appropriate UV filter. Controls are tube 1 which contains hemocyanin, and tube 2 which contains monodansylcadaverine at a concentration of $4\mu\text{mol}$.

For TGase 6_L, the coupling procedure was performed using a heterobifunctional crosslinker (sulfo-GMBS). This is one of the most widely used methods to couple a peptide to a carrier protein via a specific amino acid residue in the peptide and to conserve peptide antigenicity (Lee et al., 1985). Amino groups (lysine side chains) on the surface of the carrier protein are activated by reaction with sulfo-GMBS which adds maleimido groups to the lysine side chains. When the sulfo-GMBS-activated carrier protein is mixed with the peptide containing reduced sulfhydryl groups (such as cysteine side chains), the maleimido groups react with the sulfhydryl groups to form stable thioether bonds linking the peptide to the carrier protein. In the case of TGase 6_L, the peptide did not contain a 'naturally' occurring cysteine, and therefore a cysteine residue was added to the N-terminus of the peptide sequence. This is in accordance with the recommendation that if the peptide is found in the C-terminus of the protein, the cysteine should be at the N-terminus of the peptide.

Using this method, the TGase 6_L peptide was coupled to keyhole limpet hemocyanin. The peptide conjugates were separated from low molecular weight contaminants by

... ..

... ..

... ..

... ..



... ..

molecular sieve chromatography using PD-10 columns. The fractions were then analysed by measuring OD at 280nm and the absorbance profile is shown in Fig. 4.10.

For the TGase 6_L coupling experiment, hemocyanin was also coupled to mercapto-4-methyl-coumarin instead of the protein as a control. In this case, the coumarin component produces a strong absorbance at 358nm.

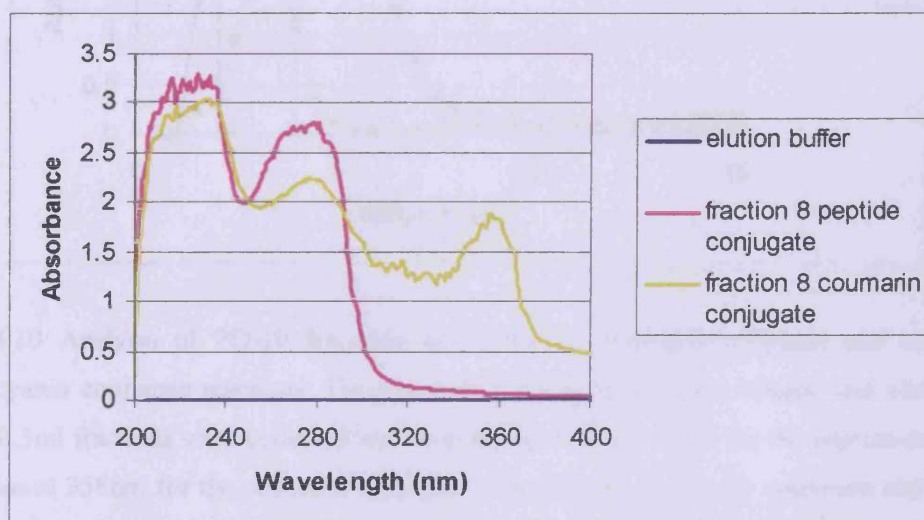


Fig. 4.9 Analysis of PD-10 fractions of TGase 6_L. Elution buffer (fraction 1), peptide-conjugate (fraction 8) for TGase 6_L coupled to hemocyanin, and fraction 8 of the control experiment were analysed by a wavelength scan ranging from OD 200nm to 400nm.

In the absorbance wavelength scan (Fig. 4.9), a peak can be seen at 358nm in the PD-10 fractions containing the hemocyanin indicating that the label (mercapto-4-methyl-coumarin) has coupled to the carrier protein. From analysis of the fractions for the peptide-conjugate and the control conjugate (Fig. 4.10), we decided to pool fractions 4-8 for use as an antigen.

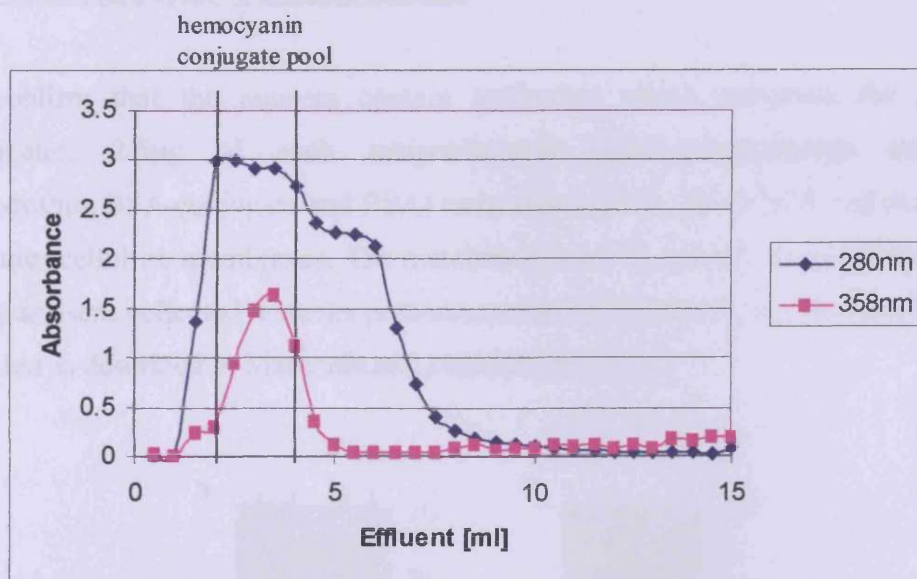


Fig. 4.10 Analysis of PD-10 fractions of TGase 6_L peptide-hemocyanin and coumarin-hemocyanin conjugate reactions. The reactions were applied to the column and eluted with PBS. 0.5ml fractions were collected and analysed for OD at 280nm for the peptide-conjugate and also at 358nm for the coumarin conjugate. The OD280 profile for coumarin and peptide conjugate was similar and therefore only one is shown. The vertical black lines indicate the fractions which contain the peptide-conjugate.

For subsequent analysis of the sera, the peptide coupling experiments were repeated but the peptides were coupled to BSA instead of hemocyanin. The fractions collected from the PD-10 column were analysed at OD280nm and based on this, fractions 7-9 were pooled for each peptide-conjugate (results not shown).

4.3.4 Production of antisera

Hemocyanin-peptide conjugate was sent to MicroPharm Ltd. for the production of antisera. The protein concentration of the peptide conjugates was determined by a BCA assay. Goats were immunised with 1mg of peptide-conjugate (mixed with Freund's complete adjuvant) and then re-immunised with 0.5mg peptide-conjugate (mixed with Freund's incomplete adjuvant) every 4 weeks over a 12 week period. Small blood samples were taken pre-immunisation and 6 weeks post-immunisation. A bleed (300ml) was taken after 10, 14 and 15.5 weeks. Blood was left to coagulate and serum harvested by centrifugation.

4.3.5 Characterisation of TGase 6 antisera

To confirm that the antisera contain antibodies which recognise the peptide-conjugates, 2.5 μ g of each antigen/control (hemocyanin-peptide conjugate, hemocyanin, BSA-conjugate and BSA) were separated by SDS-PAGE and transferred onto nitrocellulose membranes. The membranes were incubated with a 1:200 dilution of the antisera collected 6 weeks post-immunisation (sample 2) and antibody binding revealed as described in Materials and Methods (Fig. 4.11).

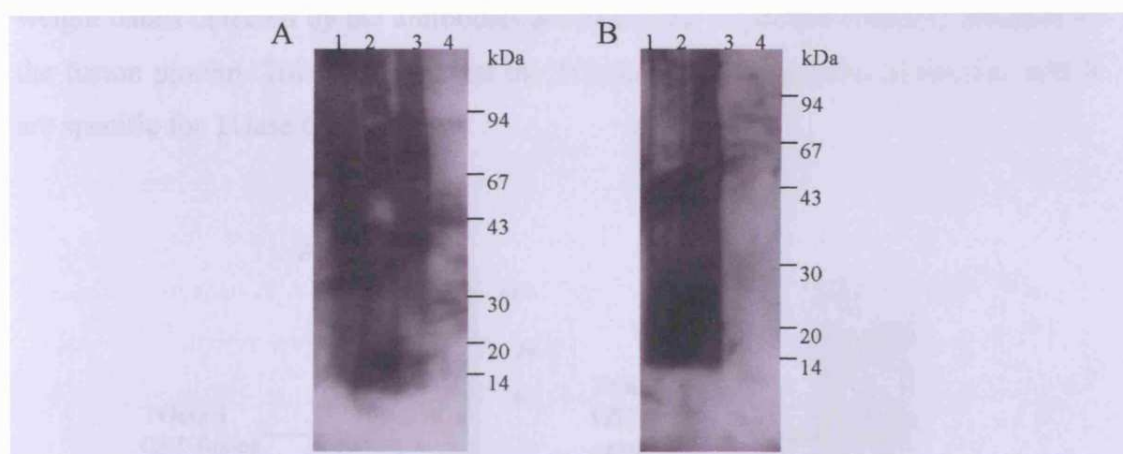


Fig. 4.11 Characterisation of sera against peptide conjugates. Panel A shows an immunoblot performed with TGase 6_s sera on TGase 6_s peptide-hemocyanin (lane 1), hemocyanin (lane 2), TGase 6_s peptide-BSA (lane 3), and BSA (lane 4). Panel B shows an immunoblot performed with TGase 6_L sera on TGase 6_L peptide-hemocyanin (lane 1), hemocyanin (lane 2), TGase 6_L peptide-BSA (lane 3), and BSA (lane 4). 2.5 μ g of each antigen/control was separated by SDS-PAGE and transferred to nitrocellulose membrane. The membranes were calibrated with low molecular weight standards

While this gave promising results for the TGase 6_s antiserum, as the antiserum detected the peptide-BSA conjugate, the results were inconclusive for the TGase 6_L antiserum. However, this does not give a clear indication of whether the sera contains antibodies specific to the peptide alone or only the peptide in its conjugated form. To determine that the TGase 6_L antiserum contains antibodies which recognise TGase 6, 10 μ g of TGase 6 GST-fusion protein containing extract and TGase 7 GST-fusion protein containing extract (provided by S. Rosser-Davies) were analysed by SDS-PAGE and Western blotting as above.

For TGase 6_L, a prominent band was detected at 55kDa in the lane containing the TGase 6 GST-fusion protein extract (Fig. 4.12). This band corresponds to the expected size of the GST-fusion protein. A band of this size was not detected in the TGase 7 fusion protein extract indicating that the TGase 6_L serum is specific for TGase 6.

To confirm that the band detected at 55kDa is the GST-fusion protein, the membrane was stripped and reprobed with an anti-GST primary antibody. The anti-GST antibody detected the same band at 55kDa in the TGase 6_L serum. The low molecular weight bands detected by the antibodies are likely the respective cleavage products of the fusion protein. This indicates that the TGase 6_L serum contains antibodies which are specific for TGase 6.

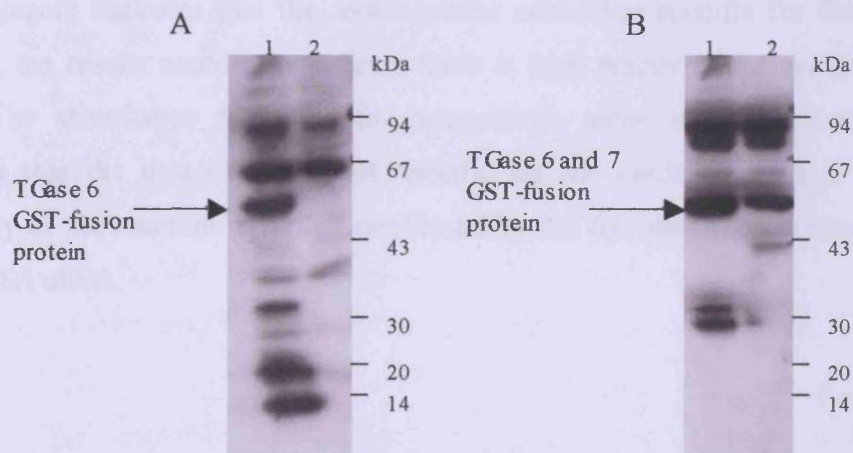


Fig. 4.12 Western blot analysis of the TGase 6_L serum with GST-fusion proteins for TGase 6 and TGase 7. Lane 1 represents 10 μ g TGase 6 GST-fusion protein extract, and lane 2, 10 μ g TGase 7 GST-fusion protein extract separated by SDS-PAGE and transferred to nitrocellulose membrane. The membrane was calibrated with low molecular weight standards. A shows an immunoblot performed with TGase 6_L sera and B, an immunoblot performed with an anti-GST antibody.

The TGase 6_S antisera could not be analysed in a similar way. The short form of TGase 6 has a frameshift resulting in termination of the protein just after the first C-terminal β -barrel domain (with the altered sequence serving as the antigen) and as the

GST-fusion protein is generated using the two C-terminal β -barrel domains the TGase 6_s antibodies will not recognise it.

To characterise the TGase 6_s serum and to further characterise the TGase 6_L serum, pre-immunisation (sample 1) and 6 weeks post-immunisation (sample 2) sera samples were used to perform an ELISA. Microtitre plates were coated with the respective antigen (hemocyanin-peptide conjugate, hemocyanin, BSA-peptide conjugate and BSA) in the appropriate wells. A dilution series of the sera samples ranging from 1:50 to 1:12800 was prepared and analysed (Fig. 4.13 and Fig. 4.14).

For TGase 6_s, the ELISA results show that there is a high titre of antibodies against the peptide as the absorbance measured is highest for the hemocyanin-peptide conjugate and the BSA-peptide conjugate. The strong reactivity of the serum with the BSA conjugate indicates that the sera contains antibodies specific for the peptide. However, the results could also indicate there is high reactivity against the linkage region. The absorbance measured for hemocyanin alone is low in comparison, indicating that the titre of antibodies specific for the carrier protein is minimal. Specificity of the reaction is further confirmed by the fact that there is no reactivity against BSA alone.

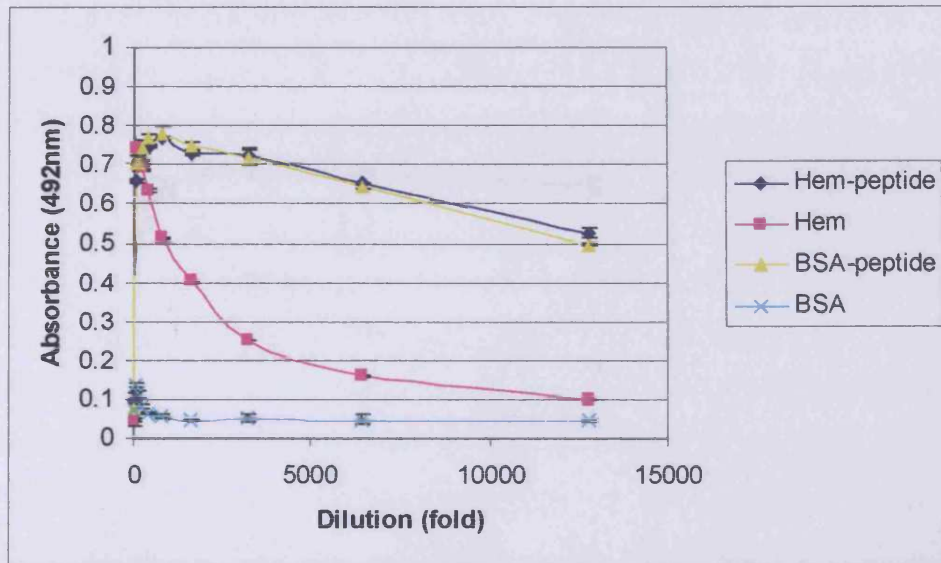


Fig. 4.13 ELISA analysis to determine specificity and titre of antibodies in TGase 6_s serum. ELISA plates were coated with 5 µg/ml antigen (hemocyanin-peptide conjugate, hemocyanin, BSA-peptide conjugate, and BSA), incubated with the antiserum diluted 1:50 to 1:12800 in 1% BSA/TBS, and antibody binding revealed with peroxidase conjugated secondary antibodies and colour development using 5-amino-2-hydroxy benzoic acid as a substrate. The results are given as the mean +/- SD.

For TGase 6_L, the ELISA results show that the absorbance measured is highest for the hemocyanin-peptide conjugate and hemocyanin alone. This indicates that the serum contains a high titre of antibodies specific to the carrier protein, hemocyanin. The absorbance measured for the BSA-peptide conjugate is less but as there was no reactivity with BSA alone, indicates the presence of antibodies specific for the peptide.

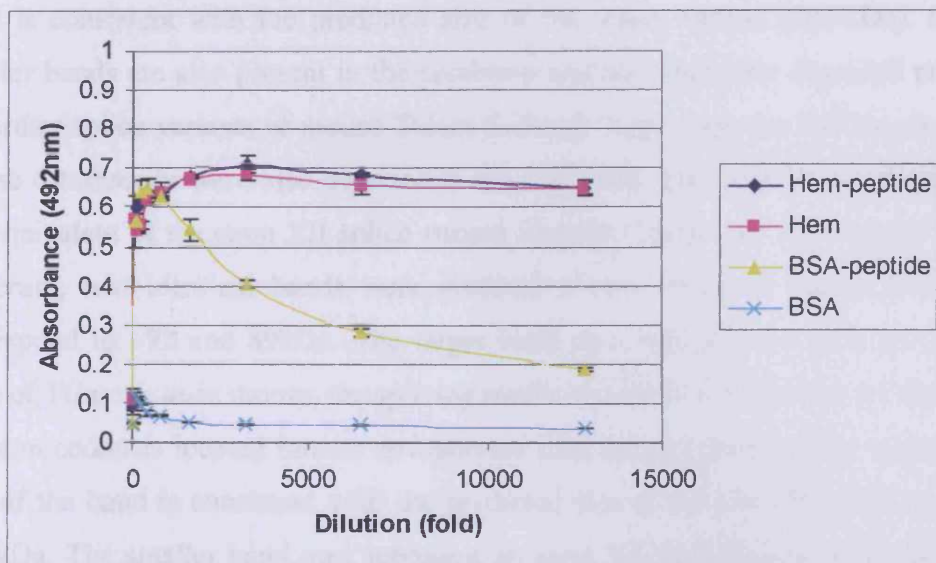


Fig. 4.14 ELISA analysis to determine specificity and titre of antibodies in TGase 6_L serum. ELISA plates were coated with 5 µg/ml antigen (hemocyanin-peptide conjugate, hemocyanin, BSA-peptide conjugate, and BSA), incubated with the antiserum diluted 1:50 to 1:12800 in 1% BSA/TBS, and antibody binding revealed with peroxidase conjugated secondary antibodies and colour development using 5-amino-2-hydroxy benzoic acid as a substrate. The results are given as the mean +/- SD.

In summary, the experiments have determined that the antisera contain antibodies specific for the peptides and can be used to analyse the expression of TGase 6_S and TGase 6_L.

4.3.6 Determination of TGase 6 expression in mouse brain

Cerebellum, medulla, cerebrum and spinal cord tissue were dissected from the brain of an adult balb-c mouse and each tissue homogenised in a Tris-buffered solution containing 8M Urea and 1% SDS. 10 µg of each tissue extract was separated by SDS-PAGE and transferred onto nitrocellulose membranes. The membranes were incubated with a 1:200 dilution of TGase 6_S or TGase 6_L sera (Fig. 4.15). With the TGase 6_L serum, a band of ~79kDa was detected in the tissue from spinal cord, cerebrum and medulla. This corresponds to the predicted size of the TGase 6_L transcript. A slightly smaller band of ~72kDa was detected in all tissues. This may correspond to an exon III splice variant of TGase 6 found in mouse as the size of the

band is consistent with the predicted size of the splice variant (70.9kDa). Several smaller bands are also present in the cerebrum and may represent degraded products or further splice variants of mouse TGase 6. Bands larger than the full-length mouse TGase 6 transcript were also detected in the cerebrum. These bands are likely to be the equivalent of the exon XII splice variant identified in human. As with the TGase 6_S serum, two identical bands were detected in the cerebrum tissue. The bands correspond to ~92 and 89kDa. The larger band may represent an exon XII spliced form of TGase 6, as in mouse, the splicing results in a frameshift similar to human but the stop codon is located further downstream than in the human splice variant. The size of the band is consistent with the predicted size of the exon XII splice variant, 92.7kDa. The smaller band may represent an exon XII spliced form with additional splicing, for example, an exon III/exon XII splice variant. This finding may indicate that two specific alternatively spliced TGase 6 isoforms are present only in this area of the brain and may suggest that these two isoforms have a specific role in the cerebrum. However, it must also be taken into consideration that the antiserum has not been affinity purified and some of the bands detected may represent unspecific binding. To improve the specificity of the experiment, it should be repeated using affinity purified sera.

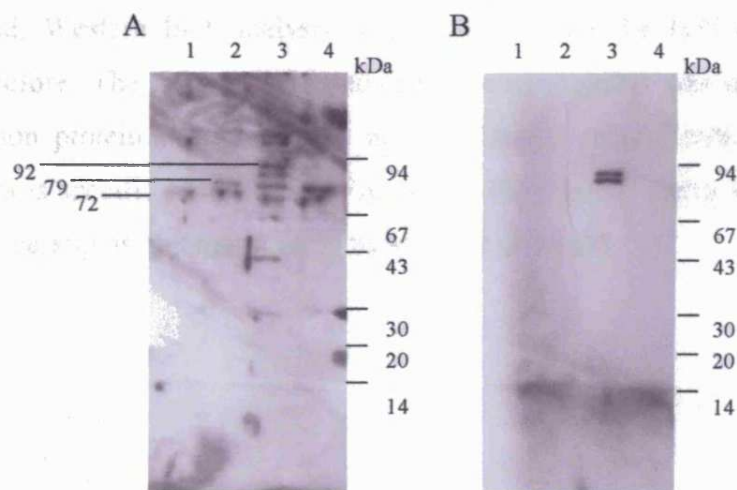


Fig. 4.15 Detection of TGase 6 in extracts of mouse brain tissue. Immunoblots performed with TGase 6_L (A) and TGase 6_S (B) sera on 10 μ g tissue extracted from the cerebellum (lane 1), medulla (lane 2), cerebrum (lane 3), and spinal cord (lane 4) separated by SDS-PAGE and transferred to nitrocellulose membrane. Mr standards are indicated on the right. The molecular weights of the TGase 6 variants present were estimated from Rf values.

4.3.7 Antibody purification

The next stage was to purify the antisera in order to isolate the antibodies specific for TGase 6_S and TGase 6_L and this was carried out by Sally Rosser-Davies.

For TGase 6_S, ECH Sepharose 4B gel matrix was incubated with EDC and sulfo-NHS to activate the matrix. The peptide was added, reacted overnight and the matrix subsequently blocked with 0.1M ethanolamine. The matrix was equilibrated in TBS and serum (TGase 6_S) was incubated with the column matrix for 30 min at 4°C. After washing, bound antisera was eluted with 3M KSCN and dialysed into TBS. From 1ml of serum approximately 1ml of specific antibodies at a concentration of 860µg/ml could be purified.

For TGase 6_L, thiol Sepharose 4B was reacted with TGase 6_L peptide. Antibodies to TGase 6_L peptide were purified as described for TGase 6_S. From 1ml of serum approximately 1ml of specific antibodies at a concentration of 960µg/ml could be purified.

To determine if the purification was successful and that the specificity of the antisera had increased, Western blot analysis was performed on the GST-fusion protein extracts as before. The TGase 6 GST-fusion protein (55kDa) was detected in the TGase 6 fusion protein extract but not in the TGase 7 extract indicating that the TGase 6_L sera is specific for TGase 6 (Fig. 4.16). Purification of the sera appears to have been successful as no other protein bands were detected.

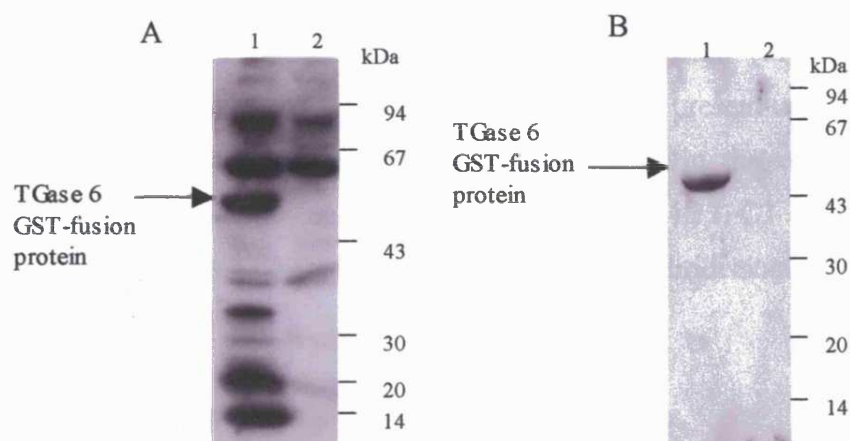


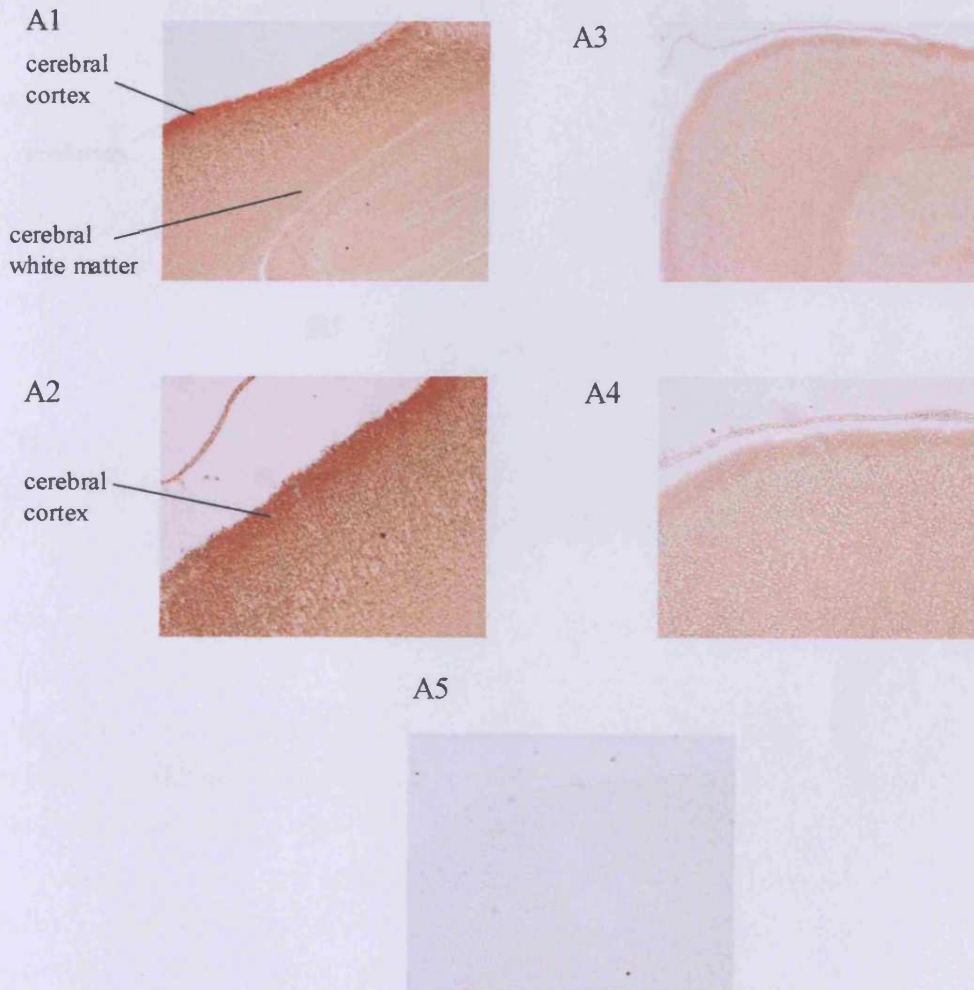
Fig. 4.16 Analysis of affinity purified TGase 6_L antibodies against GST-fusion protein extracts. 10µg of bacterial extract containing fusion protein was separated by SDS-PAGE and transferred to nitrocellulose membrane. Panel A represents immunoblotting using the antiserum before affinity purification, and panel B represents immunoblotting using affinity purified antibodies. Lane 1 contains TGase 6 GST-fusion protein extract, and lane 2, TGase 7 GST-fusion protein extract. Mr standards are indicated on the right.

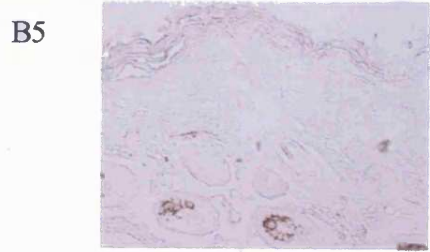
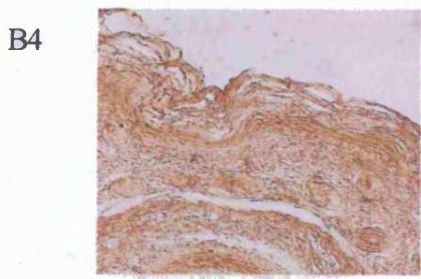
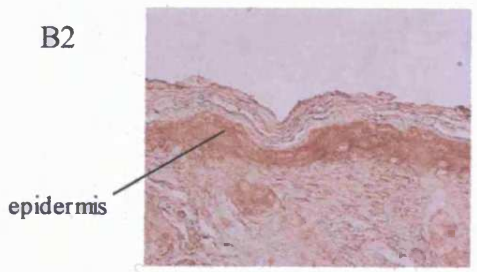
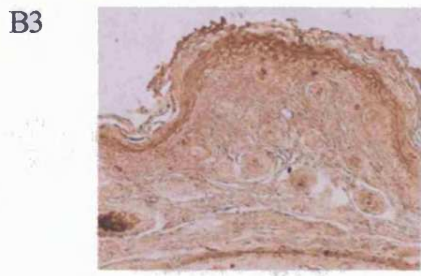
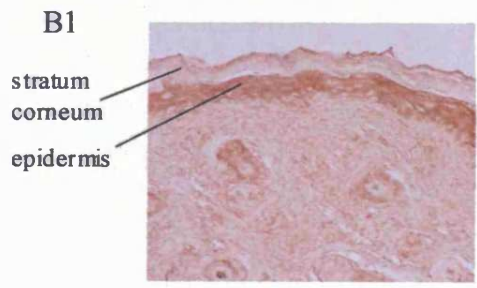
4.3.8 Immunohistochemistry

To confirm and extend the results obtained using *in situ* hybridisation in chapter 3, the proteins were localised on 5µM sagittal tissue sections of newborn mouse and mouse embryos at gestation days 11, 13 and 16 by immunohistochemistry. The sections were incubated for with purified antibodies for TGase6_L at a concentration of 1µg/ml. Antibody binding was visualised with peroxidase conjugated secondary antibodies and AEC as a peroxidase substrate.

In newborn mouse, TGase 6 is localised in the upper layers of the cerebral cortex (Fig. 4.17, A1 and A2). In skin (Fig. 4.17, B1 and B2), TGase 6 was specifically expressed in the epidermis and corresponding keratinising cells in hair follicles. TGase 6 was also present in the chondrocytes of cartilage and the developing vertebrae where vasculature invasion into the cartilage and bone formation occurs. (Fig. 4.17, C1 and C2). This localisation is consistent with the expression pattern determined by *in situ* hybridisation.

We have also determined the distribution of TGase 5 in newborn mouse. TGase 5 staining was detected throughout the cerebral cortex of the brain (Fig. 4.17, A3 and A4). In skin (Fig. 4.17, B3 and B4), as expected TGase 5 is present in the suprabasal cell layers of the epidermis. Staining was also detected in the chondrocytes of cartilage, and also in the bone cells of the vertebrae (Fig. 4.17, C3 and C4). These results are also consistent with those deduced from *in situ* hybridisation.





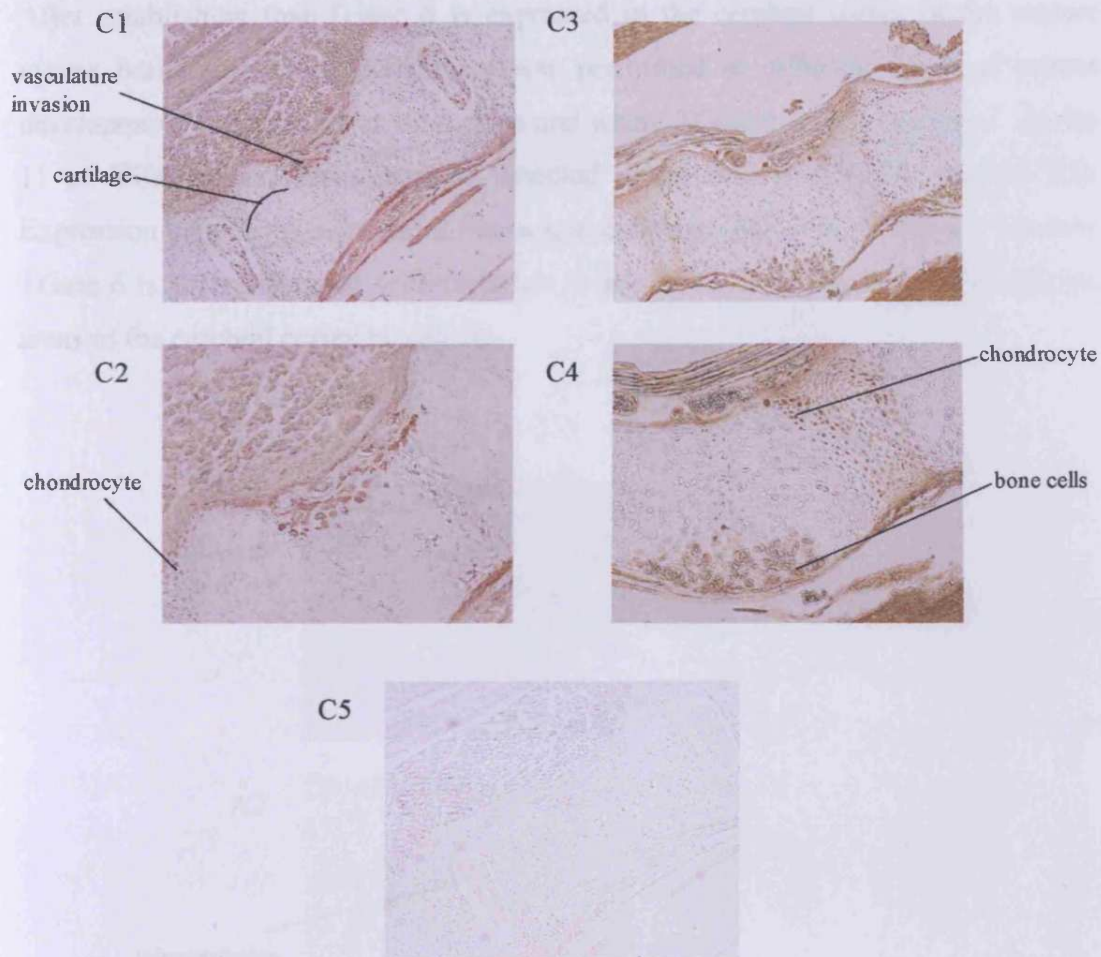
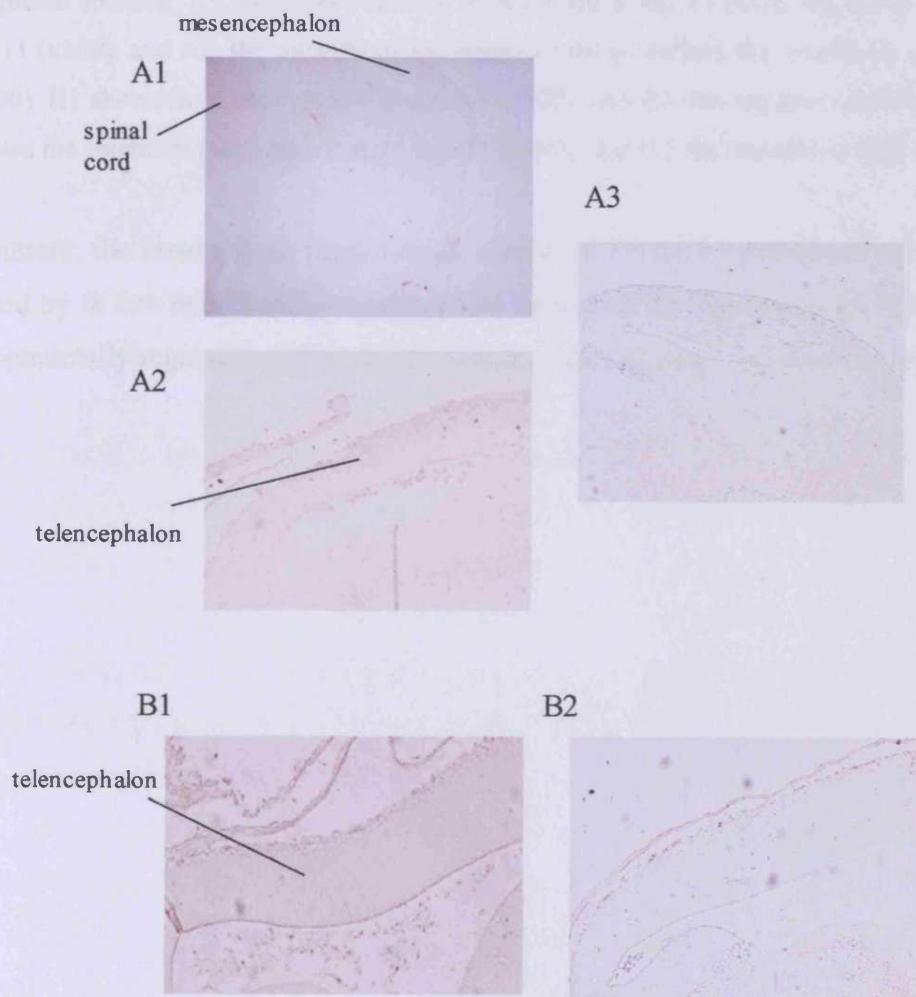


Fig. 4.17 Immunostaining of sagittal newborn mouse tissue sections with affinity purified antibodies to TGase 6_L and TGase 5. Expression was detected using an AEC development solution. A1 (x100) and A2 (x200) shows staining in the cerebral cortex with antibodies to TGase 6_L, A3 (x100) and A4 (x200) shows staining in the cerebral cortex with antibodies to TGase 5, and A5 (x100), the corresponding negative control (lacking the specific Ig to TGase 6 or TGase 5). B1 (x400) and B2 (x600) shows staining in the epidermis of the skin with antibodies to TGase 6_L, B3 (x200) and B4 (x400) shows staining in the epidermis of the skin with antibodies to TGase 5, and B5, the corresponding negative control (x400). C1 (x100) and C2 (x200) shows staining in the vertebrae with antibodies to TGase 6_L, C3 (x100) and C4 (x200) shows staining in the vertebrae with antibodies to TGase 5, and C5, the corresponding negative control (x100).

Preliminary experiments were also performed with affinity purified TGase 6_S antibodies. However the results proved inconclusive and further optimisation of the conditions would be required.

After establishing that TGase 6 is expressed in the cerebral cortex in the mature mouse brain, immunohistochemistry was performed at different stages of mouse development to determine at what stage and where TGase 6 is first expressed. At day 11 no TGase 6 expression can be detected in the brain (Fig. 4.18, A1 and A2). Expression appears to increase as the mouse develops (Fig. 4.18, B and C), whereby TGase 6 is absent from the telencephalon at day 13 but can be visualised in specific areas of the cerebral cortex by day 16.



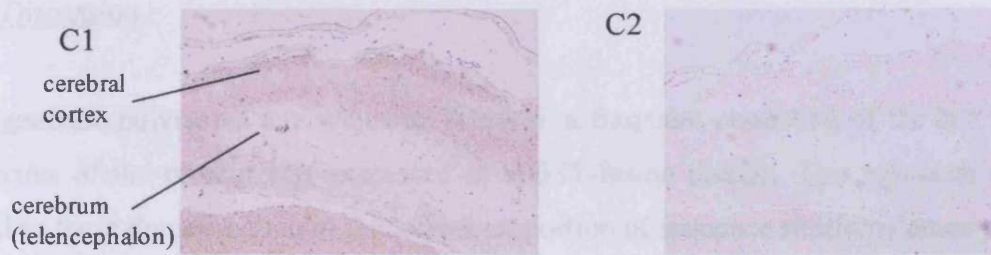


Fig. 4.18 Immunostaining of sagittal mouse embryo sections at day 11, 13 and 16, with affinity purified antibodies to TGase 6_L. Expression of TGase 6_L was detected using an AEC development solution. A1 shows an overview of the brain at day 11 (x40), A2, telencephalon at day 11 (x100), and A3, the corresponding negative control (lacking the specific Ig to TGase 6) (x100). B1 shows the telencephalon at day 13 (x100), and B2, the negative control (x100). C1 shows the cerebrum (telencephalon) at day 16 (x100), and C2, the negative control (x100).

In summary, the results from immunolocalisation of TGase 6 corroborate the results obtained by *in situ* hybridisation analysis and show that the expression of TGase 6 is developmentally regulated and occurs in conjunction with neuronal development.

4.4 Discussion

To generate polyclonal antibodies to TGase 6, a fragment consisting of the last two domains of the protein was expressed as a GST-fusion protein. This approach was used as these domains contain the lowest proportion of sequence similarity among the TGase family members, and expression as a GST-fusion protein facilitates purification. Unfortunately, initial attempts at purification of the GST-fusion protein were unsuccessful and a new approach had to be taken. Peptides for TGase 6_L, and TGase 6_S were designed, synthesised and rendered antigenic by coupling to a carrier protein, hemocyanin. ELISA and Western blotting techniques were used to demonstrate the presence of antibodies to the peptides in the antisera. These antibodies were purified for localisation of the proteins on tissue sections.

The results for TGase 6 protein expression are in accordance with the *in situ* hybridisation results whereby TGase 6 expression was observed in the brain, particularly in the cerebellum and cerebral cortex. The sera were used in Western blotting to analyse the expression of TGase 6_L and TGase 6_S in extracts from different regions of the mouse brain. Expression was detected in the spinal cord, cerebrum and medulla for the full-length TGase 6, and various splice variants were present in all tissues. The band of ~72kDa may correspond to the exon III splice variant that was identified through PCR. Other bands may correspond to previously unidentified splice variants. Two bands, corresponding to ~92 and 89kDa were also identified in the cerebrum. The larger band may represent an exon XII spliced form of TGase 6, as in mouse, the splicing results in a frameshift which results in the stop codon being located further downstream than in the human splice variant. This is further confirmed by the fact that expression for TGase 6_S was only detected in the cerebrum and the two bands detected correspond to ~92 and 89kDa bands which were also identified with the TGase 6_L serum. The 89kDa band may represent an exon XII spliced form with additional splicing, possibly of the exon III. Two additional high molecular weight forms are present, the identity of which remains unknown at this time.

Immunohistochemical studies confirmed the *in situ* hybridisation results. TGase 6 expression was again not only detected in the brain, but also in skin, and developing long bones including the vertebrae. The immunohistochemical analysis also

confirmed that the expression of TGase 6 is developmentally regulated. The expression in the cerebral cortex coincided with neurogenesis. The expression of TGase 5 was also consistent with the *in situ* hybridisation experiments, and also previous findings (Grenard et al., 2001), whereby the level of expression in brain was low and not restricted to distinct layers in the cerebral cortex.

Based on the finding that TGase 6 is expressed in these areas of the brain, it may be hypothesised that it has a functional role within the brain. It would be invaluable to determine if the expression of TGase 6 is altered in diseased brain, and if so, whether TGase 6 contributes to neurodegenerative diseases. It is of interest that TGase 6_S was only detected in the cerebrum, as the cerebral cortex of the cerebrum is one of the areas damaged in the brains of AD and HD sufferers. This may indicate that the alternatively spliced form of TGase 6 has a specific role in the cerebrum.

5.0 Generation of full-length TGase 6 cDNA

5.1 Introduction

Based on the conservation of critical residues for enzyme function and domain folding, it may be predicted that TGase 6 is an active enzyme. Also, due to the finding that TGase 6 shows the highest sequence similarity to TGase 3, it may be hypothesised that like TGase 3, TGase 6 has three Ca²⁺ binding sites, is able to bind and hydrolyse GTP, and that the transamidation activation mechanism occurs by a similar process. However, this remains to be demonstrated experimentally. Therefore, the aim of this chapter was to generate the full-length TGase 6 cDNA from the amplified PCR fragments in order to recombinantly express the enzyme and to perform experiments to determine if any of these hypotheses are valid.

5.2 Materials and Methods

5.2.1 PCR amplification

Poly(A⁺) RNA was prepared from about 10⁶ cells or 10µg total RNA using the MicroFast Track Kit as described in section 2.2.4. The poly(A⁺) RNA was reverse transcribed into DNA in a total volume of 20µl using SuperScript II as described in section 2.2.5.2. Fragments of TGase 6 cDNA were amplified by PCR using oligonucleotides:

fP1, fP2, fP3, fP4, rP5,

rP13 5'TACTTGCCCTGCCACTGTCCT

rP14 5'AGGCTTCCACGCCGAACAGC

rP17 5'AATCCACCCCTCTCAGTCCC

in various combinations with two different types of DNA polymerase. PCRs were carried out with either: 1.25 units AmpliTaq Gold DNA polymerase and 2.0µl cDNA/1.0µl PCR product in a total volume of 50µl as described in section 2.2.8; or 2.6 units ExpandTM High Fidelity PCR System enzyme mix (Boehringer Mannheim) and 2.0µl cDNA/1.0µl PCR product in a total of 50µl of supplied 1x Expand HF buffer with 2.5mM MgCl₂, 0.2mM dNTPs and 5µM of each primer. Forty cycles were carried out in a GeneAmp 9600 thermal cycler as described in section 2.2.8 with the PCR reactions containing Expand High Fidelity enzyme requiring a hot start. Amplified products were analysed on 1% agarose gels.

5.2.2 Gel extraction

5.2.2.1 Gel extraction using a QIAquick gel extraction kit

The DNA fragments were excised from the agarose gel with a clean, sharp scalpel. The gel slices were weighed in a microcentrifuge tube and the DNA extracted as described in section 2.2.9.

5.2.2.2 Gel extraction using a GeneClean kit

The DNA fragment was excised from the agarose gel with a clean, sharp scalpel and the DNA extracted using a GeneClean Kit (Qbiogene). The gel slices were weighed and 3 volumes of NaI to 1 volume of gel was added. The mixture was incubated at 55°C for 5 min, with vortexing of the tube three times during the incubation period. 10µl Glassmilk suspension (a specially prepared aqueous suspension of proprietary silica matrix) was added to the mixture and incubated for 5 min RT with vortexing every minute. This allows the DNA to bind to the silica matrix. The Glassmilk/DNA complex was pelleted by centrifugation at 16,000xg for 5 sec and the supernatant removed. The pellet was resuspended in 400µl New Wash (a concentrated solution of NaCl, Tris and EDTA to which H₂O and ethanol are added), centrifuged at 16,000xg, for 5 sec and the supernatant removed. This washing procedure was repeated twice. The pellet was air-dried for 5 min at RT for removal of any residual ethanol. For elution of the DNA, the pellet was resuspended in 10µl pre-warmed H₂O, incubated at 55°C for 2 min and the supernatant collected after centrifugation at 16,000xg for 30 sec. This step was repeated so the DNA was eluted in a total volume of 20µl H₂O.

5.2.3 Sequencing

Sequencing of each PCR product or cloned DNA fragment was performed using the ABI Prism™ dRhodamine Terminator Cycle Sequencing Ready reaction kit and the reactions analysed on an ABI Prism 310 genetic analyser as described in section 2.2.10.1.

5.2.4 Ligation of PCR product into vector

To ligate the PCR product into the pCRII vector a TA cloning kit Dual Promoter (Invitrogen) was used as described in section 2.2.16.1.2.

To ligate a PCR product digested with a restriction enzyme into the appropriately cleaved pCRII vector ligation reaction was set up as follows: an equal proportion of PCR product or cDNA fragment to pCRII vector fragment in 1x ligation buffer were

mixed with 1 unit T4 DNA ligase (Promega) and incubated at 14.5°C overnight. The reaction was placed on ice for immediate use or stored at -20°C.

5.2.5 Transformation

Transformation was performed either using frozen One Shot competent TOP10F' cells (Invitrogen) or Max Efficiency DH5 α TM Competent cells (Life Technologies) as described in section 2.2.16.1.3.

100 μ l and 200 μ l aliquots of the transformation reaction were plated onto LB plates containing either 30 μ g/ml kanamycin and 40mg X-Gal for TOP10F' cells or 50 μ g/ml Ampicillin and 40mg X-Gal for DH5 α cells. The plates were left for 5 min to ensure the liquid is absorbed, then inverted and incubated at 37°C overnight.

5.2.6 Isolation of plasmid DNA

Single, well-isolated white colonies were picked off the appropriate LB agar plate using sterile pipette tips. Each pipette tip was ejected into 3ml of LB medium containing 30 μ g/ml Kanamycin or 50 μ g/ml Ampicillin depending on the competent cells used in the transformation, to inoculate the medium. The cultures were incubated at 37°C overnight in a shaking incubator at 225rpm. A Wizard Plus SV Miniprep DNA Purification System was used to isolate the plasmid DNA as described in section 2.2.16.1.5.

5.2.7 Restriction endonuclease digests

To determine the size of the insert ligated into the vector a restriction digest using EcoRI restriction endonuclease was performed. 1 μ l plasmid DNA, 0.1 μ g acetylated BSA, 1x Buffer H (90mM Tris-HCl, pH7.5, 50mM NaCl, 10mM MgCl₂, 1mM DTT), 1 unit EcoRI (supplied in 1mM DTT, 0.1mM EDTA, 0.15% Triton X-100, 0.5mg/ml BSA and 50% glycerol) (Promega) and H₂O to a final volume of 10 μ l were incubated at 37°C for 1 hr. The samples were then analysed on 1% agarose gels. To avoid re-

ligation of the vector, the digested vector was dephosphorylated with calf intestinal alkaline phosphatase (CIP) (10 units) (NE Biolabs) for 30 min at 37°C.

To digest plasmid DNA required for ligation with another digested plasmid DNA, reactions were set up with plasmid DNA, the appropriate restriction enzyme, the appropriate supplied buffer, and 0.1µg acetylated BSA. The reactions were incubated at 37°C for 2 hrs, analysed on 1% agarose gels and gel extracted.

5.3 Results

5.3.1 Amplification of TGase 6 cDNA

To obtain the full-length cDNA of TGase 6, PCR was carried out with primers flanking the coding sequence. Primers fP1 and rP17 were used in combination with H69 cDNA as a template and Expand High Fidelity enzyme. This enzyme was used as it is especially optimised to amplify products up to 12kb and has an increased fidelity of DNA synthesis compared to *Taq* DNA polymerase. The PCR reaction was analysed on a 1% agarose gel but no PCR product had been amplified. The procedure was repeated but there was only weak amplification of PCR products and none corresponded to the expected size of the full-length cDNA.

The next approach was to amplify large sections of the cDNA by PCR. Overlapping sections of PCR products could then be used in further PCR reactions to amplify a larger section spanning the two smaller sections. PCR was performed with H69 cDNA and primer combinations fP1 and rP14, fP3 and rP14, and fP4 and rP14 with AmpliTaq Gold DNA polymerase or Expand High Fidelity enzyme. The PCR products were analysed by agarose gel electrophoresis which revealed that no products corresponding to the expected size had been amplified. 1µl of each PCR product was used in further PCR reactions in an attempt to reamplify any products that had been amplified but not to an extent for visualisation with ethidium bromide. No products of an expected size were obtained.

Due to the lack of success using PCR to join the respective fragments, it was decided to sub-clone three PCR products (which by overlapping spanned the full-length of the TGase 6 cDNA) to generate a full-length cDNA clone (Fig. 5.2). One of these fragments originated from the work described in Chapter 2 (Fig. 2.1) and the other two were generated by PCR using H69 cDNA in a combination with primers fP1 and rP13, and fP5 and rP17. The two new products generated were of the expected size (Fig. 5.1) and their identity further confirmed by sequencing.

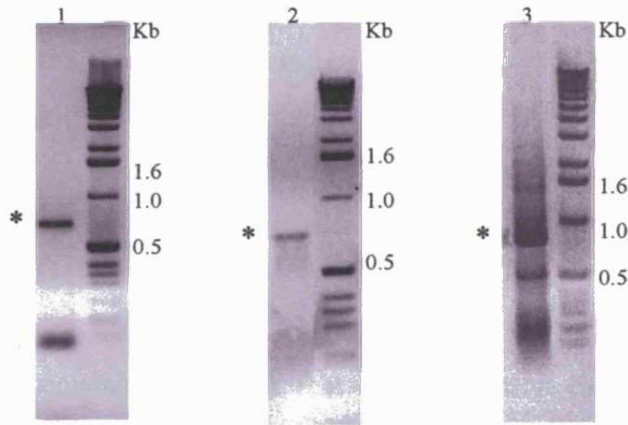


Fig. 5.1 Amplification of TGase 6 from H69 cDNA. Lane 1 represents products of PCR reaction with primers fP1 and rP13; lane 2, products of PCR reactions with fP4 and rP5; and lane 3, products amplified with fP5 and rP17. PCR products were analysed by electrophoresis in 1% agarose gels calibrated with the 1kb ladder. * indicates the PCR products which were gel extracted and sequenced.

By restriction endonuclease digestion with the enzymes indicated in Fig. 5.2, it will be possible to, ligate together the three products to span the full-length cDNA (Fig. 5 2). For simplicity, the cDNA amplified with fP1 and rP13, with fP4 and rP5, and with fP5 and rP17 will be referred to as insert 1, 2 and 3, respectively, in the following.

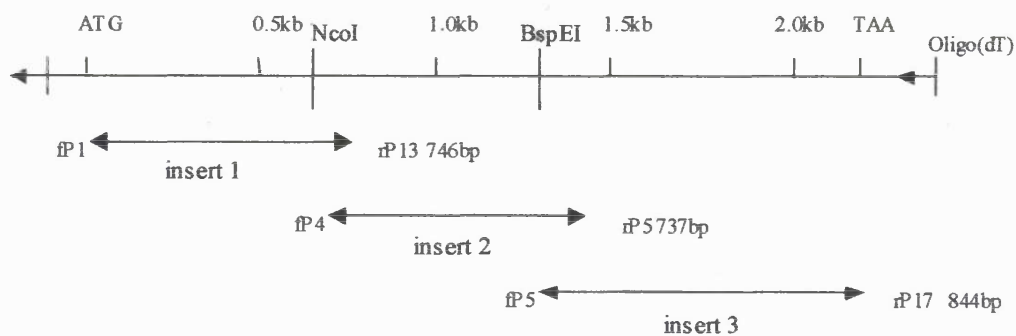


Fig. 5.2 Strategy for subcloning of three PCR products to obtain the full-length TGase 6 cDNA. The top line represents the cDNA for TGase 6 with the start and stop codons indicated. Below is an outline of the PCR products, showing the overlapping regions. The relevant restriction sites are indicated in the top line.

Each PCR product was ligated into the pCRII vector using TA cloning, transformed into *E. coli* and plasmid DNA isolated from 3ml LB broth overnight cultures. An

excess of 50 colonies were obtained for insert 1, 2, and 3 respectively. 10 colonies were picked from each plate and cultures grown overnight in LB medium containing the relevant antibiotic for selection. The plasmid DNA was extracted using the Wizard Miniprep DNA purification System and restriction digests using EcoRI were carried out to identify clones which contained a plasmid with an insert of the correct size (Fig. 5.3).

Each plasmid was sequenced to verify that the DNA insert contained no mutations and to determine the orientation of the insert. Clone 4 was chosen for insert 1, clone 3 for insert 2 and clone 3 for insert 3. Sequencing with M13 and T7 primers revealed that insert 1 and insert 2 were in the reverse orientation and insert 3 was in the forward orientation.

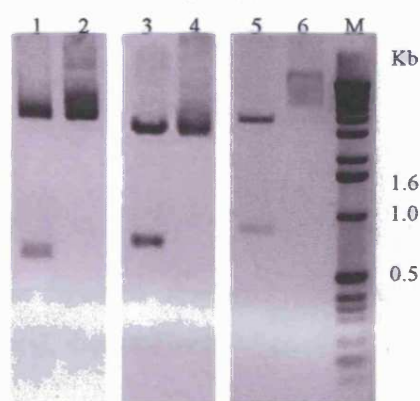


Fig. 5.3 EcoRI endonuclease digestions on plasmid DNA containing insert 1, 2 and 3, respectively. Lanes 1 and 2 show plasmid DNA containing insert 1 isolated from clone 4, digested and undigested, respectively. Lanes 3 and 4 show plasmid DNA containing insert 2 isolated from clone 2, digested and undigested, respectively. Lanes 5 and 6 show plasmid DNA containing insert 3 isolated from clone 3, digested and undigested, respectively. Restriction digests were analysed by electrophoresis on 1% agarose gels, stained with ethidium bromide and calibrated with a 1kb ladder.

It was at this stage that exon I of TGase 6 was identified and amplified by PCR. To splice this exon onto exon II, insert 1 was cleaved from the pCRII vector by restriction with EcoRI. A PCR was performed with the PCR product amplified with primers fP5'2 and rP1 (see section 2.3.2) and the excised insert 1 as a template, and fP5'2 and rP13 as primers. By using two overlapping DNA fragments as a template in

combination with the outermost forward and reverse primer in a reaction, the aim was to amplify across the two DNA fragments and obtain a joined product. The PCR product obtained was gel purified and sequenced. The PCR product (referred to as insert 1.1) was then ligated into pCRII, transformed and plasmid DNA was extracted from mini-cultures. Colonies containing the correct insert were identified by restriction with EcoRI and clone 2 was chosen to use for further analysis (Fig. 5.4). The plasmid DNA was sequenced which confirmed the absence of mutations and revealed that the insert (now referred to as insert 1.1) was in reverse orientation in the pCRII vector.



Fig. 5.4 EcoRI endonuclease digestion on plasmid DNA containing insert 1.1. Lanes 1 and 2 show plasmid DNA of clone 2, digested and undigested, respectively. Restriction digests were analysed as above.

The first step in obtaining the full-length cDNA was to join insert 1.1 and insert 2. For this purpose, a partial digest with NcoI was to be carried out on insert 1.1 followed by digestion with XbaI (Fig. 5.5). First, the appropriate dilution of NcoI for the partial digest had to be determined. A series of dilutions ranging from 1:2 to 1:128 was carried out on the plasmid DNA and from the results it was decided to perform the partial digest with a 1:64 dilution of NcoI. The partial NcoI digest was carried out first, the enzyme was heat inactivated and then a complete digest with XbaI was carried out. The digest was analysed by agarose gel electrophoresis and the 745bp band corresponding to the section of insert indicated (Fig. 5.5) was gel extracted (Fig. 5.6).

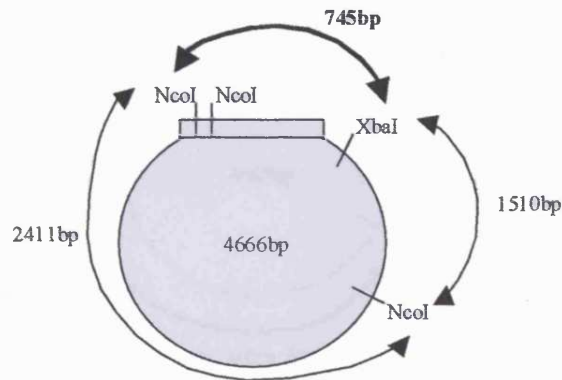


Fig. 5.5 Schematic showing the cleavage sites for NcoI and XbaI endonucleases on insert 1.1.

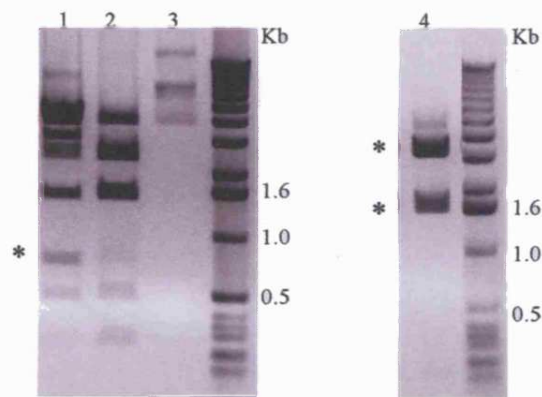


Fig. 5.6 Endonuclease digestion using NcoI and XbaI on plasmid DNA containing inserts 1.1 and 2. Lane 1 shows a partial NcoI restriction digest followed by digestion with XbaI on plasmid DNA containing insert 1.1. Lane 2 shows insert 1.1 fully digested with both NcoI and XbaI. Lane 3 shows the insert undigested. Lane 4 shows a NcoI restriction digest followed by digestion with XbaI on plasmid DNA containing insert 2. Restriction digests were analysed as above. * indicates the digest products gel extracted.

For insert 2, a complete digest with both restriction endonuclease enzymes will yield two fragments, a 1510bp fragment and a 3100bp fragment that are needed (Fig. 5.7). The aim was to ligate these two fragments with the fragment from insert 1.1.

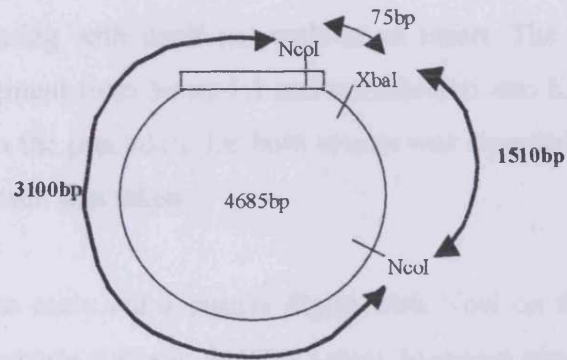


Fig. 5.7 Schematic showing the cleavage sites for NcoI and XbaI on insert 2.

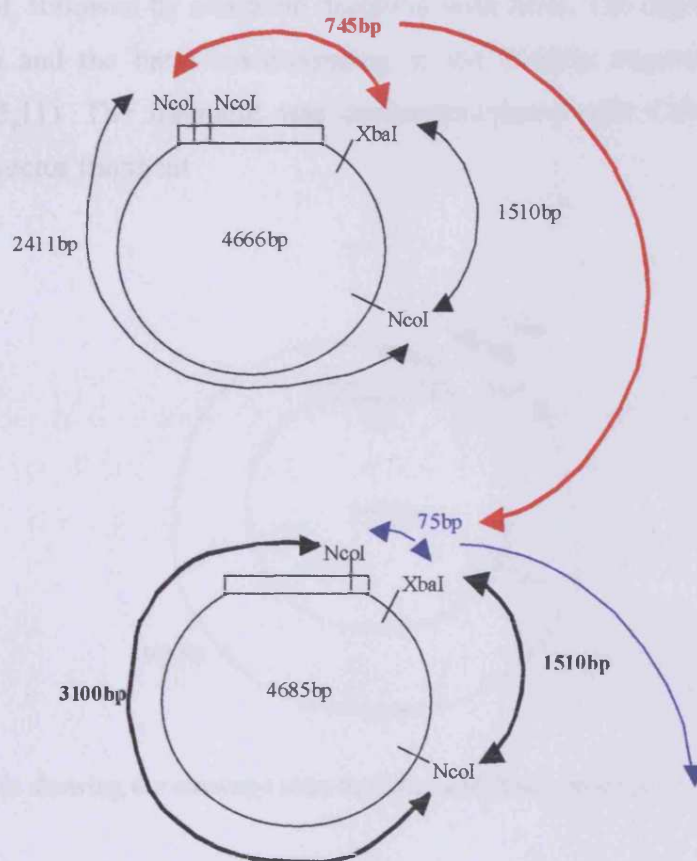


Fig. 5.8 Schematic showing the strategy for the joining of insert 1.1 and insert 2. The 745bp fragment containing insert 1.1 will be ligated with the 3100bp and 1510bp fragments of the pCRII vector containing insert 2.

After digestion, the reaction was analysed by agarose gel electrophoresis and the bands corresponding to the 1510bp and 3100 fragments were excised (Fig. 5.6, lane 4). The 3100bp fragment was dephosphorylated with CIP prior to ligation to prevent

the vector from ligating with itself i.e. without an insert. The two fragments were ligated with the fragment from insert 1.1 and transformed into *E. coli* TOP10F' cells. No colonies grew so the procedure for both inserts was repeated. Again, no colonies grew so a new approach was taken.

The new aim was to carry out a partial digest with NcoI on the vector containing insert 2 in order to obtain a single 4610bp vector fragment after complete digestion with XbaI (Fig. 5.9). The appropriate dilution of NcoI for the partial digest was determined as described previously. The partial digest was carried out with a 1:24 dilution of NcoI, followed by complete digestion with XbaI. The digest was analysed on agarose gel and the band corresponding to the 4610bp fragment excised and purified (Fig. 5.11). The fragment was dephosphorylated with CIP to prevent re-ligation of the vector fragment

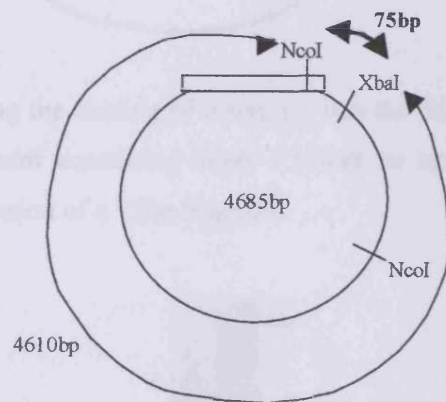


Fig. 5.9 Schematic showing the cleavage sites for NcoI and XbaI on insert 2.

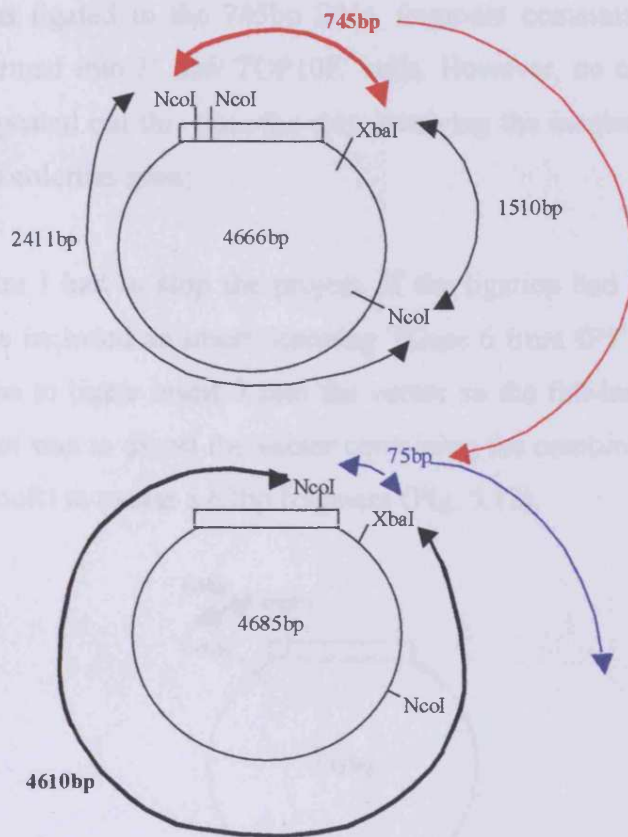


Fig. 5.10 Schematic showing the cloning of insert 1.1 into the pCRII vector containing insert 2. The 745bp DNA fragment containing insert 1.1 will be ligated into the pCRII vector containing insert 2 after excision of a 75bp fragment.



Fig. 5.11 Endonuclease digestion of plasmid containing insert 2 with NcoI and XbaI. A partial NcoI restriction was performed on the plasmid followed by complete digestion with XbaI (Lane 1). Restriction digests were analysed by electrophoresis on a 1% agarose gel. * indicates the product gel extracted.

The fragment was ligated to the 745bp DNA fragment containing insert 1.1 (Fig. 5.10) and transformed into *E. coli* TOP10F' cells. However, no colonies grew. The procedure was repeated but this time the step involving the incubation with CIP was omitted. Again no colonies grew.

It was at this point I had to stop the project. If the ligation had worked the pCRII vector would have included an insert spanning TGase 6 from fP5'2 to rP5. The next step would then be to ligate insert 3 into the vector so the full-length construct was generated. The aim was to digest the vector containing the combined insert 1.1 and 2 with BspEI and EcoRI to excise a 63bp fragment (Fig. 5.12).

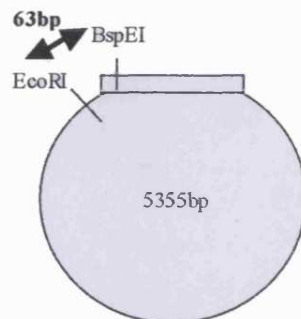


Fig.5.12 Schematic showing the cleavage sites for EcoRI and BspEI on pCRII containing insert 1.1 and insert 2.

The vector containing insert 3 would be digested with BspEI and EcoRI (Fig. 5.13) and the resulting 829bp DNA fragment containing insert 3 ligated into the cleaved vector containing insert 1.1 and 2 (Fig. 5.14).

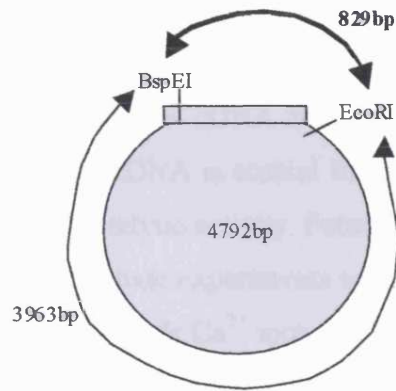


Fig. 5.13 Schematic showing the cleavage sites for EcoRI and BspEI in the pCRII vector containing insert 3.

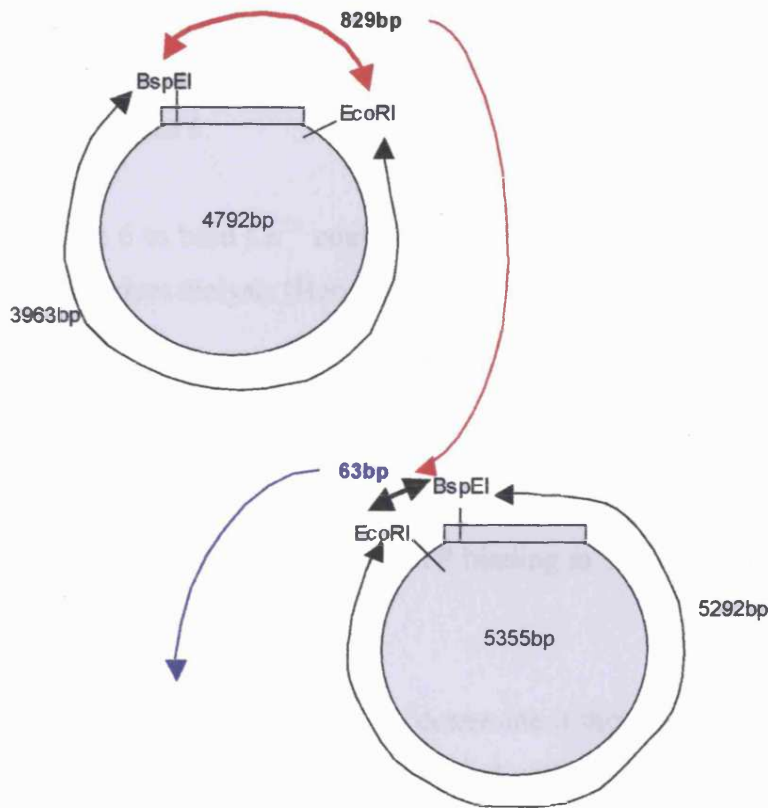


Fig. 5.14 Schematic showing the cloning of insert 3 into the pCRII vector containing insert 1.1 and 2. The 829bp DNA fragment containing insert 3 will be ligated into the pCRII vector containing insert 1.1 and 2 after excision of the 63bp fragment.

5.4 Discussion

Unfortunately, the full-length TGase 6 cDNA was not obtained during my PhD. Obtaining the full-length TGase 6 cDNA is crucial for expression of the enzyme for verification that the enzyme has catalytic activity. Future work to be carried out when the full-length clone is obtained include experiments to determine: 1) if TGase 6 is an active transamidase; 2) if TGase 6 binds Ca^{2+} ions and how many; 3) and if TGase 6 can bind and hydrolyse GTP.

In order to verify that TGase 6 is an active enzyme, a TGase activity assay could be performed by which the TGase 6 activity is measured by incorporation of [^{14}C] putrescine into casein (Folk and Cole, 1966). It has been demonstrated that casein is a substrate for several TGases, but for this to be a viable experiment it would also need to be a substrate for TGase 6.

The ability of TGase 6 to bind Ca^{2+} could be established by a calcium-binding assay which utilises equilibrium dialysis (Bergamini, 1988).

To deduce if TGase 6 can bind GTP, a GTP-binding assay could be carried out by which TGase 6 could be incubated with radioactively labelled GTP and the amount of incorporated radioactivity measured. Also, to determine if GTP binds directly to TGase 6, the thermodynamic properties of GTP binding to TGase 6 can be measured by isothermal titration calorimetry.

If GTP binding is established in TGase 6, to determine if this inactivates the enzyme, increasing concentrations of GTP can be incubated with TGase 6 and the enzymatic activity related to that in the absence of nucleotides. Also, the GTP hydrolysing activity of TGase 6 could be deduced by incubation of TGase 6 with ^{32}P -labelled GTP and then measuring the amount of ^{32}P released from [γ - ^{32}P] GTP over time.

In order to obtain an accurate structure of TGase 6, the enzyme needs to be crystallised. The crystal structure could then be solved by molecular replacement. This would enable Ca^{2+} -binding sites to be located, and any structural rearrangements

occurring due to Ca^{2+} could be resolved. The critical residues involved in Ca^{2+} -binding could also be identified. It could also be demonstrated that if TGase 6 can bind GTP, where the binding site is located and the residues involved in binding.

Molecular modelling studies could be performed as a complementary approach to understanding how the substrates interact with the enzyme. Modelling is a useful technique as it can predict the modes of inclusion, the stoichiometry of the complex and the complex efficiency. Commercially available software uses molecular mechanics and dynamic simulations, and docking programmes can also be used for qualitative purposes. However, modelling studies are not completely reliable and could only provide a possible theory.

6.0 Analysis of patient sera

6.1 Introduction

Recently, work has been carried out on the hypothesis that certain neurological illnesses relate to gluten sensitivity, so-called gluten ataxia (Hadjivassiliou et al., 1996, Hadjivassiliou et al., 1998, Hadjivassiliou et al., 2003). It has previously been determined that TGase 2 plays a role as an antigen in celiac disease as discussed in chapter 1. The aim of this chapter was to determine if antibodies to TGase 6 are present in sera of patients with the neurological manifestation, thereby explaining the distinct phenotype.

6.2 Materials and Methods

6.2.1 Cell extracts

H69 cells were extracted using Triton X-100 (Sigma). Briefly, cells were harvested by centrifugation at 1,500xg. The cell pellet was washed twice in pre-warmed PBS and was extracted in 0.25M sucrose, 1% Triton X-100 on ice. The extract was cleared from particulate material by centrifugation at 15,000xg for 10 min at 4°C. The supernatant harvested and stored at -20°C for further use.

HCA2 fibroblast extracts were kindly provided by M. Langley. These cell lines were generated as described by Stephens *et al.* (Stephens et al., 2004). Briefly, stably transfected cell lines were established harbouring constructs with the coding sequence of human TGase 2 (in the sense and antisense orientation) or the empty vector as a control. The extracts were prepared as described for the H69 cells.

6.2.2 SDS-PAGE

Cell extracts from mock, sense and antisense TGase 2 transfected HCA2 cells and H69 cells were normalised to a concentration of 200µg/ml. 10µl of each sample was mixed with an equal volume of 2x sample buffer and heated to 100°C for 5 min. SDS-PAGE was carried out as described in section 4.2.2.

6.2.3 Western blotting

Cell extracts separated on SDS-polyacrylamide gels were transferred onto nitrocellulose membranes as described in section 4.2.3. Membranes were blocked with 5% non-fat dry milk in TBS/0.05% Tween 20 for 1 hour at RT. Membranes were then incubated with the serum sample (patient samples kindly provided by Dr. M. Hadjivassiliou) (1:200 dilution in 5% milk in TBS/0.05% Tween 20) or CuB7402 monoclonal antibody to TGase 2 (Neomarkers) (1:100 dilution in 5% milk in TBS/0.05% Tween 20) overnight at 4°C. The membranes were washed in TBS/0.05% Tween 20 for 15 min at RT, followed by three washes for 5 min at RT. Membranes were then incubated with peroxidase-conjugated anti-human IgG (Neomarkers)

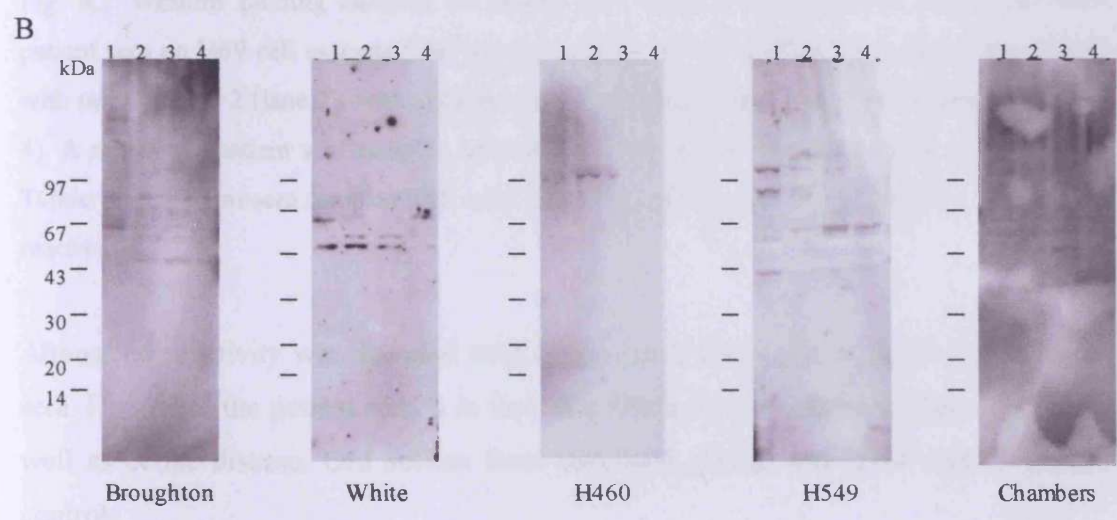
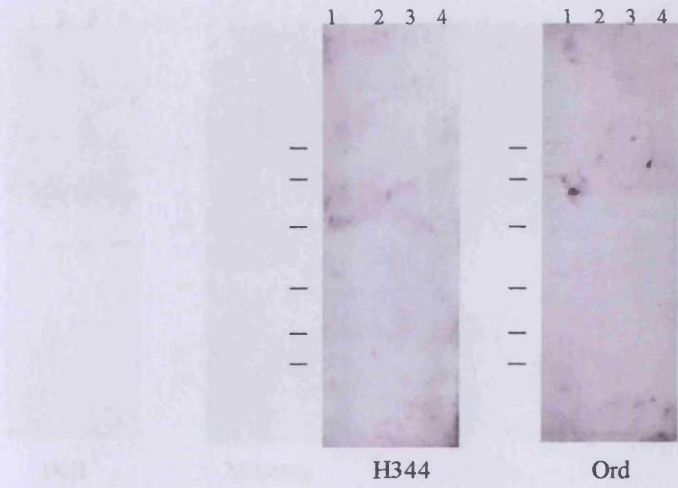
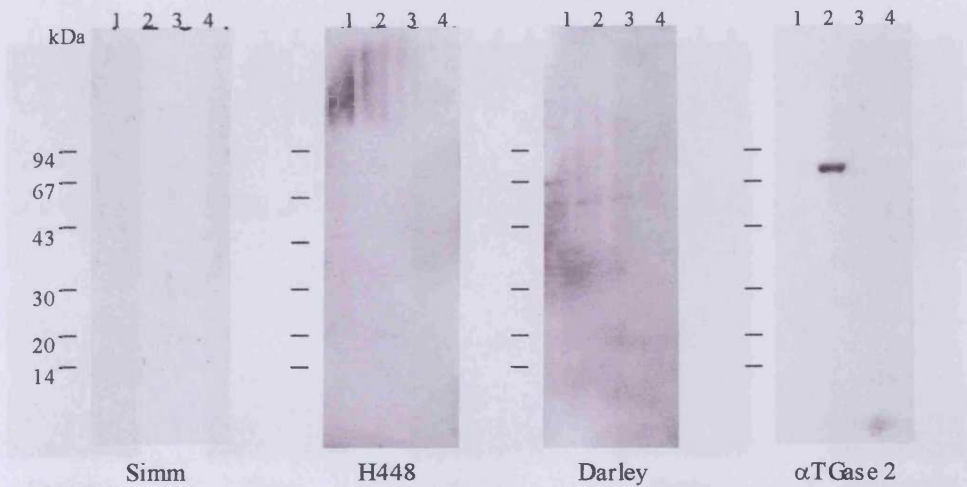
(1:1000 dilution in 5% milk in TBS/0.05% Tween 20) for 1 hr at RT. The membranes were washed as before, rinsed in distilled H₂O and the immunological reaction revealed using the enhanced chemiluminescence (ECL) kit.

6.3 Results

6.3.1 Analysis of patient sera samples

To determine if antibodies to TGase 6 are involved in the case of gluten ataxia, patient sera were analysed for immunoreactivity. As no purified TGase 6 protein was available, cells extracts were used. Cells were extracted from mock, sense and antisense TGase 2 transfected HCA2 cells and from H69 cells which are known to express TGase 6. The cell extracts were normalised to a concentration of 200µg/ml, separated by SDS-PAGE and transferred onto a membrane. The membranes were then incubated with patient serum samples (provided by Dr. M. Hadjivassiliou, neurological consultant, Sheffield) at a dilution of 1:200. 19 patient sera samples from either controls, gluten ataxia sufferers, celiac disease sufferers or suffers of both were tested. The tests were carried out without knowing the patient history which was released after completion of assessment of the immunoreactivity profile.

The blots were grouped into those which show almost no reactivity (Fig. 6.1 A), weak reactivity (Fig. 6.1 B), and strong reactivity (Fig. 6.1 C). The membrane previously probed with sera for patient Darley was stripped and re probed with a TGase 2 specific antibody. As expected a band was detected in lane 2 corresponding to the cell extract of HCA2 fibroblasts stably transfected with sense TGase 2.



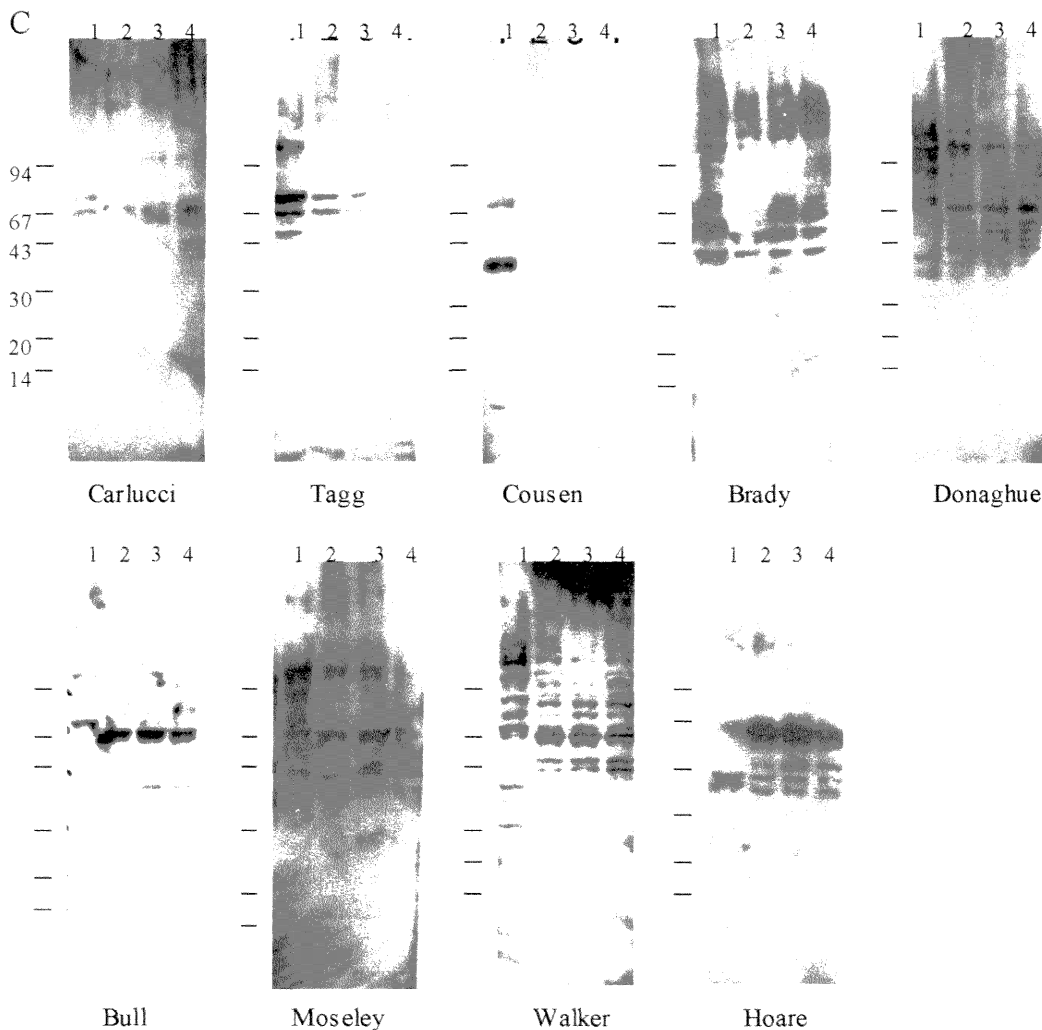


Fig. 6.1 Western blotting analysis of patient sera samples. Immunoblots performed with patient sera on H69 cell extracts (lane 1), cell extracts of HCA2 fibroblasts stably transfected with sense TGase 2 (lane 2), with antisense TGase 2 (lane 3), and with a mock construct (lane 4). A represents patient sera samples with no reactivity and also shows a positive control for TGase 2, B, patient sera samples with weak reactivity, and C, patient sera samples with strong reactivity.

Almost no reactivity was detected with Simm, Ord, H448, Darley and H344 patient sera. History of the patient sera is in fact, that Darley and Simm have gluten ataxia as well as celiac disease. Ord suffers from unrelated ataxia, and H344 and H448 are controls.

Weak reactivity was detected with Chambers, Broughton, White, H460 and H549. Broughton and White both have a band present at ~65kDa in the neuronal-like H69

cells. No other similarities could be identified. The history of the sera is that Broughton and White have gluten ataxia, Chambers has only celiac disease, and H460 and H549 are normal controls.

Strong reactivity was detected with Carlucci, Tagg, Bull and Hoare, all reacting strongly with an overlapping set of proteins. In particular, they all detect a band at ~65kDa. Brady, Donoghue, Mosely and Walker appear to have widespread reactivity. The serum of Cousen recognises proteins in the H69 cell extract which are absent in the other cell line. All, with the exception of Hoare, appear to recognise a band at ~71kDa in the cell extracts. The history of the patient sera is that Carlucci, Tagg, Cousen, Brady and Donoghue have gluten ataxia. Bull and Mosely have celiac disease but no ataxia, Walker and Hoare have ataxia from another cause.

6.4 Discussion

It has been hypothesised that gluten sensitivity can be classified as a neurological illness. Neuropathological data from patients with gluten ataxia (Hadjivassiliou et al., 2001) have indicated an inflammatory process with T-cell invasion of mainly the cerebellum resulting in the decline of Purkinje cells. Patients with gluten ataxia also have autoantibodies against brain structures, such as the cortex and the cerebellum. Patients with gluten sensitivity also display various organ involvement, for example, enteropathy, or ataxia/neuropathy (Hadjivassiliou et al., 1998), and it has been suggested that this may be due to TGases. TGase 2 is the antigen identified by antibodies for endomysium, a specific marker of gluten sensitive enteropathy. The fact that TGase 2 deamidates gliadin proteins has led to the hypothesis that the process may generate neoepitopes that could have a role in the immune pathogenesis of other diseases.

As immunoblotting of a cell extract is a very crude approach, it is difficult to reach a firm conclusion. It is expected that several bands would be visible on the western blots, however, some should be more common and stronger. For example, a clear reactivity with TGase 2 would have been expected for celiac patients. The samples sent from patients are collected at baseline and therefore should be in the “active” phase. Generally, all patients with an enteropathy would be expected to have anti-TGase 2 antibodies (correspondance with M. Hadjivassiliou). Sera from patients Chambers, Bull, and Moseley should all be positive for anti-TGase 2 antibodies as they all have celiac disease. However, the levels of antibodies will differ from person to person. Patients with gluten ataxia are by definition positive for antigliadin antibodies and have ataxia. Approximately 60% of the patients will also have anti-TGase 2 antibodies, but at much lower levels than those with enteropathy. About 30% of patients with gluten ataxia will also have an enteropathy. However, the results may have been more conclusive if an IgA secondary antibody had been used instead of an IgG antibody as the antibodies found in patients are IgA antibodies. Another problem in reaching a conclusion from these results is that epitope spreading may occur in autoimmune conditions whereby the initial selective reactivity may spread to a series

of other epitopes. This could account for the widespread reactivity detected with some patient sera samples.

The reactivity with a 65kDA protein is likely to be a protein found with gluten ataxia and celiac disease sera which is thought to represent glutamic acid decarboxylase (GAD), which is proposed to be an important antigen in gluten sensitivity (Hadjivassiliou et al., 2001). The molecular weight of TGase 6 is ~79kDA. However, there are a number of splice variants which makes a firm conclusion of the reactivity of a sera sample with TGase 6 difficult. As antibodies have now been generated for TGase 6, a positive control could be carried out whereby the blots are probed with the TGase 6 antibodies and this may enable identification of bands detected in a more specific manner.

One approach that could be taken now is to identify the epitopes that are reacting on the immunoblots in a selection of neuronal cell lines. This technique is currently being used by M Hadjivassiliou *et al.* whereby two dimensional gels are run and the immunoreactive spots are identified by mass spectrometry. If the current hypothesis that gluten sensitivity has neurological manifestations is true, then the patients sera should have antibodies with an affinity for one of the TGases, and this affinity should be higher than for TGase 2. Antibodies to TGase 6 could be identified more specifically by ELISA once recombinant TGase 6 becomes available. However, identification of an increased reactivity may be difficult as this may be confused by the fact that there is an overlap whereby some patients will also have an enteropathy. Further studies will need to be carried out in order to determine the link between gluten sensitivity and neurological illness.

7.0 Discussion

7.1 Determination of TGase 6 primary structure

In this study, a cDNA encoding a novel member of the transglutaminase gene family, TGase 6, was isolated from small cell carcinoma cell line H69. This confirmed that the TGM6 gene is a transcribed gene and not a pseudogene. Two related transcripts were obtained which encoded proteins of 708 and 629 amino acids with a molecular mass of 79,466Da and 70,671Da, respectively. The structure of the gene encoding TGase 6 (TGM6) can be inferred from comparison of the cDNA to the available genomic sequence. TGase 6 consists of 13 exons and 12 introns and the size of the exons is comparable to the other members of the TGase family. The full-length amino acid sequence of the novel human TGase gene product was determined from the cDNA sequence. The mouse cDNA sequence for TGase 6 was also isolated and TGase 6 appears to be highly conserved between human and mouse as the identity on the amino acid sequence level is 85%. This points to a similar function of the gene products in man and mouse.

7.2 Three-dimensional structure of TGase 6

The three-dimensional structures of four TGases have now been solved, i.e. human factor XIIIa (Yee et al., 1994), TGase 2 (Liu et al., 2002), TGase 3 enzymes (Ahvazi et al., 2002, 2003) and a fish enzyme (fTGase, equivalent to mammalian TGase 2, Nogouchi et al., 2001). All comprise of four domains that are similar in organisation: an amino terminal β -sandwich domain; a catalytic core domain which contains the conserved active site triad Cys, His, and Asp residues; a β -barrel 1 domain; and a β -barrel 2 domain at the carboxy terminus.

To propose a three-dimensional structure of TGase 6 and estimate whether the enzyme has similar properties to other TGases, the amino acid sequence of mouse TGase 6 was compared to those of human TGase 3, human fXIIIa and human TGase 2 using CLUSTALW (Higgins et al., 1994). Mouse TGase 6 showed the highest similarity to human TGase 3 (PDB accession no. 1L9N), particularly within the

catalytic domain. Based on this finding, a three dimensional model of mouse TGase 6 was generated with SwissModel (Guex et al., 2000) using the x-ray derived coordinates of chain A of activated human TGase 3 as a template. This was used to estimate whether key residues for enzyme activation, the transamidation activation mechanism, calcium binding and GTP binding are in a similar conformation in TGase 6 when compared to TGase 3 using the functional analysis of the TGase 3 structure based on the work of Ahvazi *et al* (Ahvazi et al., 2002, 2003, 2003a, 2004).

Based on a sequence and structural comparison of TGase 6 and TGase 3, in TGase 6 (Fig. 7.2), the N-terminal β -sandwich domain contains the first 136 amino acids. The catalytic core domain extends from residues Asn137 to Lys492 and harbours the active site Cys274 residue in the corresponding position to the active Cys272 of TGase 3. The β -barrel 1 and 2 domains span residues 493-632 and 633-706, respectively. Residues 462-491 correspond to the flexible solvent exposed loop (which is not resolved in the x-ray structure), which also joins the last α -helical region of the catalytic core domain to the first β -strand of barrel 1 domain (Fig. 7.1). This exposed region does not contain a corresponding serine residue to the Ser469 of TGase 3, which in TGase 3 is the cleavage site for proteolytic activation. However, the possibility of TGase 6 being proteolytically cleaved can not be ruled out as yet, as other residues within the loop may be sites for cleavage by specific proteases.

```

human TG6L      FGVEASGRRIRWIRRAGGRCLWRDDLLEPAT-----
mouse TG6L      LSVEAWGRRRRIRRASVRGVWRDDLLEPVT-----
human TG3L      -----KPNTPFAATSSMGLETEEQ-----
mouse TG3L      -----KPNASFGATSSRNPEGEDK-----
human TG1L      -----SKPN-VYANRGS-----
mouse TG1L      -----SKPN-VYATRDS-----
human TG2L      -----KLAEKE-----
mouse TG2L      -----KLAEKE-----
human TG4L      -----SSER-----EHRRPV-----
mouse TG4L      -----SDDKLNSRTL-----
human TG5L      ----KARSFHGSQRGAELQPSRPTSLSQDSPRSLHTPSLRPSD
mouse TG5L      ----QATRSQGPHQANSNPFSSVPPRHNSARSPDSPSLQPSD
human TG7L      -----LGPQRA----SLPFLDLLESGGLR-----
mouse TG7L      -----LEPRRN----ASSLLDLLGSENLK-----
human F13AL     -----AKKPLNTEGVMKS-----
mouse F13AL     -----AKKTLNTEGVVKS-----
human EP42L     -----MEREKDNGIRPPSLET-----
mouse EP42L     -----LKLKGDNGMCPPSCEP-----

```

Fig. 7.1 Sequence comparison of the flexible solvent exposed loop which joins the last α -helical region of the catalytic core domain to the first β -strand of barrel 1 in each TGase.

100

100

100

100

100

100

100

100

100

100

100

100

100

100

100

100

Conserved residues are highlighted in red, and the cleavage site for proteolytic activation of TGase 3 is highlighted blue.

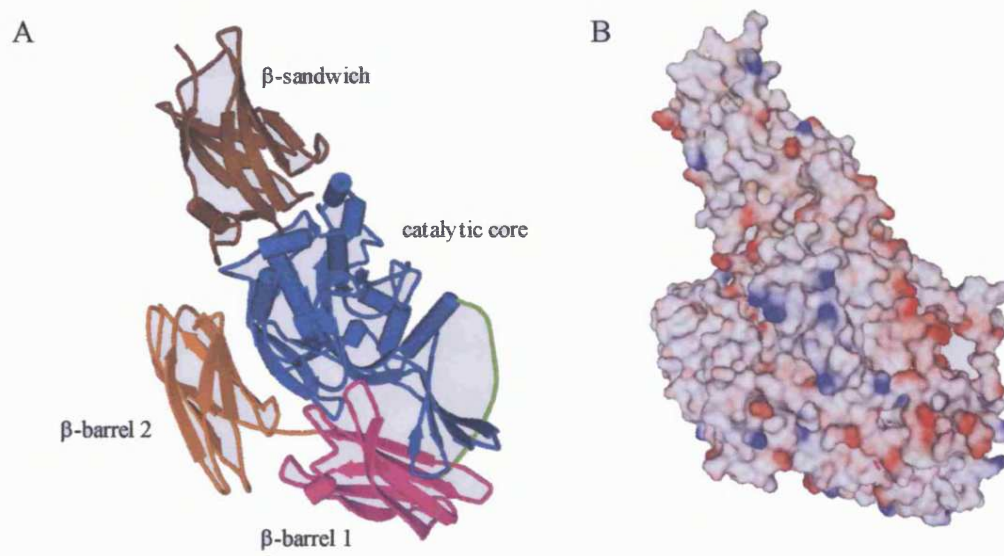


Fig. 7.2. A. Ribbon image of the predicted structure of TGase 6. The four domains are the N-terminal β -sandwich (brown), the catalytic core (blue), the β -barrel 1 (purple) and β -barrel 2 (orange). The connecting loop between the catalytic core and β -barrel 1 which is not resolved in x-ray crystallography is shown in green. B. The Connolly surface is shown for the TGase 6 structure colour coded corresponding to the electrostatic potential (blue: positive charges, red: negative charges).

7.2.1 Active site structure

In TGase 3, the catalytic triad active site residues, Cys272, His330 and Asp353 are at the base of a cavity which is bound by the catalytic core and β -barrel 1 domains, and are buried within a hydrophobic pocket. Based on the sequence alignment of TGase 3 and TGase 6, the catalytic triad active site residues for TGase 6 are Cys274, His333 and Asp356, and are also located in the same region.

Based on biochemical data (Lorand and Conrad, 1984, Folk and Chung, 1985) and the interpretation of the TGase 3 structures (Ahvazi et al., 2002, 2003) it is assumed that the sulphhydryl group of Cys272 forms a thiolate-imidazolium ion pair with His330. The second imino nitrogen atom of the His330 ring forms a hydrogen bond with the

terminal oxygen atom of Asp353. As indicated by the model, the catalytic triad residues of TGase 6 are located in corresponding positions and conformations to the TGase 3 residues, and therefore it can be proposed that this also occurs in TGase 6 (Fig. 7.3). There is evidence that two tryptophan residues (Trp236 and Trp327 in TGase 3), whose indole rings are buried near the surface are involved in the enzyme reaction mechanism (Pedersen et al., 1994). It has been proposed that as a suitable glutamyl substrate approaches, an oxyanion intermediate is formed with Trp236, which then breaks down to release NH_3 and form a thiol-acyl intermediate. This is then attacked by the ϵ -amine of a lysyl substrate to form another tetrahedral oxyanion intermediate with Trp327, which results in the crosslinked product. From the sequence alignment, the two tryptophan residues are located in corresponding positions in TGase 6, Trp238 and Trp330 (Fig. 7.4), and could also have a similar involvement in the enzyme reaction mechanism. Although the relative positions of the tryptophan residues are conserved in all TGase structures studied it has been observed that the charge distribution in the region of the proposed glutamyl substrate binding site is different (Weiss et al., 1998). Such differences may explain the different substrate specificities between the enzymes.

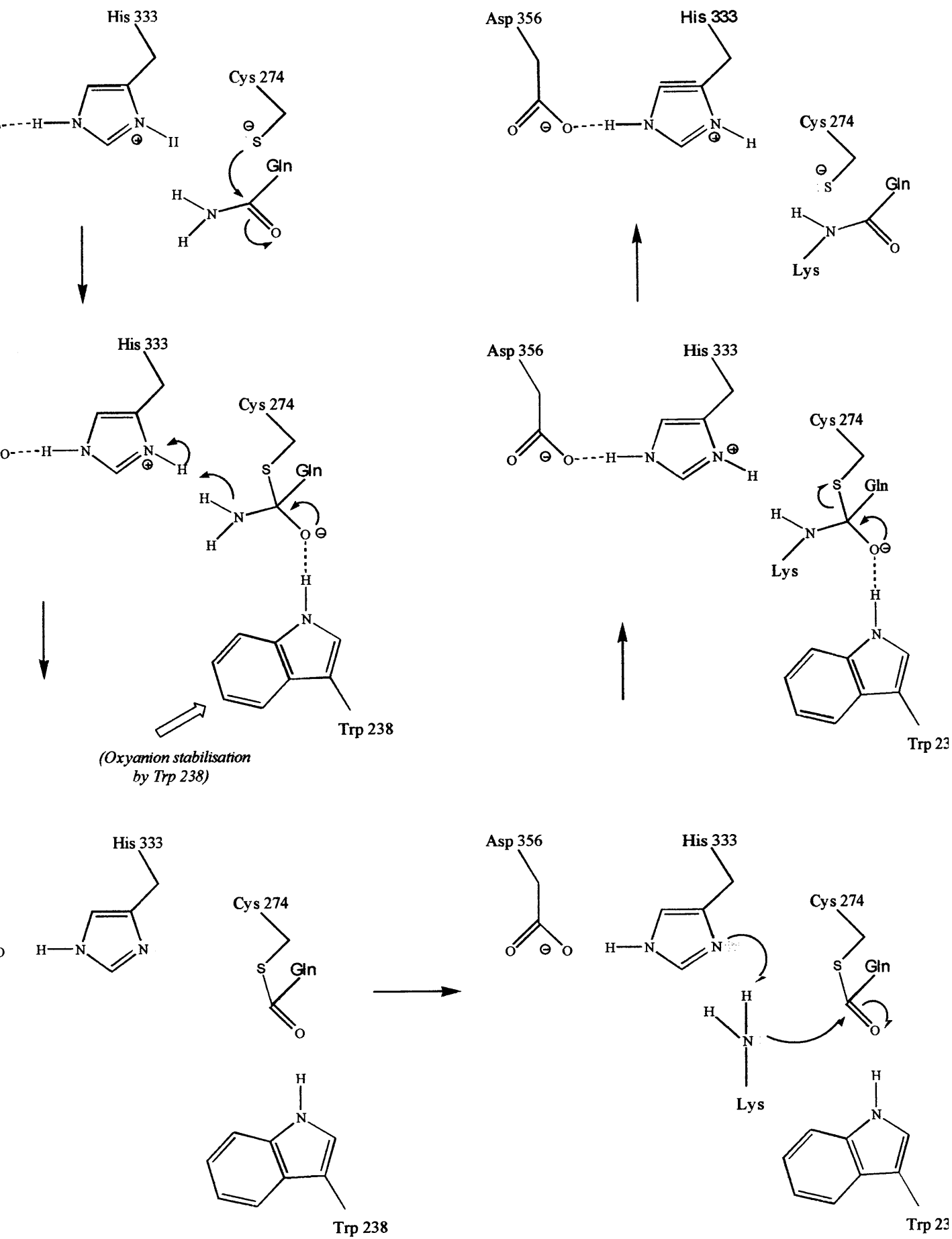


Fig. 7.3 Reaction pathway and a proposed mechanism for TGase 6-catalysed transamidations, based on the proposed TGase 2 reaction mechanism. Active site Cys274 and His333 form a thiolate-imidazolium ion pair. Binding of the glutamine (Gln) enables the thiolate to attack the γ -carboxamide group in the glutamine residue. The oxyanion intermediate is stabilised by H-bonding to Cys274 and Trp238. Nucleophilic attack by the amino group of the second substrate, lysine, leads to formation of a second oxyanion intermediate which is again stabilised by H-bonding interactions to Cys274 and Trp238. In the last step, the crosslinked product is released and TGase 6 regains its original conformation.

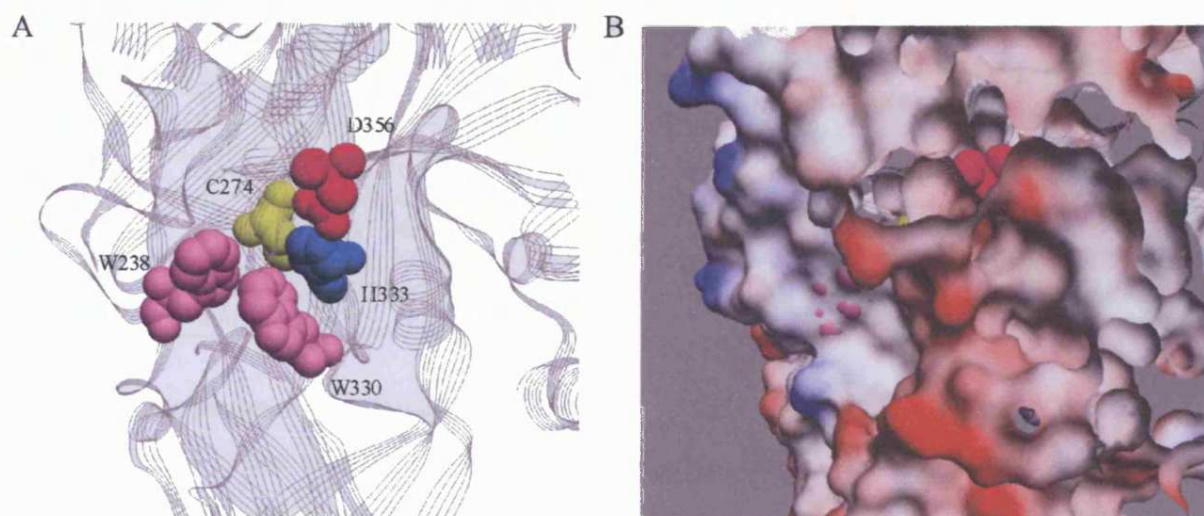


Fig. 7.4 View of the key residues involved in the active site predicted from the model of TGase 6. A. The catalytic triad residues Cys274 (yellow), His333 (blue) and Asp356 (red) are shown in CPK style. The two key tryptophan residues, Trp238 and Trp330 are shown in pink and CPK style. B. The Connolly surface (transparent) shows that the active site residues are buried and inaccessible.

7.2.2 Function of *cis* peptide bonds

Non-proline *cis* peptide bonds are rather rare (Stewart et al., 1990) and found in only 0.03% of the peptide bonds in solved protein structures (Weiss et al., 1998, Jabs et al., 1999). In the TGase 3 structures (Ahvazi et al., 2002, 2003) three peptide bonds have been identified that have *cis* conformation, two of which are non-proline *cis* peptide bonds. One is positioned at Arg268-Tyr269 in the β -strand and the loop just before the α -helix that contains the active site Cys272 residue. The second is located at Asn383-Phe384 in a loop joining two α -helices of the core domain positioned above the active site region. Additionally, Gly367-Pro368 is a proline *cis* peptide bond and

...the ... of ...
...the ... of ...
...the ... of ...
...the ... of ...



...the ... of ...
...the ... of ...

...the ... of ...
...the ... of ...

...the ... of ...
...the ... of ...

...the ... of ...
...the ... of ...

...the ... of ...
...the ... of ...

is found alongside the Asn383-Phe384 *cis* peptide bond. These bonds are also found in the same positions in the other solved TGase structures (Weiss et al., 1998) and sequence alignments have shown that their locations and flanking sequences are conserved throughout the TGase family, as shown in Table 7.1.

Table 7.1. Sequence alignment of human and mouse TGase 6 with other members of the TGase family in the regions of the *cis* peptide bonds. The residues connected by a *cis* peptide bond are coloured red. Adapted from (Weiss et al., 1998).

| Sequence | Region 1 | Region 2 | Region 3 |
|---------------|--------------------------------|--------------------------------|------------------------------|
| Mouse TGase 6 | 263 KGRYKPV KY GQCWVFAG | 378 REGDVHL AH DGPFVFAE | 364 EGMFRC GP ASVTAIR |
| Human TGase 6 | 263 KGRYKPV KY GQCWVFAG | 378 REGDVHL AH DGPFVFAE | 364 EGVFRC GP ASVTAIR |
| Human TGase 3 | 261 KSGFSPV RY GQCWVFAG | 375 REGDVQL NF DMPFIFAE | 361 QGVFQCGP ASVIGVR |
| Human FXIIIa | 304 SSE.NPV RY GQCWFAG | 418 KHGHVCF QF DAPFVFAE | 404 DGMYRCGP ASVQAIR |
| Human TGase 1 | 367 RTG.YSV PY GQCWFAG | 481 KNGLVY MKY DTGFIFAE | 367 SGIFCC GP CSVESIK |
| Human TGase 2 | 266 NHGCQRV KY GQCWVFAG | 380 KEGDLST KY DAPFVFAE | 365 EGTGCC GP VVRAIK |
| Human TGase 4 | 258 NTK.QAV CF GQCWFAG | 367 RKGDI FIVY DTRFVFSE | 353 QGVFCCGP SPLTAIR |
| Human TGase 5 | 266 ATGCQPV RY GQCWVFAG | 381 KEGEVDL NY DTPFVFMS | 367 NGVYCC GP ASVRAIK |
| Human TGase 7 | 268 ARGGQPV KY GQCWVFAS | 383 REGDVHL AY DTPFVYAE | 369 SGIFCC GP ASVKAIR |

It is thought that the stabilisation of the energetically unfavourable *cis* peptide bonds in proteins is due to considerable hydrogen bonding in addition to hydrophobic side chain interactions with adjacent residues (Jabs et al., 1999). Analysis of the available structural data revealed that the *cis* bonds in TGase 3 and other isoforms are in fact extensively bonded with their adjacent residues. This complex network of bonds in TGase 3 is proposed to draw together the two *cis* peptide bonds and the catalytic triad residues suggesting that a possible function of the associations is to maintain the active site residues in a fixed orientation in the event of a substrate approaching. This would support the proposition by Weiss *et al.* that the three *cis* peptide bonds are important for the stabilisation, activation and/or mechanism of action of the enzymes (Weiss et al., 1998).

From sequence alignment and the structural model, three *cis* peptide bonds were identified in TGase 6. The first non-proline *cis* peptide bond Arg268-Tyr269 in TGase 3 is replaced by Lys270-Tyr271 in TGase 6. Fig. 7.5A shows that although the arginine is replaced by a lysine in TGase 6, superposition results in the same orientation of the two residues and the formation of a *cis* bond between Lys270 and Tyr271. These two residues also form a *cis* peptide bond in TGase 2. The other non-

THE
MAGAZINE
OF THE
ROYAL SOCIETY

OF THE
ROYAL SOCIETY
OF LONDON

1880

THE
MAGAZINE
OF THE
ROYAL SOCIETY

OF THE
ROYAL SOCIETY
OF LONDON

proline *cis*-peptide bond, Asn383-Phe384 in TGase 3 is replaced by Ala385-His 386 in TGase 6. Superposition of the residues (Fig. 7.5B) shows that the orientation of the different residues is the same and thus, these residues are connected to a *cis* peptide bond. The proline *cis* peptide bond Gly337-Pro368 in TGase 3 is also located in a similar position in TGase 6, Gly370-Pro371, and superposition shows that the residues (Fig. 7.5C) all have the same orientation and a *cis* bond is found in the model. Therefore, it can be proposed that like the other structurally solved TGases, TGase 6 also contains three *cis* peptide bonds and these bonds may play a similar role in this enzyme.

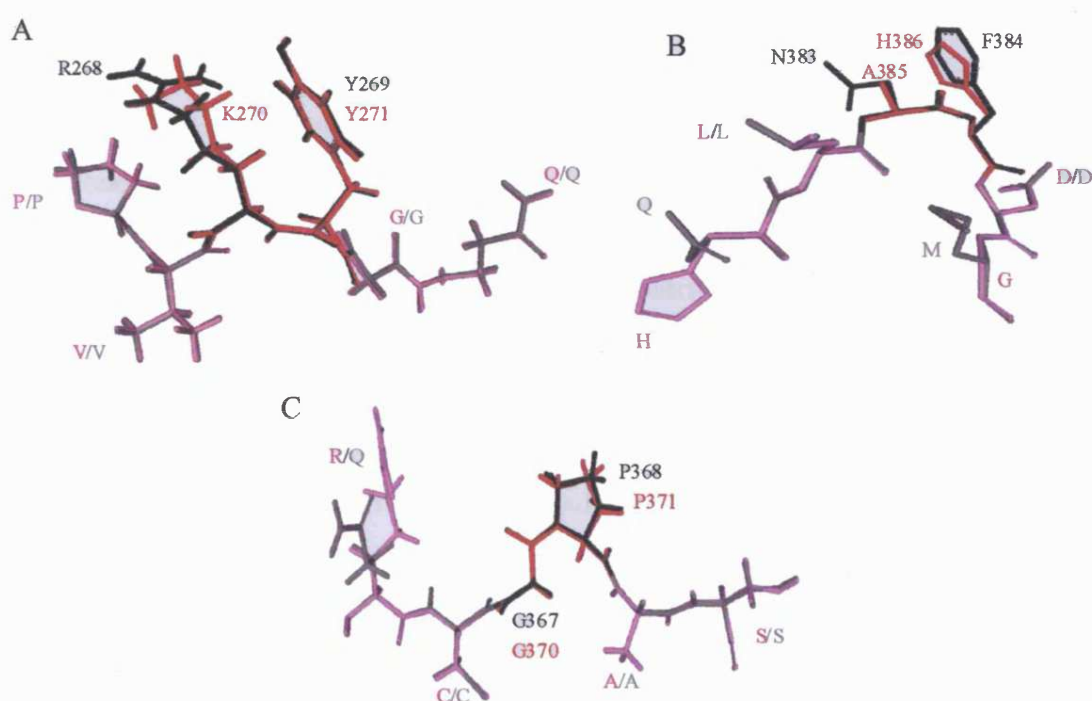


Fig. 7.5 Superposition of TGase 6 and TGase 3 residues involved in *cis* peptide formation. TGase 6 residues are shown in red (involved in *cis* peptide bond formation)/pink (adjacent to *cis* peptide bond) and TGase 3 residues are shown in black (involved in *cis* peptide bond formation) /grey (adjacent to *cis* peptide bond), and are shown in stick representation. A. Superposition of TGase 6 residues, Lys270-Tyr271, with TGase 3 residues, Arg268-Tyr269. B. Superposition of TGase 6 residues, Ala385-His386, with TGase 3 residues, Asn383-Phe384. C. Superposition of TGase 6 residues (red) Gly370-Pro371, with TGase 3 residues, Gly367-Pro368.

Consequently, while *cis* peptide bonds are energetically unfavourable, those found in TGases are tightly bonded together with adjacent structural motifs. However, the role of *cis* peptides is still unclear. Nevertheless, the fact that *cis* peptide bonds are rare in

... ..

... ..



... ..

... ..

protein structures and those found in TGases are in such positions as near to the active site indicates they may have a functional role and this is supported by the fact that the *cis* peptide bonds are conserved within the TGase family.

7.2.3 Calcium binding sites

The necessity of Ca^{2+} ions for TGase reactions is well established (Folk and Chung, 1985, Lorand et al., 1987, Melino et al., 1998). Nevertheless, the number of ions, their position in the three-dimensional structures of TGases and the role(s) in the activation and function are not yet known. The three-dimensional structures of the zymogen and activated forms of human fXIIIa enzyme have been determined (Yee et al., 1994, 1996, Weiss et al., 1998). The activated enzyme contains a single Ca^{2+} ion positioned near residues 473-490 in the catalytic core domain (Fox et al., 1999). Binding of a Ca^{2+} ion at this site does not markedly alter the structure and does not cause activation as such, but may in some way contribute to the subsequent substrate binding. It has been proposed that human TGase 2 can bind up to six Ca^{2+} ions (Bergamini, 1998), but their position within the structure remains unknown. It has been demonstrated that the zymogen form of human TGase 3 can bind one Ca^{2+} ion, but when proteolytically activated, it can bind three Ca^{2+} ions (Ahvazi et al., 2002).

7.2.3.1 Ca^{2+} ion binding site 1

The Ca^{2+} binding site 1 in the zymogen form of TGase 3 consists of residues Asn224-Asn229 (Ahvazi et al., 2002). The Ca^{2+} ion assumes six coordinations by making direct contacts with the main chain carbonyl oxygen atoms of Ala221, Asn224, Asn226 and Asn229, the carbonyl side chain oxygen of Asn224 and a water molecule. The positions of this Ca^{2+} ion site remains in the activated TGase 3, however, changes occur, whereby the Ca^{2+} ion is heptacoordinated by making direct contacts with the main chain carbonyl oxygen atoms of Ala221 (oxygen atom O), Asn224 (O) and Asn226 (O), the carbonyl side chain oxygen of Asn224 (OD1) and Asp228 (OD1 and OD2) and a water molecule. A loop consisting of Ile223-Val231 containing Asn229 has moved away and Asp228 coordinates with the ion instead. Data has indicated that the Ca^{2+} ion is now buried due to the movement of the Asp228 side chain.

From sequence alignment, it can be proposed that TGase 6 can also bind a Ca^{2+} ion at this site. Superposition of the residues involved in Ca^{2+} binding, Ala223, Asn226, Asn228 and Asp230 for TGase 6 shows that these residues are in the same orientation as the corresponding residues of TGase 3. Fig 7.6A shows that the oxygen atoms of the residues are in a position where binding to a Ca^{2+} ion could take place, Ala223 (O), Asn226 (O), Asn228 (O) and Asp230 (OD1 and OD2).

It has been suggested that this Ca^{2+} ion may be needed to maintain the correct three-dimensional structure of the active site region of TGase 3 and due to the changes in its coordination with residue side chains following activation, its presence may be essential for activity. However, the binding of only one Ca^{2+} ion is not sufficient for activity. Due to the conservation of the residues involved in Ca^{2+} binding at this site, it could be proposed that TGase 6 also binds a Ca^{2+} ion at this site.

7.2.3.2 Ca^{2+} ion binding site 2

One of the two newly acquisitioned Ca^{2+} binding sites in the activated TGase 3 is located near the end of the catalytic core domain adjacent to the loop which joins the barrel 1 domain, incorporating residues Asn430-Glu448 (Ahvazi et al., 2002). The binding site corresponds to the only identified Ca^{2+} binding site in human fXIIIa (Fox et al., 1999). The Ca^{2+} ion forms a heptacoordinated conformation making direct contacts with the carbonyl side chain oxygens of Asn393 (OD1), Glu443 (OE1 and OE2) and Glu448 (OE1), the main chain carbonyl oxygen atom of Ser415 (O) and direct links to two water molecules. Binding of the Ca^{2+} ion results in only minor changes in shape which do not result in any obvious effects on the conformation of residues near the active site.

Sequence alignment of TGase 3 and TGase 6 indicates that residues, Asn396, Glu445 and Glu450 are conserved in TGase 6, whereas, Ser415 is replaced by Thr417. Superposition of the two structures (Fig. 7.6B) shows that this does not alter the orientation of the oxygen atoms required for coordination with the Ca^{2+} ion; Asn396 (OD1), Glu445 (OE1 and OE2), Glu450 (OE1) and Thr417 (O) are in a position to coordinate a Ca^{2+} ion at this site. Therefore it can be proposed that TGase 6 can also bind a Ca^{2+} ion at this site. However, data obtained for TGase 3 indicates that a loop

of residues 462-471 in the zymogen obstructs binding of Ca^{2+} ions at site 2 and 3 and is removed on cleavage of Ser469. It is not yet known if TGase 6 is proteolytically cleaved, thus it cannot be determined if a Ca^{2+} ion could obtain access to this site, or if it is possible that the loop in TGase 6 does not obstruct entry to the site and cleavage may not be required.

7.2.3.3 Ca^{2+} ion binding site 3

The third Ca^{2+} ion binding site in the activated TGase 3 enzyme is unique. It is found close to a loop region incorporating residues Asp320-Ser325, prior to the catalytic triad residue His330 in the active site (Ahvazi et al., 2002). The exact position of this loop has changed in the activated TGase 3 when compared to the zymogen. In the activated enzyme, the loop has shifted and is directly bonded to the Ca^{2+} ion via the carbonyl side chain oxygen of Asp324, resulting in the opening of a channel through the protein. The Ca^{2+} ion is heptacoordinated with direct contacts with the carbonyl side chain atoms of Asp301 (OD1), Asp303 (OD2) and Asp324 (OD2), the side chain atoms of Asn305 (OD1 and ND2), the main chain carbonyl oxygen of Ser307 (O) and a water molecule. Sequence alignment with TGase 6 shows that all the residues are conserved within TGase 6. Superposition of the two structures (Fig. 7.6C) demonstrates that the orientation of the residues is the same and that the orientation of the oxygen atoms in TGase 6, Asp303 (OD2), Asp305 (OD2), Asp327 (OD2), Asn307 (OD1 and ND2) and Ser309 (O) is consistent with TGase 3. This suggests that TGase 6 could be capable of binding a Ca^{2+} ion at this site.

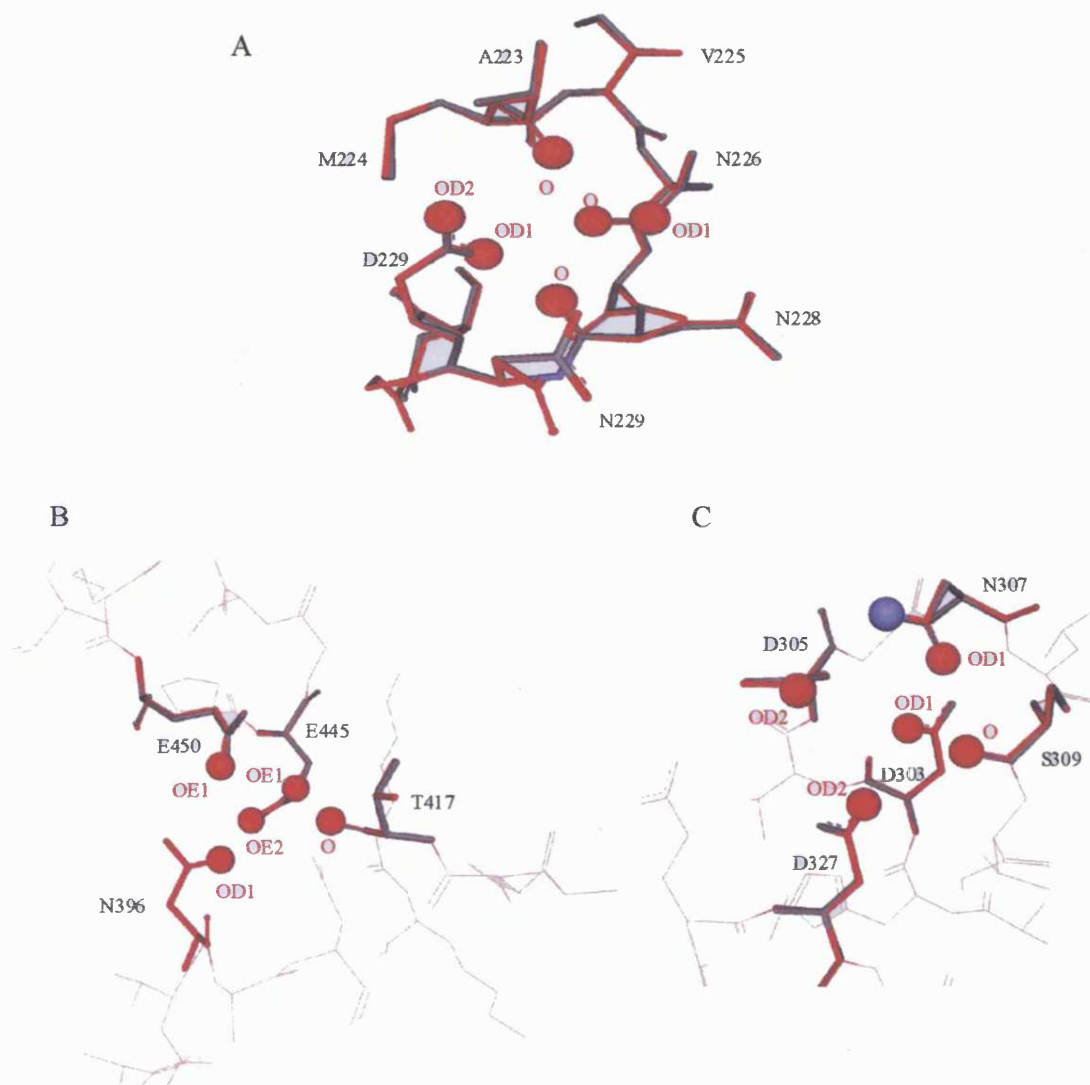


Fig. 7.6 Superposition of the key TGase 6 and TGase 3 residues involved in Ca^{2+} binding. The TGase 6 residues (red) and TGase 3 residues (grey) are shown in stick representation. The oxygen atoms involved in coordination with the Ca^{2+} ion are coloured red and labelled and the nitrogen atom is coloured blue. The $\text{C}\alpha$ atoms of adjacent residues are shown in red. A shows a superposition of the residues involved in Ca^{2+} binding residues at site 1, B shows Ca^{2+} binding residues at site 2, and C, Ca^{2+} binding residues at site 3.

Trp236 and Trp327 in TGase 3 are uncovered by the movement of the loop and the opening of the channel. The corresponding tryptophan residues in human fXIIIa, TGase 2 and fTGase are hypothesised to have several important roles in the reactions of the enzymes. The two residues are buried and need to be exposed. Shifting of their side chains would enable direct access of substrates to the buried active site, nevertheless, it has not as yet been demonstrated how these two residues become

exposed, enabling the conformations of the side chains to change allowing enzyme reaction. Also, it is thought that the two tryptophan residues play a role in oxyanion intermediate formation with the acyl glutamine donor first, followed by the acceptor lysine substrate (Pedersen et al., 1994). Due to the close identity of sequences, it was suggested that these events should also occur in TGase 3, specifically as the Ca^{2+} ion-induced formation of the channel and the resultant exposure of the two tryptophan residues will assist the movements of their side chains as appropriate substrates approach. Furthermore, since TGase 6 shows the highest similarity to TGase 3, it is possible that these events could also occur in TGase 6.

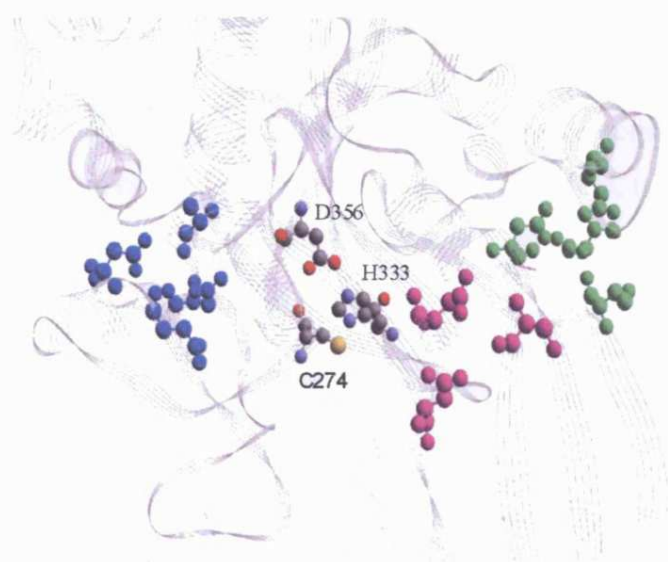


Fig. 7.7 View of the key residues involved in the Ca^{2+} ion binding sites and the active site of TGase 6. The residues involved in Ca^{2+} binding at site 1 (blue), site 2 (green) and site 3 (purple) are shown as ball and stick. The catalytic triad residues, Cys274, His333, and Asp359 are also shown.

In summary, it is suggested that the binding of three Ca^{2+} ions, particularly at site 3, enables the approach of substrate and enables activation of the enzyme. It is also feasible that these changes may regulate the substrate specificity of TGase 3. Based on the sequence alignment and similar structural features of TGase 6, it can be proposed that TGase 6 could also bind three Ca^{2+} ions (Fig. 7.7), thereby activating the enzyme. However, until it is determined whether the enzyme is proteolytically cleaved to enable access to the second and third sites, and if the two key tryptophan residues become exposed, it can only remain a hypothesis.

7.2.4 Transamidase activation mechanism

The main problem in comprehending the transamidase function of the TGases is determining how the buried active site becomes accessible to substrates. In all TGase structures reported so far, the active site Cys residue forms a hydrogen bond with a Tyr residue (Tyr525 in TGase 3) in barrel 1, which prevents access of the active site to solvent, and consequently to substrates. An initial suggestion predicted that substrates should approach the enzyme, shifting certain areas on the surface to expose the two Trp residues and breaking the hydrogen bonds concealing the active site Cys residue. Data on the structure of the active form of TGase 3 (Ahvazi et al., 2002) has revealed that the side chains of the Trp residues are exposed after the binding of Ca^{2+} ions in sites two and three, and consequently, one stage in making the enzyme accessible for substrates can be explained. It has also been demonstrated that in TGase 3, a tunnel appears that coincides with its activation. Therefore, it was speculated that the tunnel could provide a point of entry for the two substrates.

To determine how the two substrates may position themselves in the active site of TGase 3, molecular modelling and docking studies were used (Ahvazi and Steinert, 2003a).

7.2.4.1 Glutamine pocket

Studies by Ahvazi *et al.* (Ahvazi and Steinert, 2003a) indicated that the glutamine substrate should approach TGase 3 by moving under the loop Gly238-Pro246 to reach Cys272, and by shifting the loop, access to Trp236 will be gained. The movement of the loop region leads to the formation of the thiol intermediate/oxyanion complex with Trp236. Breakage of the hydrogen bond between Tyr525 and Cys272 enables displacement of a loop of sequences containing Tyr525 in the β -barrel domain which leads to exposure of the Cys residue at the active site. This movement can only take place once the TGase 3 enzyme has been proteolytically cleaved at Ser469. In TGase 3, the exposed hydrophobic pocket is lined by residues Arg247, Phe329, Trp236, Gln271, Trp273, Trp327 and Asn328.

Sequence alignment of TGase 3 with TGase 6 shows that all of the residues are conserved within TGase 6, with the exception of Arg247 in TGase 3, which is replaced by Lys249 in TGase 6. However, there are significant substitutions leading to important differences in residue interactions with regards to the other TGase enzymes: Arg247 in TGase 3 is a tyrosine in fTGase, a serine in fXIIIa and a methionine in TGase 2 (and a lysine in TGase 6). In TGase 3, Arg247 forms a hydrogen bond with the carbonyl side chain of Asp566 found in the β -barrel 2 domain. In TGase 6, it needs to be determined whether Lys249 can form a hydrogen bond with the equivalent Asp579, or if a hydrogen bond is actually required. Furthermore, Val331 in TGase 3 is particularly different from fXIIIa, fTGase and TGase 2 which all have a Cys residue at the equivalent location. In fXIIIa and fTGase, the loss of a hydrogen bond between Tyr372 in fXIIIa and Tyr247 in fTGase, and the active site Cys residue would result in the formation of a disulphide bond that would result in inhibition of the enzymes (Noguchi et al., 2001). It has been proposed that the disulphide bond is broken by the approaching glutamine substrate thereby activating the enzyme. In contrast, TGase 3 becomes a functional enzyme when it is proteolytically processed and therefore, the Val331 residue in TGase 3 may be a critical residue that regulates enzyme activity with regards to the glutamine substrate. Val334 is conserved in TGase 6 and may have a similar important role. The reported TGase structures show that the tryptophan residue (Trp236 in TGase 3) is suitably positioned to form a hydrogen bond with the main chain nitrogen atom of Cys272, thereby stabilising the acyl-enzyme intermediate. As the tryptophan residue is conserved in TGase 6, it could be assumed that similar bonding occurs in this enzyme.

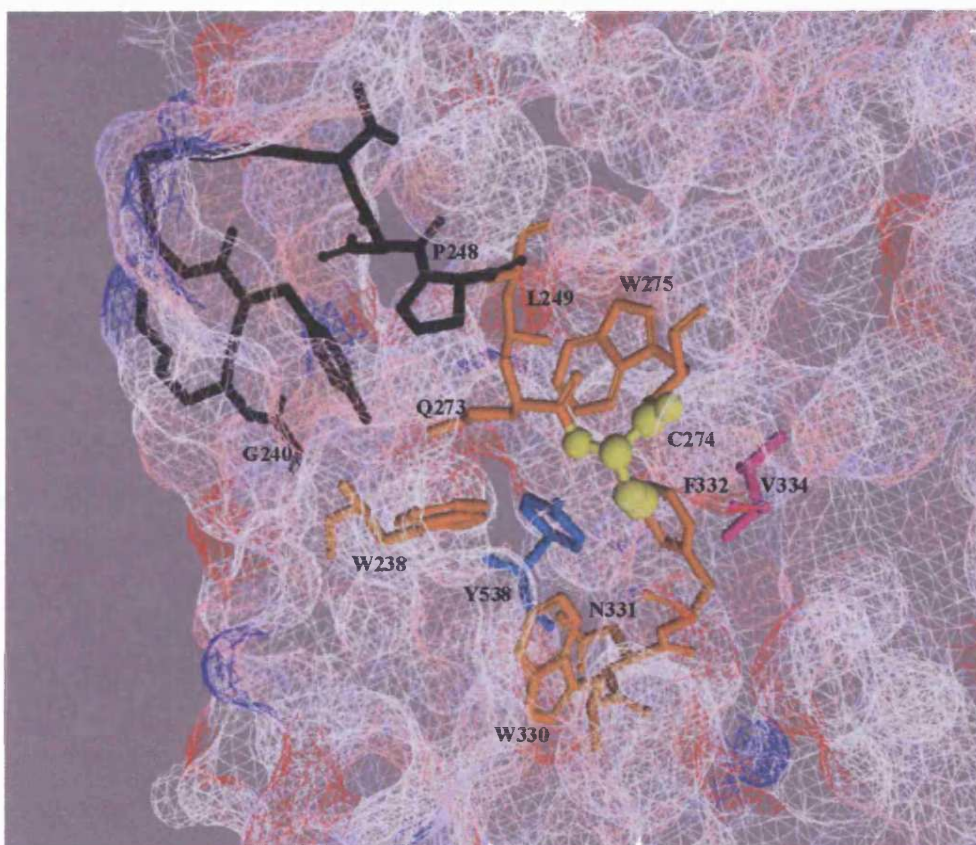


Fig. 7.8 View of the key residues of TGase 6 located in the glutamine pocket. The active site Cys274 (yellow) is shown as ball and stick. The residues proposed to be involved in coordination with a glutamine substrate (orange) and the two key tryptophan residues, Trp238 and Trp330 (orange) are shown as stick representation. The valine residue (Val333 in TGase 3, and possibly TGase 6) which is proposed to have a key role in TGase 3, and possibly TGase 6, is pink, and the loop Gly240-Pro248 (black), which the substrate is thought to move under in order to reach the active site Cys residue are shown as stick. The Tyr residue (blue) (Tyr538 in TGase 6) whose hydrogen bond with Cys274 needs to be broken to expose Cys274 is shown in stick. The image is shown with the Connolly surface overlaid.

7.2.4.2 Lysine pocket

In TGase 3, the lysine pocket is lined with the non-catalytic His300 residue and residues Trp327, Phe275, Glu358 and Glu391. Molecular docking shows that the lysine substrate can only approach the thioacyl intermediate by moving over the Ca^{2+} ion at site 1 and along the longest loop (Ala354-Gln365) (Ahvazi and Steinert, 2003a). Sequence alignment indicates that all these residues are conserved within TGase 6, His302, Trp330, Phe277, Glu361 and Glu394. In the solved structures of



The image shows a dark, textured surface, possibly a book cover or a photograph of a textured surface. The colors are predominantly dark purple and black, with some lighter, possibly metallic or reflective, areas. The texture is somewhat grainy and uneven, suggesting a physical object rather than a digital graphic.

Below the image, there is a block of text. The text is very blurry and difficult to read, but it appears to be a paragraph of text. The text is arranged in several lines, with some words appearing to be in a different font or style than others. The overall appearance is that of a scanned document or a photograph of a page of text.

TGase 3, His300 is hydrogen bonded to Glu358. From the predicted structure of TGase 6 it appears that His302 and Glu361 are positioned too far apart for hydrogen bonding to occur between the residues (Fig. 7.9). These differences could contribute specificity to each of the TGase enzymes with regards to the acceptance of lysine substrates. Additionally, in TGase 3, Trp327 is positioned opposite the Trp236 residue involved in oxyanion formation (Yee et al., 1994). It is thought that these residues therefore produce the tetrahedral oxyanion intermediate and stabilise both the glutamine and lysine substrates. From the predicted structure of TGase 6, it is shown that the two tryptophan residues are also positioned opposite each other (Fig. 7.9) and could contribute in a similar way. Following the reaction of the lysine with the acyl-enzyme intermediate, the crosslinked product is released.

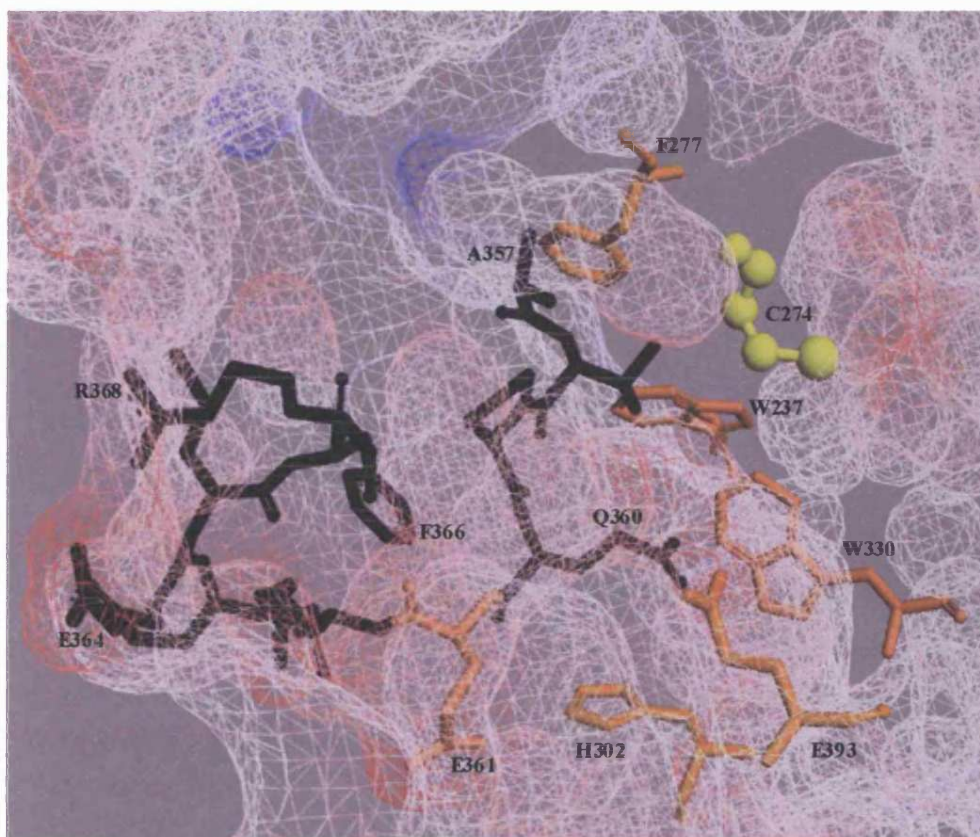


Fig. 7.9 View of the key residues of TGase 6 located in the lysine pocket. The active site Cys274 (yellow) is shown as ball and stick. The residues proposed to be involved in coordination with a lysine substrate (orange) and the loop Ala357-Arg368 (black), which the lysine substrate has to pass over, are shown as stick representation. The image is shown with a Connolly surface overlay.

Table 7.2. Tabulated drawing of the key residues involved in the two parts of the TGase reaction. The residues have been divided into three groups, those which determine the glutamine pocket, the lysine pocket or are common to the region. Residues that are different are highlighted blue and those which are conserved yet make different interactions are highlighted red. Adapted from (Ahvazi et al., 2003)

| Transglutaminase | TGase3 | TGase 6 | fXIIIIa | fTGase | TGase 2 |
|--------------------|--------|---------|---------|--------|---------|
| Glutamine pocket | Arg247 | Lys249 | Ser290 | Tyr247 | Met252 |
| | Asp566 | Asp579 | Glu601 | His566 | Asn559 |
| | Phe329 | Phe332 | Tyr372 | Phe331 | Phe334 |
| | Gln271 | Gln273 | Gln313 | Gln271 | Gln276 |
| | Trp273 | Trp275 | Trp315 | Trp272 | Trp278 |
| | Trp327 | Trp330 | Trp330 | Trp329 | Trp332 |
| | Asn328 | Asn331 | Asn371 | Asn330 | Asn333 |
| | Trp236 | Trp238 | Trp279 | Trp236 | Trp241 |
| Lysine pocket | His300 | His302 | His342 | His300 | His305 |
| | Glu358 | Glu361 | Glu401 | Glu360 | Glu363 |
| | Glu391 | Glu394 | Glu434 | Glu393 | Glu396 |
| | Trp327 | Trp330 | Trp370 | Trp329 | Trp332 |
| | Phe275 | Phe277 | Phe317 | Phe275 | Phe280 |
| | Pro356 | Pro359 | Pro399 | Pro358 | Pro361 |
| | Trp236 | Trp238 | Trp279 | Trp236 | Trp332 |
| Active site pocket | Val331 | Val334 | Cys374 | Cys333 | Cys336 |
| | Trp332 | Trp335 | Trp375 | Trp334 | Trp337 |
| | Leu352 | Leu355 | Leu395 | Leu354 | Leu357 |
| | Thr355 | Thr358 | Thr398 | Thr357 | Thr360 |
| | Phe387 | Phe390 | Phe430 | Phe389 | Phe392 |
| | Tyr525 | Tyr538 | Tyr560 | Tyr515 | Tyr516 |

In summary, molecular modelling and docking studies for TGase 3 propose that the binding of the two additional Ca^{2+} ions converts the enzyme to the active form on the approach of the first glutamine substrate. The studies indicate that the energy required for the following reaction cycles may be provided by the binding energy of a specific glutamine substrate or by the residual bonding energy from the release of NH_3 from the previous reaction cycle. The studies propose that the Ca^{2+} ions have a role by creating a channel and by stabilising and providing energy which regulates access of the lysine substrate into the active site. Due to a high degree of conservation between TGase 3 and TGase 6 with regards to the residues important for the transamidation reaction mechanism, it could be proposed that TGase 6 undergoes a similar mechanism to that of TGase 3. However, until similar molecular modelling and

1. The first part of the document discusses the importance of maintaining accurate records of all transactions and activities. It emphasizes the need for transparency and accountability in financial reporting.

2. The second part of the document outlines the various methods and techniques used to collect and analyze data. It includes a detailed description of the experimental procedures and the statistical tools employed.

3. Results and Discussion

The results of the study are presented in this section. The data shows a clear trend of increasing activity over time, which is consistent with the theoretical model proposed. The statistical analysis confirms the significance of the observed differences between the two groups.

4. The final part of the document concludes the study and provides recommendations for future research. It suggests that further exploration of the underlying mechanisms is needed to fully understand the observed phenomena.

5. Conclusion

In conclusion, this study has provided valuable insights into the relationship between the variables under investigation. The findings support the hypothesis that the proposed model accurately describes the observed behavior. The results have important implications for the field and warrant further investigation.

docking studies are performed with a solved TGase 6 structure this remains a hypothesis.

7.2.5 GTP binding

It has been assumed that of all the TGases, only TGase 2 is able to bind GTP and is able to hydrolyse it to GDP (Lee et al., 1989, Im et al., 1990, Iismaa et al., 2000). GTP binding by TGase 2 inhibits calcium binding leading to the inhibition of TGase transamidation activity. In contrast, calcium binding inhibits GTP binding and leads to activation of the enzyme (Achyuthan et al., 1989). However, it is still unknown how calcium and GTP binding are able to reciprocally regulate the activity of the enzyme. Recent work by Ahvazi *et al.* (Ahvazi et al., 2004) has demonstrated that GTP binding is in fact not limited to the TGase 2 enzyme. Not only does their data demonstrate that TGase 3 interacts with GTP, but also that TGase 3 is able to hydrolyse GTP to GDP.

The GTP-binding site in TGase 3 was found in a deep pocket produced between the core domain and the β -barrel 1 domain. The pocket comprises of residues Asn168 and Arg169 from the core domain, and Lys485, Lys487, Val488, Met491, Leu429 (1st β -strand), Arg587, Asp588, Ile590, Leu591 and Asp592 (last β -strand) from β -barrel 1. All the recognition contacts to the guanine base involve the major groove face of the ring, comprising of N-1, O-6 and N-7 groups. Ile590 forms direct hydrogen bonds with O-6 and N-1 atoms of the guanine base through its backbone nitrogen and oxygen, respectively. O-6 is also involved in hydrogen bonding via a water-mediated interaction with the backbone nitrogen of Asp588. Arg587 makes a side chain hydrogen bond with the N-7 group of the guanine ring. The GTP pocket of TGase 3 determines one site of local structural rearrangement in comparison to the TGase 3 structure in the absence of a bound guanine nucleotide.

As TGase 6 has a high sequence identity to TGase 3, a sequence and structural comparison was made to determine if TGase 6 might also be able to bind GTP. Sequence alignment revealed several differences between the two enzymes with regards to the residues located in the GTP binding pocket. The equivalent residues for

TGase 6 were Lys170 and His171 from the core domain (Asn168 and Arg169 in TGase 3), Lys498, Lys500, Val505, Pro504 and Pro505 from the first β -strand (Lys485, Lys487, Val488, Met491, Leu429 in TGase 3), Lys600, Asp601, Thr603, Leu604, and Glu605 from the last β -strand in barrel 1 (Arg587, Asp588, Ile590, Leu591, and Asp592 in TGase 3). Therefore, Ile590 in TGase 3, which forms hydrogen bonds with the O-6 and N-1 atoms of the guanine base is replaced by Thr603 in TGase 6. The O-6 is also hydrogen bonded to the nitrogen of Asp588 in TGase 3, and this residue, Asp601 is conserved in TGase 6. However, superposition of the two enzyme structures (Fig. 7.10) shows that the residues have a different orientation indicating that the residues in TGase 6 may not be able to make the same bonds as the residues in TGase 3. Arg587, which makes a side chain hydrogen bond to the N-7 group of the guanine ring in TGase 3 is replaced by Lys600 in TGase 6, and superposition of the two residues shows that the NH_2 group of Arg587 is in a different orientation to Lys600, again indicating that similar hydrogen bonding may not take place in TGase 6. In TGase 3, Arg169 stacks directly over the guanine ring, however this residue is His171 in TGase 6. Superposition of the two residues shows that the orientation of the residues is different and the ring of His171 is not in a position to stack over the guanine ring (Fig. 7.10). In TGase 3, the other side of the guanine ring moiety is sandwiched by the side chains of Val488, Leu492, Ile589 and Leu591 residues. In TGase 6, Val501 is conserved and in the same orientation; Leu492 is replaced by Pro505 which is in a different orientation; Ile589 is replaced by Thr603 which appears to be in the same orientation; and Leu591 is conserved as Leu604, however superposition reveals that the side chains are pointing in different directions (Fig. 7.10).

In the TGase-GTP complex, the α and β -phosphates are directed towards the side chain of Lys485, Lys487 and Arg587, and the γ -phosphate towards the side chain of Asn168. In TGase 6, the two lysine residues, Lys498 and Lys500 are conserved and appear to be in the same orientation; Arg587 is replaced by Lys600; and Asn168 is replaced by Lys170, and superposition shows that the side chain are different (Fig. 7.10). Interestingly, the two lysine residues in TGase 3, Lys485 and Lys487, which may have the role of a 'lysine finger' thought to be essential for stabilisation of the transition states for GTP hydrolysis are conserved in TGase 6. Also, the Asn168 in

TGase 3, predicted to facilitate in the hydrolysis reaction is not conserved in TGase 6 but is replaced with Lys170.

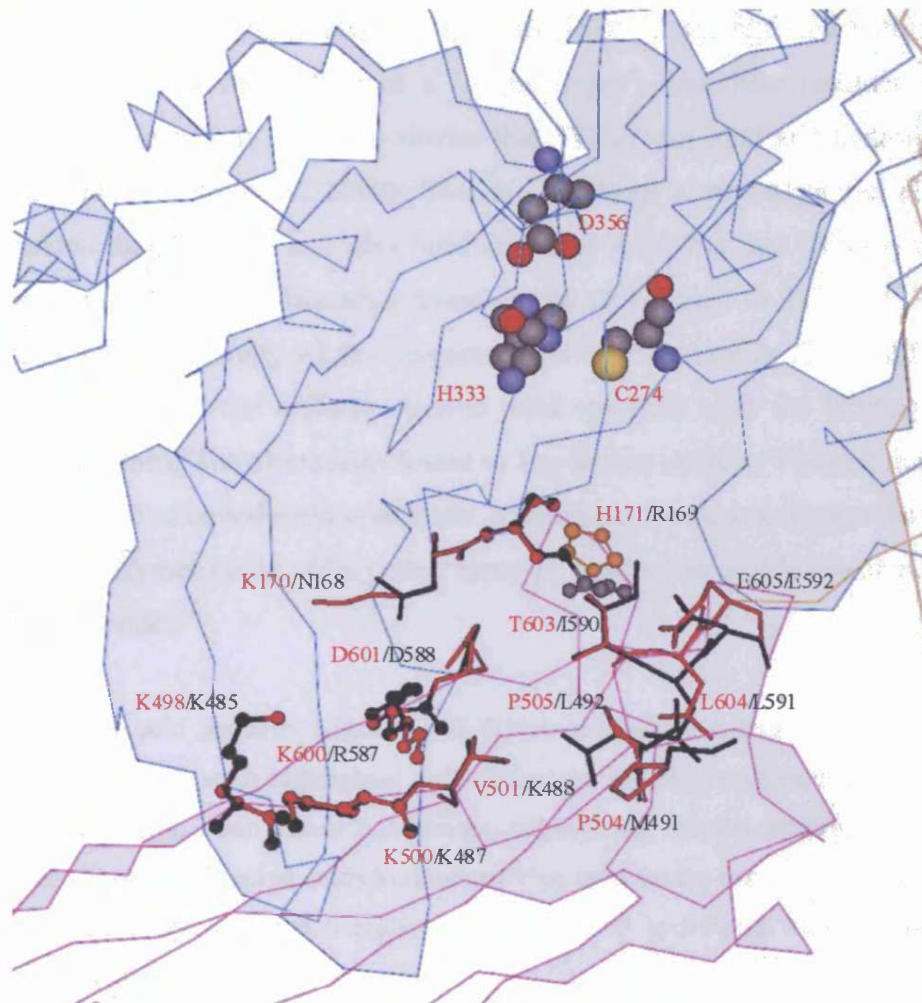


Fig. 7.10 Superposition of the key residues found in the guanine nucleotide-binding site pocket in the TGase 6 and TGase 3 complex. The TGase 6 residues (red) and the TGase 3 residues (black) are shown as stick. The active site residues for TGase 6, Cys274, His333 and Asp356 (coloured by atom) are shown in CPK. The domains are coloured as in Fig. 7.2 and the C α backbone is shown in ca-wire representation. The overlay images indicate the orientation differences in certain residues in both structures in the guanine nucleotide-binding site pocket.

In summary, it is of interest that the sequence analysis of TGase 2 with TGase 3 could not predict that TGase 3 would be able to bind and hydrolyse GTP. Additionally, even though TGase 3 and TGase 6 have a close sequence similarity it is not possible to predict if TGase 6 is also able to bind and hydrolyse GTP. Whereas TGase 2 uses a Phe residue for stacking over the guanine ring (Liu et al., 1999), TGase 3 uses an Arg

residue and based on sequence alignment, a His residue is found in this position in TGase 6. Furthermore, the two Arg residues, referred to as an 'arginine finger' needed for coordination of the γ -phosphate of GTP by TGase 2 are replaced with two Lys residues in TGase 3, referred to as a 'lysine finger', and these residues are also conserved in TGase 6. It has been shown that TGase can bind and hydrolyse both GTP and ATP (Iismaa et al., 2000). Studies by Ahvazi *et al.* (Ahvazi et al., 2004) demonstrated that TGase 3 can also bind GTP and ATP, but that the enzyme binds GTP more effectively. Studies also revealed that in contrast to TGase 2, TGase 3 shows little ATPase activity when compared to its GTPase activity. These differences could be a result of the different amino acid residues used for binding by the respective enzymes. The alterations found in key amino acids in TGase 2 and TGase 3, which are able to nonetheless contribute comparable function indicates that several other TGase enzymes could utilise other types of residues to enable interactions with guanine nucleotides.

Table 7.3 Amino acid sequence alignment of TGases around the guanine nucleotide-binding site pocket. The amino acids highlighted red indicate the Arg/Phe residues that stack over the guanine ring in TGase 2 and TGase 3 structures, respectively, and the corresponding residues in the other TGases. The amino acids highlighted blue indicate the two basic residues that are required for stabilisation of the transition states for GTP hydrolysis. Table adapted from (Ahvazi et al., 2004).

| | | | | | | | | | | | | | | | |
|-------|-----|--------|------|------|------|-----|-----|---|---|---|-----|---------|---------|-----|-----|
| TGM6 | 167 | GVEKH | IRAQ | GWN | YGQ | 181 | - | - | - | - | 495 | ITGKFKV | LEPP | 505 | |
| TGM2 | 169 | GS | AKF | IKN | IPWN | FQG | 183 | - | - | - | - | 472 | MAMRIRV | QSM | 482 |
| TGM3 | 165 | GSTNR | IGM | IGWN | FQG | 179 | - | - | - | - | 482 | IIGK | LKVAG | ML | 492 |
| FXIII | 215 | GEVND | IKTR | SWS | YSG | 229 | - | - | - | - | 518 | VDMDFE | VENAV | 528 | |
| TGM1 | 277 | GTEA | QIG | ERTW | NYG | 291 | - | - | - | - | 578 | VAMQ | VEAQ | DAV | 588 |
| TGM4 | 168 | GAARS | IKCK | PWN | FQG | 182 | - | - | - | - | 472 | LHMSV | QSDD | VL | 482 |
| TGM5 | 170 | GSKN | WIR | PCPW | NYG | 184 | - | - | - | - | 506 | VSLKF | KLLD | PP | 516 |
| TGM7 | 171 | GHER | FITS | WPWN | YGQ | 185 | - | - | - | - | 489 | AQLQ | LHLAR | IP | 499 |
| B4.2 | 168 | GTADCI | QAES | WDFG | | 182 | - | - | - | - | 474 | LYLLL | KAPSS | L | 484 |

Therefore, it may be possible that TGase 6 is able to bind and hydrolyse GTP by utilising different residues to those used by TGase 3. However, based on the residues located in the equivalent positions in TGase 3 by sequence alignment and superposition of those key residues, it does not seem likely that this is a function of TGase 6.

7.2.6 Conclusion

In summary, based on the close sequence similarity between the activated human TGase 3 and mouse TGase 6, the residues involved in enzyme activation, the transamidation activation mechanism, Ca^{2+} binding and GTP binding in TGase 6 were hypothesised. The key residues involved in the active site are conserved in TGase 6, Cys274, His333 and Asp356, and the two key tryptophan residues, Trp238 and Trp330 are found in a similar position and could therefore have a similar role to that found for the other members of the TGase family. From sequence and structural comparison between TGase 6 and TGase 3, three proposed *cis*-peptide bonds were identified in TGase 6 and due to the similarity of their location may have a similar function to that predicted for the other TGases. Also based on sequence and structural alignment, it could be hypothesised that TGase 6 is also able to bind three Ca^{2+} ions. However it needs to be determined whether Ca^{2+} ions can obtain access to the second and third sites. There is a high degree of conservation between TGase 6 and TGase 3 with regards to the residues associated with substrate binding in the glutamine and lysine pockets. It can therefore be proposed that TGase 6 has a similar transamidase activation mechanism to TGase 3, and that any differences observed could account for substrate specificity for each enzyme. Due to the different orientations of the residues involved in GTP binding for TGase 6 when compared to TGase 3, it seems unlikely that TGase 6 is involved in GTP binding. However, due to the fact that TGase 2 and TGase 3 can bind and hydrolyse GTP, yet have alterations in key amino acids involved in the process, GTP binding being a function of TGase 6 can not be ruled out. At this stage, these conclusions remain only a hypothesis and actual experimental work will need to be carried out in order to confirm these predictions.

7.3 Expression of TGase 6

The initial expression pattern of TGase 6 obtained from Northern dot blot analysis indicated that TGase 6 is present in the central nervous system. *In situ* hybridisation revealed that TGase 6 is expressed in the brain, particularly within the neurons of the cerebral cortex and the Purkinje cells of the cerebellum, and that the expression is developmentally regulated and occurs in conjunction with neuronal development. Immunohistological studies using antibodies generated against the peptide for TGase

6 which is common to all splice variants, confirmed the expression of TGase 6 within the brain, particularly in the cerebral cortex. Western blotting analysis on extracted tissue from various regions of mouse brain with the sera against TGase 6_L and TGase 6_S indicated that TGase 6 is expressed in all areas of the brain, namely in the spinal cord, cerebrum, medulla, and cerebellum. However a total of six different variants could be detected and these were expressed in a tissue-dependent manner. Expression of TGase 6_S, the splice variant lacking β -barrel 1 was only detected in the cerebrum. However, two different variants of TGase 6_S were present. Further work is needed to identify the biological significance of this complexity of molecular variants.

Based on the finding that TGase 6 is expressed in the main areas of the brain it may be hypothesised that it has a functional role within the brain. It is unclear at present whether this function relates to its structure, transamidating activity or possibly, GTP-binding. Equally, the normal function of other TGase isoforms expressed in the brain is only poorly understood. However, many studies have suggested that one of the processes responsible for the generation of abnormal inclusions in neurodegenerative diseases is TGase crosslinking. TGase catalysed isodipeptide bonds have been identified in abnormal inclusions in PD, AD, PSP and HD. Due to the finding that TGase 6 is expressed in the neurons of the brain, it may be possible that TGase 6 contributes to the generation of abnormal inclusions. It is also interesting that TGase 6 is predominantly expressed in the cerebral cortex and cerebellum, two areas of the brain greatly affected in AD and HD. It is of interest that TGase 6_S was only detected in the cerebrum, as the cerebral cortex of the cerebrum is one of the areas damaged in the brains of AD and HD sufferers. This may indicate that the alternatively spliced form of TGase 6 has a specific role in the cerebrum, which may be associated with neurodegenerative disease.

Studies on neurodegenerative diseases have mainly focused on TGase 2 as it is found in the neurons of the brain, its expression levels are increased significantly in specific areas of diseased brain and TGase 2 catalysed isopeptide bonds have been isolated from diseased brain. As TGase 6 shows significant expression in the brain, it would be invaluable to determine if its expression is also increased in diseased brain, particularly if the expression increases in specific areas of the brain affected by neurodegenerative diseases. Using quantitative PCR, our lab has determined that

various TGases are up-regulated in a tissue-specific manner in TGase 2 null mice suggesting an overlap in function among certain gene products. TGase 6 and TGase 2 are part of the same subgroup of gene products which are closely related and may therefore have comparable properties and substrate specificities. Based on its expression pattern and close relation to TGase 2, TGase 6 is a possible candidate to also contribute to the underlying TGase-related pathological changes.

7.4 Future directions

Based on the identification of novel TGases in mammals and other organisms, and the interest of the biotechnological industries in TGases, it seems fair to speculate that research into TGases will continue and may incorporate a broad range of research areas (Lorand and Graham, 2003).

New TGases, including a number from plants, will undoubtedly be identified and this may lead to the identification of new TGase activities, for example, different reaction mechanisms, or different regulatory mechanisms. The identification of new TGases and their functions, from lower organisms is likely to lead to various applications ranging from the food to the medical industry. For example, finding that TGase catalysed reactions have an important role in the growth, development and survival of nematodes may provide suitable biochemical targets to control conditions caused by these organisms in humans and animals. Therefore, it is probable that inhibition of TGase-catalysed reactions may provide parasitic-selective and less toxic agents to control and eliminate related diseases.

The applications for fibrin sealant in surgery will probably be expanded, as will the development and production of TGase-catalysed biocompatible polymers. The use of recombinant TGases in biotechnology may well be replaced by enzymes which are isolated from inexpensive natural sources and as the potential sources of TGase enzymes continues to increase, so will their potential applications. Therefore, TGases may not only be utilised for the production of food and in healthcare, but additionally, it may be incorporated to a greater extent into novel hair, skin and nail treatments, and in the manufacturing of clothing.

As discussed previously, TGases may have a role in neurodegenerative disorders by participating in the formation of various protein precipitates. Pharmaceutical interference to reduce brain TGase activity may therefore be beneficial. Several companies are working on the development of isoform specific inhibitors.

Relating to basic research, little is known about human TGases 5, 6 and 7. Another important factor which needs to be investigated is the extent to which the functions of TGase 2 can be replaced by other TGases. Also, information about protein sequences and genes of TGases from lower species is limited at the present time.

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**, 1901-1906

Achyuthan, K.E., Mary, A., Bhaerman, R., and Grenberg, C.S. (1989) Enzymatic activity of guinea pig liver transglutaminase. *Mol. Cell Biochem.* **85**, 57-65

Achyuthan, K.E., Rowland, T.C., Birckbichler, P.J., Lee, K.N., Bishop, P.D., and Achyuthan, A.M. (1996) Hierarchies in the binding of human factor XIII, factor XIIIa, and endothelial cell transglutaminase to human plasma fibrinogen, fibrin and fibronectin. *Mol. Cell Biochem.* **162**, 43-49

Aeschlimann, D., Kaupp, O., and Paulsson, M. (1995) Transglutaminase-catalysed matrix cross-linking in differentiating cartilage: identification of osteonectin as a major glutaminyl substrate. *J. Cell Biol.* **129**, 881-892

Aeschlimann, D., Koeller, M.K., Allen-Hoffman, 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**, 3542-3460

Aeschlimann, D., and Paulsson, M. (1991) Crosslinking of laminin-nidogen complexes by tissue transglutaminase. *J. Biol. Chem.* **271**, 16451-16454

Aeschlimann, D., and Paulsson, M. (1994) Transglutaminases: protein crosslinking enzymes in tissues and biological fluids. *Thromb. Haemostasis* **71**, 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**, 1-27

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.* **279**, 23834-23841

Ahvazi, B., Boeshans, K.M., Idler, W., Baxa, U., Steinert, P.M., and Rastinejad, F. (2004) Structural basis for the coordinated regulation of transglutaminase 3 by guanine nucleotides and calcium/magnesium. *J. Biol. Chem.* **279**, 7180-7192

Ahvazi, B., Boeshans, K., Jang, S.I., Kalinin, A.E., and Steinert, P.M. (2002a) Transglutaminases in skin biology. *Minerva Biotech.* **14**, 165-169

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**, 2055-2067

Ahvazi, B., and Steinert, P.M. (2003a) A model for the reaction mechanism of the transglutaminase 3 enzyme. *Exp. Mol. Medicine* **35**, 228-242

Akimov, S.S., Krylov, D., Fleishmann, L.F., and Belkin, A.M. (2000) Tissue transglutaminase is an integrin-binding adhesion coreceptor for fibronectin. *J. Cell Biol.* **148**, 825-838

Alston, T.A., Mela, L., and Bright, H.J. (1977) ϵ -Nitropropionate, the toxic substance of *Indigofera*, is a suicide inactivator of succinate dehydrogenase. *Proc. Natl. Acad. Sci. U.S.A.* **74**, 3767-3771

Ambrus, A., Banyai, I., Weiss, M.S., Hilgenfeld, R., Keresztesy, Z., Musheck, L., and Fesus, L. (2001) Ca^{2+} binding of transglutaminases: a ^{43}Ca NMR study combined with surface polarity analysis. *J. Biol. Struct. Dyn.* **19**, 59-74

Anderson, R.P., Degano, P., Godkin, A.J., Jewell, D.P., and Hill, A.V. (2000) *In vivo* antigen challenge in celiac disease identifies a single transglutaminase-modified peptide as the dominant A-gliadin T-cell epitope. *Nat. Med.* **6**, 337-342

Ando, M., Kunii, S., Tetematsu, T., and Nagata, Y. (1993) Selective alterations in transglutaminase activity of rat superior cervical ganglia in response to neurotransmitters, high potassium and sialic acid-containing compounds. *Brain Res.*, **604**, 64-68

Anwar, R., and Miloszewski, K.J. (1999) Factor XIII deficiency. *Br. J. Haematol.* **107**, 468-484

Appelt, D.M., Kopen, G.C., Boyne, L.J., and Balin, B.J. (1996) Localization of transglutaminase in hippocampal neurons: implications for Alzheimer's disease. *J. Histochem. Cytochem.* **44**, 1421-1427

Arima, K., Veda, K., Sunohara, N., Harari, S., Izumiyama, Y., Tonzuka-Vehara, H., and Kawai, M. (1998) Immunoelectron microscopic demonstration of NACCP/ α -synuclein epitopes on the filamentous component of Lewy bodies in Parkinson's disease and in dementia with Lewy bodies. *Brain Res.* **808**, 93-100

Asahina, T., Kobayashi, T., Okada, Y., Itoh, M., Yamashita, M., and Inamoto, Y. (1998) Morphological interactions of human first trimester placental villi cocultured with decidual explants. *Horm. Res.* **50**, 37-45

Asahina, T., Kobayashi, T., Okada, Y., Goto, J., and Terao, T. (2000) Maternal blood coagulation factor XIII is associated with the development of cytotrophoblastic shell. *Placenta* **21**, 388-393

Balklava, Z., Verderio, E., Adams, J., and Griffin, M. *Abstract 21. Procs. 6th Int. Conf. Transglut. And Protein crosslinking.* Lyon, Sept 2000

Baeuerle, P.A. (1991) The inducible transcription factor NF-kappaB: regulation by distinct protein subunits. *Biochim. Biophys. Acta.* **1072**, 63-80

Beal, M.F. (1998) Mitochondrial dysfunction in neurodegenerative diseases. *Biochim. Biophys. Acta.* **1346**, 211-223

Beal, M.F., Hyman, B.J., and Koroshetz, W. (1993) Do defects in mitochondrial energy metabolism underlie the pathology of neurodegenerative diseases? *Trends Neurosci.* **16**, 125-131

Beck, E., Duchert, F., and Ernst, M. (1961) The influence of fibrin stabilizing factor on the growth of fibroblasts *in vitro* and wound healing. *Thromb. Diath. Haemorrh.* **6**, 485-491

Benedetti, L., Grignani, F., Scicchitano, B.M., Jetten, A.M., Diverio, D., Lo Coco, F., Avvisati, G., Gambacorti-Passerini, C., Adamo, S., Levin, A.A., Pelicci, P.G., and Nervi, C. (1996) Retinoid-induced-differentiation of acute promyelocytic leukaemia involves PML-RAR α -mediated increase of type II transglutaminase. *Blood* **87**, 1939-1950

Bereron, C., Pollanen, M.S., Weyer, L., and Lang, A.E. (1997) Cortical degeneration in progressive supranuclear palsy. A comparison with cortical basal ganglionic degeneration. *J. Neuropathol. Exp. Neurol.* **56**, 726-734

Bergamini, C.M. (1998) GTP modulates Ca²⁺ binding and cation-induced conformational changes in erythrocyte transglutaminase. *FEBS Lett.* **239**, 255-258

Berget, S.M. (1984) Are U4 small nuclear ribonucleoproteins involved in polyadenylation. *Nature* **309**, 179-182

Bernassola, M., Boumis, G., Corazzari, M., Bertini, G., Citro, G., and Knight, R.A., (2002) Role of transglutaminase 2 in glucose tolerance: knockout mice studies and a putative mutation in a MODY patient. *FASEB J.* **16**, 1371-1378

Bowness, J., Venditti, M., Tarr, A., and Taylor, J. (1994) Increase in ϵ (γ -glutamyl)lysine crosslinks in atherosclerotic plaques. *Atherosclerosis* **111**, 247-253

Brouillet, E., and Hantraye, P. (1995) Effects of chronic MPTP and 3-nitropropionic acid in non human primates. *Curr. Opinion Neurol.* **8**, 469-473

Brouillet, E., Hantraye, P., Ferrante, R.J., Dolan, R., Leroy-Willig, A., Kowall, N.W., and Beal, M.F. (1995) Chronic mitochondrial energy impairment produces selective striatal degeneration and abnormal choreiform movements in primates. *Proc. Natl. Acad. Sci.* **92**, 7105-7109

Browne, S.E., Bowling, A.C., MacGarvey, U., Baik, M.J., Berger, S.C., Mugit, M.M., Bird, E.D., and Beal, M.F. (1999) Oxidative damage and metabolic dysfunction in Huntington's disease: selective vulnerability of the basal ganglia. *Ann. Neurol.* **41**, 646-653

Burgoyne, R.D., and Weiss, J.L. (2001) The neuronal calcium sensor family of Ca²⁺-binding proteins. *Biochem. J.* **353**, 1-12

Butterworth, J., Yates, C.M., and Reynold, G.P. (1985) Distribution of phosphate-activated glutamine succinic dehydrogenase, pyruvate dehydrogenase and γ -glutamyl transpeptidase in postmortem brain from Huntington's disease and agonal cases. *J. Neurol. Sci.* **67**, 161-171

Candi, E., Oddi, S., Paradisi, A., Terrinoni, A., Ranalli, M., Teofoli, P., et al. (2002) Expression of transglutaminase 5 in normal and pathological human epidermis. *J. Invest Dermatol.* **119**, 670-677

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**, 35014-35023

Candi, E., Paradisi, A., and Melino, G. (2003) *Proceedings of the workshop on protein cross-linking inneuronal inclusions, Transglutaminases in neurodegeneration.* May 22-25, Hungary.

Carmel, G., Mager, E.M., Binder, L.I., and Kuret, J. (1996) The structural basis of monoclonal antibody Alz50's selectivity for Alzheimer's disease pathology. *J. Biol. Chem.* **271**, 32789-32795

Carrell, N.A., Erickson, H.P., McDonagh, J. (1989) Electromicroscopy and hydrodynamic properties of factor XIII subunits. *J. Biol. Chem.* **83**, 551-556

Casadio, R., Polverini, E., Mariani, P., Spinozzi, F., Carsurghi, F., Fontana, A., Polverino de Laureto, P., Matteucci, G., and Bergamini, C.M. (1999) The structural basis for the regulation of tissue transglutaminase by Ca²⁺ ions. *Eur. J. Biochem.* **262**, 672-679

Chan, P.H., and Fishman, R.A. (1980) Transient formation of superoxide radicals in polyunsaturated fatty acid-induced brain swelling. *J. Neurochem.* **35**, 1004-1007

Chandrashekar, R., Tsuyi, N., Morales, T., Ozols, V., and Mehte, K. (1998) An Erp60 like protein from the filarial parasite *Dirofilaria immitis* has both transglutaminase and protein disulfide isomerase activity. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 531-536

Chun, W., Lesort, M., Tucholski, J., Faber, P.W., MacDonald, M.E., Ross, C.A., and Johnson, G.V. (2001) Tissue transglutaminase selectively modifies proteins associated with truncated mutant huntingtin in intact cells. *Neurobiol. Dis.* **8**, 391-404

Citron, B.A., SantaCruz, K.S., Davies, P.J.A., and Festoff, B.W. (2001) Intron-exon swapping of transglutaminase mRNA and neuronal Tau aggregation in Alzheimer's disease. *J. Biol. Chem.* **276**, 3295-3301

Clayton, D.F., and George, J.M. (1998) The synucleins: a family of proteins involved in synaptic function, plasticity, neurodegeneration and disease. *Trends Neurosci.* **21**, 249-254

Cohen, C.M., Dotimas, E., and Korsgren, C. (1993) Human erythrocyte membrane protein band 4.2 (pallidin). *Semin, Haematol.* **30**, 119-137

Coles, C.J., Edmondson, D.E., and Singer, T.P. (1979) Inactivation of succinate dehydrogenase by 3-nitropropionate. *J. Biol. Chem.* **254**, 5161-5167

Collighan, R., Cortez, J., and Griffin, M. (2002) The biotechnological applications of transglutaminases. *Minerva Biotec.* **14**, 143-148

Cook, R.D. (1974) Calcium-induced dissociation of human plasma factor XIII and the appearance of catalytic activity. *Biochem. J.* **141**, 683-691

Cooper, A.J.L., Jeitner, T.M., Gentile, V., and Blass, J.P. (2002) Cross linking of polyglutamine domains catalysed by tissue transglutaminase is greatly favored with pathological length repeats: Does transglutaminase activity play a role in (CAG)_n/Q_n-expansion diseases with special reference to the role of tissue transglutaminase and selective vulnerability. *Neurochem. Int.* **40**, 53-67

Cordella-Miele, E., Miele, L., Beninati, S., and Mukherjee, A.B. (1993) Transglutaminase-catalyzed incorporation of polyamines into phospholipase A2. *J. Biochem.* **113**, 164-173

Crawford, S.E., Nara, K., and Rifkin, D.B. (1998) Thrombospondin-1 is a major activator of TGF- β 1 *in vitro*. *Cell* **93**, 1159-1170

Credo, R.B., Curtis, C.G., and Lorand, L. (1978) Ca²⁺-related regulatory function of fibrinogen. *Proc. Natl. Acad. Sci. U.S.A.* **75**, 675-682

Cummings, C.J., and Zoghbi, H.Y. (2000) Fourteen and counting: unravelling trinucleotide repeat diseases. *Hum. Mol. Gen.* **9**, 909-916

Curtis, C.G., Brown, K.L., Credo, R.B., Domanik, R.A., Gray, A., Stenberg, P., and Lorand, L. (1974) Kinetics of transamidating enzymes. Production of thiol in the reactions of thiol esters with fibrinogenase. *Biochemistry* **13**, 3257-3262

Das, T., Baek, K.J., Gray, C., and Im, M-J. (1993) Evidence that the Gh protein is a signal mediator from α 1-adrenoceptor to a phospholipase C. *J. Biol. Chem.* **268**, 27398-27405

Davies, S.W., Turmaine, M., Cozens, B.A., DiFiglia, M., Sharp, A.H., Ross, C.A., Scherzinger, E., Wanker, E.E., Mangiarini, L., and Bates, G.P. (1997) Formation of intranuclear inclusions underlies the neurological dysfunction in mice transgenic for the HD mutation. *Cell* **90**, 537-548

Dedeoglu, A.D., Kubilus, J.K., Jeitner, T.M., Matson, S.A., Bogdanov, M., Kowall, N., Matson, W.R., Cooper, A.J.L., Ratan, R.R., Beal, M.F., Hersch, S.M., and Ferrante, R.J. (2002) Therapeutic effects of cystamine in a murine model of Huntington's disease. *J. Neurosci.* **22**, 8942-8950

Delacourte, A., and Defossez, A. (1986) Alzheimer's disease: Tau proteins, the promoting factors of microtubule assembly, are major components of paired helical filaments. *J. Neurol. Sci.* **76**, 173-186

De Laurenzi, V., and Melino, G. (2001) Gene disruption of tissue transglutaminase. *Mol. Cell Biol.* **21**, 148-155

Dietrich, W., Ehnis, T., Bauer, M., Donner, P., Volta, U., Riecken, E.O., and Schuppan, D. (1997) Identification of tissue transglutaminase as the autoantigen of celiac disease. *Nat. Med.* **3**, 797-801

DiFiglia, M., Sapp, E., Chase, K.O., Davies, S.W., Bates, G.P., Vonsattel, J.P., and Aronin, N. (1997) Aggregation of huntingtin in neuronal inclusions and dystrophic neuritis in brain. *Science* **277**, 1990-1993

Dubbink, H.J., Cleutjens, K.B., van der Korput, H.A., Trapman, J., and Romijn, J.C. (1999) An Sp1 binding site is essential for basal activity of the human prostate-specific transglutaminase gene (TGM4) promoter. *Gene* **240**, 261-267

Dubbink, H.J., Verkaik, N.S., Faber, P.W., Trapman, J., Schroder, F.H., and Romijn, J.C. (1996) Tissue specific and androgen-regulated expression of human prostate-specific transglutaminase. *Biochem. J.* **315**, 901-908

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 organisation, tissue-specific expression, and promoter characterisation. *Genomics* **51**, 434-444

Duyao, M.P., Harper, A., Hetherington, M., Lawton, Y., Trotter, H., and Bates, G.P. (1993) Inactivation of the mouse Huntington's disease gene homolog *hdh*. *Science* **269**, 407-410

Egbring, R., Kroniger, A., and Seitz, R. (1996) Factor XIII deficiency: pathogenic mechanisms and clinical symptoms. *Semin. Thromb. Haemostasis* **22**, 419-425

El-Agnaf, O.M., Jakes, R., Curran, M.D., Middleton, D., Ingenito, R., Bianchi, E., Pessi, A., Neill, D., and Wallace, A. (1998) Aggregates from mutant and wild-type α -synuclein proteins and NAC peptide induce apoptotic cell death in human neuroblastoma cells by formation of β -sheet and amyloid-like filaments. *FEBS Lett.* **440**, 71-75

Facchiano, F., Benfanati, F., Valtorta, F., and Luini, A. (1993) Covalent modification of synapsin 1 by a tetanus-toxin-activated transglutaminase. *J. Biol. Chem.* **268**, 4588-4591

Feany, M.B., and Dickson, D.W. (1996) Evidence that phospholipase C δ 1 is the effector in the Gh (transglutaminase II)-mediated signalling. *Ann. Neurol.* **40**, 139-148

Feng, J-F., Rhee, S.G., and Im, M-J. (1996) Evidence that phospholipase C- δ 1 is the effector in the Gh (transglutaminase II)-mediated signalling. *J. Biol. Chem.* **271**, 16451-16454

Fesus, L., Thomazy, V., and Falus, A. (1987) Induction and activation of tissue transglutaminase during programmed cell death. *FEBS Lett.* **224**, 104-108

Flament, S., Delacourte, A., Verny, M., Hauw, J.J., and Javoy-Agid, F. (1991) Abnormal tau proteins in progressive supranuclear palsy. *Acta. Neuropathol. Exp. Neurol.* **81**, 591-596

Folk, J.E., and Chung, S.I. (1985) Transglutaminases. *Methods Enzymol.* **113**, 358-375

Folk, J.E., and Cole, P.W. (1966) Transglutaminase: mechanistic features of the active site as determined by kinetic and inhibitor studies. *Biochim. Biophys. Acta.* **122**, 244-264

Folk, J.E., and Finlayson, J.S. (1977) The $\epsilon(\gamma\text{-glutamyl})\text{lysine}$ crosslink and the catalytic role of transglutaminase. *Adv. Protein Chem.* **31**, 1-133

Forloni, G., Bertani, I., Calella, A.M., Thaler, F., and Invernizzi, R. (2000) α -synuclein and Parkinson's disease: selective neurodegenerative effect of α -synuclein fragment on dopaminergic neurons *in vitro* and *in vivo*. *Ann. Neurol.* **47**, 632-640

Fox, B.A., Yee, V.C., Pedersen, L.C., Le Trong, I., Bishop, P.D., Stenkamp, R.E., and Teller, D.C. (1999) Identification of the calcium binding site and a novel ytterbium site in blood coagulation factor XIII by x-ray crystallography. *J. Biol. Chem.* **274**, 4917-4923

Fraij, B.M., and Gonzales, R.A. (1997) Organization and structure of the human tissue transglutaminase gene. *Biochim. Biophys. Acta.* **1354**, 65-71

Fromm, J.R., Hilman, R.E., Caldwell, E.E.O., Weiler, J.M., and Linhardt, R.J. (1995) Differences in the interaction with arginine and lysine and the importance of these basic amino acids in the binding of heparin to the acidic fibroblast growth factor. *Arch. Biochem. Biophys.* **323**, 279-285

Furie, B., and Furie, B.C. (1998) The molecular basis of blood coagulation. *Cell* **53**, 505-518

Gentile, V., Davies, P.J.A., and Baldini, A. (1994) The human tissue transglutaminase gene maps on chromosome 20q12 by *in situ* fluorescence hybridisation. *Genomics* **20**, 295-297

Gentile, V., Sepe, C., Calvani, M., Melone, M.A., Cotrufo, R., Cooper, A.J.L., Blass, J.P., and Peluso, G. (1998) Tissue transglutaminase-catalysed formation of high-molecular weight aggregates *in vitro* is favored with long polyglutamine domains: a possible mechanism contributing to CAG-triplet disease. *Arch. Biochem. Biophys.* **352**, 314-321

Gentile, V., Thomazy, V., Piacentini, M., Fesus, L., and Davies, P.J.A. (1992) Expression of tissue transglutaminase in Balb-C 3T3 fibroblasts: effects on cellular morphology and adhesion. *J. Cell Biol.* **119**, 463-474

Gibson, G.E., Park, L.C., Sheu, K.F., Blass, J.P., and Calingasan, N.Y. (2000) The α -ketoglutarate dehydrogenase complex in neurodegeneration. *Neurochem. Int.* **36**, 97-112

Gibson, G.E., Zhang, H., Toral-Barza, L., Szolosi, S., Tofel-Grehl, B. (1996) Calcium stores in cultured fibroblasts and their changes with Alzheimer's disease. *Biochim. Biophys. Acta.* **1316**, 71-77

Gilad, G.M., and Varon, L.E. (1985) Transglutaminase activity in rat brain: characterization, distribution, and changes with age. *J. Neurochem.* **45**, 1522-1526

Gilad, G.M., Varon, L.E., and Gilad, V.H. (1985) Calcium-dependent transglutaminase of rat sympathetic ganglion in development and after nerve injury. *J. Neurochem.* **44**, 1385-1390

Gill, D.L., Ghosh, T.K., and Mullaney, J.M. (1989) Calcium signalling: mechanisms in endoplasmic reticulum activated by inositol 1,4,5 triphosphate and GTP. *Cell. Calcium* **10**, 363-374

Goedert, M., Spillantini, M.G., Cairns, N.J., and Crowther, R.A. (1992) Tau proteins of Alzheimer paired helical filaments: abnormal phosphorylation of all six brain isoforms. *Neuron* **8**, 159-168

Golbe, L.I., Davies, P.H., Schoenberg, B.S., and Duvoisin, R.C. (1988) Prevalence and natural history of progressive supranuclear palsy. *Neurology* **38**, 1031-1034

Green, H. (1993) Human genetic diseases due to codon reiteration: relationship to an evolutionary mechanism. *Cell* **74**, 955-956

Greenberg, C.S., Achyuthan, K.E., Borowitz, M.J., and Schuman, M.A. (1987) The transglutaminase in vascular cells and tissues could provide an alternative pathway for fibrin stabilisation. *Blood* **70**, 702-709

Greenberg, C.S., Birckbichler, P.J., and Rice, R.H. (1991) Transglutaminases: multifunctional cross-linking enzymes that stabilize tissues. *FASEB J.* **5**, 3071-3077

Grenard, P., Bates, M.K., and Aeshlimann, 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**, 33066-33078

Grenard, P., Bresson-Hadni, S., El Alaoui, S., Chevaller, M., Vuitton, D.A., and Ricard-Blum, S. (2001a) Transglutaminase-mediated cross-linking is involved in the stabilization of extracellular matrix in human liver fibrosis. *J. Hepatol.* **35**, 367-375

Goebel, H.H., Heipertz, R., Scholz, W., Iqbal, K., and Tellez-Nagel, I. (1978) Juvenile Huntington chorea: clinical ultrastructural and biochemical studies. *Neurology* **28**, 23-31

Groenen, P.J., Seccia, M., Smulders, R.H., Gravela, E., Cheesemman, K.H., Bloemendal, H., and de Jong, W.W. (1993) Exposure of beta H-crystallin to hydroxyl radicals enhances the transglutaminase-susceptibility of its existing amine-donor and amine-acceptor sites. *Biochem. J.* **295**, 399-404

Guex, N., Schwede, T., and Pietsch, M.C. (2000) Protein tertiary structure modelling in Coligan, J.E., Dunn, B.M., Ploegh, H.L., Speicher, D.W., and Wingfield, P.T. *Current Models in Protein Science*, Unit 2.8, John Wiley & Sons, Inc., New York,

Guidetti, P., *et al.* (2001) Early degenerative changes in transgenic mice expressing mutant huntingtin involve dendritic abnormalities but no impairment of mitochondrial energy production. *Exp. Neurol.* **1169**, 340-350

Haass, C., Koo, E.H., Mellon, A., Hung, A.Y., and Selkoe, D.J. (1992) Targeting of cell-surface β -amyloid precursor protein to lysosome: alternative processing into amyloid-bearing fragments. *Nature* **357**, 500-503

Hadjivassiliou, M., Grunewald, R.A., Chattopadhyay, A.K., Davies-Jones, G.A.G., Gibson, A., Jarrat, J.A., Kandler, R., Lobo, A.J., Powell, T., and Smith, C.M.L. (1998) Clinical, radiological, neurophysiological and neuropathological characteristics of gluten ataxia. *The Lancet* **352**, 1582-1585

Hadjivassiliou, M., Grunewald, R.A., Sharrack, B., Sanders, D., Lobo, A., Wood, N., Williamson, C., Woodroffe, N., and Davies-Jones, G.A.B. (2003) Gluten ataxia in perspective: epidemiology, genetic susceptibility and clinical characteristics. *Brain* **126**, 685-691

Hadjivassiliou, M., Gibson, A., Davies-Jones, G.A.B., Lobo, A.J., Stephenson, T.J., and Milford-Ward, A. (1996) 'Does cryptic gluten sensitivity play a part in neurological illness?' *The Lancet* **347**, 369-371

Hadjivassiliou, M., Grunewald, R.A., Lawden, M., Davies-Jones, G.A.B., Powell, T., and Smith, C.M.L. (2001) Headache and CNS white matter abnormalities associated with gluten sensitivity. *Neurology* **56**, 385-388

Hand, D., Campoy, F.J., Clark, S., Fisher, A., and Haynes, L.W. (1993) Activity and distribution of tissue transglutaminase in association with nerve muscle synapses. *J. Neurochem.* **61**, 1064-1072

Hensley, K., Hall, N., Subramaniam, R., Cole, P., Harris, M., Aksenov, M., Gabbita, S.P., Wu, J.F., and Carney, J.M. (1995) Brain regional correspondance between Alzheimer's disease histopathology and biomarkers of protein oxidation. *J. Neurochem.* **65**, 2146-2156

Higgins, D., Thompson, J., Gibson, T., Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994) Improving the sensitivity of progressive multiple sequence alignment through sequencing weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673-4680

Hodson, J.G., Agopyan, N., Gutekunst, C.A., Leavitt, B.R., LePiannee, F., Singaraja, R., Smith, D.J., Bissada, N., McCutcheon, K., Nasir, J., Jamot, L., Li, Z-J., Stevens, M.E., Rosemond, E., Roder, J.C., Phillips, A.G., Rubin, E.M., Hersch, S.M., and Hayden, M.R. (1999) A YAC mouse model for Huntington's disease with full length mutant huntingtin, cytoplasmic toxicity, and selective striatal neurodegeneration. *Neuron* **23**, 181-192

Huntington's Disease Collaborative research Group (1993) A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* **72**, 971-983

Hwang, K-C., Gray, C.D., Sivasubramanian, N., and Im, M.J. (1995) Interaction site of GTP binding Gh (transglutaminase II) with phospholipase C. *J. Biol. Chem.* **270**, 27058-27062

Iannicola, C., Moreno, S., Oliverio, S., Nardacci, R., Ciofi-Luzzatto, A., Piacentini, M. (2000) Early alterations in gene expression and cell morphology in a mouse model of Huntington's disease. *J. Neurochem.* **75**, 830-839

Ichinose, A., Botterus, R.E., and Davie, E.W. (1990) Structure of transglutaminases. *J. Biol. Chem.* **265**, 13411-13414

Ichinose, A., and Davie, E.W. (1998) Characterization of the gen for the a-subunit of human factor XIII (plasma trasnglutaminase), a blood coagulation factor. *Proc. Natl. Acad. Sci. U.S.A.* **85**, 5829-5833

Ichinose, A., Izumi, T., Hashiguchi, T. (1996) The normal and abnormal genes of the a and b subunits in coagulation factor XIII. *Semin. Thromb. Haemostasis* **22**, 385-391

Ichinose, A., McMullen, B.A., Fujikawa, K., and Davie, E.W. (1986) Amino acid sequence of the b-subunit of human factor XIII, a protein composed of ten repetitive segments. *Biochemistry* **25**, 4633-4638

Iismaa, S.E., Wu, M.J., Nanda, N., Church, W.B., and Graham, R.M. (2000) GTP binding and signalling by Gh/transglutaminase II involves distinct residues in a unique GTP-binding pocket. *J. Biol. Chem.* **275**, 18259-18265

Ikura, K., Nasu, T., Yokota, H., Tsuchiya, Y., Sasaki, R., and Chiba, H. (1988) Amino acid sequence of guinea pig liver transglutaminase from its cDNA sequence. *Biochemistry* **27**, 2898-2905

Ikura, K., Takahata, K., and Sasaki, R. (1993) Cross-linking of a synthetic partial-length (1-28) peptide of the Alzheimer beta/A4 amyloid protein by transglutaminase. *FEBS Lett.* **326**, 109-111

Im, M.J., Riek, R.P., and Graham, R.M. (1990) A novel guanine nucleotide binding protein coupled to the α 1-adrenergic receptor. II. Purification, characterization and reconstitution. *J. Biol. Chem.* **265**, 18952-18960

Jabs, A., Weiss, M.S., and Hilgenfeld, R.J. (1999) Non-proline *cis* peptide bonds in proteins. *J. Mol. Biol.* **286**, 291-304

Jarnik, M., Simon, M.N., and Steven, A.C. (1998) Cornified cell envelope assembly: a model based on electron microscopic determinations of thickness and projected density. *J. Cell Sci.* **111**, 1015-1060

Jensen, P.H., Sorensen, E.S., Petersen, T.E., Gliemann, J., and Rasmussen, L.K. (1995) Residues in the synuclein consensus motif of the α -synuclein fragment, NAC participate in transglutaminase-catalysed cross-linking to Alzheimer-disease amyloid β A4 peptide. *Biochem. J.* **310**, 91-94

Jeong, J.M., Murthy, S.N., Radek, J.T., and Lorand, L. (1995) The fibronectin-binding domain of transglutaminases. *J. Biol. Chem.* **270**, 5654-5658

Johnson, G.V., Cox, T.M., Lockhart, J.P., Zinnerman, M.D., Miller, M.L., and Powers, R.E. (1997) Transglutaminase activity is increased in Alzheimer's disease brain. *Brain Res.* **751**, 323-329

Johnson, K., Hashimoto, S., Lotz, M., Pritzher, K., and Terkeltaub, R. (2001) Interleukin-1 induces pro-mineralizing activity of cartilage tissue transglutaminase and factor XIIIa. *Am. J. Pathol.* **159**, 149-163

Johnson, T.S., Skill, N.J., El Nahas, A.M., Oldroyd, S.D., Thomas, G.L., Douthwaite, J.A., Haylor, J.L., and Griffin, M. (1999) Transglutaminase transcription and antigen translocation in experimental renal scarring. *J. Am. Soc. Nephrol.* **10**, 2146-2157

Jurgensen, K., Aeshlimann, D., Cavin, V., Genge, M., and Hunziker, E.B. (1997) A new biological glue for cartilage-cartilage interfaces: tissue transglutaminase. *J. Bone Joint Surg. Am.* **79**, 185-193

Kahlem, P., Green, H., and Djian, P. (1998) Transglutaminase action imitates huntington's disease: selective polymerisation of huntingtin containing expanded polyglutamine. *Mol. Cell* **1**, 595-601

Kanaji, T., Ozaki, H., Takao, T., Kawajin, H., Ide, H., Motoki, M., and Shimomishi, Y. (1993) 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**, 1776-1781

Kang, J., Lemaire, H.G., Unterbeck, A., Salbaum, J.M., Masters, C.L., Grzescik, K.H., Multhaup, G., Beyreuther, K., and Muller-Hill, B. (1987) The precursor of Alzheimer's disease amyloid A4 protein resembles a cell-surface receptor. *Nature* **325**, 733-736

Karges, H.E., and Metzner, H. (1996) Therapeutic factor XIII preparations and perspectives for recombinant factor XIII. *Semin. Thromb. Haemostasis* **22**, 427-436

Karpuj, M.V., Becker, M.W., Springer, J.E., Chabas, D., Youssef, S., Pedotti, R., Mitchell, D., and Steinman, L. (2002) Prolonged survival and decreased abnormal movements in transgenic model of Huntington disease, with administration of the transglutaminase inhibitor cystamine. *Nat. Med.* **8**, 143-149

Karpuj, M.V., Garren, H., Slunt, H., Price, D.L., Gusella, J., Becker, M.W., and Steinman, L. (1999) Transglutaminase aggregates huntingtin into nonamyloidogenic polymers, and its enzymatic activity increases in Huntington's disease brain nuclei. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 7388-7393

Kaytor, M.D., and Warren, S.T. (1999) Aberrant protein deposition and neurological illness. *J. Biol. Chem.* **274**, 37507-37510

Kida, M., Souri, M., Yamamoto, M., Saito, H., and Ichinose, A. (1999) Transcriptional regulation of cell type-specific expression of the TATA-less A subunit gene for human coagulation factor XIII. *J. Biol. Chem.* **274**, 6138-6147

Kim, M., Lee, H.S., LaForet, G., McIntyre, C., Martin, E.J., Chang, P., Kim, T.W., Williams, M., Reddy, P.H., Tagle, D., Boyce, F.M., Won, L., Heller, A., Aronin, N., and DiFlagia, M. (1999a) Mutant huntingtin expression in clonal striatal cells: dissociation of inclusion formation and neuronal survival by caspase inhibition. *J. Neurosci.* **19**, 964-973

Kim, S.Y., Chung, S.I., and Steinert, P.M. (1995) Highly active soluble processed forms of the transglutaminase 1 enzyme in epidermal cells. *J. Biol. Chem.* **270**, 18026-18035

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**, 12682-12690

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 transglutaminase 1 and 2 in Alzheimer's disease. *J. Biol. Chem.* **274**, 30715-30721

Kim, S.Y., Jeitner, T.M., and Steinert, P.M. (2002) Transglutaminases in disease. *Neurochem. Int.* **40**, 85-103

Kim, I.G., Lee, S.C., Lee, J.H., Yang, J.M., Chung, S.I., and Steinert, P. (1994) Structure and organisation of the human transglutaminase 3 gene: evolutionary relationship to the transglutaminase family. *J. Invest. Dermatol.* **103**, 137-142

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**, 21971-21978

Kim, I.G., McBride, O.W., Wang, M., Kim, S.Y., Idler, W.W., and Steinert, P.M. (1992) Structure and organisation of the human transglutaminase 1 gene. *J. Biol. Chem.* **267**, 7710-7717

Knight, C.R., Rees, R.C., and Griffin, M. (1991) Apoptosis: a potential role for cytosolic transglutaminase and its importance in tumour progression. *Biochim. Biophys. Acta.* **1096**, 312-318

Komori, T. (1999) Tau-positive glial inclusions in progressive supranuclear palsy, corticobasal degeneration and Pick's disease. *Brain Pathol.* **9**, 663-679

Korsgren, C., and Cohen, C.M. (1991) Organisation 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. Sc. U.S.A.* **88**, 4840-4844

Korsgren, C., Lawler, J., Lambert, S., Speicher, D., and Cohen, C.M. (1990) Complete amino acid sequence and homologies of human erythrocyte membrane protein band 4.2. *Biochemistry* **87**, 613-617

Kuemmerle, S., Gutekanst, C.A., Klein, A.M., Li, X.J., Li, S.H., Beal, M.F., Hersch, S.M., and Ferrante, R.J. (1999) Huntingtin aggregates may not predict neuronal death in Huntington's disease. *Ann. Neurol.* **46**, 8420849

Kunio, G.S., Tssyganskaya, M., Zhu, J., Liu, S.L., Nagy, L., Thomazy, V., Davies, P.J., and Zern, M.A. (1998) TNF- α modulates expression of the tissue transglutaminase gene in liver cells. *Am. J. Physiol.* **274**, G240-G245

Kuraishi, C., Sakamoto, J., Yamazaki, K., Susa, Y., Kukara, C., and Soeda, T. (1997) Production of restructures meat using microbial transglutaminase without salt or cooking. *J. Food Sci.* **62**, 488

Kuraishi, C., Yamazaki, K., and Susa, Y. (2001) Transglutaminase: its utilisation in the food industry. *Food Rev. Intern.* **17**, 221-246

Lassman, H., Bancher, C., Breitschopf, H., Wegiel, J., Bobinski, M., Jellinger, K., and Wisniewski, H.M. (1995) Cell death in Alzheimer's disease evaluated by DNA fragmentation *in situ*. *Acta. Neuropath.* **89**, 35-41

Lee, K.N., Birckichler, P.J., and Pattersen, M.K.Jr. (1989) GTP hydrolysis by guinea pig liver transglutaminase. *Biochem. Biophys. Res. Commun.* **169**, 1370-1375

Lee, J.H., Jang, S.I, Yang, J.M., Markova, N., and Steinert, P (1996) The proximal promoter of the human transglutaminase 3 gene. Stratified squamous epithelial-specific expression in cultured cells is mediated by binding of Sp1 and ets transcription factors to a proximal element. *J. Biol. Chem.* **271**, 4561-4568

Lee, A.C.J., Powell., J.E., Tregear, G.W., Niall, H.D., and Stevens, V.C. (1985) A method for preparing beta-hCG COOH peptide-carrier conjugates of predictable composition. *Mol. Immunol.* **17**, 749-756

Lee, M.W., and Severson, D.L. (1994) Signal transduction in vascular smooth muscle: diacylglycerol second messengers and PKC action. *Am. J. Physiol.* **267**, 659-678

LeMosy, E.K., Erickson, H.P., Beyer, W.F., Radek, J.T., Jeong, J.M., Murthy, S.N.P., and Lorand, L. (1992) Visualization of purified fibronectin-transglutaminase complexes. *J. Biol. Chem.* **267**, 7880-7885

Lesort, M., Attanavanich, K., Zhang, J., and Johnson, G.V. (1998) Distinct nuclear localization and activity of tissue transglutaminase. *J. Biol. Chem.* **273**, 11991-11994

Lesort, M., Chun, W., Johnson, G.V., and Ferrante, R.J. (1999) Tissue transglutaminase is increased in Huntington's disease brain. *J. Neurochem.* **73**, 2018-2027

Lesort, M., Chun, W., Tucholski, J., and Johnson, G.V. (2002) Does tissue transglutaminase play a role in Huntington's disease? *Neurochem. Int.* **40**, 37-52

Lesort, M., Tucholski, J., Miller, M.L., and Johnson, G.V. (2000) Tissue transglutaminase: a possible role in neurodegenerative diseases. *Prog. Neurobiol.* **61**, 439-463

Lippert, H. (1989) *Anatomie. Text und Atlas*. Urban & Schwarzenberg, Germany

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. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 2743-2747

Lorand, L., Barnes, N., Bruner-Lorand, J.A., Hawkins, M., and Michalska, M. (1987) Inhibition of protein cross-linking in Ca²⁺-enriched human erythrocytes and activated platelets. *Biochemistry* **26**, 308-313

Lorand, L., and Conrad, S.M. (1984) Transglutaminases. *Mol. Cell Biochem.* **58**, 9-35

Lorand, L., and Graham, R.M. (2003) Transglutaminases: Crosslinking enzymes with pleiotropic functions. *Nature* **4**, 140-156

Lorand, L., Jeong, J.M., Radek, J.T., and Wilson, J. (1993) Human plasma factor XIII: Subunit interactions and activation of zymogen. *Methods Enzymol* **222**, 22-35

Lozier, J., Tukahashi, N., and Putnam, F.W. (1984) Complete amino acid sequence of human plasma β_2 -glycoprotein I. *Proc. Natl. Acad. Sci. U.S.A.* **81**, 3640-3644

Lu, S., Saydak, M., Gentile, V., Stein, J., and Davies, P.J. (1995) Isolation and characterization of the human tissue transglutaminase gene promoter. *J. Biol. Chem.* **270**, 9748-9756

Makarova, K.S., Aravind, L., and Koonin, E.V. (1999) A superfamily of archaeal, bacterial and eukaryotic proteins homologous to animal transglutaminases. *Protein Sci.* **8**, 1714-1719

Mangiarini, L., and Sathasivam, K. (1996) Exon 1 of the HD gene with an expanded CAG repeat is sufficient to cause a progressive neurological phenotype in transgenic mice. *Cell* **87**, 493-506

Marekov, L.M., and Steinert, P.M. (1998) Ceramides bound to structural proteins of the human foreskin epidermal cornified cell envelope. *J. Biol. Chem.* **273**, 17763-17770

Mariani, P., Carsughi, F., Spinozzi, F., Romanzetti, S., Meier, G., Casadio, R., and Bergamini, C. (2000) Ligand-induced conformational changes in tissue transglutaminase: Monte Carlo analysis of small-angle scattering data. *Biophys. J.* **78**, 3240-3251

Mark, R.J., Blanc, E.M., and Mattson, M.P. (1996) Amyloid beta-peptide and oxidative cellular injury in Alzheimer's disease. *Mol. Neurobiol.* **12**, 211-224

Martinez, J., Chalupowicz, D.G., Roush, R.K., Sheth, A., and Barsigian, C. (1994) Transglutaminase-mediated processing of fibronectin by endothelial cell monolayers. *Biochemistry* **33**, 2538-2545

Martini, F. (1989) Fundamentals of anatomy and physiology. Prentice Hall, New Jersey

Mastroberardino, P.G., Iannocula, C., Nardacci, R., Bernassola, F., De Laurenzi, V., Melino, G., Pavone, F., Oliverio, S., Fesus, L., and Piacentini, M. (2002) 'Tissue' transglutaminase ablation reduces neuronal death and prolongs survival in a mouse model of Huntington's disease. *Cell Death Differ.* **9**, 873-880

Matheis, G., and Whitaker, J.R. (1987) A review: enzymatic cross-linking of proteins applicable to foods. *J. Food Chem.* **11**, 311-312

Mattson, M.P. (2000) Apoptosis in neurodegenerative disorders. *Nat. Rev. Mol. Cell Biol.* **1**, 120-129

Mattson, M.P., Chan, S.L., and Camandola, S. (2001) Presenilin mutations and calcium signalling defects in the nervous and immune systems. *Bioessays* **23**, 3724-3732

Mattson, M.P., Cheng, B., Davies, D., Bryant, K., Lieberburg, I., and Rydel, R.E. (1993) Beta-amyloid peptides destabilize calcium homeostasis and render human cortical neurons vulnerable to excitotoxicity. *J. Neurosci.* **12**, 376-389

McLauchlan, J., Gaffney, D., Whitton, J.L., and Clementa, J.B. (1985) The consensus sequence YGTGTTY located downstream from the AATAAA signal is required for efficient formation of mRNA 3' termini. *Nucleic Acids Res.* **13**, 1347-1368

Mearns, B., Nanda, N., Michalick, J., Iismaa, S., and Graham, R. (2002) Impaired wound healing and altered fibroblast cytoskeletal dynamics in G_h knockout mice. *7th International Conference on TG and Protein Crosslinking Reactions*. September 14-17 2002, Ferrara, Italy

Medveder, A., Saunders, N., Matsuura, H., Christokhina, A., and Jetten, A. (1999) Regulation of the transglutaminase 1 gene. Identification of DNA elements involved in its transcriptional control in tracheobronchial epithelial cells. *J. Biol. Chem.* **274**, 3887-3896

Mehta, K., Chandrashekar, R., and Rao, U.R. (1996) Transglutaminase-catalysed incorporation of host proteins in *Brugia malayi* microfilariae. *Mol. Biochem. Parasitol.* **76**, 105-114

Mehta, K., Rao, R.U., and Chandrashekar, R. (2002) Transglutaminases of the lower organisms. Across the link between life and death. *Minerva Biotec.* **14**, 129-134

Mehta, K., Rao, U.R., Vickery, A.C., and Fesus, L. (1992) Identification of a novel transglutaminase from the filarial parasite *Brugia malayi* and its role in growth and development. *Mol. Biochem. Parasitol.* **53**, 1-15

Melino, G., Annichiarico-Petruzzelli, M., Piredda, L., Candi, E., Gentile, V., Davies, P.J., and Piacentini, M. (1994) Tissue transglutaminase and apoptosis: sense and antisense transfection studies with human neuroblastoma cells. *Mol. Cell Biol.* **14**, 6584-6596

Melino, G., De Laurenzi, V., Catani, M.V., Terrinoni, A., Ciani, B., Candi, E., Marekov, L., and Steinert, P.M. (1998) The cornified envelope: a model of cell death in the skin. *Results Probl. Cell Differ.* **24**, 175-212

Mende-Mueller, L.M., Toreff, T., Hwang, S., Chesselet, M., and Hook, V.Y.H. (2001) Tissue-specific proteolysis of Huntingtin (htt) in human brain: evidence of enhanced levels of N- and C-terminal htt fragments in Huntington's disease striatum. *J. Neurosci.* **21**, 1830-1837

Mezey, E., Dehejia, A., Harta, G., Papp, M.I., Polymeropoulos, M.H., and Brownstein, M.J. (1998) Alpha synuclein in neurodegenerative disorders: murder or accomplice? *Nat. Med.* **4**, 755-757

Miller, M.L., and Johnson, G.V.W. (1995) Transglutaminase cross-linking of the tau protein. *J. Neurochem.* **65**, 1760-1770

Mimotopes (2001) Antipeptide antibodies. Mimotopes Ltd, France

Mirza, A., Liu, S.L., Frizell, E., Zhu, J., Maddukuri, S., Martinez, J., Davies, P., Schwarting, R., Norton, P., and Zern, M.A. (1997) A role for tissue transglutaminase in hepatic injury and fibrogenesis and its regulation by NF- κ B. *Am. J. Physiol.* **272**, G281-G288

Monczak, Y., Trudel, M., Lamph, W., and Miller, J.W. (1997) Induction of apoptosis without differentiation by retinoic acid in PLB-985 cells requires the activation of both RAR and RXR. *Blood* **90**, 3345-3355

Mosesson, M.W. (1992) The roles of fibrinogen and fibrin in hemostasis and thrombosis. *Semin. Haematol.* **29**, 177-188

Mosher, D.F. (1976) Cross-linking of fibronectin to collagenous proteins. *J. Biol. Chem.* **251**, 1639-1645

Mount, S.M. (1982) A catalogue of splice junction sequences. *Nucleic Acids Res.* **10**, 459-472

Murthy, S.N., Iismaa, S.E., 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**, 2738-2742

Nakaoka, H., Perez, D.M., Baek, K.J., Das, T., Husain, A., Misono, K., Im, M.J., and Graham, R.M. (1994) Gh: a GTP-binding protein with transglutaminase activity and receptor signalling function. *Science* **264**, 1593-1596

Nanda, N., Iismaa, S.E., Owens, W.A., Husain, A., Mackay, F., and Graham, R.M. (2001) Targeted inactivation of Gh/tissue transglutaminase II. *J. Biol. Chem.* **276**, 20673-20678

Natsuka, S., Takubo, R., Seki, R., and Ikura, K. (2002) Molecular cloning and expression of *Caenorhabditis elegans* Erp57-homologue with transglutaminase activity. *J. Biochem.* **130**, 731-735

Nemes, Z., and Steinert, P.M. (1999) Bricks and mortar of the epidermal barrier. *Exp. Mol. Med.* **31**, 5-19

Noguchi, K., Ishikawa, K., Yokoyama, K., Ohtsuka, T., Nio, N., and Suzuki, E. (2001) Crystal structure of red sea bream transglutaminase. *J. Biol. Chem.* **276**, 12055-12059

Nunes, I., Gieizes, P.E., Metz, C.N., and Rifkin, D.B. (1997) Latent transforming growth factor β binding protein domains involved in activation and transglutaminase-dependent cross-linking of latent transforming growth factor- β . *J. Cell Biol.* **136**, 1151-1103

Parmentier, L., Lakhdar, H., Blanchet-Barden, C., Marchand, S., Dubertret, L., and Weissenbach, J. (1996) Mapping of a second locus for lamellar ichthyosis to chromosome 2q33-35. *Hum. Mol. Genet.* **5**, 555-559

Pastuszko, A., Wilson, D.F., and Erecinska, M. (1986) A role for transglutaminases in neurotransmitter release by rat brain synaptosomes. *J. Neurochem.* **46**, 499-508

Pedersen, L.C., Yee, V.C., Bishop, P.D., Le Trong, I., Teller, D.C., and Stenkamp, R.E. (1994) Transglutaminase factor XIII uses proteinase-like catalytic triad to crosslink macromolecules. *Protein Sci.* **3**, 1131-1135

Peng, X., Zhang, Y., Zhang, H., Graner, S., Williams, J.F., Levitt, M.L., and Lokshin, A. (1999) Interaction of tissue transglutaminase with nuclear transport protein importin- α 3. *FEBS Lett.* **446**, 35-39

Perry, M.J.M., Mahoney, S.A., and Haynes, L.W. (1995) Transglutaminase C in cerebellar granule neurons: regulation and localization of substrate cross-linking. *Neuroscience* **65**, 1063-1076

Perutz, M.F., Johnson, T., Suzuki, M., and Finch, J.T. (1994) Glutamine repeats as polar zippers: Their possible role in inherited neurodegenerative diseases. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 5355-5358

Peterson, C., and Goldman, J.E. (1986) Alterations in calcium content and biochemical processes in cultured skin fibroblasts from aged and Alzheimer donors. *Proc. Natl. Acad. Sci. U.S.A.* **83**, 2758-2762

Piacentini, M., Farrace, M.G., Matarrese, P., Ciccocanti, F., Falasca, L., Rodolfo, C., Giammarioli, A.M., Verderio, E., Griffin, M., and Malorni, W. (2002) Transglutaminase overexpression sensitises neuronal cell lines to apoptosis by increasing mitochondrial membrane potential and cellular oxidative stress. *J. Neurochem.* **81**, 1061-1072

Piacentini, M., Martinet, N., and Beninati, S., Folk, J.E. (1998) Free and protein-conjugated polyamines in mouse epidermal cells. Effect of high calcium and retinoic acid. *J. Biol. Chem.* **263**, 3790-3794

Qian, R.Q., and Glanville, R.W. (1997) Alignment of fibrillin molecules in elastic microfibrils is defined by transglutaminase-derived cross-links. *Biochemistry* **36**, 15841-15847

Raghunath, M., Hopfner, B., Aeschlimann, D., Luthi, U., Meuli, M., Altermatt, S., Gobet, R., Bruckner-Tuderman, L., and Steinmann, B. (1996) Cross-linking of the dermo-epidermal junction of skin regenerating from keratinocyte autografts. Anchoring fibrils are a target for tissue transglutaminase. *J. Clin. Invest.* **98**, 1174-1184

Raghunath, M., Unsold, C., Kubitscheck, U., Bruckner-Tuderman, L., Peters, R., and Meuli, M. (1998) The cutaneous microfibrillar apparatus contains latent transforming growth factor- β -binding protein (LTBP.1) as a repository for latent TGF β 1. *J. Invest. Dermatol.* **111**, 559-564

Radek, J.T., Noonan, S., Nagy, L., Davies, P.J., and Stein, J.P. (1993) Affinity of human erythrocyte transglutaminase for a 42-kDa gelatin-binding fragment of human plasma fibronectin. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 3152-3156

Rasmussen, L.K., Sorensen, E.S., Petersen, T.E., Gliemann, J., and Jensen, P.H. (1994) Identification of glutamine and lysine residues in Alzheimer amyloid β A β peptide responsible for transglutaminase-catalysed homopolymerization and cross-linking to alpha 2M receptor. *FEBS Lett.* **338**, 161-166

Rice, R., Mehrpouyan, M., Qin, Q., Phillips, M., and Lee, Y. (1996) Identification of phosphorylation sites in keratinocyte transglutaminase. *Biochem. J.* **320**, 547-550

Richards, R.J., Masek, L.C., and Brown, R.F. (1991) Biochemical and cellular mechanisms of pulmonary fibrosis. *Toxicol. Pathol.* **19**, 526-539

Riepe, M.W., Hori, N., Ludolph, A.C., and Carpenter, D.O. (1995) Failure of neuronal ion exchange, not potentiated excitation, causes excitotoxicity after inhibition of oxidative phosphorylation. *Neuroscience* **64**, 91-97

Riess, O., Jakes, R., and Kruger, R. (1998) Genetic dissection of familial Parkinson's disease. *Mol. Med. Today* **4**, 438-444

Ritter, S.J., and Davies, P.J. (1998) Identification of a transforming growth factor- β 1/bone morphogenetic protein 4 (TGF- β 1/BMP4) response element within the mouse tissue transglutaminase gene promoter. *J. Biol. Chem.* **273**, 12798-12806

Rogers, J., Webster, S., Lue, L.F., Brachova, L., Civin, W.H., Emmerling, M., Shivers, B., Walker, D., and McGreer, P. (1996) Inflammation and Alzheimer's disease pathogenesis. *Neurobiol. Aging* **17**, 681-686

Roop, D.R. (1995) Defects in the barrier. *Science* **267**, 474-475

Rugh, R. (1668) *The Mouse, Its Reproduction and Development*. Oxford University Press, New York

Russell, L.J., DiGiovanna, J.J., Rogers, G.R., Steinert, P.M., Hashem, N., Compton, J.G., and Bale, S.J. (1995) Mutations in the gene for transglutaminase 1 in autosomal recessive lamellar. TGM1 mutations in lamellar ichthyosis. *Nat. Genet.* **9**, 279-283

Saito, M., Asakura, H., Yoshida, T., Itoh, K., Okafuji, K., and Yoshida, T. (1990) A familial factor XIII subunit B deficiency. *Br. J. Haematol.* **74**, 290-294

Sakata, Y., and Aoki, N. (1982) Significance of cross-linking of α_2 -plasmin inhibitor to fibrin in inhibition of fibrinolysis and in hemostasis. *J. Clin. Invest.* **69**, 536-542

Sardy, M., Odenthal, U., Karpati, S., Paulsson, M., and Smyth, N., (1999) Recombinant human tissue transglutaminase. Elisa for the diagnosis of gluten-sensitive enteropathy. *Clin. Chem.* **45**, 2142-2149

Sarkar, N.K., Clarke, D.D., and Waelsch, H. (1957) An enzymatically catalysed incorporation of amines into proteins. *Biochim. Biophys. Acta.* **25**, 451-452

Saudou, F., Finkbeiner, S., Devys, D., and Greenberg, M.E. (1998) Huntingtin acts in the nucleus to induce apoptosis but death does not correlate with the formation of intranuclear inclusions. *Cell* **95**, 55-66

Schmidt, G., Selzer, J., Lerm, M., and Aktories, K., (1998) The Rho-deamidating cytotoxic necrotizing factor 1 from *Escherichia coli* possesses transglutaminase activity. Cysteine 866 and histidine 881 are essential for enzyme activity. *J. Biol. Chem.* **73**, 13669-13674

Selkoe, D.J. (1998) The cell biology of β -amyloid precursor protein and presenilin in Alzheimer's disease. *Trends Cell Biol.* **8**, 447-453

Seubert, P., Golde, T.E., Eckman, C.B., and Younkins, S.B. (1992) Isolation and quantification of soluble Alzheimer's β -peptide from biological fluids. *Nature* **359**, 325-327

Shainoff, J.R., Urbanic, D.A., and DiBello, P.M. (1991) Immuno-electrophoretic characterizations of the cross-linking of fibrinogen and fibrin by factor XIIIa and tissue transglutaminase. *J. Biol. Chem.* **266**, 6249-6437

Siebenlist, K.R., Meh, D.A., and Mosesson, M.W. (1996) Plasma factor XIII binds specifically to fibrinogen molecules containing γ chains. *Biochemistry* **35**, 10448-10453

Singer, S.M., Zainelli, G.M., Norlund, M.A., Lee, J.M., and Muma, N.A. (2002) Transglutaminase bonds in neurofibrillary tangles and paired helical filament tau early in Alzheimer's disease. *Neurochem. Int.* **40**, 17-30

Sisoda, S.S., Koo, E.H., Beyreuther, K., Unterbeck, A., and Price, D.L. (1990) Evidence that β -amyloid protein in Alzheimer's disease is not derived from normal processing. *Science* **248**, 492-495

Slife, C.W., Dorsett, M.D., Bouquett, G.T., Register, A., Taylor, E., and Conroy, S. (1985) Subcellular localization of a membrane-associated transglutaminase activity in rat liver. *Arch. Biochem. Biophys.* **241**, 329-336

Spillantini, M.G., Schmidt, M.L., Lee, V.M., Trojanowski, J.Q., Jakes, R., and Goedert, M. (1997) Alpha-synuclein in Lewy bodies. *Nature* **388**, 839-840

Spring, J. (1997) Vertebrate evolution by interspecific hybridisation- are we polyploid? *FEBS Lett.* **400**, 2-8

Steinert, P.M., Chung, S.I. And Kim, S.Y. (1996) Inactive zymogen and highly active proteolytically processed membrane-bound forms of the transglutaminase 1 enzyme in human epidermal keratinocytes. *Biochem. Biophys. Res. Commun.* **221**, 101-106

Steinert, P.M., Dyer, P.Y., and Rogers, G.E. (1971) The isolation of non-keratin protein filaments from inner root sheath cells of the hair follicles. *J. Invest. Dermatol.* **56**, 49-54

Steinert, P.M., Kim, S.Y., and Marekov, L.M. (1996a) The transglutaminase 1 enzyme is variably acylated by myristate and palmitate during differentiation in epidermal keratinocytes. *J. Biol. Chem.* **271**, 26242-26250

Steinert, P.M., and Marekov, L.N. (1997) Direct evidence that involucrin is a major early isopeptide cross-linked component of the keratinocyte cornified cell envelope. *J. Biol. Chem.* **272**, 2021-2030

Stephens, P., Grenard, P., Aeschlimann, P., Langley, M., Blaim, E., Kipling, D., Thomas, D., and Aeschlimann, D. (2004) Crosslinking and G-protein functions of Transglutaminase 2 contribute differentially to fibroblast wound healing responses. *J. Cell Sci.* **117**, 3389-3403

Stewart, D.E., Sarker, A., and Wampler, J.E. (1990) Occurrence and role of *cis* peptide bonds in protein structures. *J. Mol. Biol.* **214**, 253-260

Sung, L.P.A., Chien, S., Fan, Y.S., Lin, C.C., Lambert, K., Zhu, L.Y., Lam, J.S., and Chang, L.S. (1992) Human erythrocyte protein 4.2: Isoform expression, differential splicing, and chromosomal assignment. *Blood* **79**, 2763-2770

Suto, N., Ikura, K., and Sasaki, R. (1993) Expression induced by interleukin-6 of tissue-type transglutaminase in human hepatoblastoma HepG2 cells. *J. Biol. Chem.* **268**, 7469-7473

Suto, N., Ikura, K., Shinagawa, R., and Sasaki, R. (1993a) Identification of promoter region of guinea pig liver transglutaminase gene. *Biochim. Biophys. Acta.* **1172**, 319-322

Takahashi, N., Takahashi, Y., and Putnam, F.W. (1986) Primary structure of blood coagulation factor XIIIa (fibrinolytic transglutaminase) from human placenta. *Proc. Natl. Acad. Sci. U.S.A.* **83**, 8019-8023

Tetzlaff, W., Gilad, V.H., Leonard, C., Bisby, M.A., and Gilad, G.M. (1998) Retrograde changes in transglutaminase activity after peripheral nerve injuries *Brain Res.* **445**, 142-146

Thomazy, V., and Fesus, L. (1989) Differential expression of tissue transglutaminase in human cells. An immunohistochemical study. *Cell Tissue Res.* **255**, 215-224

Tortora, G.J., and Grabowski, S.R. (2000) Principles of Anatomy and Physiology, 9th Edition. John Wiley and Sons, New York

Trottier, Y., Devys, D., Imbert, G., Saudou, F., An, I., Lutz, Y., Webber, C., and Agid, Y. (1995) Cellular localization of the Huntington's disease protein and discrimination of the normal and mutated form. *Nature Genet.* **10**, 104-110

Turmaine, M., Roza, A., Mahal, A., Mangiarini, L., Bates, G.P., and Davies, S.W. (2000) Nonapoptotic neurodegeneration in a transgenic mouse model of Huntington's disease. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 8093-8097

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

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., and Sutton, G.G. (2001) The sequence of the human genome. *Science* **291**, 1304-1351

Verderio, E., Telci, D., Okoye, A., Melino, G., and Griffin, M. (2002) Fibronectin bound tissue transglutaminase rescues cells from anoikis. *Minerva Biotechnol.* **14**, 206

Vonsattel, J.P., and DiFiglia, M. (1998) Huntington disease. *J. Neuropathol. Exp. Neurol.* **57**, 369-384

Webb, G.C., Coggan, M., Ichinose, A., Board, P.G. (1989) Localisation of the coagulation factor XIII B subunit gene (F13B) to chromosome bands 1q31-32. *Hum. Genet.* **81**, 157-160

Weiss, M.S., Metzner, H.J., and Hilgenfeld, R. (1998) Two non-proline cis peptide bonds may be important for factor XIII function. *FEBS Lett.* **423**, 291-296

Willemijn, L., Keaveny, A.P., Offner, G.D., Bootle, E., and Nunes, D.P. (2002) Specificity of tissue transglutaminase explains cereal toxicity in coeliac disease. *J. Exp. Med.* **195**, 643-649

Wirhns, O., Weickert, S., Majtenyi, K., Kahle, P.J., Okochi, M., Haass, C., Multhaup, G., Beyreuther, K., and Bayer, T.A. (2000) Lewy body variant of Alzheimer disease: α -synuclein in dystrophic neuritis of A β plaques. *Neuro. Report* **11**, 3737-3741

Wozniak, M., and Hruska, K.A. (2000) Mechanically strained cells of the osteoblast lineage organize their extracellular matrix through unique sites of $\alpha\beta_3$ -integrin expression. *J. Bone Miner. Res.* **15**, 1731-1745

Yamanish, K., Inazawa, J., Liew, F.M., Nonomura, K., Ariyama, T., Yasuno, H., Abe, T., Doi, H., Hirano, J., and Fukushima, S. (1992) Structure of the gene for human transglutaminase 1. *J. Biol. Chem.* **267**, 17858-1786

Yang, J.M., Ahn, K.S., Cho, M.O., Yoneda, K., Park, K.B., and Lee, E.S., (2001) Novel mutations of the transglutaminase 1 gene in lamellar ichthyosis. *J. Invest. Dermatol.* **117**, 214-218

Yee, V.C., Le Trong, I., Bishop, P.D., Pedersen, L.C., Stenkamp, R.E., and Teller, D.C. (1996) Structure and function studies of factor XIIIa by x-ray crystallography. *Semin. Thromb. Haemostasis* **22**, 377-384

Yee, V.C., Pedersen, L.C., Bishop, P.D., Stenkamp, R.E., and Teller, D.C. (1995) Structural evidence that the activation peptide is not released upon thrombin cleavage of factor XIII. *Thromb. Res.* **78**, 389-397

Yee, V.C., Pedersen, L.C., Le Trong, I., Bishop, P.D., Stenkamp, R.E. and Teller, D.C. (1994) Three-dimensional structure of a transglutaminase: human blood coagulation factor XIII. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 7296-7300

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

Zemaitaitis, M.O., Lee, J.M., Troncoso, J.C., and Muma, N.A. (2000) Transglutaminase-induced cross-linking of tau proteins in progressive supranuclear palsy. *J. Neuropathol. Exp. Neurol.* **59**, 983-989

Zhang, J., Guttman, R.P., and Johnson, G.V. (1998) Tissue transglutaminase is an in situ substrate of calpain: regulation of activity. *Neurochem.* **71**, 240-247

Zhang, W., Johnson, B.R., Suri, D.E., Martinez, J., and Bjornsson, T.D. (1998a) Immunohistochemical demonstration of tissue transglutaminase in amyloid plaques. *Acta. Neuropathol.* **96**, 395-400

Appendix 1

1: AW245863. 2822690.5prime NI...[gi:6588856] Links

LOCUS AW245863 513 bp mRNA linear EST 07-JAN-2000
DEFINITION 2822690.5prime NIH_MGC_7 Homo sapiens cDNA clone IMAGE:2822690
5',

mRNA sequence.

ACCESSION AW245863

VERSION AW245863.1 GI:6588856

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 513)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Other_ESTs: 2822690.3prime

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: DCTD/DTP cDNA Library Preparation: Ling

Hong/Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E.

Consortium (LLNL) DNA Sequencing by: Berkeley MGC sequencing

project Clone distribution: MGC clone distribution information can

be found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/bbrp/image/image.html Base Calling / Quality

Scores: PHRED from University of Washington Genome Center. Vector

Trimming: cross_match from University of Washington Genome Center

PHRAP suite. Poly-T Identification: patMatch.pl from Berkeley

Drosophila Genome Project. University of Washington Genome Center:

<http://www.genome.washington.edu>

Plate: LLCM10 row: B column: 3

High quality sequence stop: 493.

FEATURES Location/Qualifiers

source 1..513

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:2822690"

/tissue_type="small cell carcinoma"

/cell_line="MGC3"

/lab_host="DH10B (phage-resistant)"

/clone_lib="NIH_MGC_7"

/note="Organ: lung; Vector: pOTB7; Site_1: XhoI; Site_2:

EcoRI; cDNA made by oligo-dT priming. Directionally

cloned into EcoRI/XhoI sites using the following 5'

adaptor: GGCACGAG(G). Size-selected >500bp for average

insert size 1.8kb. Library constructed by Ling Hong in

the laboratory of Gerald M. Rubin (University of

California, Berkeley) using ZAP-cDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies)."

ORIGIN

1 atgcctggga aattaattcc gagccagcag ggcggcgacc tggctgttcg gcgtggaagc

61 ctctggaagg agaatctgga tccgcagggc tgggggtcgc tgtctctggc gtgacgacct
121 cctggagcct gccaccaagc ccagcatcgc tggcaagttc aaggctgctag agcctcccat
181 gctggggccac gacctgagac tggccctgtg cttggccaac ctcacctccc gggcccagcg
241 ggtgaggggtc aacctgagcg gtgccaccat cctctatacc cgcaagccag tggcagagat
301 cctgcatgaa tcccacgccg tgaggctggg gccgcaagaa gagaagagaa tcccaattac
361 aatatcttac tctaagtata aagaagacct gacagaggac aagaagatcc tgttggctgc
421 catgtgcctt gtcaccaaag gagagaagct tctggtggag aaggacatta ctctagagga
481 cttcatcacc atcaaggttc tgggccagc cat

Appendix 2

Table 1

| Primer name | Primer sequence |
|-------------|---------------------------------|
| fp1 | 5' ATCAGAGTCACCAAGGTGGAC |
| fp2 | 5' CAGCCTCAGTCACCGCCATCCGC |
| fp3 | 5' CAGGCTTTCCTCTCACCGCAAACAC |
| fp4 | 5' TCTACGTCACCAGGGTCATCAGTGC |
| fp5 | 5' CATCACTGACCTCTACAAGTATCC |
| fp6 | 5' CATCCTCTATACCCGCAAGCC |
| fp7 | 5' AGGGCTGGGGGTCGCTGTC |
| fp8 | 5' GGATCCCCGAGGGCTGGGGGTCGCTGTC |
| fp5' 1 | 5' GGGAAACCAGTGCAGGATCCATG |
| fp5' 2 | 5' GGAGTCCAGCTGGCCTTCACATG |
| fp5' 3 | 5' GCTATTACACCACATATGGCAG |
| fp5' 4 | 5' GAAGGACCCACTTGCCAGATG |
| fp5' 5 | 5' CAAGTCTTCCAGGTCCAGAGG |
| fp5' 6 | 5' CCACATGGAAAGGAACCAAG |
| fp5' 7 | 5' CAGTTGATATTCTTGAAAGGGAAG |
| fp5' 8 | 5' GATTGGACAAAGGAGAGATGAAC |
| fp5' 9 | 5' TGGCTCAGCAGAAGGAAGGGC |

| Primer name | Primer sequence |
|-------------|---------------------------------|
| rP1 | 5' TCACAGTCCAGGGTCTGCTCAG |
| rP2 | 5' GATACTTGTAGGTCAGTGATG |
| rP3 | 5' AGAAACACATCGTCCTCTGCACACC |
| rP4 | 5' CGTACTTGACTGGCTTGTACCTGCC |
| rP5 | 5' GCCTGTTACCCGCTTGCTGT |
| rP6 | 5' ACGGCGTGGGATTCATGCAGG |
| rP7 | 5' AGGTTGAGGCAGGATTAAGTGAAGCCTC |
| rP8 | 5' GATGTCTGGAACACAGCTTTGG |
| rP9 | 5' GCGTGAGGCTGAACGACTGG |
| rP10 | 5' CAGCTCAGGGCAGGGGTACTC |
| rP11 | 5' AATCCACCCCTCTCAGTCCCT |
| rP12 | 5' CCACCATCAGCGCAGAGTCC |
| rP13 | 5' TACTTGCCCTGCCACTGTCCT |
| rP14 | 5' AGGCTTCCACGCCGAACAGC |
| rP15 | 5' GCCTTGCTGTACACCTGCCTC |
| rP16 | 5' GTGATGTAGTCGGCGTTGACCTC |
| rP17 | 5' AATCCACCCCTCTCAGTCCC |
| rP18 | 5' ATGATGCCGCTGTCGCTGAG |
| rP19 | 5' GAATTCAATCCACCCCTCTCAGTCCCT |
| rP5' 1 | 5' ACCGTCAGGACAGTGTGGC |

Table 2

| Primer name | Primer sequence |
|-------------|--------------------------|
| mfP1 | 5' GAGGACGATGTGTTTCTGGAC |
| mfP2 | 5' TAGTCACCAGGAGGACCCAG |
| mfP3 | 5' TACAGAAGTGGTTCAAGGG |
| mfP4 | 5' GCCTCGGTCACTGCCATCC |
| mfP5 | 5' CTAAGTGTGGAAGCCTGGG |
| mfP6 | 5' CCCTGGAGGACTTCATCACC |
| mfP7 | 5' CAGCAGTGGTAGGAGTGACAG |
| mfP8 | 5' ATTACCCTTGCTGTGAACTGG |
| mfP9 | 5' AGAACACAATCACCGTCAGCC |

| Primer name | Primer sequence |
|-------------|--------------------------|
| mrP1 | 5' TGTCTGTTCCCTCTTTGCTGC |
| mrP2 | 5' CTCTGCTGTCTGGAATACGGC |
| mrP3 | 5' ATGCGTGAGGTCCTGTCTCC |
| mrP4 | 5' CAGTTCAGAGCAAGGGTAATC |
| mrP5 | 5' AGGGTGAGACTGAGTAACTGG |
| mrP6 | 5' GTAATCTTGGGTGTGGTGGG |
| mrP7 | 5' ATCCTGGTGACATAGACTGGG |
| mrP8 | 5' GAGTCCAGAAACACATCGTCC |
| mrP9 | 5' CAGGTCCTCCAGGGTTCGG |
| mrP10 | 5' TCCGCAAACACGAAGGGGC |
| mrP11 | 5' ATCCGCCTCCTCCTCCCC |
| mrP12 | 5' TTCTTCTAGTGGCCCCAGC |
| mrP13 | 5' TTCTTTTCTTCTACTGGCCCC |
| mrP14 | 5' AGAACACAATCACCGTCAGCC |

Table 3

| Primer name | Primer sequence |
|--------------|------------------------------|
| TGase 5 mfP1 | 5' AGCCCCCTGGTTCCATTCTGG |
| TGase 5 mrP1 | 5' GGCTTTGTGATGGGGTTTGAGGACG |

Table 4

| TGase | Forward primer sequence | Reverse primer sequence | Product size (bp) |
|---------|-----------------------------|-------------------------------|-------------------|
| TGase 1 | 5' CTCGGACTCTGTGACCATGC | 5' CAACTGCTGCTCCCAGTAACG | 182 |
| TGase 2 | 5' CGAATCCTCTACGAGAAGTACAGC | 5' CAGTTTGCGGTTTTGCTTGG | 177 |
| TGase 3 | 5' GCAGCCTCAAGATTGATGTGC | 5' AGACATCAATGGGCAGCATGG | 157 |
| TGase 4 | 5' CCCTGTGTTTCTGTACTCTGC | 5' CAATGAACTTCTGGGATTCC | 252 |
| TGase 5 | 5' AGCAGCAGAGAGTTCTCATTGG | 5' CCTTGATGTCTTTAAACCTATTGCTC | 135 |
| TGase 6 | 5' GGCAGTGACTCTCGTGTGGAC | 5' TCACGGTGTGGGATTCACGC | 171 |
| TGase 7 | 5' GGGAGTGGCCTCATCAATGGG | 5' CCTTGACCTCACTGCTGCTGA | 126 |
| hFXIII | 5' GACCAATGAAGAAGATGTTCCGT | 5' GAAGGGCGTCTTTGAATCTGCAC | 169 |
| hB4.2 | 5' CAGGAGGAGTAAGGGGAATTGG | 5' TGCAGGCTACTGGAATCCACAG | 120 |

Appendix 3

1: AF540969. Homo sapiens tran...[gi:33331029] [Links](#)

LOCUS AF540969 2292 bp mRNA linear PRI 11-NOV-2003

DEFINITION Homo sapiens transglutaminase y (TGM6) mRNA, complete cds;
 alternatively spliced.

ACCESSION AF540969

VERSION AF540969.1 GI:33331029

KEYWORDS .

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 2292)

AUTHORS Grenard,P., Bates,M.K. and Aeschlimann,D.

TITLE 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

JOURNAL J. Biol. Chem. 276 (35), 33066-33078 (2001)

MEDLINE 21413858

PUBMED 11390390

REFERENCE 2 (bases 1 to 2292)

AUTHORS Thomas,H. and Aeschlimann,D.

TITLE A novel transglutaminase expressed in the nervous system

JOURNAL Unpublished

REFERENCE 3 (bases 1 to 2292)

AUTHORS Thomas,H. and Aeschlimann,D.

TITLE Direct Submission

JOURNAL Submitted (09-AUG-2002) Matrix Biology & Tissue Repair Research
 Unit, Dental School, University of Wales College of Medicine, Heath
 Park, Cardiff CF14 4XY, United Kingdom

FEATURES Location/Qualifiers

source 1..2292

 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /chromosome="20"
 /map="20q11"
 /cell_line="H69"
 /cell_type="small cell carcinoma"
 /tissue_type="lung"

gene 1..2292

 /gene="TGM6"

CDS 62..2182

 /gene="TGM6"

 /EC_number="2.3.2.13"

 /note="transglutaminase type 6, TGy, TG6; catalyzes
 R-glutamyl-peptide, amine-gamma-glutamyl transferase
 reaction; alternatively spliced"

 /codon_start=1

 /product="transglutaminase y"

 /protein_id="AAQ10751.1"

 /db_xref="GI:33331030"

/translation="MAGIRVTKVDWQSRNGAAHHTQEYPCPELVVRRGQSFSLTLEL
SRALDCEEILIFTVETGPRASEALHTKAVFQTSELERGEGWTAAREAQMEKTLTVSLA
SPPSAVIGRYLLSIRLSSHRKHSNRRLGEFVLLFNPWCAEDDVFLASEEERQEYVLS
SGIIFRGVEKH IRAQGWNYGQFEEDILNICLSILDRSPGHQNNPATDVSCRHNPIYVT
RVISAMVNSNDRGVVQGWQKYGGGTSPLHWRGSAVAILQKWLKGRYKPVKYGQCWV
FAGVLCTVLRCLGIATRVSNFNSAHDTDQNLSDKYVDSFGRTLEDLTEDSMWNFHV
WNESEWAFARQDLGPSYNGWQVLDATPQEESEGVFRCGPASVTAIREGDVHLAHDGPFVF
AEVNADYITWLWHEDESREVRYSNTKKIGRCISTKAVGSDSRVDITDLYKYPEGSRKE
RQVYSKAVNRLFGVEASGRRIWIRRAGGRCLWRDDLLEPATKPSIAGKFKVLEPPMLG
HDLRLALCLANLTSRAQRVRVNLSGATILYTRKPVAEILHESHAVRLGPQEEKRIPIT
ISYSKYKEDLTEDKKILLAAMCLVTKGEKLLVEKDITLEDFITIKVLGPAMVGVAVTV
EVTVVNPLIERVKDCALMVEGSGLLQEQLSIDVPTLEPQERASVQFDITPSKSGPRQL
QVDLVSPHFPDIKGFVIVHVATAK"

ORIGIN

| | | | | | | |
|------|-------------|------------|-------------|------------|------------|-------------|
| 1 | actgtcctga | cggtgcacac | actgctgtgt | ggaggaacag | aggagtccag | ctggccttca |
| 61 | catggcagg | atcagagtca | ccaaggtgga | ctggcagcgg | tcgaggaatg | gcgctgcca |
| 121 | ccacaccag | gagtaccct | gccctgagct | ggtggttcgc | agggccag | cgttcagcct |
| 181 | cacgtggag | ctgagcagag | ccctggactg | tgaggagatc | ctcatcttca | cggtggagac |
| 241 | aggaccccg | gcttctgagg | ccctccacac | caaagctgtg | ttccagacat | cggagctgga |
| 301 | gcgggtgag | ggctggacag | cagcaagga | ggctcagatg | gagaaaactc | tgaccgtcag |
| 361 | tctcgccag | cctcccagtg | ctgtcattgg | ccgctacctg | ctgagcatca | ggcttctctc |
| 421 | tcaccgcaa | cacagcaacc | ggaggctggg | cgagtttgtt | ctccttttca | acctatgggtg |
| 481 | tgagaggac | gatgtgtttc | tggcctcaga | ggaggagaga | caggagtacg | tgctcagcga |
| 541 | cagcggcatc | atcttccgag | gcgtggagaa | gcacatacga | gcccagggct | ggaactacgg |
| 601 | gcagtttgag | gaggacatcc | tgaacatctg | cctctccatc | ctggatcgaa | gccccgggtca |
| 661 | caaaaacaac | ccagccaccg | acgtgtcctg | ccgccacaac | cccatctacg | tcaccaggg |
| 721 | catcagtgcc | atggtgaaca | gcaacaacga | ccgaggtgtg | gtgcaaggac | agtggcagg |
| 781 | caagtacggc | ggcggcacca | gcccgcgtgca | ctggcgcggc | agcgtggcca | ttctgcagaa |
| 841 | gtggctcaag | ggcaggtaca | agccagtcaa | gtacggccag | tgctgggtct | tcgccggagt |
| 901 | cctgtgcaca | gtcctcaggt | gcttggggat | agccacacgg | gtcgtgtoca | acttaactc |
| 961 | agcccacgac | acagaccaga | acctgagtg | ggacaaatac | gtggactcct | tcgggcggac |
| 1021 | cctggaggac | ctgacagaag | acagcatgtg | gaatttccat | gtctggaatg | agagctggt |
| 1081 | tgcccggcag | gacctaggcc | cctcttataa | tgctggcag | gttctggatg | ccacccccca |
| 1141 | ggaggagagt | gaaggtgtgt | tccggtgagg | cccagcctca | gtcaccgcca | tccgcgagg |
| 1201 | tgatgtgcac | ctggctcag | atggcccctt | cgtgtttgcg | gaggtcaacg | ccgactacat |
| 1261 | cacctggctg | tggcacgagg | atgagagccg | ggagcgtgta | tactcaaaca | cgaagaagat |
| 1321 | tgggagatgc | atcagcacca | aggcgggtgg | cagtgactcc | cgctgggaca | tactgacct |
| 1381 | ctacaagtat | ccggaagggt | cccggaaaga | gaggcaggtg | tacagcaagg | cggtgaacag |
| 1441 | gctgttcggc | gtggaagcct | ctggaaggag | aatctggatc | cgcagggctg | gggtctcgtg |
| 1501 | tctctggcgt | gacgacctcc | tggagcctgc | caccaagccc | agcatcgctg | gcaagttcaa |
| 1561 | ggtgctagag | cctcccatgc | tgggccacga | cctgagactg | gccctgtgct | tggccaacct |
| 1621 | cacctcccgg | gcccagcggg | tgagggtcaa | cctgagcggg | gccaccatcc | tctatacccg |
| 1681 | caagccagtg | gcagagatcc | tgcatgaatc | ccacgccgtg | aggctggggc | cgcaagaaga |
| 1741 | gaagagaatc | ccaattaca | tatcttactc | taagtataaa | gaagacctga | cagaggacaa |
| 1801 | gaagatcctg | ttggctgcca | tgtgccttgt | caccaaagga | gagaagcttc | tggtggagaa |
| 1861 | ggacattact | ctagaggact | tcataccat | caaggttctg | ggcccagcca | tggtgggagt |
| 1921 | ggcagttaca | gtggaagtga | cagtagtcaa | ccccctcata | gagagagtga | aggactgtgc |
| 1981 | gctgatgggtg | gagggcagcg | gccttctcca | ggaacagctc | agcatcgacg | tgctacctc |

2041 ggagcctcag gagagggcct cagtccagtt tgacatcacc ccctccaaaa gtggccaag
2101 gcagctgcag gtggaccttg taagccctca cttcccggac atcaagggtt ttgtgatcgt
2161 ccatgtggcc actgccaagt gatggatcat gagggactga gaggggtgga ttggcccct
2221 gtcctcctcc tgcccattct ttgtctcttc cacatgggag ccaggaggcc tcagttaatc
2281 ctgcctcaac ct

Appendix 4

1: [AF540970](#). Homo sapiens tran...[gi:33331031] Links

LOCUS AF540970 2158 bp mRNA linear PRI 11-NOV-2003

DEFINITION Homo sapiens transglutaminase y short form (TGM6) mRNA, complete cds; alternatively spliced.

ACCESSION AF540970

VERSION AF540970.1 GI:33331031

KEYWORDS .

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 2158)

AUTHORS Thomas,H. and Aeschlimann,D.

TITLE A novel transglutaminase expressed in the nervous system

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 2158)

AUTHORS Thomas,H. and Aeschlimann,D.

TITLE Direct Submission

JOURNAL Submitted (09-AUG-2002) Matrix Biology & Tissue Repair Research

Unit, Dental School, University of Wales College of Medicine, Heath

Park, Cardiff CF14 4XY, United Kingdom

FEATURES Location/Qualifiers

source 1..2158

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/chromosome="20"

/map="20q11"

/cell_line="H69"

/cell_type="small cell carcinoma"

/tissue_type="lung"

gene 1..2158

/gene="TGM6"

CDS 62..1939

/gene="TGM6"

/EC_number="2.3.2.13"

/note="transglutaminase type 6 (d12), TGy(d12), TG6(d12);

catalyzes R-glutaminyI-peptide, amine-gamma-glutamyl

transferase reaction; alternatively spliced; lacks

sequence from exon 12 encoding for the second beta-barrel

domain"

/codon_start=1

/product="transglutaminase y short form"

/protein_id="AAQ10752.1"

/db_xref="GI:33331032"

/translation="MAGIRVTKVDWQSRNGAAHHTQEYPCPELVVRRGQSFSLTLEL

SRALDCEEILIFTVETGPRASEALHTKAVFQTSELERGEGWTAAREAQMEKTLTVSLA

SPPSAVIGRYLLSIRLSSHRKHSNRRLGEFVLLFNPWCAEDDVFLASEEERQEYVLS

SGIIFRGVEKH IRAQGWNYGQFEEDILNICALSILDRSPGHQNNPATDVSCRHNPIYVT

RVISAMVNSNDRGVVQGWQGGKYGGGTSPLHWRGVSVAIIQKWLKGRYKPVKYQCWV
FAGVLCTVLRCLGIATRVSNFNSAHDTDQNLSDKYVDSFGRTLEDLTEDSMWNFHV
WNESWFARQDLGPSYNGWQVLDATPQEESEGVFRCGPASVTAIREGDVHLAHDGPFVF
AEVNADYIITWLWHEDESRRVYSNTKKIGRCISTKAVGSDSRVDITDLYKYPEGSRKE
RQVYSKAVNRLFVEASGRRIWIRRAGGRCLWRDLDLLEPATKPSIAGKFKVLEPPMLG
HDLRLALCLANLTSRAQRVRVNLGATILYTRKPVAEILHESHAVRLGPEEKRIPIIT
ISYSKYKEDLTEDKKILLAAMCLVTKGEKLLVEKDITLEDFITIKRAYPGASGEGLSP

V"

ORIGIN

```
//      1 actgtcctga cggtgcacac actgctgtgt ggaggaacag aggagtccag ctggccttca
61 catggcaggg atcagagtca ccaaggtgga ctggcagcgg tcgaggaatg gcgctgcca
121 ccacaccag gagtaccct gccctgagct ggtggtcgc aggggccagt cgttcagcct
181 cacgctggag ctgagcagag ccctggactg tgaggagatc ctcatcttca cggtgagag
241 aggaccccg gcttctgagg ccctccacac caaagctgtg ttccagacat cggagctgga
301 gcggggtgag ggctggacag cagcaagga ggctcagatg gagaaaactc tgaccgtcag
361 tctcgccagc cctcccagtg ctgtcattgg ccgctacctg ctgagcatca ggctttcctc
421 tcaccgcaaa cacagcaacc ggaggctggg cgagtgtgtt ctcttttca acccatggtg
481 tgcagaggac gatgtgttcc tggcctcaga ggaggagaga caggagtacg tgctcagcga
541 cagcggcatc atcttccgag gcgtggagaa gcacatacga gcccagggct ggaactcagg
601 gcagtttgag gaggacatcc tgaacatctg cctctccatc ctggatcgaa gccccggtca
661 caaaacaac ccagccaccg acgtgtcctg ccgccacaac cccatctacg tcaccaggg
721 catcagtgcc atgggtgaaca gcaacaacga ccgaggtgtg gtgcaaggac agtggcagg
781 caagtacggc ggcggcacca gcccgctgca ctggcgcggc agcgtggcca ttctgcagaa
841 gtggctcaag ggcaggtaca agccagtcaa gtacggccag tgctgggtct tcgcccaggt
901 cctgtgcaca gtctcaggtt gcttggggat agccacacgg gtcgtgtcca acttcaactc
961 agcccacgac acagaccaga acctgagtggt ggacaaatac gtggactcct tcgggaggac
1021 cctggaggac ctgacagaag acagcatgtg gaatttccat gtctggaatg agagctggtt
1081 tgcccggcag gacctaggcc cctcttaciaa tggctggcag gttctggatg ccaccccca
1141 ggaggagagt gaaggtgtgt tccggtgagg cccagcctca gtcaccgcca tccgagagg
1201 tgatgtgcac ctggctcacg atggcccctt cgtgtttgag gaggtaacg ccgactacat
1261 cacctggctg tggcacgagg atgagagccg ggagcgtgta tactcaaaca cgaagaagat
1321 tgggagatgc atcagcacca agcggtggg cagtactcc cgcgtggaca tcaactgact
1381 ctacaagtat ccggaagggg cccgaaaga gaggcaggtg tacagcaagg cgggtaacag
1441 gctgttcggc gtggaagcct ctggaaggag aatctggatc cgcagggctg ggggtcgtg
1501 tctctggcgt gacgacctcc tggagcctgc caccaagccc agcatcgtg gcaagttcaa
1561 ggtgctagag cctcccatgc tggccacga cctgagactg gccctgtgct tggccaacct
1621 cacctcccgg gcccagcggg tgagggtcaa cctgagcggg gccaccatcc tctatacccg
1681 caagccagtg gcagagatcc tgcagatcc ccacgctg aggctggggc cgcaagaaga
1741 gaagagaatc ccaattacaa tatcttactc taagtataaa gaagacctga cagaggacaa
1801 gaagatcctg ttggctgcca tgtgccttgt caccaaagga gagaagcttc tggtgagaa
1861 ggacattact ctagaggact tcatccat caagcgtgcc taccctggag cctcaggaga
1921 gggcctcagt ccagtttgac atcaccocct ccaaaagtgg cccaaggcag ctgcaggtgg
1981 acctgtgaag ccctcacttc ccggacatca agggctttgt gatcgtccat gtggcactg
2041 ccaagtgatg gatcatgagg gactgagagg ggtgatttg gccctgtcc tctcctgccc
2101 cattctttgt ctctccaca tgggagccag gaggcctcag ttaatcctgc ctcaacct
```

Appendix 5

1: [BY357397](#). BY357397 RIKEN fu...[gi:26586885] Links

IDENTIFIERS

dbEST Id: 16035305
EST name: BY357397
GenBank Acc: BY357397
GenBank gi: 26586885

CLONE INFO

Clone Id: 2700032M19 (3')
DNA type: cDNA

PRIMERS

PolyA Tail: Unknown

SEQUENCE

```
TGCACTTAGAGGATTTCCGGCGTAGGAGTGGACCCCTCTCAAGATGCCTTTGAAACATGGGGTTGAGGGCTG  
CGCTACAGGAGGAACAGGCCAGGAAGGTCACAAGACTAATGCTAAGCGGACCAGGGAGAATAGGAGGAG  
CTAATGGTTATTATGGTCAGCCGTTCTTACCATTTCGCATCCCTCAAGTACATG
```

Entry Created: Dec 12 2002

Last Updated: Dec 12 2002

COMMENTS

Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H., Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R., Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y. Direct Submission

Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences *Mamm. Genome*. 12, 673-677 (2001)

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. *Genome Res*. 10 (10), 1617-1630 (2000)

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. *Genome Res*. 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. *Genome Res*. 11 (2), 281-289 (2001)

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

LIBRARY

Lib Name: RIKEN full-length enriched, 11 days embryo whole body
Organism: Mus musculus

Strain: C57BL/6J
Tissue type: whole body
Develop. stage: 11 days embryo

SUBMITTER

Name: Yoshihide Hayashizaki
Lab: Laboratory for Genome Exploration Research Group, RIKEN
Genomic Sciences Center(GSC), Yokohama Institute
Institution: The Institute of Physical and Chemical Research (RIKEN)
Address: 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045,
Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
E-mail: genome-res@gsc.riken.go.jp,
URL:<http://genome.gsc.riken.go.jp/>

CITATIONS

Medline UID: 22354683

Title: Analysis of the mouse transcriptome based on functional
annotation of 60,770 full-length cDNAs

Authors: Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H.,
Kondo, S., Nikaido, I., Osato, N., Saito, R., Suzuki, H.,
Yamanaka, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y.,
Nogami, A., Schonbach, C., Gojobori, T., Baldarelli, R., Hill
, D.P., Bult, C., Hume, D.A., Quackenbush, J., Schriml, L.M.,
Kanapin, A., Matsuda, H., Batalov, S., Beisel, K.W., Blake, J.A.,
Bradt, D., Brusic, V., Chothia, C., Corbani, L.E., Cousins, S.,
Dalla, E., Dragani, T.A., Fletcher, C.F., Forrest, A., Frazer
, K.S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A.,
Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson
, I.J., Jarvis, E.D., Kanai, A., Kawaji, H., Kawasaki, Y.,
Kedzierski, R.M., King, B.L., Konagaya, A., Kurochkin, I.V., Lee
, Y., Lenhard, B., Lyons, P.A., Maglott, D.R., Maltais, L.,
Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K.
, Okido, T., Pavan, W.J., Pertea, G., Pesole, G., Petrovsky, N.,
Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S., Ravasi, T.,
Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M.,
Sandelin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K.
, Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D.,
Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y.,
Watanabe, Y., Wells, C., Wilming, L.G., Wynshaw-Boris, A.,
Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y.
, Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T.,
Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T.,
Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A.
, Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I.,
Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A.,
Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S., Rogers
, J., Birney, E., Hayashizaki, Y.

Citation: Nature 420: 563-573 2002

Appendix 6

1: BB612839. BB612839 RIKEN fu...[gi:16453665] [Links](#)

IDENTIFIERS

dbEST Id: 10035201
EST name: BB612839
GenBank Acc: BB612839
GenBank gi: 16453665

CLONE INFO

Clone Id: 4732422L03 (5')
DNA type: cDNA

PRIMERS

PolyA Tail: Unknown

SEQUENCE

```
GATGTTCTTCATTAATGTGGATCTGGCTGGAGATCCTGACGAACTACACTCTGCTGCCCA
GGGAGCAGGCGTCCATCTGGCTGTTGCATGGCAGGGGTCAGAATCACCAAGGTCGACTGG
CAGCATTCGAAAAATGGTGCTGCCACCACACCCAAGATTACCCTTGCTCTGAACTGGTG
GTTCGAAGAGGCCAGTTATTTCAGTCTCACCTGGATCTGAGTAGAGTCCTGGACAGCGAG
GAGGCCCTCATCTTCACCGTGGAGACAGTTTGAAGAGGACATCCTGAACATCTGCCTCTC
CATTCTGGACCGAAGTCCTAGTCACCAGGAGGACCCAGCCACCGATGTGTCCACCGCCA
TGACCCAGTCTATGTACCAGGATCATCAGTGCCATGGTGAACAGCAACAATGACCGAGG
TGTGGTACAAGGCCAGTGGCAGGGCAAATATGGAGGGTGGTTACCCCAAACCCACTGAAC
TGGGCGTGGCAGTGTGGCCATTCCCTTAACAGNAAAGTGGGTTTCAAAGGGGGAGGTAC
AAGCCAGTCAAATATGGCCAGTGTGGGTCTTCGCTGGAGTCATGTGCACAGTCCTCAGG
TGCCCTGGGGATCGCAACCAGGGTGGTGG
```

Entry Created: Oct 26 2001

Last Updated: Oct 26 2001

COMMENTS

Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and
Hayashizaki,Y.

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. *Genome Res.* 10 (10), 1617-1630 (2000)

wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. *Genome Res.* 10 (11), 1757-1771 (2000)

Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. *Genome Res.* 11 (2), 281-289 (2001)

Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.

Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences. *Mamm. Genome*. 12, 673-677 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

LIBRARY

Lib Name: RIKEN full-length enriched, 10 day neonate skin

Organism: Mus musculus

Strain: C57BL/6J

Sex: mixed

Tissue type: skin

Develop. stage: 10 days neonate

Lab host: DH10B

R. Site 1: Sall

R. Site 2: BamHI

Description: cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5'GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5'GAGAGAGAGATTCTCGAGTTAATTAATTAATCCCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I

SUBMITTER

Name: Yoshihide Hayashizaki

Lab: Laboratory for Genome Exploration Research Group, RIKEN
Genomic Sciences Center(GSC), Yokohama Institute

Institution: The Institute of Physical and Chemical Research (RIKEN)

Address: 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045,
Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

E-mail: genome-res@gsc.riken.go.jp

URL:<http://genome.gsc.riken.go.jp/>

CITATIONS

Title: RIKEN Mouse ESTs (Arakawa,T., et al. 2001)

Authors: Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T.,
Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J.,
Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A.,
Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C.,
Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A.,
Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A.,
Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M.,
Hayashizaki,Y.

Year: 2001

Status: Unpublished

Appendix 7

1: [BB630356](#). BB630356 RIKEN fu...[gi:16467348] [Links](#)

IDENTIFIERS

dbEST Id: 10048884
EST name: BB630356
GenBank Acc: BB630356
GenBank gi: 16467348

CLONE INFO

Clone Id: A030011P16 (5')
DNA type: cDNA

PRIMERS

PolyA Tail: Unknown

SEQUENCE

```
ATTAATGTGGATCTGGCTGGAGATCCTGACGAACTACACTCTGCTGCCAGGGAGCAGGC
GTCCATCTGGCTGTTGCATGGCAGGGGTCAGAATCACCAAGGTCGACTGGCAGCATTCTGA
AAAATGGTCTGCCACCACACCCAAGATTACCCTTGCTCTGAACTGGTGGTTCGAAGAG
GCCAGTTATTCAGTCTCACCCCTGGATCTGAGTAGAGTCTGGACAGCGAGGAGGCCCTCA
TCTTCACCGTGGAGACAGGACCTCACGCATCCGAGGCTCTCCACACCAAAGCCGTATTCC
AGACAGCAGAGCTAGAAAATAGGGGACACTTGGACAGCTGCAAAAAGAGGAACAGACAGAGA
ACACAATCACCGTCAGCCTCTCGAGCCCTTCCAATGCAGTCATTGGCCGATACCTGCTGA
GTGTCAGGCCCTCTTCCCGCAGGAAACACAGTGACCGAAAGCTGGGCCAGTTTATTCTCC
TGTTCAATCCCTGGTGCCAGAGGACGATGTGTTTCTGGACTCAGAGGAAGAGAGGCAGG
AGTACGTGCTTAACGACAGTGGCGTCATCTTCCGAGGCGTGAGAAAGCACATCCGAGCCC
AGGGCTGGAACACTACGGGCAGTTTGAAAAAGACATCCTGAACATCTGCCTTTC
```

Entry Created: Oct 26 2001

Last Updated: Oct 26 2001

COMMENTS

Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes.

Genome Res.. 10 (10), 1617-1630 (2000) wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y. RIKEN integrated sequence analysis (RISA) system--384-formatsequencing pipeline with 384 multicapillary sequencer.

Genome Res.. 10 (11), 1757-1771 (2000)Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P.,Sugahara,Y. and Hayashizaki,Y.

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library.

Genome Res.. 11 (2), 281-289 (2001) Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.

Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp>)for further details.e mouse tissues.

LIBRARY

Lib Name: RIKEN full-length enriched, 6 days neonate skin

Organism: Mus musculus

Tissue type: skin

Develop. stage: 6 days neonate

Lab host: DH10B

R. Site 1: Sall

R. Site 2: BamHI

Description: cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5'GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 185.0. Second strand cDNA was prepared with the primer adapter of sequence [5'GAGAGAGAGATTCTCGAGTTAATTAATTAATCCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I.

SUBMITTER

Name: Yoshihide Hayashizaki

Lab: Laboratory for Genome Exploration Research Group, RIKEN
Genomic Sciences Center(GSC), Yokohama Institute

Institution: The Institute of Physical and Chemical Research (RIKEN)

Address: 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045,
Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

E-mail: genome-res@gsc.riken.go.jp

URL: <http://genome.gsc.riken.go.jp/>

CITATIONS

Title: RIKEN Mouse ESTs (Arakawa,T., et al. 2001)

Authors: Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M., Hayashizaki,Y.

Year: 2001

Status: Unpublished

Appendix 8

1: [AK037229](#). *Mus musculus* 6 da...[gi:26332075] [Links](#)

LOCUS AK037229 3311 bp mRNA linear HTC 03-APR-2004
DEFINITION *Mus musculus* 6 days neonate skin cDNA, RIKEN full-length enriched library, clone:A030011P16 product:similar to PROTEIN-GLUTAMINE GLUTAMYLTRANSFERASE E3 PRECURSOR (EC 2.3.2.13) (TGASE E3) (TRANSGLUTAMINASE 3) [*Mus musculus*], full insert sequence.

ACCESSION AK037229

VERSION AK037229.1 GI:26332075

KEYWORDS HTC; CAP trapper.

SOURCE *Mus musculus* (house mouse)

ORGANISM *Mus musculus*

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1

AUTHORS Carninci,P. and Hayashizaki,Y.

TITLE High-efficiency full-length cDNA cloning

JOURNAL *Meth. Enzymol.* 303, 19-44 (1999)

MEDLINE [99279253](#)

PUBMED [10349636](#)

REFERENCE 2

AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.

TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes

JOURNAL *Genome Res.* 10 (10), 1617-1630 (2000)

MEDLINE [20499374](#)

PUBMED [11042159](#)

REFERENCE 3

AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.

TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer

JOURNAL *Genome Res.* 10 (11), 1757-1771 (2000)

MEDLINE [20530913](#)

PUBMED [11076861](#)

REFERENCE 4

AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.

TITLE Functional annotation of a full-length mouse cDNA collection

JOURNAL *Nature* 409, 685-690 (2001)

REFERENCE 5

AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.

TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

JOURNAL *Nature* 420, 563-573 (2002)

REFERENCE 6 (bases 1 to 3311)

AUTHORS Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T., Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.

TITLE Direct Submission

JOURNAL Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)

COMMENT cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site for further details.

URL:<http://genome.gsc.riken.jp/>

URL:<http://fantom.gsc.riken.jp/>

FEATURES Location/Qualifiers

source 1..3311
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:A030011P16"
/db_xref="taxon:10090"
/clone="A030011P16"
/tissue_type="skin"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="6 days neonate"

CDS 78..2198
/note="unnamed protein product; putative similar to PROTEIN-GLUTAMINE GLUTAMYLTRANSFERASE E3 PRECURSOR (EC 2.3.2.13) (TGASE E3) (TRANSGLUTAMINASE 3) [Mus musculus] (SWISSPROT|Q08189, evidence: FASTY, 48.9%ID, 99.8%length, match=2109)"
/codon_start=1
/protein_id="BAC29768.1"
/db_xref="GI:26332076"
/translation="

MAGVRI TKVDWQH SKNGAAHHTQDYPCSELVVRGQLFSLTLDL

SRVLDSEEALIFTVETGPHASEALHTKAVFQTAELEIGDTWTAAKEEQTENTITVLSL

SPSNAVIGRYLLSVRPSSRRKHSRDLGQFILLFNPWCPEDDVFLDSEERQEYVLND
SGVIFRGVEKHIRAQGWNYGQFEEDILNICLSILDRSPSHQEDPATDVSHRHDPVYVT
RIISAMVNSNDRGVVQGWQKYGGGTNPLNWRGSAVAILQKWFKGRYKPVKYGQCWV
FAGVMCTVLRCLGIATRVSFNSAHDTDGNLSVDKYVDSYGRITLEDLTDSDMWNFHV
WNESWFARQDLGPSYDGWQVLDATPQEESSEGMFRCGPASVTAIREGDVHLAHDGPFVF
AEVNADYITWLWHEDKRRERVYSDTKKIGRCISTKAVGSDSRVDITGLYKYPEGSRKE
RQVYSKAVKKLLSVEAWGRRRRIRRASVRGVWRDDLLEPVTKPSITGKFKVLEPPVLG
QDLKLALCLTNLTARAQRVRVNVSGATILYTRKPVAEILRESHTVKLGPLEEKKIPVT
ISYSQYKGDLTEDKKILLAAMCLVSKGEKLLVEKDITLEDFITIKVLGPAVVGVTVTV
EVLVINPLSESVKDCVLMVEGSGLLQGQLSIEVPSLQPQEKALIQFNITPSKSGPRQL
QVDLVSSQFPDIKGFVIIHVATAK"

polyA signal 3297..3302

/note="putative"

polyA site 3311

/note="putative"

ORIGIN

1 attaatgtgg atcgtgctgg agatcctgac gaactacact ctgctgccca gggagcaggc
61 gtccatctgg ctgtgcatg gcagggtgca gaatcaccaa ggtcactgag cagcattcga
121 aaaatggtgc tccccaccac acccaagatt acccttgctc tgaactggtg gticgaagag
181 gccagttatt cagtctcacc ctggatctga gtagagtctt ggacagcgag gaggccctca
241 tcttcaccgt ggagacagga cctcacgcat cggaggctct cacacacaaa gccgtattcc
301 agacagcaga gctagaataa ggggacactt ggacagctgc aaaagaggaa cagacagaga
361 acacaatcac cgtcagcctc tcgagccctt ccaatgcagt cattggccga tacctgctga
421 gtgtcaggcc ctctccccc aggaaacaca gtgaccgaaa gctgggccag ttattctcc
481 ttttcaatcc ctgtgcccc gaggacgatg ttttctgga ctcagaggaa gagaggcagg
541 agtactgtct taacgacagt ggcgtcatct tccgaggcgt ggagaagcac atccgagccc
601 agggctggaa ctacggcgag ttgaagagg acatcctgaa catctgcctc tccattctg
661 accgaagtcc tagtcaccag gaggaccag ccaccgatgt tcccaccgc catgaccag
721 tctatgtcac caggatcac agtgccatg tgaacagcaa caatgaccga ggtgtgtac
781 aaggccagtg gcagggcaaa tatggaggtg gtaccaacc actgaactgg cgtggcagt
841 tggccatcct acagaagtgg ttcaaggca ggtacaagcc agtcaaatat ggccagtgt
901 gggctctcgc tggagtcag tgcacagtc tcagggtcct ggggatcga accagggtg
961 tgtccaactt caaticagcc cagcacaccg atggcaatct gagtgtgac aaatatgtg
1021 acttattagg ccgaacctg gaggacctga cagaagacag catgtggaac ttccagctt
1081 ggaatgagag ctggtttgca cgacaggacc tgggcccac ctatgatgac tggcaggtc
1141 tggatgccac tcccgaagag gagagcgaag gtatgttccg gtgtgtctc cctcgggca
1201 ctgcatccg tgaaggtag gtgcatctag cccatgatgg cccctcgtc ttgaggagg
1261 tcaatgcaga ctacatcacc tggctatgac acgaggacaa gagggggag cgtgtgtact
1321 cggacaccaa gaagatcggc aggtgatca gtaccaaggc agtgggcagt gactctctg
1381 tggacattac aggcctctac aagtatccag aagggtcccg gaaggagagg caggtgtaca
1441 gcaaagctgt gaagaaactg ctaagtgtgg aagcctgggg gaggaggagg cggatccgca
1501 gggccagcgt gcgcggtgtc tggcgcgacg acctcctgga accggtcacc aaaccagca
1561 tcaatggcaa gttcaaatg ctggagccac cgggtctcgg ccaggacctg aagctggctc
1621 tgtgctgac caacctcacc gccagggcc agcgggtccg agtcaatgtg agcggcgcca
1681 ccatctcta taccgcaag cctgtggcag agatcctgcg tgaatccac accgtgaagc
1741 tggggccact agaagaaaaa aagatcccag tcacaatctc ttactctcag tacaaggag
1801 acctgacaga agacaagaag atcctgttgg ctgcatgtg cctgtcagt aaggcgaga
1861 agctctctgt ggagaaggac attaccctgg aggacttcat caccatcaag gtgctggggc
1921 cagcagtgtt aggagtgaca gtcactgtgg aagtcttggg gatcaacccc ctctggaga
1981 gcgtaagga ctgtgtgctg atggtagagg gcagtggcct tctccaagga cagtaagta
2041 ttgagggtcc cagcttcag cctcaggaga aggccttcat tcaattcaac ataaccctt
2101 ccaagagtgg cccagggcag ctgcaggtgg acctgtcag ctacagttc ccagacatca

2161 agggctttgt gatcatccac gtggctacag ccaagtgatg ggccacgagg aactggggag
2221 gactttgtcc atctgtctgt ttctccata taagagacag gggctcagg ttggctctgc
2281 ttctgtcttt ggctccatg agctcggca aggtccttcc cagtgcattg atggaggtgc
2341 tggactcagt aaccctcac cctcaggaca agtggctgat gtaagagcag agggaccctc
2401 caactctgt caccttctg ctgggagagt ggcattgcc agatttctga gagccctggc
2461 ccaagagctc agccagctca gcccaggtaa ctcaccactt ctgccccttc tatcatctct
2521 tcaaatgtcc tgaccacatt tggacccaag aattctcaa gccggagcat cccattata
2581 gaggagcctt tctgaagcca taagacaag cagaacaag gttgagaacc aagtggctcc
2641 ccaactcatg ttctaaaag agcactggat tttaagcaac aggagattct tcctatgtg
2701 gaaacactga ctctggattt ggtgatgga gctaaaactg cctgagaaca gctgaagaag
2761 acaaaggagt tctcccccac actgggcagc atggatatac tccagctctg gccatcctgg
2821 agttcacagg actgcccag atcactcagc caticcttct tggaaactaga atggcttga
2881 cggtgccgat cctacacttg gagtcaaagc aggagcatgg atgcctggaa tcaggaccag
2941 agtagacaca acaaggccat gaagaatag ccactagata tgaattcag cactgaat
3001 tttagacaga cattaaat aagctacag ttaagaacat taacaaaca acaggtgcat
3061 ctgcaactca gtgcctctg gatacaaatg acggtgaggg aattgcagag atggctcagc
3121 tgtaagagt actttctgat ctctagagg ggtgagttt ggtcccagc actggtgtca
3181 gacagctcac atactataac tccaactcca gggctccaac atttctggtc tctgtggca
3241 cccacacta cctgctgcat gtacaccctc ctccaaaca catgcaatia aaataaata
3301 aatcctttt t

Appendix 9

1: [AA961594.oq80d09.s1 NCI_CG...\[gi:3133758\]](#) Links

IDENTIFIERS

dbEST Id: 1697949
EST name: oq80d09.s1
GenBank Acc: AA961594
GenBank gi: 3133758

CLONE INFO

Clone Id: IMAGE:1592657 (3')
Source: NCI
DNA type: cDNA

PRIMERS

Sequencing: -40m13 fwd. ET from Amersham
PolyA Tail: Unknown

SEQUENCE

```
TGTGTCTGGAGCCATCACAAAGGGGAAATGGATCTGGGACACGTGGAGCCTCCCTGCGGGA
AGGAGCATCTTTATACGTGGAGTGGCAGATGGCCTCCTGTTCCAGGGAAGATGCTTACAG
TGTTCCCTTCACTTGTGGCAGATGCATAATAGGAGAAAGGAACCCGCATGATGAGAGGAG
GGATCCCTGGTCAGTCGCTCCATAAGGATTTGGGCTATGAATGCAGATTCTCACAGAGTGA
CAGAGCTAAAAAATCCAGGGGGGTACCATGCTTCTCAAATATTTCTACCCAAGTGTCTC
AAAGGGAAAGGAAGGTGAGCGCAGATCCCTGGCTCAAGGACAGGCTGTCTTCACCCAGTG
ACCAGGGATATGGGGAATCTCAGATGAGAATGGAAATCCAAAGCTAAGGCCAGATCCATC
TGGGAC
```

Quality: High quality sequence stops at base: 403

Entry Created: May 15 1998

Last Updated: May 15 1998

COMMENTS

Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Stratagene, Inc.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing
Center
Clone distribution: NCI-CGAP clone distribution information
can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

LIBRARY

Lib Name: NCI_CGAP_Kid6
Organism: Homo sapiens
Sex: mixed
Organ: kidney
Tissue type: kidney tumor
Lab host: SOLR (kanamycin resistant)

Vector: Bluescript SK-

R. Site 1: EcoRI

R. Site 2: XhoI

Description: Cloned unidirectionally. Primer: Oligo dT. Pooled kidney tumors. 5' adaptor sequence: 5' GAATTCGGCACGAG 3' 3' adaptor sequence:

5'CTCGAGTTTTTTTTTTTTTTTTTTT 3' Average insert size: 1.0 kb.

SUBMITTER

Name: Robert Strausberg, Ph.D.

E-mail: cgapbs-r@mail.nih.gov

CITATIONS

Title: National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

Authors: NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>

Year: 1997

Status: Unpublished

Appendix 10

1: AL833804. Mouse DNA sequenc...[gi:32187965] Links

LOCUS AL833804 200521 bp DNA linear ROD 24-JUN-2003

DEFINITION Mouse DNA sequence from clone RP23-20A6 on chromosome 2, complete sequence.

ACCESSION AL833804

VERSION AL833804.16 GI:32187965

KEYWORDS HTG.

SOURCE *Mus musculus* (house mouse)

ORGANISM *Mus musculus*

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 200521)

AUTHORS Griffiths,C.

TITLE Direct Submission

JOURNAL Submitted (24-JUN-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

COMMENT On Jun 24, 2003 this sequence version replaced gi:32131092.

Sequence from the Mouse Genome Sequencing Consortium whole genome
shotgun may have been used to confirm this sequence. Sequence data
from the whole genome shotgun alone has only been used where it has
a phred quality of at least 30.

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: humquery@sanger.ac.uk

During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality \geq
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest, except on the rare
occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:

Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information
on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-20A6 is
from the RPCI-23 Mouse BAC Library

constructed by the group of Pieter de Jong.

For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBACe3.6.

```
FEATURES          Location/Qualifiers
source            1..200521
                 /organism="Mus musculus"
                 /mol_type="genomic DNA"
                 /db_xref="taxon:10090"
                 /chromosome="2"
                 /clone="RP23-20A6"
                 /clone_lib="RPCI-23"
```

Appendix 11

1: [AY159126](#). *Mus musculus* tran...[gi:37724519] [Links](#)

LOCUS AY159126 3317 bp mRNA linear ROD 11-NOV-2003
DEFINITION *Mus musculus* transglutaminase y (Tgm6) mRNA, complete cds.
ACCESSION AY159126
VERSION AY159126.1 GI:37724519
KEYWORDS .
SOURCE *Mus musculus* (house mouse)
ORGANISM *Mus musculus*
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 3317)
AUTHORS Thomas,H., Aeschlimann,P. and Aeschlimann,D.
TITLE A novel transglutaminase expressed in the nervous system
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 3317)
AUTHORS Aeschlimann,D., Thomas,H. and Aeschlimann,P.
TITLE Direct Submission
JOURNAL Submitted (04-OCT-2002) Matrix Biology & Tissue Repair Research
Unit, Dental School, University of Wales College of Medicine, Heath
Park, Cardiff CF14 4XY, United Kingdom
FEATURES Location/Qualifiers
source 1..3317
/organism="Mus musculus"
/mol_type="mRNA"
/strain="BALB/c"
/db_xref="taxon:10090"
/chromosome="2"
/map="74-78cM"
/tissue_type="brain"
gene 1..3317
/gene="Tgm6"
CDS 78..2198
/gene="Tgm6"
/EC_number="2.3.2.13"
/function="catalyzes R-glutaminyI-peptide
amine-gamma-glutamyl transferase reaction"
/note="transglutaminase type 6; TGy; TG6; similar to
sequence deposited in GenBank Accession Number AF540969"
/codon_start=1
/product="transglutaminase y"
/protein_id="AAO17786.1"
/db_xref="GI:37724520"
/translation="MAGVRI TKVDWQH SKNGAAHHTQDYPCSELVVRGQLFSLTLDDL
SRVLDSEEALIFTVETGPHASEALHTKAVFQTAELEIGDTWTAKEEQTENTITVSLL
SPSNAVIGRYLLSARPSSRRKHS DRKLGQFILLFNPWCPEDDVFLDSEEERQEYVLND
SGVIFRGVEKH IRAQGWN YGQFEEDILN ICLSILDRSPSHQEDPATDVSHRHDPVYVT
RIISAMVNSNDRGVVQGWQGGKYGGGTNPLNWRGSVAI LQKWFKGRYKPKVYGGQCWW

FAGVMCTVLRCLGIATRUVSNFNNSAHDTDGNLSVDKYVDSYGRITLEDLTEDSMWNFHV
 WNESWFARQDLGPSYDGWQVLDATPQEESEGMFRCPASVTAIREGDVHLAHDGPFV
 AEVNADYITWLWHEDKRRRERVYSDTKKIGRCISTKAVGSDSRVDITGLYKYPEGSRKE
 RQVYSKAVKKLLSVEAWGRRRRIRRASVRGVWRDDLLEPVTKPSITGKFKVLEPPVLG
 QDLKLALCLTNLTARAQRVRVNVSGATILYTRKPVAEILRESHTVKLGPLLEEKIPVT
 ISYSQYKGDLTEDKKILLAAMCLVSKGEKLLVEKDITLEDFITIKVLGPAVVGVTVTV
 EVLVINPLSESVKDCVLMVEGSGLLQGLSIEVPSLQPQEKALIQFNITPSKSGPRQL
 QVDLVSSQFPDIKGFVIIHVATAK"

polyA_signal 3291..3296
 /gene="Tgm6"

ORIGIN

| | | | | | | |
|------|-------------|------------|------------|------------|------------|------------|
| 1 | attaatgtgg | atctggctgg | agatcctgac | gaactacact | ctgctgcca | gggagcaggc |
| 61 | gtccatctgg | ctggttgcag | gcaggggtca | gaatcaccaa | ggtcgactgg | cagcattcga |
| 121 | aaaatggtgc | tgcccaccac | acccaagatt | acccttgctc | tgaactggtg | gttcgaagag |
| 181 | gccagttatt | cagtctcacc | ctggatctga | gtagagtcc | ggacagcgag | gaggccctca |
| 241 | tcttcaccgt | ggagacagga | cctcacgcat | ccgaggctct | ccacaccaa | gccgtattcc |
| 301 | agacagcaga | gctagaaata | ggggacactt | ggacagctgc | aaaagaggaa | cagacagaga |
| 361 | acacaatcac | cgtcagcctc | ttgagccctt | ccaatgcagt | cattggccga | tacctgctga |
| 421 | gtgccaggcc | ctcttcccgc | aggaaacaca | gtgaccgaaa | gctgggccag | tttattctcc |
| 481 | tgttcaatcc | ctgggtgcca | gaggacgatg | tgtttctgga | ctcagaggaa | gagaggcagg |
| 541 | agtacgtgtc | taacgacagt | ggcgtcatct | tccgaggcgt | ggagaagcac | atccgagccc |
| 601 | agggctgaaa | ctacgggcag | tttgaagagg | acatcctgaa | catctgcctc | tccattctgg |
| 661 | accgaagtcc | tagtcaccag | gaggaccag | ccaccgatgt | gtcccaccgc | catgaccag |
| 721 | tctatgtcac | caggatcatc | agtgccatgg | tgaacagcaa | caatgaccga | ggtgtggtac |
| 781 | aaggccagtg | gcagggcaaa | tatggaggtg | gtaccaaccc | actgaactgg | cgtggcagtg |
| 841 | tggccatcct | acagaagtgg | ttcaagggca | ggtacaagcc | agtcaaatat | ggccagtgtc |
| 901 | gggtcttcgc | tggagtcatg | tgcacagtcc | tcaggtgcct | gggatcgca | accaggggtg |
| 961 | tgtccaactt | caactcagcc | cacgacaccg | atggcaatct | gagtgtggac | aaatatgtgg |
| 1021 | actcttatgg | ccgaaccctg | gaggacctga | cagaagacag | catgtggaac | ttccacgtct |
| 1081 | ggaatgagag | ctggtttgca | cgacaggacc | tgggcccatc | ctatgatggc | tggcaggttc |
| 1141 | tggatgccac | tcccgaagag | gagagcgaag | gtatgttccg | gtgtggtcct | gcctcggtea |
| 1201 | ctgccatccg | tgagggtgac | gtgcatctag | cccatgatgg | cccctcgtc | tttgcggagg |
| 1261 | tcaatgcaga | ctacatcacc | tggctatggc | acgaggacaa | gaggcgggag | cgtgtgtact |
| 1321 | cggacacca | gaagatcggc | aggtgatca | gtaccaaggc | agtgggcagt | gactctctgt |
| 1381 | tggacattac | agacctctac | aagtatccag | aagggtcccg | gaaggagagg | caggtgtaca |
| 1441 | gcaaagctgt | gaagaaactg | ctaagtgtgg | aagcctgggg | gaggaggagg | cggatccgca |
| 1501 | gggccagcgt | gcgcggtgtc | tggcgcgacg | acctcctgga | accggtcacc | aaaccagca |
| 1561 | tactggcaa | gttcaaagtg | ctggagccac | cgtgtctcgg | ccaggacctg | aagctggctc |
| 1621 | tgtgcttgac | caacctcacc | gccagggccc | agcgggtccg | agtcaatgtg | agcggcgcca |
| 1681 | ccatctctca | taccgcaag | cctgtggcag | agatcctcgc | tgaatcccac | accgtgaagc |
| 1741 | tggggccact | agaagaaaag | aagatcccag | tcacaatatc | ttactctcag | tacaaaggag |
| 1801 | acctgacaga | agacaagaag | atcctgttgg | ctgccatgtg | ccttgtcagt | aagggcgaga |
| 1861 | agctcctggt | ggagaaggac | attaccctgg | aggacttcat | caccatcaag | gtgctggggc |
| 1921 | cagcagtggt | aggagtgaca | gtcactgtgg | aagtcttggg | gatcaacccc | ctctcggaga |
| 1981 | gcgtaaagga | ctgtgtgctg | atggtagagg | gcagtggcct | tctccaagga | cagctaagta |
| 2041 | ttgagtgccc | cagcttgacg | cctcaggaga | aggccttgat | tcaattcaac | ataaccctt |
| 2101 | ccaagagtgg | ccccaggcag | ctgcaggtgg | acctgtcag | ctcacagttc | ccagacatca |
| 2161 | agggcttgg | gatcatccac | gtggctacag | ccaagtgatg | ggccacgagg | aactggggag |
| 2221 | gactttgtcc | atctgtctgt | ttctcccata | taagagacag | gggtctcagg | ttgtctctgc |
| 2281 | ttctgtcttt | ggcctccatg | agcttcggca | aggctcttcc | cagtgcattg | atggaggtgc |
| 2341 | tggactcagt | aacccttcac | cctcaggaca | agtggtgat | gtaagagcag | agggaccctc |
| 2401 | caactcttgt | cacccttctg | ctgggagagt | ggcatgtccc | agatttctga | gagccctggc |
| 2461 | ccaagagctc | agccagctca | gcccaggtaa | ctcaccactt | cttgcttttc | tatcatctct |
| 2521 | tcaaagtgtcc | tgaccacatt | tggacccaag | aattctccaa | gccggagcat | cccattata |

2581 gaggagcctt tctgaagcca taagacaaag cagaaacaag gttgagaacc aagtggcct
2641 ccaactcatg ttcttaaaag agcactggat ttaagcaac aggagattct tccctatgtg
2701 gaaacactga ctctggattt gggatgatga gctaaaactg cctgagaaca gctgaagaag
2761 acaaaggagt tctcccccac actgggcagc atggatatac tccagctctg gccatcctgg
2821 agttcacagg actgccccag atcactcagc cactccttct tggaaactaga atggcttgca
2881 cgggtgccgat cctacacttg gagtcaaagc aggacatgga tgcttggaat caggaccaga
2941 gtagacacaa caaggccatg aagaatatgc cactagatat gaaattcagc actgaaattt
3001 ctagacagac attaaaatta agctacagtt taagaacatt aaacaaaca caggtgcatc
3061 tgcaactcag tgctcttg atacaaatga cggtgagggga attgcagaga tggctcagct
3121 gttaagagta ctttctgac ttctagaggg gttgagttg gttcccagca ctggtgacg
3181 acagctcaca tactataact ccaactccag ggctccaaca tttctggtct ctgtgggcac
3241 ccacacttac ctgctgcatg tacaccctcc tcccaaacac atgcaattaa aataaaataa
3301 atcctttttt aaaatac

Appendix 12

1: [AY177607](#). *Mus musculus* tran...[gi:37729508] [Links](#)

LOCUS AY177607 3308 bp mRNA linear ROD 27-JAN-2004

DEFINITION *Mus musculus* transglutaminase y splice variant (Tgm6) mRNA,
complete sequence.

ACCESSION AY177607

VERSION AY177607.1 GI:37729508

KEYWORDS .

SOURCE *Mus musculus* (house mouse)

ORGANISM [Mus musculus](#)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 3308)

AUTHORS Thomas,H., Aeschlimann,P. and Aeschlimann,D.

TITLE A novel transglutaminase expressed in the nervous system

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 3308)

AUTHORS Aeschlimann,D., Thomas,H. and Aeschlimann,P.

TITLE Direct Submission

JOURNAL Submitted (11-NOV-2002) Matrix Biology & Tissue Repair Research

Unit, Dental School, University of Wales College of Medicine, Heath

Park, Cardiff CF14 4XY, United Kingdom

FEATURES Location/Qualifiers

source 1..3308

/organism="Mus musculus"

/mol_type="mRNA"

/strain="BALB/c"

/db_xref="taxon:10090"

/chromosome="2"

/map="74-78cM"

/tissue_type="brain"

gene 1..3308

/gene="Tgm6"

/note="transglutaminase y splice variant; alternatively

spliced exon 1; full length sequence found in GenBank

Accession Number AY159126; coding region not determined"

polyA_signal 3282..3287

/gene="Tgm6"

ORIGIN

```
1 cactctaacc ttcacttaca gaatgatcta ttagattcca gaggattttg aagaggatcc
61 tttgggtgaat tctggggggtc agaatcacca aggtcgactg gcagcattcg aaaaatgggtg
121 ctgcccacca caccacaagat tacccttgct ctgaactggt ggttcgaaga ggccagttat
181 tcagtctcac cctggatctg agtagagtcc tggacagcga ggaggccctc atcttcaccg
241 tggagacagg acctcacgca tccgaggctc tccacaccaa agccgtattc cagacagcag
301 agctagaagt aggggacact tggacagctg caaaagagga acagacagag aacacaatca
361 ccgtcagcct cttgagccct tccaatgcag tcattggccg atacctgctg agtgccaggc
421 cctcttcccg caggaaacac agtgaccgaa agctgggcca gtttattctc ctgttcaatc
481 cctggtgccc agaggacgat gtgtttctgg actcagagga agagaggcag gactacgtgc
541 ttaacgacag tggcgtcatc ttccgaggcg tggagaagca catccgagcc cagggctgga
601 actacgggca gtttgaagag gacatcctga acatctgcct ctccattctg gaccgaagtc
661 ctagtcacca ggaggaccca gccaccgatg tgtcccaccg ccatgacca gtctatgtca
721 ccaggatcat cagtgccatg gtgaacagca acaatgaccg aggtgtggta caaggccagt
781 ggcagggcaa atatggagggt ggtaccaacc cactgaactg gcgtggcagt gtggccatcc
841 tacagaagtg gttcaagggc aggtacaagc cagtcaaata tggccagtgc tgggtcttcg
901 ctggagtcac gtgcacagtc ctcagggtgc tggggatcgc aaccagggtg gtgtccaact
961 tcaactcagc ccacgacacc gatggcaatc tgagtgtgga caaatatgtg gactcttatg
```

1021 gccgaaccct ggaggacct acagaagaca gcatgtgaa cttccacgtc tggaaatgaga
1081 gctgggttgc acgacaggac ctgggcccat cctatgatgg ctggcaggtt ctggatgcca
1141 ctccccaaga ggagagcgaa ggtatgttcc ggtgtggtcc tgcctcggtc actgccatcc
1201 gtgaggggtga cgtgcatcta gcccatgatg gcccttcgt ctttgcgag gtcaatgcag
1261 actacatcac ctggctatgg cacgaggaca agaggcgga gcgtgtgtac tcggacacca
1321 agaagatcgg caggtgcatc agtaccaagg cagtgggcag tgactctcgt gtggacatta
1381 caggcctcta caagtatcca gaaggggtccc ggaaggagag gcaggtgtac agcaaagctg
1441 tgaagaaact gctaagtgtg gaagcctggg ggaggaggag gcggatccgc agggccagcg
1501 tgcgcggtgt ctggcgcgac gacctcctgg aaccggtcac caaacccagc atcactggca
1561 agttcaaagt gctggagcca ccgggtctgc gccaggacct gaagctggct ctgtgcttga
1621 ccaacctcac cgccagggcc cagcgggtcc gagtcaatgt gagcggcgcc accatcctct
1681 ataccgcaa gcctgtggca gagatcctgc gtgaatccca caccgtgaag ctggggccac
1741 tagaagaaaa gaagatccca gtcacaatat cttactctca gtacaaagga gacctgacag
1801 aagacaagaa gatcctgttg gctgccatgt gccctgtcag taagggcgag aagctcctgg
1861 tggagaagga cattaccctg gaggacttca tcaccatcaa ggtgctgggg ccagcagtg
1921 taggagtgc agtcaactgt gaagtcttgg tgatcaacc cctctcggag agcgtaaagg
1981 actgtgtgct gatggtagag ggcagtggcc ttctccaagg acagctaagt attgaggtgc
2041 ccagcttgca gcctcaggag aaggccttga ttcaattcaa cataaccct tccaagagt
2101 gccccaggca gctgcaggtg gacctgttca gctcacagtt cccagacatc aagggtcttg
2161 tgatcatcca cgtggctaca gccagtgat gggccacgag gaactgggga ggactttgtc
2221 catctgtctg tttctcccat ataagagaca ggggtctcag gttggctcctg cttctgtctt
2281 tggcctccat gagcttgggc aaggtccttc ccagtgcatt gatggaggtg ctggactcag
2341 taacccttca ccctcaggac aagtggctga tgtaagagca gagggaccct ccaactcttg
2401 tcacccttct gctgggagag tggcatgtcc cagatttctg agagccctgg cccaagagct
2461 cagccagctc agcccaggta actcaccact tcttgcttt ctatcatctc ttcaaagtgc
2521 ctgaccacat ttggacccaa gaattctcca agccggagca tccccattat agaggagcct
2581 ttctgaagcc ataagacaaa gcagaaacaa ggttgagaac caagtgttcc tccaactcat
2641 gttcttaaaa gagcaactgga ttttaagcaa caggagattc ttccctatgt ggaaacactg
2701 actctggatt tgggtgatgg agctaaaact gcctgagaac agctgaagaa gacaaaggag
2761 ttctcccca aactgggcag catggatata ctccagctct ggccatcctg gacttcacag
2821 gactgcccc gatcactcag ccactccttc ttggaactag aatggcttgc acgggtccga
2881 tctacactt ggagtcaaag caggacatgg atgcctggaa tcaggaccag agtagacaca
2941 acaaggccat gaagaatag ccactagata tgaattcag cactgaaatt tctagacaga
3001 cattaataat aagctacagt ttaagaacat taaacaaaca acaggtgcat ctgcaactca
3061 gtgcctcttg gatacaaatg acggtgaggg aattgcagag atggctcagc tgtaagagt
3121 actttctgat cttctagagg ggttgattt ggttcccagc actggtgtca gacagctcac
3181 atactataac tccaactcca gggctccaac atttctggtc tctgtgggca cccacactta
3241 cctgctgcat gtacaccctc ctcccaaaca catgcaatta aaataaata aatcctttt
3301 taaaatac

Appendix 13

1: AY177606. *Mus musculus* tran...[gi:37729471] Links

LOCUS AY177606 2955 bp mRNA linear ROD 27-JAN-2004
DEFINITION *Mus musculus* transglutaminase y short splice variant (Tgm6) mRNA,
complete cds.

ACCESSION AY177606

VERSION AY177606.1 GI:37729471

KEYWORDS .

SOURCE *Mus musculus* (house mouse)

ORGANISM *Mus musculus*

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; *Mus*.

REFERENCE 1 (bases 1 to 2955)

AUTHORS Thomas,H., Aeschlimann,P. and Aeschlimann,D.

TITLE A novel transglutaminase expressed in the nervous system

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 2955)

AUTHORS Aeschlimann,D., Thomas,H. and Aeschlimann,P.

TITLE Direct Submission

JOURNAL Submitted (10-NOV-2002) Matrix Biology & Tissue Repair Research

Unit, Dental School, University of Wales College of Medicine, Heath

Park, Cardiff CF14 4XY, United Kingdom

FEATURES Location/Qualifiers

source 1..2955

/organism="Mus musculus"

/mol_type="mRNA"

/strain="BALB/c"

/db_xref="taxon:10090"

/chromosome="2"

/map="74-78cM"

/tissue_type="brain"

gene 1..2955

/gene="Tgm6"

CDS 78..263

/gene="Tgm6"

/note="alternatively spliced product missing exons 3 and 4

which results in premature termination; full length

sequence found in GenBank Accession Number AY159126"

/codon_start=1

/product="transglutaminase y short splice variant"

/protein_id="AAO26186.1"

/db_xref="GI:37729472"

/translation="MAGVRITKVDWQHSKNGAAHHTQDYPCSELVVRGQLFSLTLDL

SRVLDSEEALIFTVETV"

polyA_signal 2929..2934

/gene="Tgm6"

ORIGIN

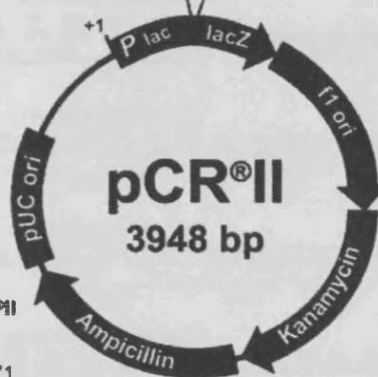
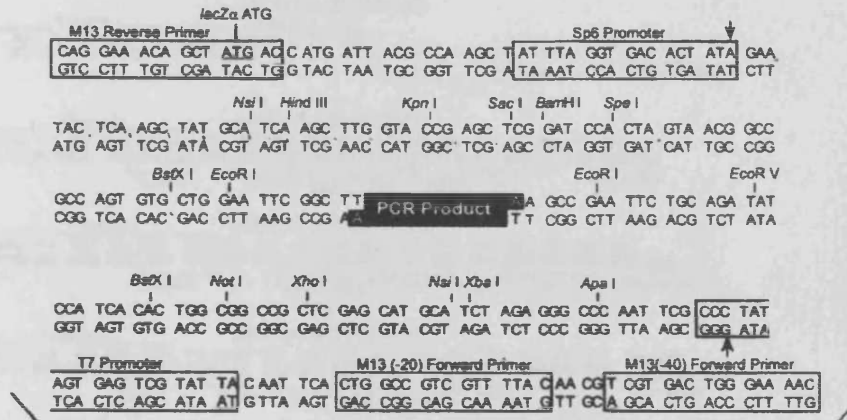
```
1  attaatgtgg atctggctgg agatcctgac gaactacact ctgctgcccc gggagcaggc
61  gtccatctgg ctgttgcatg gcaggggtca gaatcaccaa ggtcgactgg cagcattcga
121  aaaatggtgc tgcccaccac acccaagatt acccttgctc tgaactggtg gttcgaagag
181  gccagttatt cagtctcacc ctggatctga gtagagtcct ggacagcgag gaggccctca
241  tcttcaccgt ggagacagtt tgaagaggac atcctgaaca tctgcctctc cattctggac
301  cgaagtccca gtcaccagga ggaccagcc accgatgtgt cccaccgcca tgaccagtc
```

361 tatgtcacca ggatcatcag tgccatggtg aacagcaaca atgaccgagg tgtggtacaa
421 ggccagtggc agggcaaata tggaggtggt accaaccac tgaactggcg tggcagtgtg
481 gccatcctac agaagtgggt caagggcagg tacaagccag tcaaatatgg ccagtgtgtg
541 gtcttcgctg gagtcatgtg cacagtcctc aggtgcctgg ggatcgcaac caggtgtgtg
601 tccaacttca actcagccca cgacaccgat ggcaatctga gtgtggacaa atatgtggac
661 tcttatggcc gaaccctgga ggacctgaca gaagacagca tgtggaactt ccacgtctgg
721 aatgagagct ggtttgcacg acaggacctg ggccatcct atgatggctg gcaggttctg
781 gatgccactc cccaagagga gagcgaaggt atgttccggt gtggctctgc ctcggtcact
841 gccatccgtg aggggtgacgt gcatctagcc catgatggcc ccttcgtctt tgcggaggtc
901 aatgcagact acatcacctg gctatggcac gaggacaaga ggccgggagcg tgtgtactcg
961 gacaccaaga agatcggcag gtgcatcagt accaaggcag tgggcagtga ctctcgtgtg
1021 gacattacag gcctctacaa gtatccagaa ggggccgga aggagaggca ggtgtacagc
1081 aaagctgtga agaaactgct aagtgtggaa gcctggggga ggaggaggcg gatccgcagg
1141 gccagcgtgc gcggtgtctg gcgcgacgac ctccctggaac cggtcaccaa acccagcatc
1201 actggcaagt tcaaagtgct ggagccaccg gtgctcggcc aggacctgaa gctggctctg
1261 tgcttgacca acctcaccgc cagggcccag cgggtccgag tcaatgtgag cggcgccacc
1321 atcctctata cccgcaagcc tgtggcagag atcctgcgtg aatccacac cgtgaagctg
1381 gggccactag aagaaaagaa gatcccagtc acaatatctt actctcagta caaaggagac
1441 ctgacagaag acaagaagat cctgttggct gccatgtgcc ttgtcagtaa gggcgagaag
1501 ctccctgggg agaaggacat taccctggag gacttcatca ccatcaaggt gctggggcca
1561 gcagtggtag gagtgacagt cactgtggaa gtcttggtag tcaacccct ctcgagagc
1621 gtaaaggagt gtgtgctgat ggtagagggc agtggccttc tccaaggaca gctaagtatt
1681 gaggtgccca gcttgcagcc tcaggagaag gccttgattc aattcaacat aacccttcc
1741 aagagtggcc ccaggcagct gcaggtggac cttgtcagct cacagttccc agacatcaag
1801 ggctttgtga tcatccacgt ggctacagcc aagtgatggg ccacgaggaa ctggggagga
1861 ctttgtccat ctgtctgttt ctcccatata agagacaggg gtctcaggtt ggtcctgctt
1921 ctgtctttgg cctccatgag ctctcggaag gtccctccca gtgcattgat ggagggtgctg
1981 gactcagtaa cccttccacc tcaggacaag tggctgatgt aagagcagag ggaccctcca
2041 actcttgtca cccttctgct gggagagtgg catgtcccag atttctgaga gccctggccc
2101 aagagctcag ccagctcagc ccaggttaact caccacttct tgcccttcta tcatctctt
2161 aaatgtcctg accacatttg gacccaagaa ttctccaagc cggagcatcc ccattataga
2221 ggagcctttc tgaagccata agacaaagca gaaacaaggt tgagaacca gtggtcctcc
2281 aactcatggt cttaaaagag cactggattt taagcaacag gagattcttc cctatgtgga
2341 aacactgact ctggatttgg gtgatggagc taaaactgcc tgagaacagc tgaagaagac
2401 aaaggagttc tccccaaac tgggcagcat ggatatactc cagctctggc catcctggag
2461 ttcacaggac tgccccagat cactcagcca ctccctcttg gaactagaat ggcttgacg
2521 gtgccgatcc tacacttggg gtcaaagcag gacatggatg cctggaatca ggaccagagt
2581 agacacaaca aggccatgaa gaatatgcca ctagatatga aattcagcac tgaatttct
2641 agacagacat taaaattaag ctacagttta agaacattaa acaacaaca ggtgcatctg
2701 caactcagtg cctcttggat acaaatgacg gtgagggaaat tgcagagatg gctcagctgt
2761 taagagtact ttctgatctt cttagaggggt tgagtttgggt tcccagcact ggtgtcagac
2821 agctcacata ctataactcc aactccaggg ctccaacatt tctggtctct gtgggcccc
2881 acacttacct gctgcagta caccctcctc ccaaacacat gcaattaaaa taaaataaat
2941 ccttttttaa aatac

Appendix 14

The pCR®II Vector

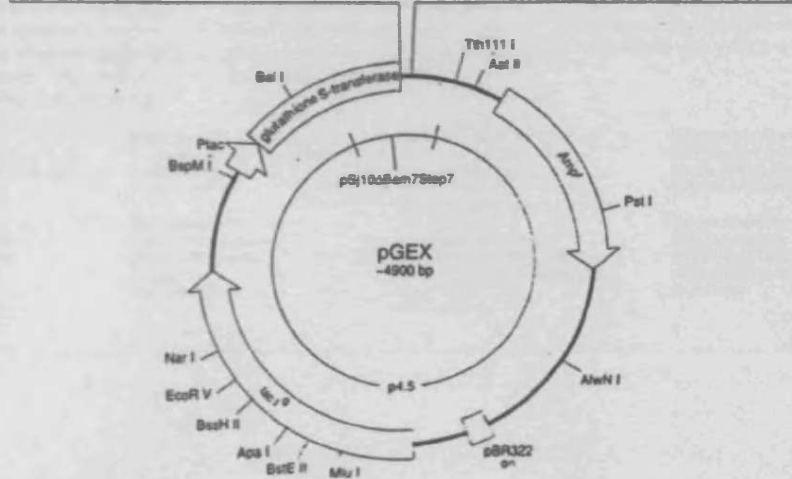
Map of pCR®II The map of the linearized vector, pCR®II, is shown below. The sequence of the multiple cloning site is shown with a PCR product inserted by TA Cloning®. The arrows indicate the start of transcription for the T7 and Sp6 RNA polymerases.



Comments for pCR®II
3948 nucleotides

- LacZ gene: bases 1-571
- Sp6 promoter: bases 239-255
- Multiple Cloning Site: bases 269-381
- TA Cloning® site: bases 335-336
- T7 promoter: bases 388-407
- F1 origin: bases 572-986
- Kanamycin resistance: bases 987-2114
- Ampicillin resistance: bases 2133-2992
- pUC origin (pMB1-derived): bases 3182-3765

Appendix 15

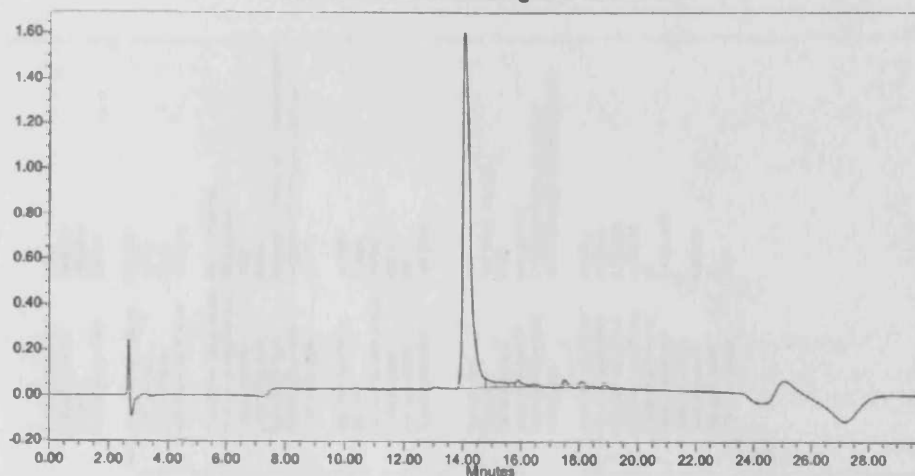


Sample Information

SampleName 72073-2b6601-4
 Vial 68
 Injection 1
 System Name SCOOPYDOO2
 Run Time 30.00 Minutes

Date Acquired 5/30/03 11:19:19 PM
 Acq Method Set STANDARD
 Processing Method Auto Integration
 Date Processed 6/2/03 5:28:42 AM
 Injection Volume 20.00 ul

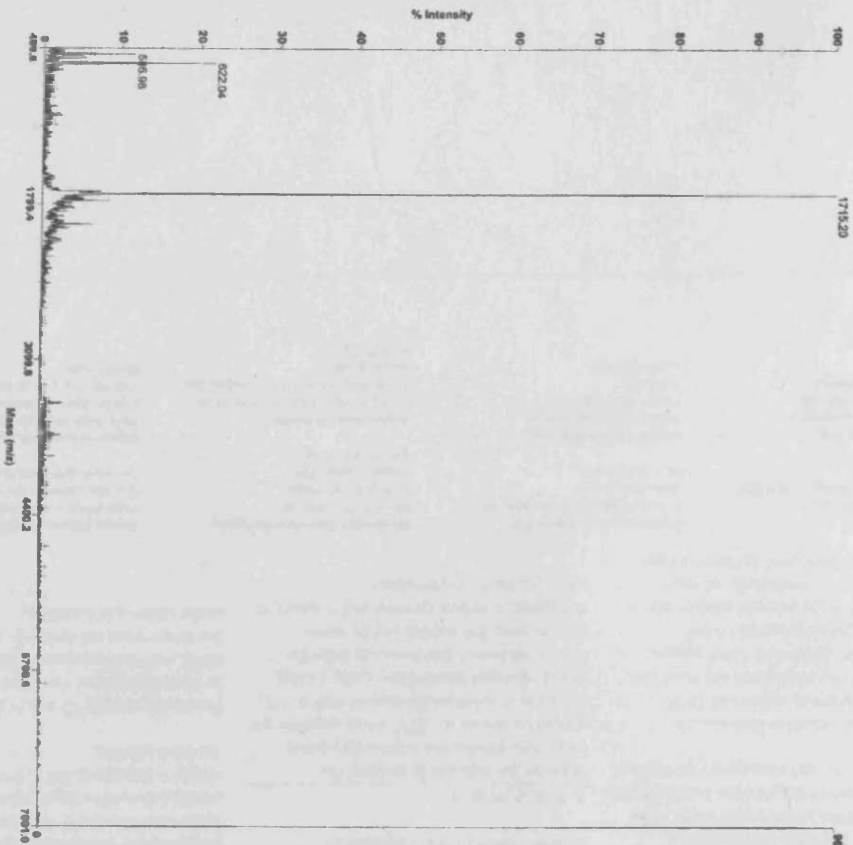
Auto-Scaled Chromatogram at 214 nm



Integration Results

| | RT | Area | Height | % Area | % Height |
|----|--------|----------|---------|--------|----------|
| 1 | 14.065 | 27839236 | 1571599 | 90.18 | 88.35 |
| 2 | 14.797 | 570498 | 38277 | 1.86 | 2.15 |
| 3 | 15.200 | 533788 | 23159 | 1.74 | 1.30 |
| 4 | 15.686 | 242236 | 18655 | 0.79 | 1.05 |
| 5 | 15.903 | 462790 | 30071 | 1.51 | 1.69 |
| 6 | 16.487 | 153976 | 9930 | 0.50 | 0.56 |
| 7 | 17.485 | 349378 | 32259 | 1.14 | 1.81 |
| 8 | 18.076 | 363227 | 23740 | 1.19 | 1.33 |
| 9 | 18.424 | 101483 | 7634 | 0.33 | 0.43 |
| 10 | 18.682 | 23074 | 4065 | 0.08 | 0.23 |

Voyager Spec #10-SM5191 = 1714.7_9631



Acquired: 17:40:00, May 30, 2003
 Sample Description: 72075-259801-4
 D:\Voyager\DATA\epub\sm5191\500\May30_0021.dat

| | |
|----------------------------|---------------------------------------|
| Mode of operation: | Linear |
| Extraction mode: | Delayed |
| Polarity: | Negative |
| Acquisition control: | Manual |
| M413 | |
| Accelerating voltage: | 20000 V |
| Grid voltage: | 94.5% |
| Guide wire 0: | 0.04% |
| Extraction delay time: | 150 msec |
| Acquisition mass range: | 500 - 7000 Da |
| Number of laser shots: | 80/spectrum |
| Laser intensity: | 1404 |
| Laser Rep Rate: | 3.0 Hz |
| Calibration type: | External - D1\Voyager\Calibrat\WV-M03 |
| Calibration matrix: | a-Cyano-4-hydroxycinnamic acid |
| Low mass gate: | 400 Da |
| Digitizer start time: | 14.888 |
| Bin size: | 2 msec |
| Number of data points: | 19695 |
| Vertical scale: | 200 mV |
| Vertical offset: | 0% |
| Input bandwidth: | 150 MHz |
| Sample well: | |
| Plate ID: | 17 |
| Serial number: | 100 WELL PLATE |
| Instrument name: | Voyager-0E |
| Plate type filename: | C:\Voyager\100 well plate.glt |
| Lab name: | GEMOSYS BIOTECHNOLOGIES |
| Absolde x-position: | |
| Absolde y-position: | 32067.9 |
| Relative x-position: | 42226.3 |
| Relative y-position: | 0.390508 |
| Shots in spectrum: | -1.19115 |
| Source pressure: | 90 |
| Mirror pressure: | 6.427e-007 |
| TIC2 pressure: | 0 |
| TIS gate width: | 0.01776 |
| TIS gate width: | 30 |
| TIS flight length: | 940 |



QUALITY ASSURANCE
DOCUMENTATION

hT66g (sheet form)

Customer Number: 101721
Peptide #: 72073-1
Lot #: D6619-002
Sequence: TIRAYPGASGEGLSLSP
Length: 15 Mgs Shipped: 10.9
Molecular Weight: 1474.6500244141 Final Purity 79
Peptide Name PEP1
Notes: N-Term:[H] C-Term:[NH2]

M_v (w/o NH₂, w/ C) 1475.6
pI: 5.66
 $A_{280}^{1\%}$ = 0.862
GRAVY -0.273

Column: Discovery Bio Wide Pore C-18, 250mmX4.6mm, 5µm
Mobile Phase: A=0.1% TFA/Water
B=0.1% TFA/Acetonitrile
Gradient: 0 to 2 min.: 100% A
2 to 20 min.: 0 to 67.5% B
Detection: 214nm
Storage Conditions: -20 C for Long Term
4 C for Short Term

Sigma-Genosys makes no claims to the peptide's ability to function in the specific application of the customer. All peptides are shipped lyophilized and as gross weight. Peptide for research use only.

Analytical HPLC and Mass Spectral analyses are enclosed.

Handling & Storage of Peptides

Resuspension

1. Hydrophilic peptides: Resuspend the peptide in distilled or deionized water. Make sure that the pH of the solution is around pH 7.0, and adjust the pH if necessary to improve solubility.
2. Hydrophobic peptides: Dissolve the peptide in a minimum volume of DMF, DMSO (if the sequence does not have Cys, Met or Trp), IPA, ethanol, or other organic solvents that can be tolerated in the experiment. Add the peptide solution slowly (dropwise) into a stirred buffer of your choice. For non-biological studies, stronger

solvents like TFA, formic acid or acetic acid may be used for initial solubilization.

Storage

Peptides are supplied lyophilized. For maximum stability, the lyophilized peptide should be stored at -20°C. Most peptides will be stable in solutions for several days at 4°C. Avoid repeated freeze/thaw cycles if using frozen solutions. Upon removal from the freezer, bring the peptide vial to room temperature before opening and weighing to avoid moisture condensation.

Peptides containing W, M, and C are prone to oxidation and should be stored in a lyophilized state under an inert gas. These peptides should be degassed and stored in aliquots at -80°C if they will be stored in solution.

Peptides containing Q and N are susceptible to deamidation and should be stored at -80°C. Additionally, peptides containing a Q at the N-terminus will cyclize to pyroglutamate under dilute acid conditions.

Sigma-Genosys
1442 Lake Front Circle
The Woodlands, Texas 77380-3800
1-800-234-5362

Sigma-Genosys Japan K.K.
777-13 Nishi-1, Shin-Ko, Ishikari
Hokkaido, Japan
81 133 73 5005

Cashmere Scientific Company
No. 2-1, 2F, Lane 40
Chang-An W. Road
Taipei, Taiwan 104
886-2-25416188

Sigma-Aldrich Canada Ltd.
2149 Winston Park Drive
Oakville, Ontario L6H 6J8
CANADA (905) 829-9500

Sigma-Genosys Ltd.
London Road, Pampisford
Cambridge CB2 4EF, UK
(+44) (0) 1223 839000

Sigma-Aldrich Pty Ltd.
Unit 2, 14 Anella Avenue
Castle Hill NSW 2154
Australia
1-800-800-097

Sigma-Aldrich Korea
Samhan Camus Annex, 10th Floor
17-26 Yoido-dong, Yungdeungpo-ku
Seoul, Korea
82 2 783 5211

Sigma-Aldrich Pte., Ltd.
102E Pasir Panjang Road
#08-01 Callink Warehouse
118529, Republic of Singapore
65 271 1089

24

64

43

20

21

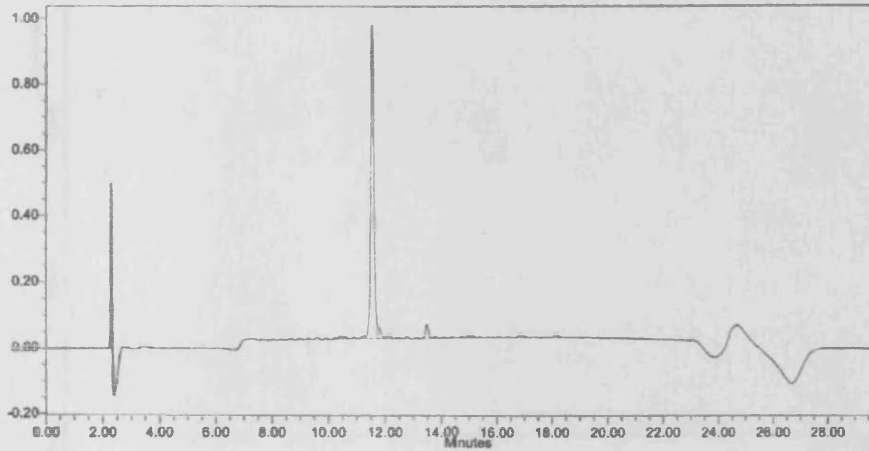
19

Sample Information

SampleName 72073-1D6619-2RR
 Vial 12
 Injection 1
 System Name SCOOPYDOO2
 Run Time 30.00 Minutes

Date Acquired 6/6/03 2:43:42 PM
 Acq Method Set STANDARD
 Processing Method Auto Integration
 Date Processed 6/6/03 3:24:21 PM
 Injection Volume 5.00 ul

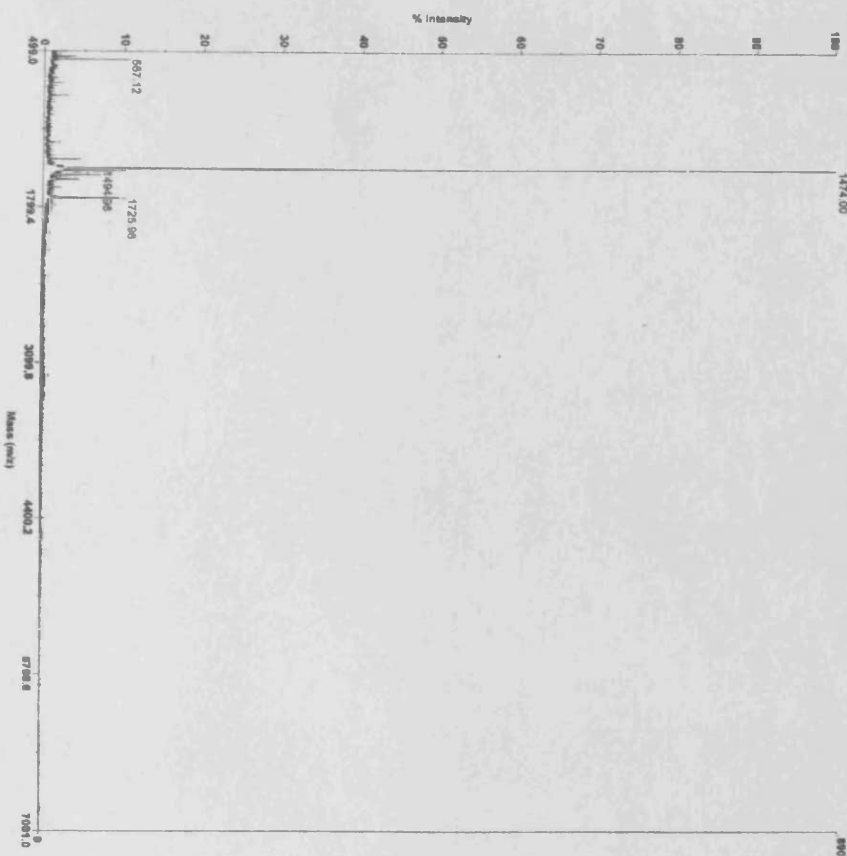
Auto-Scaled Chromatogram at 214 nm



Integration Results

| | RT | Area | Height | % Area | % Height | | RT | Area | Height | % Area | % Height |
|----|--------|-------|--------|--------|----------|----|--------|---------|--------|--------|----------|
| 1 | 8.051 | 20093 | 2404 | 0.22 | 0.21 | 11 | 10.881 | 30620 | 4055 | 0.33 | 0.35 |
| 2 | 8.360 | 24797 | 2801 | 0.27 | 0.24 | 12 | 11.237 | 67361 | 7974 | 0.73 | 0.69 |
| 3 | 8.641 | 35322 | 3873 | 0.39 | 0.33 | 13 | 11.537 | 7333793 | 950030 | 79.96 | 81.69 |
| 4 | 8.988 | 43684 | 4881 | 0.48 | 0.42 | 14 | 11.717 | 261031 | 37841 | 2.85 | 3.25 |
| 5 | 9.291 | 32021 | 4107 | 0.35 | 0.35 | 15 | 12.139 | 88462 | 10386 | 0.96 | 0.89 |
| 6 | 9.604 | 59887 | 6619 | 0.65 | 0.57 | 16 | 12.431 | 22200 | 2363 | 0.24 | 0.20 |
| 7 | 9.932 | 41588 | 4727 | 0.45 | 0.41 | 17 | 12.770 | 41532 | 5423 | 0.45 | 0.47 |
| 8 | 10.236 | 41297 | 4891 | 0.45 | 0.42 | 18 | 13.123 | 30858 | 3858 | 0.34 | 0.31 |
| 9 | 10.407 | 77498 | 9121 | 0.84 | 0.78 | 19 | 13.463 | 312301 | 40611 | 3.41 | 3.49 |
| 10 | 10.589 | 57285 | 6993 | 0.62 | 0.60 | 20 | 13.776 | 63772 | 8874 | 0.70 | 0.59 |

Voyager Spec #118 = 1474.1, 89051



Acquired: 18:53:00, June 05, 2003
 Sample Description: 72073-1DB616-2
 D:\Voyager\Data\Peptides\June05\JUNE03_0118.dak

Mode of operation: Linear
 Extraction mode: Delayed
 Polarity: Positive
 Acquisition control: Manual

Acceleration voltage: 20000 V
 G4 voltage: 94.5%
 Guide wire O: 0.04%
 Extraction delay time: 150 nsec

Acquisition mass range: 500 - 7000 Da
 Number of laser shots: 60/spectrum
 Laser intensity: 1596
 Laser Rep Rate: 3.0 Hz
 Calibration type: External - D:\Voyager\Calibrat\Appl
 Calibration method: s-Cyano-4-hydroxymannic acid
 Low mass gate: 400 Da

Dryer start time: 14.694
 Bin size: 2 nsec
 Number of data points: 16863
 Vertical scale: 200 mV
 Vertical offset: 0%
 Input bandwidth: 150 MHz

Sample well: 18
 Plate ID: 100 WELL PLATE
 Serial number: 1105
 Instrument name: Voyager-DE
 Plate type/Name: C:\Voyager\100 well plate.pl
 Lab name: GENOSYS BIOTECHNOLOGIES

Absolute x-position: 42228.7
 Absolute y-position: 42228.3
 Relative x-position: -0.787619
 Relative y-position: 0.787121
 Shots in spectrum: 60
 Source pressure: 6.018e-007
 Mirror pressure: 0
 TCD pressure: 0.01658
 TIS gate width: 30
 TIS flight length: 940

Printed: 18:53, June