

SUBSTRATE SPECIFICITY OF TRANSGLUTAMINASES FOR GLUTEN PEPTIDES

Doctoral thesis by
Siri Dørum



**Centre for Immune Regulation
Institute of Immunology
Institute of Clinical Medicine
University of Oslo
2010**



© Siri Dørum, 2010

*Series of dissertations submitted to the
Faculty of Medicine, University of Oslo
No. 1029*

ISBN 978-82-8072-548-6

All rights reserved. No part of this publication may be reproduced or transmitted, in any form or by any means, without permission.

Cover: Inger Sandved Anfinsen.
Printed in Norway: AiT e-dit AS.

Produced in co-operation with Unipub.
The thesis is produced by Unipub merely in connection with the thesis defence. Kindly direct all inquiries regarding the thesis to the copyright holder or the unit which grants the doctorate.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	3
ABBREVIATIONS	4
LIST OF PUBLICATIONS	5
INTRODUCTION	6
TRANSGLUTAMINASES.....	7
<i>The enzyme family of transglutaminases</i>	7
<i>Transglutaminase 2</i>	8
Structure and conformation.....	8
Enzymatic activity.....	8
Substrate specificity.....	8
Expression and localization.....	9
An enzyme with multiple functions.....	9
TG2 in human diseases.....	11
CELIAC DISEASE.....	12
<i>Clinical aspects</i>	12
<i>Genetic factors</i>	12
<i>Gluten – the trigger of celiac disease</i>	12
<i>Peptide binding to HLA-DQ2/DQ8</i>	13
<i>The gluten specific immune response</i>	13
THE ROLE OF TRANSGLUTAMINASE 2 IN CELIAC DISEASE.....	15
<i>TG2-mediated deamidation and transamidation of gluten peptides</i>	15
<i>Autoantibodies to TG2</i>	16
THE ROLE OF OTHER TRANSGLUTAMINASES IN GLUTEN SENSITIVE DISEASES.....	17
MASS SPECTROMETRY AND PROTEOMICS IN CELIAC DISEASE RESEARCH.....	18
<i>Specificity and enzymatic activity of TG2</i>	18
<i>Identification of T-cell epitopes in gluten</i>	18
<i>Mass spectrometry to study gluten proteolysis</i>	19
AIMS OF THE STUDY	20
SUMMARY OF PAPERS	21
METHODOLOGICAL CONSIDERATIONS	22
GENERAL DISCUSSION	25
SELECTION OF GLUTEN T-CELL EPITOPES IN CELIAC DISEASE.....	25
<i>Proteolytic stability of epitopes</i>	25
<i>Selection of epitopes by TG2</i>	25
<i>TG2-mediated deamidation in vivo</i>	26
<i>Peptide binding to HLA-DQ2.5 and HLA-DQ8</i>	27
GLUTEN SENSITIVE DISEASES AND REDUNDANCY OF TRANSGLUTAMINASES.....	28
FUTURE THERAPEUTIC STRATEGIES OF CELIAC DISEASE.....	29
<i>TG2-based therapeutic strategies</i>	29
<i>Alternative therapeutic approaches</i>	30
Degradation or removal of T-cell epitopes.....	30
Blocking of peptide presentation by HLA-DQ molecules.....	31
Peptide-based therapy.....	31
MASS SPECTROMETRY-BASED STRATEGIES AND APPLICATIONS IN FUTURE THERAPY.....	32
FINAL COMMENTS	33
REFERENCES	34

ACKNOWLEDGEMENTS

The work presented in this thesis has been carried out at the Institute of Immunology (IMMI) at Oslo University Hospital-Rikshospitalet during the period of 2006-2010. I would like to thank IMMI for providing great working facilities and a friendly research environment. My work has been funded by the Research Council of Norway (*FUGE, Functional Genomics*), for which I am most grateful.

First of all I wish to express my gratitude to my supervisor Burkhard Fleckenstein who introduced me to the fascinating world of mass spectrometry and proteomics. With his impressive in-depth knowledge and numerous creative ideas he has provided invaluable guidance. Thank you for believing in me and encouraging me.

I am also very thankful to my co-supervisor Ludvig Sollid who is an exceptional scientist and who, repeatedly, helped me see the bigger picture. Thank you for your incredibly quick feedback which always got the ball rolling.

This work has been dependent on a number of close collaborators. In particular I would like to thank Anders Holm for patiently training me in mass spectrometry and Jorunn Stamnæs a.k.a. "TG2-guru" for sharing her knowledge about the enigmatic transglutaminases. A special thank goes to Astrid Tuttunen, Maria Stensland, Marit Jørgensen and Tahira Riaz, my fellow companions at the Proteomics group through these years. Thank you for sharing the many thrills and frustrations both in science and life. The days at the lab would not have been the same without you.

I would further like to acknowledge the rest of the past and present "gutfeelers" which I have had the privilege to work with: Elin Bergseng, Lars-Egil Fallang, Shuo-Wang Qiao, Michael Bodd, Roberto Di Niro, Silja S. Amundsen, Ulrike Jüse, Rasmus Iversen, Luka Mesin, Marianne Sponheim, Patricia Stadtmüller, Sylvie Pollmann, Bjørg Simonsen, Marie K. Johannesen, Melinda Ráki, Knut E.A. Lundin, Øyvind Molberg, Eirik H. Halvorsen, Margit Brottveit, Ann-Kristin R. Beitnes, Axel Berg-Larsen, Ingrid Olsen, Stig Tollefsen and Anders Fallang. Thank you for great scientific input, for being excellent travelling partners to conferences, for the social fun and for giving me a tough time at the "ball games".

I would also like to thank my skilled co-authors for their contributions to my work; Magnus Ø. Arntzen, Christian J. Köhler and Bernd Thiede at the Biotechnology Centre of Oslo and Jorunn Stamnæs, Shuo-Wang Qiao and Anders Holm from IMMI.

Last, but not least, I would like to thank my family and my friends who mean a lot to me and who, each in their way, supported me and believed in me. A special thanks to my dear Jørgen who came into my life at the intense end of my PhD, but still stayed.

Oslo, June 2010

Siri Dørum

ABBREVIATIONS

5-BP	5-biotinamido-pentylamine
DH	dermatitis herpetiformis
ECM	extracellular matrix
GA	gluten ataxia
HMW	high molecular weight
HLA	human leukocyte antigen
LC-MS	liquid chromatography-mass spectrometry
LC-MS/MS	liquid chromatography/tandem-mass spectrometry
LMW	low molecular weight
MALDI-TOF	matrix-assisted laser desorption/ionization time-of-flight
MRM	multiple reaction monitoring
MS	mass spectrometry
PEP	prolyl endopeptidase
PTCEC	pepsin, trypsin, chymotrypsin, elastase, carboxypeptidase
Q-TOF	quadrupole- time-of-flight
TG	transglutaminase
TG2	transglutaminase 2

LIST OF PUBLICATIONS

Paper I

A quantitative analysis of transglutaminase 2-mediated deamidation of gluten peptides: implications for the T-cell response in celiac disease

S. Dørum, S.W. Qiao, L.M. Sollid, B. Fleckenstein

J. Proteome Res., 2009, 8 (4), 1748–1755

Paper II

Gluten T-cell epitope targeting by TG3 and TG6; implications for dermatitis herpetiformis and gluten ataxia

J. Starnæs, S. Dørum, B. Fleckenstein, D. Aeschlimann, L.M. Sollid

Amino Acids, 2010, Mar 19. [Epub ahead of print], PMID: 20300788

Paper III

The preferred substrates for transglutaminase 2 in a complex wheat gluten digest are peptide fragments harboring celiac disease T-cell epitopes

S. Dørum, M.Ø. Arntzen, A. Holm, S.W. Qiao, C.J. Koehler, B. Thiede, L.M. Sollid, B. Fleckenstein

Manuscript submitted to PLoS ONE

INTRODUCTION

The human immune system has developed intricate mechanisms to protect the body by discriminating between infectious agents and self. Unfortunately, in some cases these mechanisms can be bypassed and immune responses may be elicited by antigens derived from self. The loss of tolerance to autoantigens may lead to the development of autoimmune diseases.

Autoimmune diseases are chronic inflammatory diseases of unknown etiology where both genetic and environmental factors play a role. Human leukocyte antigens (HLA) have been shown to be the most important susceptibility factor for several of the autoimmune diseases, what strongly suggests the involvement of T cells. In many autoimmune diseases the self antigen causing the disease is not known. This is the case for rheumatoid arthritis, multiple sclerosis and type 1 diabetes. For celiac disease however, the trigger of the disease is known to be gluten. Interestingly, although gluten was known to be the environmental factor for years, it was not until a decade ago that a posttranslational modification of gluten mediated by an enzyme called tissue transglutaminase 2 (TG2) was discovered to be critical for the disease. The posttranslationally modified gluten peptides bind the disease associated HLA-molecules with a higher affinity than the unmodified gluten peptides, what results in a multifaceted T-cell response. Thus, posttranslational modifications of self-antigens, or in this case; of food antigens normally tolerated by the body, is one way in which novel epitopes are created that are not tolerated by the immune system. Interestingly, immune responses directed towards enzymatically modified self-antigens were also reported for other autoimmune diseases, e.g. against citrullinated proteins in rheumatoid arthritis and methylated and phosphorylated proteins in systemic lupus erythematosus. As the importance of posttranslational modifications of gluten peptides is acknowledged in celiac disease, further research in this field may be relevant also for other autoimmune diseases.

In this thesis, we have used mass spectrometry-based strategies to investigate the posttranslational gluten modifications catalyzed by the transglutaminase enzymes implicated in gluten sensitive diseases. The main focus has been on TG2 and its important role in T-cell epitope selection in celiac disease.

Transglutaminases

The enzyme family of transglutaminases

Transglutaminases (TGs: EC 2.3.2.13) are a family of structurally and functionally related enzymes expressed in almost all mammalian cells and tissues. The major function of the TGs is to covalently crosslink proteins through an acyl-transfer reaction between a peptidylglutamine and a polypeptide-bound lysine. These crosslinks are highly stable and are resistant to mechanical and proteolytic breakdown. Numerous biological processes are dependent on the crosslinking activity of TGs, ranging from wound healing and apoptosis to extracellular matrix assembly.

Nine TG genes have been identified in human, where eight of them encode active enzymes (TG1-TG7 and plasma factor XIII) [1]. The TG genes are closely related and five of the TGs are found within two gene clusters; TG5, TG7 and Band 4.2 cluster on chromosome 15q15, while TG3 and TG6 cluster at 20p13. Notably, TG2 is also localized at chromosome 20, but at 20q11.23. There is a high degree of similarity between the TG enzymes, both in their primary sequence, protein structure and regarding their catalytic mechanism. It has been suggested that the TG genes derive from a common ancestral gene related to cysteine proteases [2].

Among the eight active TGs only five possess a clear biological function (Table 1). Although each of the TGs has its own distinct targets for crosslinking and their own typical tissue distribution, they are often found in combination with each other. The enzymatic activity of the TGs is tightly regulated throughout the body, both by various cofactors and through their localization, to prevent excessive protein aggregation.

Table 1. The human transglutaminase family

Protein	Gene	Mw (kDa)	Tissue expression ^a	Localization	Function
Factor XIIIa ^b	F13A1	83	Platelets, astrocytes, dermal dendritic cells, chondrocytes, placenta, plasma, synovial fluid	Cytosolic, extracellular	Blood coagulation [3], bone growth [4]
TG1	TGM1	90	Keratinocytes, brain	Membrane, cytosolic	Cell-envelope formation [5]
TG2	TGM2	80	Ubiquitous	Cytosolic, nuclear, membrane, cell surface, extracellular	Multiple (see text) [6]
TG3 ^c	TGM3	77	Squamous epithelium, brain	Cytosolic	Cell-envelope formation [5]
TG4	TGM4	77	Prostate	Unknown	Semen coagulation in rodents [7]
TG5	TGM5	81	Ubiquitous except for the CNS and lymphatic system	Unknown	Unknown
TG6	TGM6	?	Unknown	Unknown	Unknown
TG7	TGM7	?	Ubiquitous	Unknown	Unknown
Band 4.2 ^d	EPB42	72	Red blood cells, bone marrow, fetal liver and spleen	Membrane	Membrane skeletal component [8]

^a Only tissues of high level expression are indicated. ^b Thrombin activated. ^c Protease activated. ^d No TG activity. Table 1 is modified from [9].

Transglutaminase 2

Transglutaminase 2 (TG2) was the first TG to be discovered over 40 years ago [10]. Since then it has been extensively studied and it is now the best characterized enzyme in the family. Still, the precise physiological functions of TG2 remain to be elucidated.

Structure and conformation

TG2 is comprised of four distinct domains; a N-terminal β -sandwich harboring a fibronectin binding site, a catalytic core containing the active site, and two C-terminal β -barrel domains [11]. The conformation and the enzymatic activity of TG2 is regulated by the small ligands GTP and Ca^{2+} [12]. When Ca^{2+} is present and GTP is absent, TG2 is in its catalytically active form. Binding of GTP induces a conformational change that stabilizes the enzyme in a more compact, catalytically inactive form [13]. Two crystal structures of TG2 have recently been reported, one describing TG2 in a closed, GDP bound conformation [14], and the other describing it in an open, extended conformation where the active site is occupied by a peptide-like inhibitor [15]. To date, the structure of the Ca^{2+} bound form of TG2 is unresolved.

Enzymatic activity

TG2, in common with most TGs, has a cysteine proteinase-like active site where a catalytic triad is involved in the crosslinking reaction [16]. The catalytic triad is comprised of cysteine 277, histidine 335 and aspartic acid 358. In the presence of Ca^{2+} , cysteine 277 performs a nucleophilic attack on the δ -carbon of the glutamine side chain of an acyl-donor substrate, resulting in a thioester bound enzyme-substrate intermediate and release of ammonia. Next, a nucleophilic primary amine (small biogenic amines or the ϵ -amine group of a lysine; acting as acyl-acceptor) or water attacks the enzyme-substrate complex and the enzyme, together with the transamidated or deamidated product, are released. TG2 can also exert isopeptidase activity [17] by hydrolyzing the transamidated products. This reaction also results in the generation of deamidated products.

Substrate specificity

The only amino acid that can act as an acyl-donor substrate for TG2 is glutamine. The sequence specificity of TG2 towards acyl-donor substrates has in the recent years been thoroughly investigated and was found to be strongly influenced by the primary sequence. In particular the positioning of proline residues (P) plays a crucial role. Vader *et al.* and Fleckenstein *et al.* examined the targeting of glutamine residues (Q) in peptides by guinea pig TG2, addressing the deamidation and transamidation reaction, respectively [18;19]. Glutamine residues in the QXP sequence motif were found to be targeted by TG2, while glutamine residues with a P in position +1 or +3 were not targeted. Other amino acids in positions -1, +1, +2 and +3 were found to influence the targeting to a lower extent. The same QXP sequence motif was later confirmed in studies using phage-display peptide libraries and

human TG2 [20;21]. Targeting of glutamine residues within intact proteins will in addition be influenced by the protein's secondary and tertiary structure.

A variety of compounds can serve as acyl-acceptor substrates for TG2. These are primary amines like present in the side chain of the amino acid lysine or small biogenic amines such as histamine. TG2 is much less selective towards lysine residues than towards glutamine residues, but the enzymatic reactivity will be affected by the amino acids preceding the lysine residue. While a leucine, serine, alanine or arginine residue has been shown to enhance the reactivity, a proline or glycine residue reduces the reactivity [22;23].

Expression and localization

TG2 is ubiquitously expressed in many different cells and tissues. In addition it can be up-regulated during wound healing and in response to stress signals [24;25]. The enzyme is mainly localized in the cytosol (80%) [9], but it has also been detected in the nucleus [26], mitochondria [27;28], at the cell surface [29] and in the extracellular matrix (ECM) [30]. The presence of TG2 in the ECM and on the cell surface is, however, a conundrum as TG2 lacks both stabilizing disulfide bonds and a leader sequence required to traverse from the cytosol over the plasma membrane. Secreted TG2 either localizes to the cell surface or is deposited in the ECM.

An enzyme with multiple functions

TG2 is a multifunctional enzyme which serves various biochemical functions depending on its subcellular location (Figure 1). In the cytosol, the concentration of GTP is high (200-500 μ M) while the Ca^{2+} concentration is low (<1 μ M). TG2 will therefore mainly be catalytically inactive and the enzyme can function as a G-protein. Acting as a G-protein, TG2 transmits signals from several kinds of cell surface receptors including membrane α_{1B} and α_{1D} -adrenergic receptors [31;32], TP α thromboxane A_2 receptor [32] and the oxytocin receptor [33], leading to activation of the downstream cytoplasmic target phospholipase C and inositol trisphosphate (IP3) production, which in turn results in an increased cytoplasmic Ca^{2+} concentration. Notably, TG2 has also been shown to exert protein disulfide isomerase activity in the cytosol [34]. During inflammation and apoptosis the intracellular Ca^{2+} level will rise and TG2 may be activated to exert its crosslinking activity [35].

In the ECM the Ca^{2+} concentration is sufficiently high for constitutive activation of TG2 (2.5-5 mM) [36], thus it is primarily believed to act as a crosslinking enzyme. However, a recent study reported that the majority of extracellular TG2 is inactive under normal conditions but transiently activated upon tissue injury [37]. TG2 in the extracellular space is directly involved in the assembly, remodeling and stabilization of the ECM in different tissues [35] in addition to wound healing and angiogenesis [38].

On the cell surface, TG2 associates with integrins and via its tight interaction with fibronectin it is involved in cell adhesion, cell spreading and cell migration [2;39-42]. It has also been demonstrated that cell surface TG2, when acting as a integrin co-receptor, is able to transmit signals from the extracellular to the intracellular compartment through the activation of the Rhoa/ROCK and the focal adhesion kinase signaling pathway [43]. Notably, the interaction of TG2 with integrins does not require crosslinking activity. Despite the high Ca^{2+} concentration in the extracellular space, it is unclear whether cell surface TG2 is active or whether it is held latent due to the massive associations with integrins.

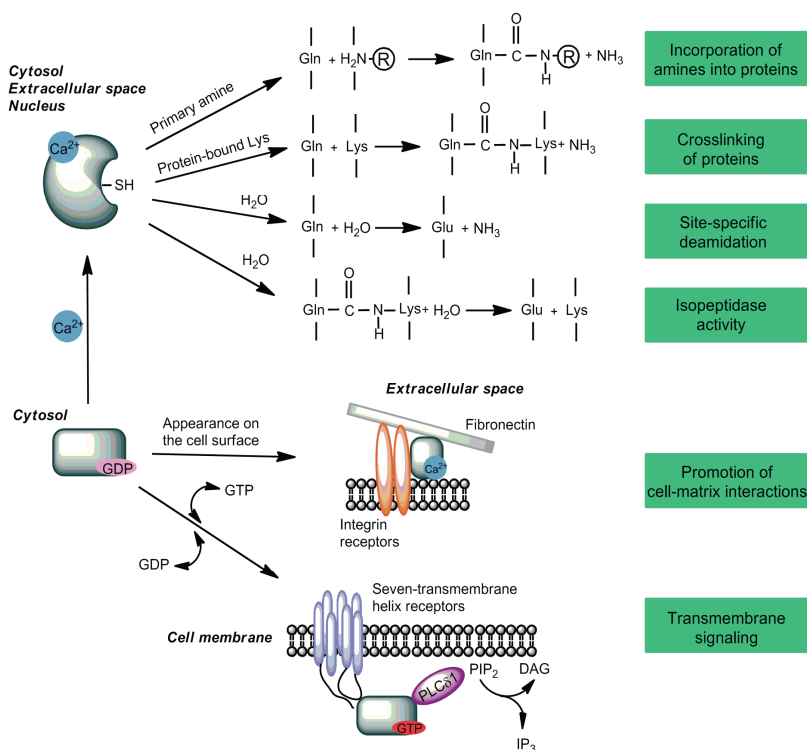


Figure 1. Biochemical activities of TG2. In the presence of Ca^{2+} , TG2 catalyzes acyl-transfer reactions between specific protein-bound glutamines and either primary amines or the amino-group of selected protein-bound lysine residues. Water can replace the primary amine what results in deamidation of the targeted glutamine residue. TG2 can also exert isopeptidase bond hydrolysis and indirectly deamidate its substrate. On the cell surface, TG2 acts as an integrin co-receptor which binds tightly to fibronectin and thereby promotes cell-matrix interactions. In the cytosol, TG2 exerts its function as a GTPase and transmits signals from seven-transmembrane helix receptors activating phospholipase C (PLC δ 1). Figure modified from [11]

TG2 in human diseases

TG2 has been implicated in the pathogenesis of various human diseases including celiac disease [44], cancer [45;46] and the neurological diseases Alzheimer's and Huntington's [47]. The functional role of the enzyme in the diseases is, however, often unknown. Only in celiac disease is the main function of TG2 well established.

Celiac disease

Clinical aspects

Celiac disease (CD) is a genetically-determined, chronic inflammatory disorder caused by the ingestion of gluten proteins in wheat and similar proteins from barley and rye. The disease is characterized by small intestinal damage with loss of absorptive villi, hyperplasia of the crypts and a massive infiltration of lymphocytes in the epithelium and in the lamina propria [48]. Clinical symptoms include diarrhea, anemia, osteoporosis, depression and infertility in addition to skin manifestations and neurological diseases [49]. Elimination of gluten from the diet leads to clinical improvement in most patients. CD is a frequent disorder, affecting about 1:100 in populations of mainly Caucasians [50;51]. Patients on a gluten-containing diet have increased levels of serum antibodies to both gluten and the enzyme TG2 [44]. These antibodies are strictly dependent on the intake of gluten. Serological tests for IgA anti-TG2 antibodies are used for the diagnosis of CD due to their high sensitivity and specificity (>98%) [52]. The final diagnosis of CD is still to be based on the finding of characteristic changes in the histology of small intestinal biopsies.

Genetic factors

CD displays a strong genetic association. The concordance rate is around 8% in first-degree relatives and about 75% in monozygotic twins [53]. The HLA genes are the key genetic risk factors in CD, and they are believed to account for about 50% of the genetic contribution [54]. More than 90% of CD patients express the HLA-DQ2.5 (DQA1*05/DQB1*02) molecule, while the majority of the remaining patients display the HLA-DQ8 (DQA1*03/DQB1*0302) molecule [55]. However, in the normal population the prevalence of HLA-DQ2 is as high as 25-30%, what suggests additional non-HLA susceptibility genes involved in CD pathogenesis. In the recent years several non-HLA candidate genes have been identified through genome-wide association studies [56-58]. Although the contribution of these genes is minimal compared to the HLA genes, they point towards a functional role in the immune system.

Gluten – the trigger of celiac disease

Gluten, the major storage protein in wheat, was identified as the trigger of CD more than 50 years ago [59]. Subsequent studies revealed that also rye, barley and possibly oat provoke the disease [60-62], although the toxicity of the latter is debated.

Wheat gluten contain hundreds of protein components that are present either as monomers or as oligo- and polymers when linked by interchain disulfide bonds [63]. The molecular weights of the proteins range from around 30.000 to more than 10 million [64]. Traditionally, gluten proteins are classified into two main groups according to their solubility in aqueous alcohol; the ethanol-soluble gliadins and the ethanol-insoluble glutenins. The gliadins are mainly monomeric proteins and they have traditionally been further separated into α -, β -, γ - and ω -gliadins according to their electrophoretic behavior [65]. Later studies on amino acid

sequences, however, revealed that the α - and β - gliadins belong to one group (α/β -type) [64]. The ethanol-insoluble glutenins, which comprise aggregated proteins, become soluble in aqueous alcohol after reduction of the disulfide bonds and the glutenin subunits can be divided into low molecular weight (LMW) and high molecular weight (HMW) glutenins. The amino acid composition of gluten is unique, with a high content of both glutamine (30%) and proline residues (10%) and with very few charged residues. Characteristic for both gliadins and glutenins are their repetitive sequence motifs rich in glutamine (Q), proline (P), phenylalanine (F) and tyrosine (Y). The repetitive motifs of the gliadins are typically QPQPFQQPYP (α -gliadin), QPQQPFP (γ -gliadin) and PQQPFPQQ (ω -gliadin) while the LMW glutenins typically contain repetitive units such as QQQPPFS [64]. As proline residues enforce strong restraints on the conformation of a peptide chain, the gluten peptides will often form a polyproline type II helix [66].

Similar alcohol-soluble fractions have been demonstrated in barley (hordeins) and rye (secalins) [67;68]. For cereals in general, the alcohol soluble fraction is termed prolamines and the alcohol insoluble fraction is termed glutelins [69]. Notably, the prolamines of barley and rye are rich in glutamine- and proline residues, while the nontoxic prolamines of rice and corn have a lower glutamine and proline content. Most T cells isolated from gluten challenged small intestinal biopsies of CD patients recognize the gliadin fraction of gluten [70].

Peptide binding to HLA-DQ2/DQ8

Gluten peptides are recognized by gluten-reactive T cells only in the context of HLA-DQ2.5 or -DQ8 molecules [70;71]. These HLA-molecules seem to be uniquely suited for presentation of gluten-derived peptides. The peptide binding motif of HLA-DQ2.5 has been thoroughly investigated and it was found to be rather different from those of other HLA class II molecules. HLA-DQ2.5 can, in contrast to other HLA-molecules, accommodate peptides with several proline residues. In addition, a preference for negative charged anchor residues has been found in positions P4, P6 and P7 of the peptide binding motif [72-74]. This preference for binding negatively charged residues is shared by HLA-DQ8, where it is found for the anchor residues in positions P1 and in P9 [75;76].

The gluten specific immune response

In CD there is a well-established adaptive immune response towards specific gluten peptides. The high amount of proline residues in gluten makes it relatively resistant to degradation by luminal and brush-border enzymes [77]. Fragments of gluten will therefore survive digestion and can be transported across the mucosal epithelium as polypeptides. Several mechanisms have been suggested to explain this peptide transport from the intestinal lumen into the lamina propria [78-80]. The enzyme TG2, which is present in the small-intestinal mucosa, will mediate an ordered deamidation of specific glutamine residues within the proteolytically stable peptides. Importantly, the affinity of the gluten peptides to the disease associated HLA-

molecules is strongly enhanced by the negative charges introduced by deamidation. Upon recognition of the HLA-peptide complex by HLA-DQ2.5- or HLA-DQ8 restricted gluten reactive CD4+ T cells, a cytokine response dominated by interferon- γ is induced [81]. Several gluten T-cell epitopes and their deamidation sites have been identified during the recent years [82-90].

It was recently reported that gluten in addition can induce an innate immune response. The "toxic" α -gliadin derived peptide p31-43 can increase the production of IL-15 resulting in the expression of NKG2D receptors on intraepithelial T cells and its ligand MICA on epithelial cells. The subsequent NKG2D-MICA interaction may result in destruction of the intestinal epithelium[91].

The role of transglutaminase 2 in celiac disease

TG2-mediated deamidation and transamidation of gluten peptides

TG2 plays an important role in the adaptive immune response in CD by deamidating the gluten peptides and thereby increasing their binding affinity to HLA-DQ2.5 and HLA-DQ8. Notably, deamidation by TG2 is also expected to slow down the off-rate so that the gluten peptide will stay on the HLA molecule for a longer period of time [92]. TG2 can target specific glutamine residues within the gluten peptides and catalyze either a deamidation or a transamidation reaction. Deamidation results in conversion of a glutamine residue to a negatively charged glutamic acid residue, whereas in the transamidation reaction primary amines are crosslinked to the side chain of a glutamine residue (Figure 2).

The *in vivo* deamidation of gliadin peptides in CD is not completely understood. TG2 expression has been demonstrated mainly in the brush border and in the subepithelial region in the small intestine [93] where the presence of primary amines would favor the transamidation reaction. However, as most gut T cells recognize the gluten peptides preferentially or only after being deamidated by TG2 [94], deamidation of gluten peptides obviously occurs *in vivo*. The ratio between transamidation and deamidation has been shown to be dependent on the concentration of amines and peptides present, and also on the pH in the environment [19]. At a lower pH (<6.5), the deamidation reaction is favored suggesting that deamidation may occur in compartments of the gut with slightly acidic pH. To date, it is not known where TG2 processes gluten peptides in the intestinal mucosa.

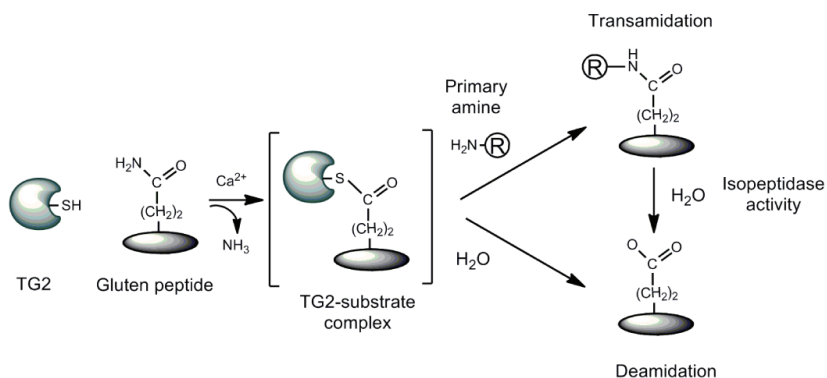


Figure 2. TG2-catalyzed deamidation and transamidation of gluten peptides. In the presence of Ca²⁺, the TG2 active site cysteine reacts with specific glutamine residue side chains within the gluten peptides. An intermediate TG2-substrate complex is formed and ammonia is released. A primary amine or water attacks the TG2 substrate complex displacing the enzyme and generating a transamidated or deamidated gluten peptide, respectively. TG2 can further hydrolyze the transamidated peptide to a deamidated peptide. Figure is modified from [95].

Autoantibodies to TG2

TG2 was identified as the autoantigen in CD in 1997 by Dieterich *et al.*, and the TG2-specific antibodies now represent a hallmark of the disease. These antibodies are found in the small intestine where they have been shown to co-localize with extracellular TG2 [96] and in the blood where they are used as diagnostic tools for CD [97]. Notably, the presence of these IgA and IgG antibodies strongly depends on the dietary intake of gluten. A hypothesis based on a hapten-carrier model has been proposed to explain the presence of the antibodies in the celiac lesion [98]. Uptake of TG2-gluten complexes by TG2-specific B cells is suggested to result in activation of CD4+ gluten-reactive T cells which will give the necessary help to the B cell for isotype switching and secretion of TG2-specific antibodies. Gluten peptides have been demonstrated to be crosslinked to TG2 either by a transient thioester bond to the active site cysteine [99], or to specific lysine residues of the enzyme via very stable isopeptide bonds [100]. As it to date is not known where the formation and the cleavage of such complexes take place, the hapten-carrier hypothesis remains to be proven.

Whether the anti-TG2 autoantibodies have a functional role or not in the pathogenesis of CD is conversely discussed. Several groups have investigated whether the autoantibodies have an effect on the activity of TG2 but contradictory results have been reported [101-103]

The role of other transglutaminases in gluten sensitive diseases

Gluten sensitive diseases (GSD) are a collection of diseases that are induced by dietary gluten and they typically present as CD. However, manifestations can also exist in the absence of the gastrointestinal symptoms typically observed in patients with CD, such as in the skin disease dermatitis herpetiformis (DH) and in the neurological disorder gluten ataxia (GA). Interestingly, other TGs than TG2 appear to be the main autoantigen in these diseases.

DH is affecting around 10% of patients with CD [104]. The disease is characterized by sub-epidermal blistering and deposition of IgA antibodies in the papillary dermis, and the symptoms usually disappear when gluten is removed from the diet. The antibody profile of DH is similar to that observed in CD patients. However, in addition to the circulating antibodies to TG2, DH patients have specific antibodies towards another TG isoform, TG3 (epidermal TG). IgA deposits in the dermal papillae have been shown to co-localize with TG3, but not with TG2 [105].

GA is a newly recognized neurological condition and it is one of the more frequent neurologic syndromes associated with CD. Most of the patients have no gastrointestinal symptoms. The clinical manifestations of GA are primarily in the brain or in the peripheral nervous system, and the disease is believed to be caused by antibody damage to Purkinje cells in the cerebellum. Thus, removal of gluten from the diet may not improve the symptoms as damage to the cerebellum may be permanent. The majority of the GA patients have circulating antibodies towards TG2. In addition, they have antibodies recognizing the novel TG isoform TG6 [106]. Post mortem examination of brain tissue showed that TG6 was present in IgA deposits in the cerebellum.

It is not clear whether TG3 and TG6 are expressed in the intestine. However, the antibody production which is dependent on gluten and the IgA isotype of the antibodies suggest an intestinal origin.

Mass spectrometry and proteomics in celiac disease research

In the last decade, the technological developments in mass spectrometry (MS) have expanded rapidly with continuous improvements in sensitivity, resolution and obtained mass accuracy. This progress has been taken advantage of to address many biochemical, enzymatic and immunological questions related to the pathogenesis of CD. MS has been applied to characterize natural ligands of HLA-DQ2 [107], to identify antigens recognized by serum antibodies of CD patients [108] or to characterize gluten proteins. The major achievements regarding our current understanding of the pathogenesis of CD have however been made in the three following fields.

Specificity and enzymatic activity of TG2

MS has been important for the characterization of the specificity of TG2. Two proteomic approaches have been used to study the influence of the primary sequence on the TG2-mediated targeting of specific glutamine residues. Fleckenstein *et al.* incubated synthetic peptide libraries with TG2 in the presence of primary amines and semi-quantified the transamidated peptides using MS [19]. The targeting by TG2 was found to be strongly influenced by the positioning of proline residues. These results were similar to findings by Vader *et al.* who analyzed deamidation within synthetic substitution analogs of gliadin peptides using MS [18].

In order to evaluate the proposed hapten-carrier hypothesis, complexes of TG2 and gliadin peptides were characterized using biochemical and mass spectrometric analysis. The TG2-gliadin complexes were found to be formed either by relatively unstable thioester bonds during catalysis, or via very stable isopeptide bonds between the gluten peptides and particular lysine residues of the enzyme. Mass spectrometric analysis of tryptic digests of the complexes revealed six acyl-acceptor lysine residues within TG2 [100].

Identification of T-cell epitopes in gluten

The first gluten T-cell epitopes were identified in 1998 by the groups of Sollid *et al.* and Koning *et al.* [83;109]. With the rapid progress in MS, several HLA-DQ2.5 and HLA-DQ8 restricted T-cell epitopes were identified in the following years [82-90]. The T-cell epitopes have been identified from gluten digests of various complexities; e.g. from single recombinant α - or γ -gliadin proteins [85;86] or from purified subgroups of gluten containing either α -gliadin, γ -gliadin or glutenin proteins [87]. The peptide digests were fractionated followed by testing in T-cell assays. Fractions containing T-cell stimulatory peptides were analyzed by MS, and identified peptides were synthesized, treated with TG2 and again tested in T-cell assays to identify the T-cell stimulatory peptides. The deamidation sites were identified by tandem MS.

Mass spectrometry to study gluten proteolysis

In order for a gluten peptide to be a T-cell epitope it must survive proteolytic digestion in the gut. Shan *et al.* analyzed proteolytic digests of recombinant α 2-gliadin to identify regions with high proteolytic stability using liquid chromatography coupled to mass spectrometry (LC-MS). They discovered a 33mer peptide that was stable to all gastric, pancreatic and brush-border enzymes [77]. TG2-treatment of this peptide followed by mass spectrometric analysis revealed that the peptide was an excellent substrate for TG2. Interestingly, the 33mer peptide contained multiple copies of the already identified α -gliadin epitopes and it was found to be efficiently recognized by T cells from all CD patients [110]. Similarly, a proteolytically stable 26mer from a recombinant γ -gliadin was identified which is a good substrate for TG2 and harbors several known γ -gliadin epitopes [111]. With the aim to degrade these proteolytically stable gluten fragments, Shan *et al.* treated gluten with peptidases capable of cleaving proline-rich peptides. LC-MS based methods were central in the assessment of degradation of the immunostimulatory peptides.

In the last decade, MS has been instrumental also in the characterization of cereal proteins. MS-based studies have made a significant contribution to the understanding of the composition and structure of the gluten proteins and for determining their molecular mass (reviewed in [112;113]).

AIMS OF THE STUDY

The focus of this thesis has been to investigate the role of TG2 in the generation of gluten T-cell epitopes in CD. Two MS-based methods were established for this purpose; a method to determine the kinetics of TG2-mediated deamidation of gluten peptides and a proteomic method to identify the best substrates of TG2 in complex digests of whole gluten. The quantification method was in addition applied to investigate whether other TG isoforms than TG2 could accommodate gluten peptides as substrates.

The specific aims of my thesis were:

- To quantify TG2-mediated deamidation of the known HLA-DQ2.5 restricted gliadin epitopes and see whether this correlates with the frequency by which the epitopes are recognized by T cells of CD patients.
- To investigate whether the TG isoforms TG3 and TG6 that are implicated in the gluten sensitive diseases DH and GA, respectively, can utilize gluten peptides as substrates and whether they can form covalent complexes with the gluten peptides.
- To identify the preferred peptide substrates of TG2 in proteolytic digests of whole gluten and address whether these peptides are identical to the known gluten T cell epitopes.

SUMMARY OF PAPERS

Paper I

In this study we established a quantitative, MS-based approach to address the kinetics of TG2-mediated deamidation of different HLA-DQ2.5 restricted gliadin epitopes. Our results demonstrated that there are large differences in the rate of deamidation for the different peptides. We observed a positive correlation between how quickly the various epitopes becomes deamidated by TG2 and the frequency by which the epitopes are recognized by T cells of CD patients. In addition, we found that TG2 deamidates the individual glutamine residues within the gluten epitopes with very different rates, what may influence whether the epitopes are better recognized in the context of HLA-DQ2.5 or HLA-DQ8.

Paper II

The TG isoforms TG3 and TG6 are considered to be the main autoantigens in the gluten sensitive disorders DH and GA, respectively. We investigated whether these enzymes can accommodate gluten peptides as substrates what may be relevant for the gluten-dependent autoantibody production. Our results show that similar to TG2, TG3 and TG6 can specifically deamidate gluten T-cell epitopes, but the enzymes differ in their fine specificities. We found that both TG3 and TG6 were able to form complexes with gluten peptides through thioester linkage although less efficiently than TG2. Only TG6 was found to generate isopeptide linked complexes. These findings support the notion that TG3 and TG6 are involved in the gluten induced autoimmune responses of DH and GA.

Paper III

This study aimed to identify the preferred peptide substrates of TG2 in a highly heterogeneous proteolytic digest of whole gluten. We established a method for enrichment of gluten peptides targeted by TG2 and identified these substrates by LC-MS/MS, database searching of a tailored gluten database and manual data interpretation. Interestingly, our results revealed that the majority of the identified preferred substrates of TG2 harbor known gluten T-cell epitopes, and moreover, we identified two novel T-cell epitopes by this approach. Our findings demonstrate that the selective forces exerted by TG2, together with those created by gastrointestinal proteolysis, are key factors to determine the repertoire of gluten epitopes in CD.

METHODOLOGICAL CONSIDERATIONS

Production of recombinant human TG2

Recombinant human TG2 was expressed in BL21 competent *E.coli* with an N-terminal hexahistidine tag and purified as described by *Piper et al.* [99]. The His₆-TG2 plasmid was a kind gift from Chaitan Khosla, Stanford University. The yield and the activity of the batch of recombinant TG2 used in this thesis were found to be similar to those of previously produced batches of TG2. The deamidation activity was assessed by MS by measuring the deamidation of a gluten epitope peptide at different concentration and at different time points.

Quantification of peptide deamidation by mass spectrometry

The conventional method of quantifying TG2-mediated peptide deamidation is via a coupled assay established previously by others [114]. In this assay, the ammonium released in the TG2-catalyzed reaction serves as a substrate in a subsequent reaction catalyzed by glutamate dehydrogenase. This reaction consumes NADH which is photometrically monitored. However, as only the sum of the ammonium production is measured in this assay, it is not possible to determine the deamidation of each of the peptides when present together in a peptide mixture. Capillary electrophoresis combined with laser induced fluorescence-detection has in addition been used to measure deamidation of gluten peptides [19]. Although this method is highly sensitive, it also shows limitations when it comes to measuring of peptides in a mixture. MS, on the other hand, allows a simultaneous and quantitative analysis of deamidation for each peptide present in a mixture. In addition, the deamidation of individual glutamine residues within an individual peptide can be quantified by MS. Another advantage of the MS-based method is that it is fast and only small amounts of peptide and enzyme are required.

The conversion of a glutamine (Mw 128.1) to a glutamic acid residue (Mw 129.1) in a peptide results in a mass increase of +1 Dalton. Thus, the isotopic envelopes of the native and the deamidated peptides overlap. Quantification of deamidation is consequently based on determination of the centroid mass using the formula $m_c = \sum m_i I_i / \sum I_i$ (m_c : centroid mass, m_i : masses of individual signals in the isotopic envelope, I_i : intensity of the individual signals in the isotopic envelope) [115]. To quantify the total deamidation of a TG2-treated peptide, the calculated centroid mass of the native peptide was subtracted from the calculated centroid mass of the TG2-treated peptide giving a centroid mass shift which corresponds to the average number of deamidation (Figure 3). Importantly, the mass spectrometric response for the native and the deamidated peptide was first demonstrated to be equal, what is a prerequisite for the method applied.

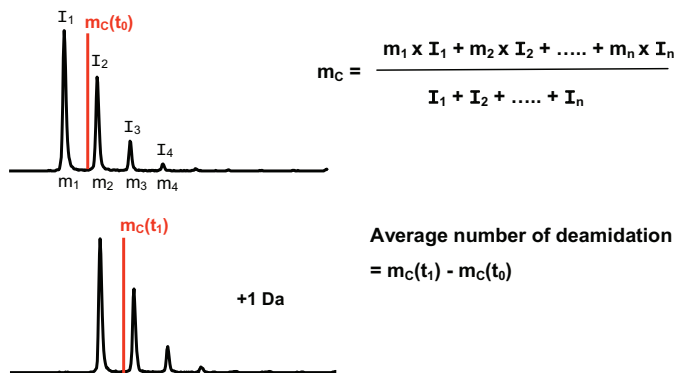


Figure 3. Quantification of TG2-mediated peptide deamidation. The centroid mass of the native peptide (upper panel) and the TG2-treated, deamidated peptide (lower panel) is calculated by the given formula $m_c = \sum m_i I_i / \sum I_i$ (m_c : centroid mass, m_i : masses of individual signals in the isotopic envelope, I_i : intensity of the individual signals in the isotopic envelope). The calculated centroid mass of the native peptide is subsequently subtracted from the calculated centroid mass of the TG2-treated peptide what gives the total deamidation of the TG2-treated peptide (here shown as 1 Dalton e.g. corresponding to complete deamidation of a peptide in which only one glutamine residue is targeted).

Total deamidation of the peptides was measured by MALDI-TOF MS (matrix-assisted laser desorption/ionization time-of-flight mass spectrometry). To quantify deamidation of individual glutamine residues within the peptides, centroid masses of fragments reporting on specific glutamine residues were calculated. Due to its better resolution, those MS/MS experiments were performed on a Q-TOF (quadrupole-time-of-flight) instrument.

Enrichment of TG2 peptide substrates from a digest of whole gluten

Wheat gluten digested with PTCEC (pepsin, trypsin, chymotrypsin, elastase and carboxypeptidase) yields an extremely complex mixture of peptides and tens of thousands of different peptides may be present in the final digest. In order to identify substrate peptides of TG2 from such a complex mixture, a method for enrichment of these substrates is needed. We established an approach which is based on the transamidation reaction catalyzed by TG2. To selectively label and enrich the peptides targeted by TG2, a biotinylated primary amine; 5-biotinamido-pentylamine (5-BP), was used. TG2 covalently crosslinks 5-BP to distinct glutamine residues in the gluten peptide substrates what will increase the mass of the peptides by 311.17 Dalton per incorporation. The transamidated and hence biotinylated peptides were enriched from the digest using magnetic streptavidin beads. Enriched gluten peptides were eluted off the beads by adding an excess of biotin.

To ensure that only the best substrates among the gluten peptides would become transamidated, small amounts of 5-BP and short incubation times with TG2 were chosen.

These conditions are probably not reflecting the *in vivo* situation where primary amines likely are present in ample amounts and the concentration of TG2 is the limiting factor. However, in both cases TG2-mediated reaction takes place in a competitive environment suggesting that the best peptide substrates of TG2 identified by our approach are also targeted by TG2 *in vivo*.

Identification of enriched TG2 peptide substrates by LC-MS

The enriched TG2 peptide substrates were analyzed by both a nano-LC Q-TOF mass spectrometer and a nano-LC LTQ-Orbitrap mass spectrometer. A database consisting of all known sequences of common wheat (*Triticum aestivum*) was constructed using the Sequence Retrieval System (SRS) at the European Bioinformatics Institute. The obtained LC-MS/MS (liquid chromatography/tandem-mass spectrometry) data were searched against the *Triticum aestivum* database. Transamidation by 5-BP was implemented in the in-house Mascot search engine as a modification of glutamine residues and was selected as a variable modification together with deamidation (NQ) and pyro-glutamate formation (N-terminal Q). All reported hits were manually interpreted to validate the identification. This was especially important considering the very similar sequences of the different gluten proteins present in the wheat gluten digest. In addition, to resemble the *in vivo* digestion, we used a mixture of different enzymes, PTCEC, with different cleavage specificities (trypsin, cleavage after K, R; elastase, cleavage after A, G, S, V; chymotrypsin, not very specific, prefers to cleave at F, W, Y, M, L; pepsin, broader specificity, reported to be similar to that of chymotrypsin; carboxypeptidase A, non-specific, prefers aromatic and long side-chain amino acids [111]). Thus, for database searching we selected "none" in the enzyme settings.

T-cell proliferation assay

T-cell recognition of synthetic gluten derived peptides was tested using a ³H-thymidine incorporation assay. T-cell proliferation is measured upon recognition of the peptide-HLA complex by the T-cell receptor, where irradiated HLA-DQ2.5 homozygous EBV-transformed B lymphoblastoid cells are used as antigen presenting cells. A high correlation between T-cell proliferation and other effector functions has been reported [81]. It should however be noted that the method gives a bias towards T cells that proliferates well upon activation which means that gluten specific T cells with a poorer proliferation could be neglected.

GENERAL DISCUSSION

Selection of gluten T-cell epitopes in celiac disease

Wheat gluten is an extremely complex mixture of proteins. Only a few of the thousands of peptides generated by gastrointestinal digestion of gluten proteins serve as T-cell epitopes. Several factors are influencing what gluten peptides are eventually presented to the gluten-specific T cells; proteolytic stability of the peptides, their ability to be a TG2 substrate and their affinity to HLA-DQ2.5 and/or HLA-DQ8. Importantly, these factors are acting together.

Proteolytic stability of epitopes

Resistance to gastrointestinal proteolysis is a prerequisite for a gluten peptide to be selected as a T-cell epitope. The high amount of proline residues in gluten will affect its susceptibility to digestion as only a limited number of mammalian proteases are able to cleave the amide bond N-terminally of proline residues [66]. Interestingly, it was previously discovered that the gliadin-derived T-cell epitopes were not randomly spread in gliadin but rather clustered in regions rich in proline residues [86].

A proteolytically stable α -gliadin derived 33mer peptide that contains several copies of the known α -gliadin epitopes is considered to be the immunodominant peptide in CD. Similarly, a γ -gliadin derived proteolytically stable 26mer peptide that harbors several known γ -gliadin epitopes has been identified. This suggests the existence of also other potential immunogenic proteolytically resistant peptides in gluten. Shan *et al.* thus performed a computational prediction of all peptides generated from wheat, rye and barley when digested with the major pancreatic enzymes. Their results revealed that long, proteolytically stable peptides existed in all of these cereals [111]. However, 50% of the identified peptides did not contain any of the known T-cell epitopes suggesting that the T-cell epitope repertoire in gluten has not yet been thoroughly mapped or that other factors than proteolytic stability play a role in the selection of epitopes.

Selection of epitopes by TG2

The epitopes require deamidation by TG2 to increase their binding affinity to HLA-DQ2.5 or HLA-DQ8. The high frequency of proline residues in the epitopes not only prevents intestinal proteolysis but also plays a crucial role regarding the specificity of TG2 as the positioning of the proline residues determines whether a glutamine residue is targeted or not.

To date, several TG2-dependent T-cell epitopes have been identified. The majority of these derive from the α - and γ -gliadins and they are restricted to HLA-DQ2.5. Notably, a few TG2-independent gluten epitopes have been reported [87;116]. Interestingly, it seems as the epitopes show a hierarchy in their recognition by patient derived T cells. While the α -gliadin epitopes in general are recognized by T cells from almost all patients, the γ -gliadin epitopes

are infrequently recognized. This hierarchy could be explained by the rates by which TG2 deamidates the different epitopes. In paper I we addressed this hypothesis and we demonstrated large variations in the rate of deamidation of the different known HLA-DQ2.5 restricted epitopes. We found that the α -gliadin epitopes in general were faster deamidated by TG2 than the γ -gliadin epitopes. Our results thus suggest that there is a correlation between the level of deamidation and the frequency of gluten epitope specific T cells observed in patient biopsies.

Two studies have aimed to identify the peptide substrates of TG2 in the gliadin fraction of gluten [117;118]. However, these studies were carried out with purified gliadin or with recombinant α -gliadin. So far, no studies have aimed to identify the peptide substrates of TG2 in whole wheat gluten, probably due to its extreme complexity. Considering the abundance of proline and glutamine residues in gluten, one would expect the presence of numerous potential substrates of TG2. Results in paper I did however show that the velocity by which TG2 deamidates the different epitope peptides differs significantly. It may thus be that *in vivo* only a limited number of gluten peptides will be substrates of TG2. In paper III we developed a method to enrich and identify the preferred substrates of TG2 under the competitive conditions which exist when a highly complex digest of whole wheat gluten is applied. Strikingly, we found that the majority of the best peptide substrates of TG2 harbored complete or partial sequences of T-cell epitopes. Thus, after enzymatic gluten degradation TG2 is not active on a broad scale of different gluten peptides but shows a high preference for those that carry T-cell epitopes.

Results from paper III also demonstrated the important role of proteolytic stability in the selection of T-cell epitopes. Those T-cell epitopes for which we observed only partial sequences in the identified favored TG2 substrates are known to be infrequently recognized by T cells from CD patients. In contrast, the epitopes that were observed with their complete 9mer core region are frequently recognized by T cells of CD patients. Thus, the hierarchy of T-cell responses observed among the patients is clearly affected by both gastrointestinal proteolysis and deamidation by TG2.

TG2-mediated deamidation *in vivo*

Although we and several other groups have put much emphasis in investigating the TG2-catalyzed deamidation of gluten peptides, it is still not clear where and when the gluten peptides encounter active TG2 *in vivo* and thus become deamidated. TG2 is mainly expressed in the subepithelial region of the mucosa and to a lesser extent in the brush border [93]. In the subepithelial region the pH is neutral and ample amounts of primary amines are likely to be present. Although the transamidation reaction most likely is the favored reaction in this environment, deamidation has been shown to occur in the presence of primary amines [19]. The deamidation reaction has however been demonstrated to significantly increase

when the pH is lowered [19]. In the proximal small intestine the pH is around 6.6 [119]. One may therefore envisage that TG2 encounters the gluten peptides in the brush border where deamidation could be the favored reaction. It has also been suggested that TG2 is endocytosed and remains active in the early endosomes where there is sufficient Ca^{2+} [19]. Integrin associated TG2 could thus be implicated in receptor-mediated endocytosis of gluten peptides. Another possibility is that gluten complexed to surface immunoglobulin on TG2-specific B cells (hapten-carrier hypothesis) is endocytosed and subsequently becomes deamidated, either by direct deamidation by TG2 or indirectly by isopeptide bond hydrolysis of the TG2-gluten complex.

Peptide binding to HLA-DQ2.5 and HLA-DQ8

The final determinant in the selection of gluten epitopes is peptide binding to HLA-DQ2 or HLA-DQ8 (Figure 4). In order to generate a T-cell response the deamidated gluten peptides must have a sufficient binding affinity to these HLA-molecules. Both the positioning of proline residues and the positioning of the deamidated glutamine residues will affect this binding affinity. Previously it was shown that some gliadin epitopes were recognized by T cells in the context of HLA-DQ2.5 or HLA-DQ8 when bound in exactly the same register, but they required different glutamine residues to be deamidated [89]. This finding was supported in paper I. We found that TG2 deamidated the individual glutamine residues within the epitopes with very different efficiencies and thus generated epitopes with a HLA-DQ2.5 or HLA-DQ8 signature. This “deamidation pattern” appears to determine whether some epitopes are better recognized in the context of HLA-DQ2.5 or HLA-DQ8.

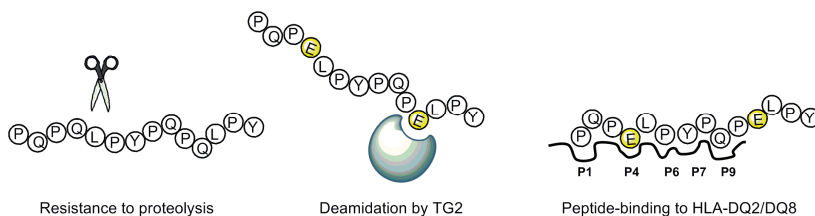


Figure 4. Selection of gluten T-cell epitopes. *Gluten peptides that are rich in proline residues (P) will survive intestinal degradation. TG2 can deamidate specific glutamine residues (Q) within these proteolytically stable peptides and the positioning of deamidated glutamine residues and proline residues will affect the peptides' binding affinity to HLA-DQ2.5. Figure modified from [120].*

In conclusion, our results shed light on the crucial role of TG2 in epitope selection in CD. It may be tempting to speculate whether other TGs play a similarly important role in the epitope selection of gluten sensitive diseases other than celiac disease.

Gluten sensitive diseases and redundancy of transglutaminases

Other TGs than TG2 seem to be implicated in the gluten sensitive diseases DH and GA. The main autoantigens in DH and GA are now considered to be TG3 and TG6, respectively. In addition to specific antibodies towards these autoantigens, both DH and GA patients have antibody populations recognizing TG2 and antibodies which are cross-reactive with TG2 and TG3/TG6. The production of autoantibodies to TG3, and most likely also to TG6, is dependent on gluten intake. This may suggest that TG3 and TG6, similarly to TG2, are able to create hapten-carrier complexes thereby driving their own antibody production. In paper II we wanted to shed light on the enzymes' potential redundancy in gluten sensitive diseases by comparing the biochemical properties of these closely related enzymes using gluten peptides as model substrates. We showed that gluten peptides indeed are substrates for TG3 and TG6. Both enzymes were found to form thioester linked complexes with the peptides, although to less extent than TG2. However, in an *in vivo* setting the capability of TG3 and TG6 to create complexes with gluten may be of importance. A prerequisite for an involvement of TG3 and TG6 in these gluten sensitive diseases is of course that TG3 and TG6 will encounter gluten peptides. Preliminary results by D. Aeschlimann (personal communication) indicate that the enzymes are expressed in the intestine.

Surprisingly, it was reported that TG2 knock-out mice have no clear spontaneous abnormalities [121;122]. It has therefore been speculated whether the activity of other TGs could compensate for the lack of TG2. In fact, other TGs including TG3 have been reported to be upregulated in TG2 knock-out mice [1]. The hypothesis of redundancy is further strengthened by our results, demonstrating that different TG isoforms can perform similar tasks.

Notably, TG3 and TG6 may display a stronger sequence preference towards other gluten peptides than the TG2-dependent gluten T-cell epitopes tested in this paper. An interesting approach would thus be to use the methodology established in paper III to identify the best gluten peptide substrates of TG3 and TG6.

Future therapeutic strategies of celiac disease

At present, the only treatment of CD is a life-long exclusion of gluten from the diet. As gluten is present in various food products total elimination is both difficult and inconvenient and quality of life is reduced for many patients. Alternative therapies are thus highly desired. However, it is relatively safe to be on a gluten-free diet what must be taken into account when developing novel therapies. With respect to the selection process of gluten T-cell epitopes, several interesting targets can be considered. Our results have underlined the crucial role of TG2 in that selection what suggests that therapeutic approaches based on the inhibition of TG2 could be beneficial for CD patients.

TG2-based therapeutic strategies

Two *ex vivo* studies have demonstrated the potential of TG2 inhibitors as therapeutic agents for CD. Both Molberg *et al.* [94] and Maiuri *et al.* [123] showed that inhibition of endogenous TG2 in small intestinal biopsies from CD patients with cystamine and 2-[(2-oxopropyl)thio]imidazolium inhibitor L682777, respectively, could prevent gluten peptide deamidation and consequently led to reduced T-cell activation.

To date, a variety of potential inhibitors of the active site of TG2 have been developed. They can be divided into competitive amine inhibitors, reversible inhibitors and irreversible inhibitors depending on their inhibition mechanism [95]. The best characterized TG2 inhibitors are competitive amine inhibitors such as putrescine, cystamine, spermidine, histidine, monodansylcadaverine and 5-(biotinamido)pentylamine [95]. These compounds are commercially available, chemically stable and non-toxic (exception: cystamine) [124;125]. Competitive amine inhibitors inhibit TG2 activity by competing with poly-peptide bound lysine residues and other primary amines in the transamidation reaction, thus TG2-mediated transamidation may continue to occur in the presence of these inhibitors. *In vivo* however, these compounds should not be considered as selective for TG2 [126]. Reversible inhibitors block the access of substrates to the active site without covalently modifying the enzyme. Among the reversible inhibitors are the cofactors GTP and GDP which will close the active confirmation of TG2 and the metal ion Zn^{2+} which competes for the same site as Ca^{2+} in the enzyme [127-129]. Irreversible TG2 inhibitors covalently modify cysteine 277 in the active site of the enzyme. The irreversible inhibitors 3-halo-4,5-dihydroisoxazoles are one of the most studied classes. KCC009, a synthetic compound containing halo-dihydroisoxazole is particularly promising. It has been demonstrated to inhibit intestinal TG2 when given orally and it shows low toxicity in mice. The compound in addition has a short serum half-life what limits its exposure to other organs [130]. A main drawback of these compounds are however their low solubility in buffers at physiological pH [95].

Since TG2 is a multifunctional enzyme involved in numerous biological processes, its systemic inhibition can be damaging. Although TG2 knock-out mice have no obvious

abnormalities they seem to have a systemic triggering of apoptosis [131]. Hence, the activity of any potential TG2 inhibitor should be restricted to the gut. An optimal TG2 inhibitor should in addition fulfill other criteria including low toxicity, high proteolytic stability and good solubility. Importantly, the inhibitor should be metabolized in the liver to avoid systemic effects in the body. To minimize possible side effects of unspecific inhibition on other TGs, isoform specific inhibitors will be required.

TG2-mediated deamidation of gluten peptides is assumed to occur extracellularly in the gut mucosa. A potential danger of inhibiting extracellular TG2 is therefore to impair TG2's activity with respect to e.g. tissue formation and wound healing. However, as proteins of the ECM are excellent substrates of TG2, one may envisage that inhibition of most of the TG2 present in the ECM mainly would affect the generation of deamidated gluten peptides. An appealing strategy would be to develop TG2 inhibitors capable of selectively inhibit the deamidation reaction what would limit potential side-effects as the transamidation reaction would not be affected.

Alternative therapeutic approaches

Degradation or removal of T-cell epitopes

Enzymatic degradation of gluten is also considered as an attractive strategy for oral therapy. To date, several prolyl endopeptidases (PEPs) capable of cleaving proline-rich gluten peptides are studied for their potential use either alone or in combination with a glutamine specific endoprotease [132-136]. Both *in vitro* and *in vivo* (rat) experimental systems have been developed in order to address the kinetics of gluten degradation by PEPs and the minimal therapeutic dose required. The efficiency of oral therapy is however questionable as a high amount of PEP and prolonged exposure seems to be needed to completely break down all immunogenic gluten peptides. Ongoing clinical trials will show the therapeutic potential of oral enzyme supplementation.

Another enzymatic therapeutic strategy was suggested by Gianfrani *et al.* where transamidation of wheat flour using microbial TG and lysine or lysine methyl ester as amines was shown to abolish the stimulatory capacity of gliadin [137]. Recently, a clinical trial was reported in which treated CD patients were tested for their tolerance to transamidated wheat flour. Approximately half of the patients were found to be tolerant [138].

An alternative approach would be to remove or lower the level of epitopes in wheat gluten. This could potentially be achieved by developing non-toxic wheat strains through traditional breeding or by genetic engineering of wheat. However, the number and repetition of the epitope peptides in wheat makes this approach very difficult and it may also have a negative influence on the baking quality of those wheat strains. In any case, a comprehensive characterization of the immunogenic peptides in all gluten proteins is first required.

Blocking of peptide presentation by HLA-DQ molecules

Another apparent approach for therapy would be to block HLA-DQ2 or HLA-DQ8 presentation of gluten peptides to CD4+ T cells. Notably, peptide blockers are expected to have few side effects as most individuals are heterozygous for HLA molecules. Several studies have aimed to develop suitable peptide blockers based on proteolytically stable gluten peptide analogues [139-141]. As gluten peptides bind to HLA-DQ2 with a mediocre affinity, the bottleneck hereby is to generate binders with sufficiently high affinity. Recently, Jüse *et al.* identified several high-affinity HLA-DQ2 binders using soluble peptide libraries and recombinant soluble HLA-DQ2 molecules as screening tools [142]. The best peptide ligand, FWADYEEEW, bound 50-fold better than the immunodominant gluten epitope DQ2- α -I-gliadin. However, in contrast to gluten peptides most other peptides will be degraded by gastrointestinal enzymes *in vivo*. Thus, further attempts to stabilize such peptide lead structures towards proteolysis, e.g. by amidation of the C-terminus and acetylation of the N-terminus, or by including non-natural amino acids, are required to make this approach successful.

Peptide-based therapy

Many of the drugs which are currently under development as therapeutic agents for CD are thought to be a supplement to the gluten free diet. Peptide-based therapeutic vaccines on the other hand would greatly improve the quality of life for patients as such a vaccine has the potential to give a long-term qualitative change in the immune response to gluten. The aim is to recover the immunological tolerance to gluten by feeding the patients with deamidated gluten peptides in multiple low doses. The key to create an effective peptide-based therapy is therefore to identify the hierarchy of immunogenicity of gluten peptides. This is a big challenge as the intestinal T-cell responses to gluten seem to be very heterogeneous [90]. The gliadin proteins of gluten are considered to be the most immunogenic and several studies have compared the T-cell responses among HLA-DQ2.5 positive patients towards different gliadin epitopes. Considering the high sequence similarity between the proteins in gluten one would expect the presence of several immunogenic peptides also in the glutenin fraction. The TG2-modified T-cell stimulatory gliadin and glutenin peptides identified from whole wheat gluten in our study (paper III) could thus be potential candidates for a peptide-based therapy.

Interestingly, peptide-based vaccines can induce immune tolerance not only to the selected immunodominant epitopes, the tolerance may potentially also spread to other less dominant immunogenic epitopes [143]. A peptide-based vaccine, "Nexvax2", developed by the company Nexpep with Dr. Robert Anderson in the lead, is currently in a clinical trial in Australia (www.clinicaltrial.org).

Mass spectrometry-based strategies and applications in future therapy

In the last couple of years, specific detection of individual peptides present in highly heterogeneous samples by multiple reaction monitoring (MRM) has made huge progress. This technology, when performed on triple quadrupole mass spectrometers is highly sensitive and allows for quantification of selected peptides in complex mixtures down to attomole amounts [144]. MRM is thus well suited for the targeted detection of any given gluten peptide in gluten-free food. Gluten-free food is traditionally validated by several commercial enzyme-linked immunosorbent assay (ELISA) kits [145]. These are based on different monoclonal or polyclonal antibodies generated against a variety of gliadin components. Proteomic methods using LC-MS/MS aiming to qualitatively and quantitatively detect traces of gluten in gluten-free food have in addition been developed [146] where gliadin-components common for several wheat varieties have then been selected as protein markers. The large set of epitopes identified to date allows searching specifically for the presence of these immunogenic epitope peptides in gluten using MRM. The first studies have now taken this into account. Sealey-Voyksner *et al.* selected six potential immunogenic gluten peptides that ionize well in the mass spectrometer as markers for gluten content in food. A highly specific and sensitive analytical LC-MS/MS method using MRM was then developed for quantification of these gluten peptides in various food and consumer products [147].

In addition, MRM may be useful in the development of other therapeutic strategies. Wheat strains engineered to have a low level of immunogenicity or wheat where the gluten epitopes have been "masked" could similarly be analyzed by LC-MS/MS using MRM to ensure the complete removal of the epitopes. MRM may also be used to specifically monitor degradation of specific epitope peptides in cereals treated with different PEPs. Of course MRM will only detect the gluten peptides specified to be targeted by that approach, meaning that a comprehensive characterization of all epitopes in gluten proteins is required to take full advantage of that analytical technique. The epitope peptides identified in paper III may thus be taken into account when selecting peptides for targeted MRM detection. The MS-based method established in our study may also be an interesting approach to identify TG2-dependent epitopes in other cereals immunogenic for CD patients.

FINAL COMMENTS

This thesis has aimed to investigate TG catalyzed modifications of gluten peptides in order to understand the enzyme's role in the pathogenesis of gluten sensitive diseases. The "missing link" between TG2's affinity for the different gluten peptides and the observed hierarchy in gluten T-cell responses in CD patients has been established and a potential redundancy among the TGs implicated in gluten sensitive diseases has been demonstrated.

The best substrates of TG2 identified from complex gluten digest by the established MS-based method were found to mainly harbor the known T-cell epitopes. Hence, the method functioned as a rapid screening method to identify TG2-dependent T-cell epitopes. As only a very few epitopes have been identified from other cereals toxic for CD patients, the established enrichment method could further be applied to proteolytic digests of rye and barley with the aim to identify novel T-cell epitopes. The method may potentially also be used to identify the best substrates of other TGs. The finding that TG3 and TG6 similarly to TG2 can accommodate gluten peptides implies that also these TG isoforms can drive their own antibody production and may also play a role in T-cell epitope selection in the gluten sensitive diseases DH and GA. The enrichment method could thus be utilized to analyze the selection of gluten peptides by TG3 and TG6.

In general, the results presented in this thesis demonstrate the crucial role of enzymes in generating posttranslationally modified antigens which are recognized by T cells. This suggests that enzymatic modifications of self-proteins could be a more general mechanism in which T-cell epitopes are created. Future research should shed light on whether other enzymes play a similarly crucial role as TG2 in the pathogenesis of other autoimmune diseases.

REFERENCES

- [1] Grenard P, Bates MK, Aeschlimann D. 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* 2001 Aug 31;276(35):33066-78.
- [2] Yee VC, Le Trong I, Bishop PD, et al. Structure and function studies of factor XIIIa by x-ray crystallography. *Semin Thromb Hemost* 1996;22(5):377-84.
- [3] Lorand L. Factor XIII: structure, activation, and interactions with fibrinogen and fibrin. *Ann N Y Acad Sci* 2001;936:291-311.
- [4] Wozniak M, Fausto A, Carron CP, et al. Mechanically strained cells of the osteoblast lineage organize their extracellular matrix through unique sites of alphavbeta3-integrin expression. *J Bone Miner Res* 2000 Sep;15(9):1731-45.
- [5] Kalinin A, Marekov LN, Steinert PM. Assembly of the epidermal cornified cell envelope. *J Cell Sci* 2001 Sep;114(Pt 17):3069-70.
- [6] Chen JS, Mehta K. Tissue transglutaminase: an enzyme with a split personality. *Int J Biochem Cell Biol* 1999 Aug;31(8):817-36.
- [7] Williams-Ashman HG, Notides AC, Pabalan SS, et al. Transamidase reactions involved in the enzymic coagulation of semen: isolation of -glutamyl- -lysine dipeptide from clotted secretion protein of guinea pig seminal vesicle. *Proc Natl Acad Sci U S A* 1972 Aug;69(8):2322-5.
- [8] Ideguchi H, Nishimura J, Nawata H, et al. A genetic defect of erythrocyte band 4.2 protein associated with hereditary spherocytosis. *Br J Haematol* 1990 Mar;74(3):347-53.
- [9] Lorand L, Graham RM. Transglutaminases: crosslinking enzymes with pleiotropic functions. *Nat Rev Mol Cell Biol* 2003 Feb;4(2):140-56.
- [10] Sarkar NK, Clarke DD, Waelch H. An enzymically catalyzed incorporation of amines into proteins. *Biochim Biophys Acta* 1957 Aug;25(2):451-2.
- [11] Fesus L, Piacentini M. Transglutaminase 2: an enigmatic enzyme with diverse functions. *Trends Biochem Sci* 2002 Oct;27(10):534-9.
- [12] Achyuthan KE, Greenberg CS. Identification of a guanosine triphosphate-binding site on guinea pig liver transglutaminase. Role of GTP and calcium ions in modulating activity. *J Biol Chem* 1987 Feb 5;262(4):1901-6.
- [13] Monsonogo A, Friedmann I, Shani Y, et al. GTP-dependent conformational changes associated with the functional switch between Galpha and cross-linking activities in brain-derived tissue transglutaminase. *J Mol Biol* 1998 Oct 2;282(4):713-20.
- [14] Liu S, Cerione RA, Clardy J. Structural basis for the guanine nucleotide-binding activity of tissue transglutaminase and its regulation of transamidation activity. *Proc Natl Acad Sci U S A* 2002 Mar 5;99(5):2743-7.
- [15] Pinkas DM, Strop P, Brunger AT, et al. Transglutaminase 2 undergoes a large conformational change upon activation. *PLoS Biol* 2007 Dec;5(12):e327.
- [16] Pedersen LC, Yee VC, Bishop PD, et al. Transglutaminase factor XIII uses proteinase-like catalytic triad to crosslink macromolecules. *Protein Sci* 1994 Jul;3(7):1131-5.
- [17] Parameswaran KN, Cheng XF, Chen EC, et al. Hydrolysis of gamma:epsilon isopeptides by cytosolic transglutaminases and by coagulation factor XIIIa. *J Biol Chem* 1997 Apr 11;272(15):10311-7.

- [18] Vader LW, de Ru AH, van der Wal Y, et al. Specificity of tissue transglutaminase explains cereal toxicity in celiac disease. *J Exp Med* 2002 Mar 4;195(5):643-9.
- [19] Fleckenstein B, Molberg O, Qiao SW, et al. Gliadin T cell epitope selection by tissue transglutaminase in celiac disease. Role of enzyme specificity and pH influence on the transamidation versus deamidation process. *J Biol Chem* 2002 Sep 13;277(37):34109-16.
- [20] Keresztessy Z, Csosz E, Harsfalvi J, et al. Phage display selection of efficient glutamine-donor substrate peptides for transglutaminase 2. *Protein Sci* 2006 Nov;15(11):2466-80.
- [21] Sugimura Y, Hosono M, Wada F, et al. Screening for the preferred substrate sequence of transglutaminase using a phage-displayed peptide library: identification of peptide substrates for TGASE 2 and Factor XIIIa. *J Biol Chem* 2006 Jun 30;281(26):17699-706.
- [22] Grootjans JJ, Groenen PJ, de Jong WW. Substrate requirements for transglutaminases. Influence of the amino acid residue preceding the amine donor lysine in a native protein. *J Biol Chem* 1995 Sep 29;270(39):22855-8.
- [23] Groenen PJ, Smulders RH, Peters RF, et al. The amine-donor substrate specificity of tissue-type transglutaminase. Influence of amino acid residues flanking the amine-donor lysine residue. *Eur J Biochem* 1994 Mar 15;220(3):795-9.
- [24] Ientile R, Caccamo D, Griffin M. Tissue transglutaminase and the stress response. *Amino Acids* 2007 Aug;33(2):385-94.
- [25] Telci D, Griffin M. Tissue transglutaminase (TG2) - a wound response enzyme. *Front Biosci* 2006;11:867-82.
- [26] Lesort M, Attavanich K, Zhang J, et al. Distinct nuclear localization and activity of tissue transglutaminase. *J Biol Chem* 1998 May 15;273(20):11991-4.
- [27] Mastroberardino PG, Farrace MG, Viti I, et al. "Tissue" transglutaminase contributes to the formation of disulphide bridges in proteins of mitochondrial respiratory complexes. *Biochim Biophys Acta* 2006 Sep;1757(9-10):1357-65.
- [28] Krasnikov BF, Kim SY, McConoughey SJ, et al. Transglutaminase activity is present in highly purified nonsynaptosomal mouse brain and liver mitochondria. *Biochemistry* 2005 May 31;44(21):7830-43.
- [29] Upchurch HF, Conway E, Patterson MK, et al. Localization of cellular transglutaminase on the extracellular matrix after wounding: characteristics of the matrix bound enzyme. *J Cell Physiol* 1991 Dec;149(3):375-82.
- [30] Martinez J, Chalupowicz DG, Roush RK, et al. Transglutaminase-mediated processing of fibronectin by endothelial cell monolayers. *Biochemistry* 1994 Mar 8;33(9):2538-45.
- [31] Nakaoka H, Perez DM, Baek KJ, et al. Gh: a GTP-binding protein with transglutaminase activity and receptor signaling function. *Science* 1994 Jun 10;264(5165):1593-6.
- [32] Iismaa SE, Wu MJ, Nanda N, et al. GTP binding and signaling by Gh/transglutaminase II involves distinct residues in a unique GTP-binding pocket. *J Biol Chem* 2000 Jun 16;275(24):18259-65.
- [33] Park ES, Won JH, Han KJ, et al. Phospholipase C-delta1 and oxytocin receptor signalling: evidence of its role as an effector. *Biochem J* 1998 Apr 1;331 (Pt 1):283-9.
- [34] Hasegawa G, Suwa M, Ichikawa Y, et al. A novel function of tissue-type transglutaminase: protein disulphide isomerase. *Biochem J* 2003 Aug 1;373(Pt 3):793-803.
- [35] Aeschlimann D, Thomazy V. Protein crosslinking in assembly and remodelling of extracellular matrices: the role of transglutaminases. *Connect Tissue Res* 2000;41(1):1-27.
- [36] Cooper GM. *The Cell: A Molecular Approach*. 2nd ed. Washington DC: ASM Press, 2000.

- [37] Siegel M, Strnad P, Watts RE, et al. Extracellular transglutaminase 2 is catalytically inactive, but is transiently activated upon tissue injury. *PLoS ONE* 2008;3(3):e1861.
- [38] Haroon ZA, Lai TS, Hettasch JM, et al. Tissue transglutaminase is expressed as a host response to tumor invasion and inhibits tumor growth. *Lab Invest* 1999 Dec;79(12):1679-86.
- [39] Gaudry CA, Verderio E, Aeschlimann D, et al. Cell surface localization of tissue transglutaminase is dependent on a fibronectin-binding site in its N-terminal beta-sandwich domain. *J Biol Chem* 1999 Oct 22;274(43):30707-14.
- [40] Akimov SS, Krylov D, Fleischman LF, et al. Tissue transglutaminase is an integrin-binding adhesion coreceptor for fibronectin. *J Cell Biol* 2000 Feb 21;148(4):825-38.
- [41] Akimov SS, Belkin AM. Cell surface tissue transglutaminase is involved in adhesion and migration of monocytic cells on fibronectin. *Blood* 2001 Sep 1;98(5):1567-76.
- [42] Zemskov EA, Janiak A, Hang J, et al. The role of tissue transglutaminase in cell-matrix interactions. *Front Biosci* 2006;11:1057-76.
- [43] Janiak A, Zemskov EA, Belkin AM. Cell surface transglutaminase promotes RhoA activation via integrin clustering and suppression of the Src-p190RhoGAP signaling pathway. *Mol Biol Cell* 2006 Apr;17(4):1606-19.
- [44] Dieterich W, Ehnis T, Bauer M, et al. Identification of tissue transglutaminase as the autoantigen of celiac disease. *Nat Med* 1997 Jul;3(7):797-801.
- [45] Birckbichler PJ, Orr GR, Conway E, et al. Transglutaminase activity in normal and transformed cells. *Cancer Res* 1977 May;37(5):1340-4.
- [46] Barnes RN, Bungay PJ, Elliott BM, et al. Alterations in the distribution and activity of transglutaminase during tumour growth and metastasis. *Carcinogenesis* 1985 Mar;6(3):459-63.
- [47] Lesort M, Tucholski J, Miller ML, Jet al. Tissue transglutaminase: a possible role in neurodegenerative diseases. *Prog Neurobiol* 2000 Aug;61(5):439-63.
- [48] Trier JS. Celiac sprue. *N Engl J Med* 1991 Dec 12;325(24):1709-19.
- [49] Green PH, Jabri B. Coeliac disease. *Lancet* 2003 Aug 2;362(9381):383-91.
- [50] Maki M, Mustalahti K, Kokkonen J, et al. Prevalence of Celiac disease among children in Finland. *N Engl J Med* 2003 Jun 19;348(25):2517-24.
- [51] Dube C, Rostom A, Sy R, et al. The prevalence of celiac disease in average-risk and at-risk Western European populations: a systematic review. *Gastroenterology* 2005 Apr;128(4 Suppl 1):S57-S67.
- [52] Fasano A, Catassi C. Current approaches to diagnosis and treatment of celiac disease: an evolving spectrum. *Gastroenterology* 2001 Feb;120(3):636-51.
- [53] Greco L, Romino R, Coto I, et al. The first large population based twin study of coeliac disease. *Gut* 2002 May;50(5):624-8.
- [54] Sollid LM, Lie BA. Celiac disease genetics: current concepts and practical applications. *Clin Gastroenterol Hepatol* 2005 Sep;3(9):843-51.
- [55] Sollid LM, Thorsby E. HLA susceptibility genes in celiac disease: genetic mapping and role in pathogenesis. *Gastroenterology* 1993 Sep;105(3):910-22.
- [56] van Heel DA, Franke L, Hunt KA, et al. A genome-wide association study for celiac disease identifies risk variants in the region harboring IL2 and IL21. *Nat Genet* 2007 Jul;39(7):827-9.
- [57] Hunt KA, Zhernakova A, Turner G, et al. Newly identified genetic risk variants for celiac disease related to the immune response. *Nat Genet* 2008 Apr;40(4):395-402.

- [58] Adamovic S, Amundsen SS, Lie BA, et al. Association study of IL2/IL21 and FcgRIIa: significant association with the IL2/IL21 region in Scandinavian coeliac disease families. *Genes Immun* 2008 Jun;9(4):364-7.
- [59] Dicke WK, Weijers HA, van de Kamer JH. Coeliac disease. II. The presence in wheat of a factor having a deleterious effect in cases of coeliac disease. *Acta Paediatr* 1953 Jan;42(1):34-42.
- [60] Dicke WK. Coeliac disease. Investigation of the harmful effects of certain types of cereals on patients with coeliac disease. Thesis, University of Utrecht; 1950.
- [61] van de Kamer JH, Weijers HA, Dicke WK. Coeliac disease. IV. An investigation into the injurious constituents of wheat in connection with their action on patients with coeliac disease. *Acta Paediatr* 1953 May;42(3):223-31.
- [62] Dicke WK, van de Kamer JH, Weijers HA. Celiac disease. *Adv Pediatr* 1957;9:277-318.
- [63] Wrigley CW, Bietz JA. Proteins and amino acids. In: *Wheat-Chemistry and Technology*, Pomeranz, Y. St.Paul American Association of Cereal Chemistry, 1988.
- [64] Wieser H. Chemistry of gluten proteins. *Food Microbiol* 2007 Apr;24(2):115-9.
- [65] Shewry PR, Tatham AS, Kasarda DD. Cereal proteins and coeliac disease. In: *Coeliac Disease*, M.N. Marsh. Oxford: Blackwell Scientific Publications, 1992, p. 305-48.
- [66] Vanhoof G, Goossens F, De Meester I, et al. Proline motifs in peptides and their biological processing. *FASEB J* 1995 Jun;9(9):736-44.
- [67] Einhof H. Chemical analysis of rye (*Secale cereale*). *Neues Allgem J Chem* 1805;5:131-53.
- [68] Einhof H. Chemical analysis of small barley (*Hordeum vulgare*). *Neues Allgem J Chem* 1806;6:62-98.
- [69] Osborne TB. *The vegetable proteins*. 2nd ed. London: Longmans, Green and Co, 1924.
- [70] Lundin KE, Scott H, Hansen T, et al. Gliadin-specific, HLA-DQ(alpha 1*0501,beta 1*0201) restricted T cells isolated from the small intestinal mucosa of celiac disease patients. *J Exp Med* 1993 Jul 1;178(1):187-96.
- [71] Lundin KE, Scott H, Fausa O, et al. T cells from the small intestinal mucosa of a DR4, DQ7/DR4, DQ8 celiac disease patient preferentially recognize gliadin when presented by DQ8. *Hum Immunol* 1994 Dec;41(4):285-91.
- [72] Johansen BH, Vartdal F, Eriksen JA, et al. Identification of a putative motif for binding of peptides to HLA-DQ2. *Int Immunol* 1996 Feb;8(2):177-82.
- [73] van de Wal Y, Kooy YM, Drijfhout JW, et al. Peptide binding characteristics of the coeliac disease-associated DQ(alpha1*0501, beta1*0201) molecule. *Immunogenetics* 1996;44(4):246-53.
- [74] Vartdal F, Johansen BH, Friede T, et al. The peptide binding motif of the disease associated HLA-DQ (alpha 1* 0501, beta 1* 0201) molecule. *Eur J Immunol* 1996 Nov;26(11):2764-72.
- [75] Kwok WW, Domeier ML, Raymond FC, et al. Allele-specific motifs characterize HLA-DQ interactions with a diabetes-associated peptide derived from glutamic acid decarboxylase. *J Immunol* 1996 Mar 15;156(6):2171-7.
- [76] Godkin A, Friede T, Davenport M, et al. Use of eluted peptide sequence data to identify the binding characteristics of peptides to the insulin-dependent diabetes susceptibility allele HLA-DQ8 (DQ 3.2). *Int Immunol* 1997 Jun;9(6):905-11.
- [77] Shan L, Molberg O, Parrot I, et al. Structural basis for gluten intolerance in celiac sprue. *Science* 2002 Sep 27;297(5590):2275-9.

- [78] Bethune MT, Siegel M, Howles-Banerji S, et al. Interferon-gamma released by gluten-stimulated celiac disease-specific intestinal T cells enhances the transepithelial flux of gluten peptides. *J Pharmacol Exp Ther* 2009 May;329(2):657-68.
- [79] Matysiak-Budnik T, Moura IC, Arcos-Fajardo M, et al. Secretory IgA mediates retrotranscytosis of intact gliadin peptides via the transferrin receptor in celiac disease. *J Exp Med* 2008 Jan 21;205(1):143-54.
- [80] Matysiak-Budnik T, Candalh C, Dugave C, et al. Alterations of the intestinal transport and processing of gliadin peptides in celiac disease. *Gastroenterology* 2003 Sep;125(3):696-707.
- [81] Nilsen EM, Lundin KE, Krajci P, et al. Gluten specific, HLA-DQ restricted T cells from coeliac mucosa produce cytokines with Th1 or Th0 profile dominated by interferon gamma. *Gut* 1995 Dec;37(6):766-76.
- [82] van de Wal Y, Kooy YM, van Veelen PA, et al. Small intestinal T cells of celiac disease patients recognize a natural pepsin fragment of gliadin. *Proc Natl Acad Sci U S A* 1998 Aug 18;95(17):10050-4.
- [83] Sjostrom H, Lundin KE, Molberg O, et al. Identification of a gliadin T-cell epitope in coeliac disease: general importance of gliadin deamidation for intestinal T-cell recognition. *Scand J Immunol* 1998 Aug;48(2):111-5.
- [84] van de Wal Y, Kooy YM, van Veelen P, et al. Glutenin is involved in the gluten-driven mucosal T cell response. *Eur J Immunol* 1999 Oct;29(10):3133-9.
- [85] Arentz-Hansen H, Korner R, Molberg O, et al. The intestinal T cell response to alpha-gliadin in adult celiac disease is focused on a single deamidated glutamine targeted by tissue transglutaminase. *J Exp Med* 2000 Feb 21;191(4):603-12.
- [86] Arentz-Hansen H, McAdam SN, Molberg O, et al. Celiac lesion T cells recognize epitopes that cluster in regions of gliadins rich in proline residues. *Gastroenterology* 2002 Sep;123(3):803-9.
- [87] Vader W, Kooy Y, van Veelen P, et al. The gluten response in children with celiac disease is directed toward multiple gliadin and glutenin peptides. *Gastroenterology* 2002 Jun;122(7):1729-37.
- [88] Qiao SW, Bergseng E, Molberg O, et al. Refining the rules of gliadin T cell epitope binding to the disease-associated DQ2 molecule in celiac disease: importance of proline spacing and glutamine deamidation. *J Immunol* 2005 Jul 1;175(1):254-61.
- [89] Tollefsen S, Arentz-Hansen H, Fleckenstein B, et al. HLA-DQ2 and -DQ8 signatures of gluten T cell epitopes in celiac disease. *J Clin Invest* 2006 Aug;116(8):2226-36.
- [90] Camarca A, Anderson RP, Mamone G, et al. Intestinal T cell responses to gluten peptides are largely heterogeneous: implications for a peptide-based therapy in celiac disease. *J Immunol* 2009 Apr 1;182(7):4158-66.
- [91] Hue S, Mention JJ, Monteiro RC, et al. A direct role for NKG2D/MICA interaction in villous atrophy during celiac disease. *Immunity* 2004 Sep;21(3):367-77.
- [92] Fallang LE, Bergseng E, Hotta K, et al. Differences in the risk of celiac disease associated with HLA-DQ2.5 or HLA-DQ2.2 are related to sustained gluten antigen presentation. *Nat Immunol* 2009 Oct;10(10):1096-101.
- [93] Molberg O, McAdam SN, Korner R, et al. Tissue transglutaminase selectively modifies gliadin peptides that are recognized by gut-derived T cells in celiac disease. *Nat Med* 1998 Jun;4(6):713-7.
- [94] Molberg O, McAdam S, Lundin KE, et al. T cells from celiac disease lesions recognize gliadin epitopes deamidated in situ by endogenous tissue transglutaminase. *Eur J Immunol* 2001 May;31(5):1317-23.

- [95] Siegel M, Khosla C. Transglutaminase 2 inhibitors and their therapeutic role in disease states. *Pharmacol Ther* 2007 May 13.
- [96] Korponay-Szabo IR, Halttunen T, Szalai Z, et al. In vivo targeting of intestinal and extraintestinal transglutaminase 2 by coeliac autoantibodies. *Gut* 2004 May;53(5):641-8.
- [97] Sardy M, Odenthal U, Karpati S, et al. Recombinant human tissue transglutaminase ELISA for the diagnosis of gluten-sensitive enteropathy. *Clin Chem* 1999 Dec;45(12):2142-9.
- [98] Sollid LM, Molberg O, McAdam S, et al. Autoantibodies in coeliac disease: tissue transglutaminase - guilt by association? *Gut* 1997 Dec;41(6):851-2.
- [99] Piper JL, Gray GM, Khosla C. High selectivity of human tissue transglutaminase for immunoactive gliadin peptides: implications for celiac sprue. *Biochemistry* 2002 Jan 8;41(1):386-93.
- [100] Fleckenstein B, Qiao SW, Larsen MR, et al. Molecular characterization of covalent complexes between tissue transglutaminase and gliadin peptides. *J Biol Chem* 2004 Apr 23;279(17):17607-16.
- [101] Esposito C, Paparo F, Caputo I, et al. Anti-tissue transglutaminase antibodies from coeliac patients inhibit transglutaminase activity both in vitro and in situ. *Gut* 2002 Aug;51(2):177-81.
- [102] Dieterich W, Trapp D, Esslinger B, et al. Autoantibodies of patients with coeliac disease are insufficient to block tissue transglutaminase activity. *Gut* 2003 Nov;52(11):1562-6.
- [103] Kiraly R, Vecsei Z, Demenyi T, et al. Coeliac autoantibodies can enhance transamidating and inhibit GTPase activity of tissue transglutaminase: dependence on reaction environment and enzyme fitness. *J Autoimmun* 2006 Jun;26(4):278-87.
- [104] Trier JS. Diagnosis of celiac sprue. *Gastroenterology* 1998 Jul;115(1):211-6.
- [105] Sardy M, Karpati S, Merkl B, et al. Epidermal transglutaminase (TGase 3) is the autoantigen of dermatitis herpetiformis. *J Exp Med* 2002 Mar 18;195(6):747-57.
- [106] Hadjivassiliou M, Aeschlimann P, Strigun A, et al. Autoantibodies in gluten ataxia recognize a novel neuronal transglutaminase. *Ann Neurol* 2008 Sep;64(3):332-43.
- [107] Stepniak D, Wiesner M, de Ru AH, et al. Large-Scale Characterization of Natural Ligands Explains the Unique Gluten-Binding Properties of HLA-DQ2. *J Immunol* 2008 Mar 1;180(5):3268-78.
- [108] Stulik J, Hemychova L, Porkertova S, et al. Identification of new celiac disease autoantigens using proteomic analysis. *Proteomics* 2003 Jun;3(6):951-6.
- [109] van de Wal Y, Kooy Y, van Veelen PA, et al. Selective deamidation by tissue transglutaminase strongly enhances gliadin-specific T cell reactivity. *J Immunol* 1998 Aug 15;161(4):1585-8.
- [110] Marti T, Molberg O, Li Q, et al. Prolyl endopeptidase-mediated destruction of T cell epitopes in whole gluten: chemical and immunological characterization. *J Pharmacol Exp Ther* 2005 Jan;312(1):19-26.
- [111] Shan L, Qiao SW, Arentz-Hansen H, et al. Identification and analysis of multivalent proteolytically resistant peptides from gluten: implications for celiac sprue. *J Proteome Res* 2005 Sep;4(5):1732-41.
- [112] Cunsolo V, Foti S, Saletti R. Mass spectrometry in the characterization of cereal seed proteins. *Eur J Mass Spectrom (Chichester, Eng)* 2004;10(3):359-70.
- [113] Ferranti P, Mamone G, Picariello G, et al. Mass spectrometry analysis of gliadins in celiac disease. *J Mass Spectrom* 2007 Dec;42(12):1531-48.

- [114] Day N, Keillor JW. A continuous spectrophotometric linked enzyme assay for transglutaminase activity. *Anal Biochem* 1999 Oct 1;274(1):141-4.
- [115] Robinson NE, Robinson AB. Deamidation of human proteins. *Proc Natl Acad Sci U S A* 2001 Oct 23;98(22):12409-13.
- [116] Dewar DH, Amato M, Ellis HJ, et al. The toxicity of high molecular weight glutenin subunits of wheat to patients with coeliac disease. *Eur J Gastroenterol Hepatol* 2006 May;18(5):483-91.
- [117] Mazzeo MF, De Giulio B, Senger S, et al. Identification of transglutaminase-mediated deamidation sites in a recombinant alpha-gliadin by advanced mass-spectrometric methodologies. *Protein Sci* 2003 Nov;12(11):2434-42.
- [118] Mamone G, Ferranti P, Melck D, et al. Susceptibility to transglutaminase of gliadin peptides predicted by a mass spectrometry-based assay. *FEBS Lett* 2004 Mar 26;562(1-3):177-82.
- [119] Evans DF, Pye G, Bramley R, et al. Measurement of gastrointestinal pH profiles in normal ambulant human subjects. *Gut* 1988 Aug;29(8):1035-41.
- [120] Sollid LM. Coeliac disease: dissecting a complex inflammatory disorder. *Nat Rev Immunol* 2002 Sep;2(9):647-55.
- [121] De Laurenzi V, Melino G. Gene disruption of tissue transglutaminase. *Mol Cell Biol* 2001 Jan;21(1):148-55.
- [122] Nanda N, Iismaa SE, Owens WA, et al. Targeted inactivation of Gh/tissue transglutaminase II. *J Biol Chem* 2001 Jun 8;276(23):20673-8.
- [123] Maiuri L, Ciacci C, Ricciardelli I, et al. Unexpected role of surface transglutaminase type II in celiac disease. *Gastroenterology* 2005 Nov;129(5):1400-13.
- [124] Verderio E, Nicholas B, Gross S, et al. Regulated expression of tissue transglutaminase in Swiss 3T3 fibroblasts: effects on the processing of fibronectin, cell attachment, and cell death. *Exp Cell Res* 1998 Feb 25;239(1):119-38.
- [125] Karpuj MV, Becher MW, Springer JE, et al. Prolonged survival and decreased abnormal movements in transgenic model of Huntington disease, with administration of the transglutaminase inhibitor cystamine. *Nat Med* 2002 Feb;8(2):143-9.
- [126] Esposito C, Caputo I, Troncone R. New therapeutic strategies for coeliac disease: tissue transglutaminase as a target. *Curr Med Chem* 2007;14(24):2572-80.
- [127] Aeschlimann D, Paulsson M. Transglutaminases: protein cross-linking enzymes in tissues and body fluids. *Thromb Haemost* 1994 Apr;71(4):402-15.
- [128] Lai TS, Slaughter TF, Peoples KA, et al. Regulation of human tissue transglutaminase function by magnesium-nucleotide complexes. Identification of distinct binding sites for Mg-GTP and Mg-ATP. *J Biol Chem* 1998 Jan 16;273(3):1776-81.
- [129] Lorand L, Conrad SM. Transglutaminases. *Mol Cell Biochem* 1984;58(1-2):9-35.
- [130] Choi K, Siegel M, Piper JL, et al. Chemistry and biology of dihydroisoxazole derivatives: selective inhibitors of human transglutaminase 2. *Chem Biol* 2005 Apr;12(4):469-75.
- [131] Szondy Z, Sarang Z, Molnar P, et al. Transglutaminase 2^{-/-} mice reveal a phagocytosis-associated crosstalk between macrophages and apoptotic cells. *Proc Natl Acad Sci U S A* 2003 Jun 24;100(13):7812-7.
- [132] Ehren J, Moron B, Martin E, et al. A food-grade enzyme preparation with modest gluten detoxification properties. *PLoS ONE* 2009;4(7):e6313.
- [133] Gass J, Bethune MT, Siegel M, Set al. Combination enzyme therapy for gastric digestion of dietary gluten in patients with celiac sprue. *Gastroenterology* 2007 Aug;133(2):472-80.

- [134] Siegel M, Bethune MT, Gass J, et al. Rational design of combination enzyme therapy for celiac sprue. *Chem Biol* 2006 Jun;13(6):649-58.
- [135] Stepniak D, Spaenij-Dekking L, Mitea C, et al. Highly efficient gluten degradation with a newly identified prolyl endoprotease: implications for celiac disease. *Am J Physiol Gastrointest Liver Physiol* 2006 Oct;291(4):G621-G629.
- [136] Mitea C, Havenaar R, Drijfhout JW, et al. Efficient degradation of gluten by a prolyl endoprotease in a gastrointestinal model: implications for coeliac disease. *Gut* 2008 Jan;57(1):25-32.
- [137] Gianfrani C, Siciliano RA, Facchiano AM, et al. Transamidation of wheat flour inhibits the response to gliadin of intestinal T cells in celiac disease. *Gastroenterology* 2007 Sep;133(3):780-9.
- [138] Rossi M, Salvati VM, Mazzarella G. Reintroduction of gluten following wheat flour transamidation in treated adult celiac patients. 13th International Coeliac Disease Symposium 2009. Ref Type: Abstract
- [139] Siegel M, Xia J, Khosla C. Structure-based design of alpha-amido aldehyde containing gluten peptide analogues as modulators of HLA-DQ2 and transglutaminase 2. *Bioorg Med Chem* 2007 Sep 15;15(18):6253-61.
- [140] Xia J, Siegel M, Bergsens E, et al. Inhibition of HLA-DQ2-mediated antigen presentation by analogues of a high affinity 33-residue peptide from alpha2-gliadin. *J Am Chem Soc* 2006 Feb 15;128(6):1859-67.
- [141] Xia J, Bergsens E, Fleckenstein B, et al. Cyclic and dimeric gluten peptide analogues inhibiting DQ2-mediated antigen presentation in celiac disease. *Bioorg Med Chem* 2007 Oct 15;15(20):6565-73.
- [142] Juse U, van de Wal Y, Koning F, et al. Design of new high-affinity peptide ligands for human leukocyte antigen-DQ2 using a positional scanning peptide library. *Hum Immunol* 2010 May;71(5):475-81.
- [143] Verhoef A, Alexander C, Kay AB, et al. T cell epitope immunotherapy induces a CD4+ T cell population with regulatory activity. *PLoS Med* 2005 Mar;2(3):e78.
- [144] Onisko B, Dynin I, Requena JR, et al. Mass spectrometric detection of attomole amounts of the prion protein by nanoLC/MS/MS. *J Am Soc Mass Spectrom* 2007 Jun;18(6):1070-9.
- [145] Thompson T, Mendez E. Commercial assays to assess gluten content of gluten-free foods: why they are not created equal. *J Am Diet Assoc* 2008 Oct;108(10):1682-7.
- [146] Mamone G, Ferranti P, Chianese L, et al. Qualitative and quantitative analysis of wheat gluten proteins by liquid chromatography and electrospray mass spectrometry. *Rapid Commun Mass Spectrom* 2000;14(10):897-904.
- [147] Sealey-Voyksner JA, Khosla C, Voyksner RD, et al. Novel aspects of quantitation of immunogenic wheat gluten peptides by liquid chromatography-mass spectrometry/mass spectrometry. *J Chromatogr A* 2010 Feb 1.

The Preferred Substrates for Transglutaminase 2 in a Complex Wheat Gluten Digest are Peptide Fragments Harboring Celiac Disease T-cell Epitopes

Siri Dørum¹, Magnus Ø. Arntzen^{2,3,4}, Shuo-Wang Qiao¹, Anders Holm⁵, Christian J. Koehler², Bernd Thiede², Ludvig M. Sollid^{1,5} & Burkhard Fleckenstein^{1,3}

¹ *Centre for Immune Regulation, Institute of Immunology, University of Oslo*

² *The Biotechnology Centre of Oslo, University of Oslo*

³ *Proteomics Core Facility at Oslo University Hospital-Rikshospitalet*

⁴ *Proteomics Core Facility, Norwegian University of Life Sciences, Ås*

⁵ *Centre for Immune Regulation, Institute of Immunology, Oslo University Hospital-Rikshospitalet*

* Corresponding author: Siri Dørum, Centre for Immune Regulation, Institute of Immunology, University of Oslo, N-0027 Oslo, Norway, Tel. +47 23074214, E-mail: siri.dorum@rr-research.no

Abstract

Background: Celiac disease is a T-cell mediated chronic inflammatory disorder of the gut that is induced by dietary exposure to gluten proteins. CD4+ T cells of the intestinal lesion recognize gluten peptides in the context of HLA-DQ2.5 or HLA-DQ8 and the gluten derived peptides become better T-cell antigens after deamidation catalyzed by the enzyme transglutaminase 2 (TG2). In this study we aimed to identify the preferred peptide substrates of TG2 in a heterogeneous proteolytic digest of whole wheat gluten.

Methods: A method was established to enrich for preferred TG2 substrates in a complex gluten peptide mixture by tagging with 5-biotinamido-pentylamine. Tagged peptides were isolated and then identified by nano-liquid chromatography online-coupled to tandem mass spectrometry, database searching and final manual data validation.

Results: We identified 31 different peptides as preferred substrates of TG2. Strikingly, the majority of these peptides were harboring known gluten T-cell epitopes. In addition, two TG2 peptide substrates were predicted to bind to HLA-DQ2.5 but did not contain previously characterized sequences of T-cell epitopes. Both these peptides elicited T-cell responses when tested for recognition by intestinal T-cell lines of celiac disease patients, and they thus contain novel T-cell epitopes. We also found that the intact 9mer core sequences of the respective epitopes were not present in all peptide substrates. Interestingly, those epitopes that were represented by intact forms were frequently recognized by T cells in celiac disease patients, whereas those that were present in truncated versions were infrequently recognized.

Conclusion: TG2 as well as gastrointestinal proteolysis play important roles in the selection of gluten T-cell epitopes in celiac disease.

Introduction

Celiac disease is an intestinal chronic inflammatory disorder elicited by dietary wheat gluten and related proteins from barley and rye. The disease is characterized by malabsorption, flattening of mucosa and the presence of autoantibodies. The large majority of the celiac disease patients express the HLA-DQ2.5 and/or HLA-DQ8 molecules [1], and gluten specific T-cells restricted to these HLA-molecules can be isolated from intestinal biopsies of the patients [2,3]. The peptide binding motifs of HLA-DQ2.5 and HLA-DQ8 have been thoroughly investigated and both molecules show a preference for negative charges at certain anchor positions; DQ2.5 in positions P4, P6 and P7 [4-6] and DQ8 in positions P1 and P9 [7,8]. Interestingly, these restricted T cells recognize gluten peptides only or preferentially after they undergo deamidation, a modification which is catalyzed by the enzyme transglutaminase 2 (TG2).

TG2 is a calcium-dependent enzyme which targets specific glutamine residues in peptides and proteins for either transamidation (crosslinking) or deamidation [9]. In the transamidation reaction, a stable isopeptide bond is formed between a peptidylglutamine residue (acyl-donor) and the amino group of an acyl-acceptor, e.g. peptidyllysine or a primary amine molecule. Deamidation of the targeted glutamine residue results in conversion to a glutamic acid and introduces a negative charge. It has been demonstrated that both for the deamidation [10] and transamidation reactions [11] the targeting of glutamine (Q) residues in peptides is strongly influenced by the positioning of C-terminally located proline (P) residues. Whereas a Q residue in the QXP consensus sequence is targeted by TG2, Q residues in a QP or QXXP sequence motif are not.

During the recent years several T-cell epitopes of wheat gluten have been identified [12-19]. The majority of these epitopes derive from the gliadin protein fraction of gluten and is presented to T cells in the context of HLA-DQ2.5. Notably, the gliadin epitopes often cluster in regions rich in proline residues. Conceivably, the selection of the T-cell epitopes is governed by several factors. T-cell epitopes are usually longer than 9 amino acids, and protection against complete digestion in the gastrointestinal tract imposed by proline residues seems

important [20]. Further, a selective force is exerted by TG2 as these epitopes are dependent on deamidation by this enzyme. Finally, epitope selection by HLA seems of importance as both HLA-DQ2.5 and HLA-DQ8 prefer binding of peptides with negatively charged anchor residues. Importantly, all these selective forces will act in concert.

Several recent studies have pointed towards a role of TG2 in the selection of T-cell epitopes in celiac disease. Using a set of synthetic overlapping peptides covering the whole sequence of a γ -gliadin protein, those peptides that were quickly deamidated by TG2 were also recognized by T-cell lines of celiac disease patients [11]. We have also demonstrated that the known HLA-DQ2.5-restricted gliadin epitopes are substrates for TG2, but the rate by which this modification occurs differs considerably between the peptides [21]. We found a correlation between the rate of deamidation of the different epitopes and their T cell immunostimulatory capacity. Finally, the finding that proline governs the specificity of the enzyme and that gluten T-cell epitopes are rich in proline residues further supports the notion of a selective force exerted by TG2.

In this study we aimed to shed further light on the role of TG2 in the selection of gluten T-cell epitopes. Gluten is an extreme complex mixture of diverse proteins. Several hundred distinct gluten proteins are belonging to the gliadin and glutenin fractions of a single wheat variety, and therefore tens of thousands different peptides will be present in a digest of wheat flour. To mimic the selection of T-cell epitopes from this vast sea of peptides that is present in the gut after gluten ingestion, we have developed a method to identify the preferred peptide substrates for TG2 in a proteolytic digest of whole gluten. Interestingly, we found that the great majority of preferred substrates for TG2 from this heterogeneous mixture are indeed peptides that harbor known gluten T-cell epitopes.

Materials and Methods

Synthesis of peptides and purification of recombinant human TG2

Synthetic peptides were purchased from GL Biochem Ltd (Shanghai, China). Recombinant human TG2 was expressed in *Escherichia coli* with an N-terminal hexa-histidine tag and purified as described previously [22] with some minor modifications [23].

Proteolysis of gluten

Whole gluten (Bob's Red Mill, Milwaukie, Oregon) digested with pepsin, trypsin, chymotrypsin, elastase and carboxypeptidase (PTCEC gluten) was a gift from M. Bethune (Stanford University) and was prepared as follows: A solution of 15 mg/ml gluten was dissolved in 0.01 N HCL (pH ~2.1) and was incubated with 0.6 mg/ml pepsin (American Laboratories) for 60 min at 37°C. The digest was adjusted to pH 6.5 with 50 mM sodium phosphate buffer and incubated with 0.375 mg/ml trypsin (Sigma), 0.375 mg/ml chymotrypsin (Sigma), 0.075 mg/ml elastase (Sigma) and 0.075 mg/ml carboxypeptidase A (Sigma) for 120 min at 37°C. The sample was heated to 95°C for 5 min to stop proteolysis followed by centrifugation and filtration of the supernatants.

Validation of the enrichment procedure using synthetic peptides

To test the specificity of the enrichment procedure, 20 µM of the synthetic peptides DQ2- α -I (QLQFPQPQLPY) and DQ2- γ -II (GIIQPQQPAQL) was separately mixed with 20 µM 5-biotinamido-pentylamine (5-BP) in 10 µl of 100 mM Tris/HCl pH 7.4 supplemented with 2 mM CaCl₂. Both samples were incubated with 0.1 µg/µl TG2 at 37°C for 60 min before TG2 was inactivated by adding 1 µl iodoacetamide to a final concentration of 25 mM. A volume of 1 µl of each of the samples was added to 20 µl of 1 mg/ml PTCEC gluten digest. The sample was incubated with 200 µg Dynabeads M270 streptavidin (capacity: 200 pmol biotinylated peptides/1 mg beads) (Invitrogen DYNAL AS, Norway) for 30 min in PBS/0.1% SDS followed by three washing steps each with 100 µl PBS/0.1% SDS and 100 µl 20% acetonitrile/water. For elution, beads were incubated with 15 µl 70% acetonitrile/2% formic acid/0.2 mM biotin for 30 min at 37°C and the eluate was analyzed by MALDI-TOF mass spectrometry (Ultraflex II, Bruker Daltonics, Bremen, Germany).

In order to rank the peptides according to their level of transamidation, a sample containing 20 μM of each of the peptide epitopes DQ2- α -II (PQPQLPYPQPQLPY), DQ2- γ -II (GIIQPQQPAQL), DQ2- γ -III (FPQQPQQYPYQQP) and DQ2- γ -IV (FSQPQQQFPQPQ) was incubated with 0.1 $\mu\text{g}/\mu\text{l}$ TG2 and 20 μM 5-BP in 20 μl 100 mM Tris/HCl pH 7.4 with 2 mM CaCl_2 at 37°C. Aliquots of 5 μl were removed after 1, 5 and 15 min incubation time and 1 μl iodoacetamide (final concentration 25 mM) was added to inactivate TG2. For enrichment, 1 μl of each of the samples was incubated with 150 μg Dynabeads M270 streptavidin for 30 min in PBS/0.1% SDS. Washing of the beads and elution of the enriched biotinylated peptides was performed as described above. MALDI-TOF spectra were acquired for the eluates of enriched transamidated peptides.

Further, 30 μM of a chymotryptic digest of α 2-gliadin was incubated with 0.1 $\mu\text{g}/\mu\text{l}$ TG2 in 25 μl 100 mM Tris/HCl pH 7.4 supplemented with 2 mM CaCl_2 at 37°C for 60 min and the reaction was terminated by adding 1 μl iodoacetamide to a final concentration of 25 mM. The sample was purified by C_{18} ZipTips (Millipore, Billerica, MA, USA), the eluate was vacuum-dried, and the peptides were redissolved in 0.1% formic acid. Next, 300 fmol of the sample was analyzed by nano-LC coupled to a quadrupole-time-of-flight (Q-TOF) mass spectrometer (settings as described below). Similarly, 42 μM of a chymotryptic digest of α 2-gliadin was incubated under identical conditions in the presence of 20 μM 5-BP. Transamidated peptides were enriched by incubating with 150 μg Dynabeads M270 streptavidin for 30 min in PBS/0.1% SDS followed by washing of the beads as described above. Enriched peptides were eluted by adding 10 μl of 70% acetonitrile/2% formic acid/0.2 mM biotin and incubating for 30 min at 37°C. Finally, MALDI-TOF spectra were acquired.

Treatment of a PTCEC gluten digest with TG2 and enrichment of transamidated gluten peptides

A sample containing 1 mg/ml PTCEC gluten digest, 0.1 $\mu\text{g}/\mu\text{l}$ TG2 and 200 pmol 5-BP in 100 mM Tris/HCl pH 7.4 supplemented with 2 mM CaCl_2 (total volume of 20 μl) was incubated at 37°C for 30 min. After inactivation of TG2 by adding iodoacetamide (final concentration of 25

mM), the sample was incubated with 1 mg Dynabeads M270 streptavidin for 30 min in PBS/0.1% SDS and the beads were washed as described above. Enriched transamidated and biotinylated peptides were eluted from the beads by adding 25 μ l of 70% acetonitrile/2% formic acid/0.2 mM biotin and incubating for 30 min at 37°C. Finally, the eluate was vacuum-dried.

Analysis by nano-LC tandem mass spectrometry

The transamidated peptides enriched from PTCEC gluten were analyzed either by a Q-TOF hybrid mass spectrometer (MicroToF-q, Bruker Daltonics, Bremen, Germany) or an Orbitrap mass spectrometer (LTQ Orbitrap XL, Thermo Scientific, Bremen, Germany), both coupled to a nano-LC system. For nano-LC-Q-TOF analysis, dried samples were dissolved in 8 μ l 0.1% trifluoroacetic acid and a volume of 6 μ l was injected into an Agilent 1100 series nano-LC system (Agilent Technologies, Palo Alto, CA). Peptides were separated on an analytical column (150 mm \times 0.075 mm) packed with 100 Å C18 3.5 μ m particles (G&T Septeck AS, Norway). A linear gradient of 2-60% solvent B in 60 min was applied with a flow rate of 300 nL/min (solvent A: 0.1 % acetic acid/water, solvent B: 0.1% acetic acid/acetonitrile). The Q-TOF was operated in the data-dependent mode to automatically switch between MS and MS/MS acquisition. Collision energies were set between 30-50 eV depending on the charge state of the precursor ions. Data were acquired using microTOFControl v2.0 and processed using DataAnalysis v3.4. For nano-LC-Orbitrap analysis, the peptides were purified by C₁₈ ZipTips (Millipore, Billerica, MA, USA), vacuum-dried and redissolved in 10 μ l of 1% formic acid, and then 3 μ l were injected into an Ultimate 3000 nanoLC system (Dionex, Sunnyvale, CA, USA). For separation of peptides, an Acclaim PepMap 100 column (120 mm \times 0.075 mm) packed with 100 Å C18 3 μ m particles (Dionex) was used. A flow rate of 300 nL/min was employed with a solvent gradient of 7-35% B in 77 min and then from 35% to 50% B in 10 min (solvent A: 0.1% formic acid/water; solvent B: 0.1% formic acid/90% acetonitrile). The mass spectrometer was operated in the data-dependent mode to automatically switch between Orbitrap-MS and LTQ-MS/MS acquisition. Data were acquired using Xcalibur v2.5.

Identification of enriched transamidated gluten peptides

First, a database consisting of all known protein sequences derived from *Triticum aestivum* was built by extracting entries (n=4654) from the UniprotKB database (release 15.0) at the European Bioinformatics Institute (EBI) using the SRS server. The LC-MS/MS data were searched against this database using the Mascot search engine and Proteome Discoverer software version 1.0 (Thermo Fisher Scientific Inc., Waltham, MA). In Mascot the glutamine modification by 5-BP linkage (mass increase of 311.17 Da) was implemented and selected as variable modification together with deamidation of glutamine and asparagine and pyroglutamate formation. For analysis of LTQ-Orbitrap data, the mass tolerance was set as 10 ppm for the precursor ion and 0.6 Da for the fragment ions, whereas for analysis of Q-TOF data the mass tolerance was set as 0.5 Da for the precursor ion and 0.2 Da for the fragments ions. All reported hits were in addition manually interpreted. If the glutamine residues targeted by TG2 could not be unambiguously assigned based on the MS/MS spectrum, also the established rules governing the targeting of glutamine residues by TG2 were considered.^{10, 11} In a few cases, extracted ion chromatograms were generated to distinguish peptides that carry both transamidated and deamidated glutamine residues from peptides harboring only a transamidated glutamine residue.

T cell assay

The T-cell stimulatory capacity of the synthetic TG2-treated peptides SHQQQPFQYPQQPYPS (peptide #25) and PHQPQQVQPQQQPF (peptide #19) were tested in proliferation assays together with a panel of synthetic epitope peptides using T-cell lines derived from celiac disease patients as previously described.¹⁶ First, both peptides were incubated at 100 μ M with 0.1 μ g/ μ l TG2 in 100 mM Tris/HCl pH 7.4 buffer supplemented with 2 mM CaCl_2 for 120 min at 37°C. Sixty thousand HLA-DQ2 homozygous EBV-transformed B lymphoblastoid cells were irradiated (75 Gy) and incubated overnight with different concentrations of the TG2-treated peptides. On day 1, 40 000 cells from polyclonal T-cell lines generated from intestinal biopsies of celiac disease patients were added. One μ Ci ³H-thymidine was added per well on day 3 and T-cell proliferation was measured as thymidine incorporation on day 4.

Results

Validation of the procedure to identify the best TG2 substrates

Incubation of a PTCEC digest of whole wheat gluten with small amounts of the primary amine 5-BP and TG2 will result in transamidation of a limited number of gluten peptides. In order to detect and identify these TG2 peptide substrates in such a heterogeneous peptide mixture, transamidated and hence biotinylated peptides were enriched using magnetic streptavidin beads (Fig. 1). First, we tested the specificity of the enrichment method for transamidated peptides. Two synthetic peptides known to be substrates for TG2, DQ2- α -I and DQ2- γ -II, were each incubated with TG2 in the presence of 5-BP. Both samples were spiked into a PTCEC digest of gluten and streptavidin beads were added. Analysis by MALDI-TOF mass spectrometry before and after enrichment demonstrated a high specificity of the enrichment procedure as only the transamidated DQ2- α -I and DQ2- α -II peptides were detected after elution from streptavidin beads (Fig. 2). The transamidated peptides were not observed prior to enrichment.

We next tested whether the deamidation and transamidation reactions catalyzed by TG2 prefer the same gluten peptides. Previously we have shown that epitope peptides DQ2- α -II and DQ2- γ -II undergo rapid deamidation by TG2, whereas epitope peptides DQ2- γ -III and DQ2- γ -IV are poorly deamidated [21]. Equal amounts of these four peptides were mixed with a limited amount of 5-BP and incubated with TG2 for different periods of time. The transamidated peptides were enriched from the samples and MALDI-TOF spectra were acquired. The relative signal intensities of the transamidated DQ2- α -II and DQ2- γ -II epitope peptides increased much faster over time than those of the DQ2- γ -III and DQ2- γ -IV epitope peptides suggesting that the former are better substrates of TG2 (data not shown). This finding is consistent with previously reported data for deamidation of these four peptides [21]. Further, we tested which peptides were deamidated and transamidated from a heterogeneous mixture of peptides derived from digestion of a single recombinant α 2-gliadin protein by chymotrypsin. For both reactions, the 33mer peptide (Table 1) and the peptide VRVPVQLQPQNPSQQQPQEQVPL were identified as substrates. Thus we concluded that enrichment of gluten peptides by means of crosslinking with 5-BP can be used to provide

information as to which peptides in a heterogeneous mixture are preferred also for the deamidation reaction. The method can also determine which glutamine residues within the preferred substrates are targeted by the enzyme.

Identification of enriched gluten peptides by mass spectrometry

In order to identify optimal TG2 gluten peptide substrates, a PTCEC digest of whole gluten was incubated with TG2 in the presence of limiting amounts of 5-BP for 30 minutes. After specific enrichment of transamidated gluten peptides by magnetic streptavidin beads, eluted peptides were identified by LC-MS/MS analysis and subsequent database searching (Fig. 1). Altogether, 31 different peptide sequences were identified from the enriched sample of transamidated gluten peptides (Table 2). The targeted glutamine (Q) residues in these peptides were typically in the QXP-motif which is in accordance with previous results regarding TG2 specificity [10,11]. Remarkably, the majority of the identified peptides contain complete or truncated versions of known gluten T-cell epitopes (Tables 1 and 2). Among the identified nine α -gliadin derived peptides, five peptides harbor at least two of the DQ2- α -I, DQ2- α -II or DQ2- α -III epitopes. The DQ2- α -II epitope is present in all of the five peptides in one or two copies (peptides #1-5), while the DQ2- α -I and DQ2- α -III epitopes are present in peptides #3-5 and peptides #1-3, respectively. The α -gliadin derived peptide #6 harbors an incomplete DQ2- α 20-epitope. The α -gliadin derived 33mer peptide that is known to be a superior TG2 substrate was not identified. However, several isoforms (peptides #3 and #5) and truncated versions (peptides #1, #2, and #4) of this peptide were observed. The α -gliadin derived peptides #7-9 do not harbor any known T-cell epitopes. Peptide #7 corresponds to the p31-44 "toxic" peptide previously reported [24].

Ten peptides derived from γ -gliadin were identified, and eight of these peptides contain parts of the DQ2- γ -II epitope (peptides #10-17). Notably, these peptides share the same peptide core region but differ in their N- and C-terminal extensions. The DQ2- γ -VIIb/DQ8- γ -I epitope was present in one of the identified γ -derived peptides (peptide #18).

Twelve identified peptides derive from glutenin proteins. The complete DQ2-restricted glutenin-17 epitope, Glt-17, was present in three of these peptides (peptides #21, #22 and #24) while two of the peptides (peptides #20 and #23) harbored truncated versions of the epitope. Interestingly, several of the other identified TG2 peptide substrates of glutenin proteins have sequences similar to known epitopes.

Identification of two novel T-cell epitopes

For some of the identified TG2 peptide substrates that do not contain any known T-cell epitopes, sequence motifs and targeting of Q residues suggested that they could bind to HLA-DQ2.5. In an attempt to identify new T-cell epitopes, we therefore tested two of the peptides for recognition by gluten reactive, polyclonal T-cell lines generated from intestinal biopsies of HLA-DQ2.5-positive celiac disease patients. Peptide #19 (PHQPQQQVPQP**Q**QPPF, predicted HLA-DQ2.5 binding register underlined, Q residue targeted by TG2 in bold) and peptide #25 (SHQQPFP**Q**Q*YP***Q***Q*PYPS, the two predicted overlapping HLA-DQ2.5 binding registers are either underlined or in italics, Q residues targeted by TG2 in bold) were treated with TG2 and tested together with a panel of known gluten epitopes in a proliferation assay. Six of the seven gluten reactive, polyclonal T-cell lines tested did not recognize peptide #19 or peptide #25 (data not shown). However, the T-cell line BW.CD-E made a strong response to peptide #25 and a weaker response to peptide #19 in a TG2-dependent and dose-dependent fashion (Fig. 3). Except for epitopes DQ2- α -I and the DQ2- ω -17mer peptide [25], none of the other peptides was recognized by this particular T-cell line. Interestingly, no T-cell response was observed for the DQ2- γ -VI epitope which has a very similar sequence as peptide #25, indicating that there is no cross-reactivity between the new peptides and the DQ2- γ -VI epitope in this T-cell line.

Transamidation and deamidation of distinct glutamine residues within the same peptide

Four α -gliadin and one γ -gliadin derived TG2 peptide substrates were observed as derivatives carrying both a transamidated and a deamidated glutamine residue. For peptide #1 (PQPQLPYPQPQLPYPQPFP), transamidation and deamidation sites could be assigned to

Q11 and Q4 (underlined), respectively. A transamidated and deamidated derivative was also observed for an elongated version of peptide #1. In this 25mer α -gliadin peptide (peptide #3, QLQPFQQQLPYPQQLPYPQPQPF, targeted Q residues underlined), it could however not be unambiguously determined which of the two targeted glutamine residues (Q9 and Q16) was transamidated or deamidated. For peptide #8 (LQPQNPSQQQPQEQVPL), Q14 was found to be transamidated whereas Q9 was deamidated. These modifications were observed for the same glutamine residues in peptide #9 which represents an N-terminally extended derivative of peptide #8. The γ -gliadin derived peptide VQGQGIIQPQQPA (peptide #13) was identified to be transamidated at the Q4 residue and deamidated at the Q10 residue.

Discussion

In this study we have identified the preferred substrates of TG2 in proteolytic digests of whole wheat gluten. To simplify the identification we established an enrichment method which is based on the TG2-mediated transamidation reaction followed by sequence determination by mass spectrometry. We identified 31 different peptide substrates of TG2. Strikingly, the majority of these peptides harbor known gluten T-cell epitopes.

To date, more than 15 HLA-DQ2.5 and HLA-DQ8-restricted gluten derived T-cell epitopes have been identified [12-19]. For the majority of HLA-DQ2.5-restricted gliadin epitopes, the kinetics of TG2-mediated deamidation was recently determined and large variations between the peptide epitopes were observed [21]. Interestingly, there seems to be a correlation between the level of deamidation of the different epitopes and how frequently the epitopes are recognized by T-cells of celiac disease patients.

Peptide substrates of TG2 can be either transamidated or deamidated. Deamidated reaction products generated in the absence of primary amines can be directly detected by LC-MS in mixtures of low complexity, e.g. in a chymotryptic digest of a single recombinant gliadin protein. However, in our hands this approach was not feasible for the identification of the best substrates of TG2 when analyzing the PTCEC digest of whole gluten because of its extreme heterogeneity. Although the propensity for either of the two modifications may differ slightly for some peptide substrates, the substrate specificity for deamidation and transamidation overall appears very similar [23]. We confirmed this notion, as we found the same two peptides being substrates in the deamidation and transamidation reaction when testing a chymotryptic digest of a recombinant gliadin protein. Moreover, when testing substrate potency of four peptides for transamidation and deamidation, we observed the same rank order for the peptides in both reactions. We therefore decided to harness the transamidation reaction to identify the best peptide substrates in the complex peptide mixture of a PTCEC digest of whole gluten. By this approach, we were able to tremendously reduce the sample complexity by enriching for the TG2-modified peptides and thereby facilitating identification of individual peptides.

The transamidation reaction in this highly competitive environment was performed in the presence of limited amounts of 5-BP and short incubation periods. It is striking that the majority of the identified transamidated peptides contained intact known T-cell epitopes or truncated versions thereof. Five of the nine identified α -gliadin derived peptides carry more than one copy of either of the epitopes DQ2- α -I, DQ2- α -II, or DQ2- α -III, and eight of the ten identified γ -gliadin peptides harbor a part of the DQ2- γ -II epitope. These epitopes showed also the highest deamidation rates in our recent study [21], and they constituted nearly half of the gluten-TG2 substrates we identified.

The identified TG2 substrates that carry T-cell epitopes can be dichotomized according to whether they harbor complete or incomplete 9mer core binding regions. Interestingly, this classification correlates with the frequency by which the T-cell epitopes are recognized by T cells of celiac disease patients. T-cell epitopes that are frequently recognized by T cells of celiac disease patients (e.g. DQ2- α -I, DQ2- α -II/DQ2- α -III) were represented by their intact 9mer sequences whereas the T-cell epitopes that are infrequently recognized (e.g. DQ2- γ -II, gliadin- α 20, glutenin-17) were represented by partial 9mer sequences. This was particularly striking for the DQ2- γ -II epitope which was never represented by a complete 9mer, but appeared in C-terminally truncated forms in 8 distinct peptides. In keeping with our results, Shan *et al* reported that the DQ2- γ -II epitope was not among the most proteolytically stable fragments of a recombinant γ -gliadin protein [20]. These data underscore that proteolytic degradation is an important force in the selection of gluten T-cell epitopes in celiac disease.

The sequences of some of the identified TG2 peptide substrates and the observed targeting by TG2 suggested that these peptides could contain new T-cell epitopes. Two of them, (peptide #19, PHQPQQQVPQPQQPQQPF, and peptide #25, SHQQQ***FPQQPY**PQQPYPS*****, suggested 9mer core region are underlined or in italic, Q residues targeted by TG2 are given in bold) were tested for recognition by polyclonal T-cell lines from HLA-DQ2.5-positive celiac disease patients. Interestingly, both peptides induced a dose-dependent T-cell response in one of the seven lines tested. The fact that this T-cell line did not cross-react with a highly similar known epitope, the DQ2- γ -VI epitope, and otherwise displayed only weak to borderline

responses toward two out of a large panel of known T-cell epitopes tested, strongly suggests that we have identified two novel gluten epitopes.

Peptides #8 (LQPzNPS**QQPQE**QVPL) and #9 (VPVQLQPzNPS**QQPQE**QVPL) (suggested 9mer core regions binding to HLA-DQ2.5 are underlined or in italic; Q residues targeted by TG2 in bold; z indicates Q or K) of the well-investigated α 2-gliadin protein have also been reported as substrates for TG2 [26] and their sequences suggest that they bind to HLA-DQ2.5. Although peptides representative of peptide #9 have failed to stimulate any T-cell lines or T-cell clones tested so far by others [12,27], our results warrant further investigation into whether this indeed could be a T-cell epitope. Some of the remaining peptides which we identified (e.g. #28, #31) do most likely not harbor the binding motifs of HLA-DQ2.5 or HLA-DQ8 and this probably explains why they are not T-cell epitopes in celiac disease. Notwithstanding these examples, we find it striking that the great majority of the preferred TG2 substrates in a complex gluten mixture contain known T-cell epitopes.

Among the identified TG2 peptide substrates a few peptides were observed as transamidated and deamidated at the same time. These findings may have implications for the generation of TG2-specific autoantibodies in celiac disease whose appearance is strictly dependent on the dietary intake of gluten [28]. In a previously proposed hapten-carrier model, uptake of covalent complexes between TG2 and gluten peptides by TG2-specific B cells was suggested to result in activation of gluten-specific T-cells which subsequently provide help for the B cells to secrete antibodies [29]. In accordance with this hypothesis, cross-linking of gluten peptides to TG2 via generation of isopeptide bonds between glutamine residues in the peptide and lysine residues of TG2 has been described [30]. If those TG2-bound gluten peptides carry an additional deamidated glutamine residue in a T-cell epitope distant from the transamidation site, e.g. as observed for peptides #1 and #3, presentation of this epitope would only require cleavage of the peptide backbone by proteases present in the endocytic compartments while leaving the isopeptide bond intact.

In a heterogeneous gluten digest, TG2 is not targeting all the different gluten peptides, but shows a clear preference for those which are found as T-cell epitopes in celiac disease patients. This finding argues for TG2 exerting a strong pressure in the selection of T-cell epitopes in celiac disease. Our data also demonstrate the important role of proteolysis in the selection of T-cell epitopes as the T-cell epitopes that are frequently recognized by T cells of celiac disease patients more often remain intact than the epitopes that are infrequently recognized. Together, the selective forces exerted by TG2 and gastrointestinal proteolysis are key factors to determine the repertoire of gluten epitopes in celiac disease.

Acknowledgement

We thank Michael Bethune (Stanford University) for the kind gift of PTCEC gluten.

References

1. Sollid LM (2000) Molecular basis of celiac disease. *Annu Rev Immunol* 18:53-81.
2. Lundin KE, Scott H, Hansen T, Paulsen G, Halstensen TS, et al. (1993) Gliadin-specific, HLA-DQ(alpha 1*0501,beta 1*0201) restricted T cells isolated from the small intestinal mucosa of celiac disease patients. *J Exp Med* 178:187-196.
3. Molberg O, Kett K, Scott H, Thorsby E, Sollid LM, et al. (1997) Gliadin specific, HLA DQ2-restricted T cells are commonly found in small intestinal biopsies from coeliac disease patients, but not from controls. *Scand J Immunol* 46:103-109.
4. Johansen BH, Vartdal F, Eriksen JA, Thorsby E, Sollid LM (1996) Identification of a putative motif for binding of peptides to HLA-DQ2. *Int Immunol* 8:177-182.
5. van der Wal Y, Kooy YM, Drijfhout JW, Amons R, Koning F (1996) Peptide binding characteristics of the coeliac disease-associated DQ(alpha1*0501, beta1*0201) molecule. *Immunogenetics* 44:246-253.
6. Vartdal F, Johansen BH, Friede T, Thorpe CJ, Stevanovic S, et al. (1996) The peptide binding motif of the disease associated HLA-DQ (alpha 1* 0501, beta 1* 0201) molecule. *Eur J Immunol* 26:2764-2772.
7. Kwok WW, Domeier ML, Raymond FC, Byers P, Nepom GT (1996) Allele-specific motifs characterize HLA-DQ interactions with a diabetes-associated peptide derived from glutamic acid decarboxylase. *J Immunol* 156:2171-2177.
8. Godkin A, Friede T, Davenport M, Stevanovic S, Willis A, et al. (1997) Use of eluted peptide sequence data to identify the binding characteristics of peptides to the insulin-dependent diabetes susceptibility allele HLA-DQ8 (DQ 3.2). *Int Immunol* 9:905-911.
9. Folk JE. (1983) Mechanism and basis for specificity of transglutaminase-catalyzed epsilon-(gamma-glutamyl) lysine bond formation. *Adv Enzymol Relat Areas Mol Biol* 54:1-56.
10. Vader LW, de Ru AH, van der Wal Y, Kooy YM, Benckhuijsen W, et al. (2002) Specificity of tissue transglutaminase explains cereal toxicity in celiac disease. *J Exp Med* 195:643-649.
11. Fleckenstein B, Molberg O, Qiao SW, Schmid DG, von der Mulbe F, et al. (2002) Gliadin T cell epitope selection by tissue transglutaminase in celiac disease. Role of enzyme specificity and pH influence on the transamidation versus deamidation process. *J Biol Chem* 277:34109-34116.
12. Arentz-Hansen H, Korner R, Molberg O, Quarsten H, Vader W, et al. (2000) The intestinal T cell response to alpha-gliadin in adult celiac disease is focused on a single deamidated glutamine targeted by tissue transglutaminase. *J Exp Med* 191:603-612.
13. Arentz-Hansen H, McAdam SN, Molberg O, Fleckenstein B, Lundin KE, et al. (2002) Celiac lesion T cells recognize epitopes that cluster in regions of gliadins rich in proline residues. *Gastroenterology* 123:803-809.
14. Qiao SW, Bergseng E, Molberg O, Jung G, Fleckenstein B, et al. (2005) Refining the rules of gliadin T cell epitope binding to the disease-associated DQ2 molecule in celiac disease: importance of proline spacing and glutamine deamidation. *J Immunol* 175:254-261.

15. Sjoström H, Lundin KE, Molberg O, Korner R, McAdam SN, et al. (1998) Identification of a gliadin T-cell epitope in coeliac disease: general importance of gliadin deamidation for intestinal T-cell recognition. *Scand J Immunol* 48:111-115.
16. Tollefsen S, Arentz-Hansen H, Fleckenstein B, Molberg O, Raki M. et al. (2006) HLA-DQ2 and -DQ8 signatures of gluten T cell epitopes in celiac disease. *J Clin Invest* 116:2226-2236.
17. Vader W, Kooy Y, van Veelen P, de Ru A, Harris D, et al. (2002) The gluten response in children with celiac disease is directed toward multiple gliadin and glutenin peptides. *Gastroenterology* 122:1729-1737.
18. van der Wal Y, Kooy YM, van Veelen PA, Pena SA, Mearin LM. et al. (1998) Small intestinal T cells of celiac disease patients recognize a natural pepsin fragment of gliadin. *Proc Natl Acad Sci U S A* 95:10050-10054.
19. van der Wal Y, Kooy YM, Van VP, Vader W, August SA, et al. (1999) Glutenin is involved in the gluten-driven mucosal T cell response. *Eur J Immunol* 29:3133-3139.
20. Shan L, Qiao SW, Arentz-Hansen H, Molberg O, Gray GM, et al. (2005) Identification and analysis of multivalent proteolytically resistant peptides from gluten: implications for celiac sprue. *J Proteome Res* 4:1732-1741.
21. Dorum S, Qiao SW, Sollid LM, Fleckenstein B (2009) A quantitative analysis of transglutaminase 2-mediated deamidation of gluten peptides: implications for the T-cell response in celiac disease. *J Proteome Res* 8:1748-1755.
22. Piper JL, Gray GM, Khosla C (2002) High selectivity of human tissue transglutaminase for immunoactive gliadin peptides: implications for celiac sprue. *Biochemistry* 41:386-393.
23. Stamaes J, Fleckenstein B, Sollid LM (2008) The propensity for deamidation and transamidation of peptides by transglutaminase 2 is dependent on substrate affinity and reaction conditions. *Biochim Biophys Acta*.
24. de Ritis G, Auricchio S, Jones HW, Lew EJ, Bernardin JE, et al. (1988) In vitro (organ culture) studies of the toxicity of specific A-gliadin peptides in celiac disease. *Gastroenterology* 94:41-49.
25. Camarca A, Anderson RP, Mamone G, Fierro O, Facchiano A, et al. (2009) Intestinal T cell responses to gluten peptides are largely heterogeneous: implications for a peptide-based therapy in celiac disease. *J Immunol* 182:4158-4166.
26. Mazzeo MF, De Giulio B, Senger S, Rossi M, Malorni A, et al. (2003) Identification of transglutaminase-mediated deamidation sites in a recombinant alpha-gliadin by advanced mass-spectrometric methodologies. *Protein Sci* 12:2434-2442.
27. Anderson RP, Degano P, Godkin AJ, Jewell DP, Hill AV (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.
28. Sulkanen S, Halttunen T, Laurila K, Kolho KL, Korponay-Szabo IR, et al. (1998) Tissue transglutaminase autoantibody enzyme-linked immunosorbent assay in detecting celiac disease. *Gastroenterology* 115:1322-1328.
29. Sollid LM, Molberg O, McAdam S, Lundin KE (1997) Autoantibodies in coeliac disease: tissue transglutaminase--guilt by association? *Gut* 41:851-852.

30. Fleckenstein B, Qiao SW, Larsen MR, Jung G, Roepstorff P, et al. (2004) Molecular characterization of covalent complexes between tissue transglutaminase and gliadin peptides. *J Biol Chem* 279:17607-17616.
31. Marti T, Molberg O, Li Q, Gray GM, Khosla C, et al. (2005) Prolyl endopeptidase-mediated destruction of T cell epitopes in whole gluten: chemical and immunological characterization. *J Pharmacol Exp Ther* 312:19-26.

Figure legends

Figure 1: Schematic view of the established method to enrich and analyze TG2 peptide substrates. A PTCEC digest of wheat gluten was mixed with a small amount of 5-BP which served as a substrate for TG2 in a transamidation reaction. The transamidated, biotinylated peptides were enriched from the digest using magnetic streptavidin beads, eluted with an excess of biotin and analyzed by LC-MS/MS. Database searching was performed using a database made up of all entries of *Triticum aestivum* present in the Uniprot database. In addition, MS/MS spectra were manually inspected.

Figure 2: Specific enrichment of transamidated peptides. A PTCEC digest of whole gluten was spiked with the transamidated DQ2- α -I and DQ2- γ -II peptides. MALDI-TOF spectra before (upper panel) and after enrichment (lower panel) are shown. The inset shows the four signals obtained for each of the two enriched peptides that correspond to the transamidated peptide, its sodium adduct, its potassium adduct and the adduct with two sodium ions, respectively. All signals observed in the lower mass range (up to m/z 900) were derived from matrix clusters.

Figure 3: Recognition of gluten peptides #19 and #25 by T-cell line BW.CD-E.

The identified TG2 peptide substrates SHQQPFPQQPYQPYPYS (P#25) and PHQPQQVQPQQPQQPF (P#19) were tested together with a panel of synthetic epitope peptides for recognition by a T-cell line generated from a biopsy of a HLA-DQ2.5-positive celiac disease patient. The sequences of the other synthetic peptides tested were as follows: DQ2- α -I, QLQFPQPPELPY; DQ2- α -II, PQPELPYPQPQLPY; DQ2- α 33mer, LQLQFPQPPELPYPQPELPYPQPELPYPQPQQPF; DQ2- γ -I, pyroglu-PQQPQQSFPEQQR; DQ2- γ -II, GIHQPEQAQL; DQ2- γ -III, FPEQPEQPYPEQ; DQ2- γ -IV, FSQPEQEFPQPQ; DQ2- γ -VI, PEQPFPEQPEQ; DQ2- γ -VIIa, TEQPEQFPQP; DQ2- γ -VIIb, FPQPEQEFPQPQ; DQ2- γ -III/DQ2- γ -VI, LQPEQPFPEQPEQPYPEQPQ; DQ2- γ 23mer, EQPFPEQPEQPYPEQPEQFPQP; DQ2- ω 17mer, QPQQPFPQPEQFPWPQ. Note that peptide #25 and peptide #19 were tested at different concentrations. The DQ2- α 33mer peptide was tested at 2 μ M while the other peptides were tested at 10 μ M. Responses were

measured in a proliferation assay by the incorporation of ^3H -thymidine (counts per minute (CPM) $\times 10^3$). The dashed line indicates two-fold background proliferation observed with medium only. Each antigen was tested in triplicates and error bars indicate the standard error of mean. The experiment was repeated twice.

Table 1: Identified T-cell epitopes in wheat gluten and their observed T-cell response in celiac disease (CD) patients

T-cell epitopes	Peptide binding register									Observed T-cell response in CD-patients	Identified as TG2-substrates in PTCEC gluten
	1	2	3	4	5	6	7	8	9		
DQ2- α -I	P	F	Q	P	P	Q	L	P	Y	8/14 [31], 12/17 [12], 7/20 [17]	✓
DQ2- α -II	P	Q	P	L	P	Y	P	Q		14/14 [31] ^a , 10/20 [17] ^a , 11/17 [12] ^a	✓
DQ2- α -III	P	Y	Q	P	P	Q	L	P	Y	14/14 [31] ^a , 10/20 [17] ^a , 11/17 [12] ^a	✓
DQ2- γ -I	P	Q	S	F	P	Q	Q	Q		3/14 [31]	
DQ2- γ -II	I	P	Q	P	A	Q	L			4/14 [31], 5/13 [14]	✓♦
DQ2- γ -III/DQ8- γ -I	Q	Q	P	Q	P	Y	P	Q		6/14 [31] ^b	
DQ2- γ -IV	S	Q	P	Q	Q	F	P	Q		3/13 [14]	
DQ2- γ -VI	Q	Q	P	F	P	Q	Q	P	Q	6/14 [31] ^b	
DQ2- γ -VII	P	Q	P	Q	Q	F	P	Q		-	
DQ2- γ -VIIb/DQ8- γ -I	Q	Q	P	Q	P	F	P	Q		3/13 [14]	✓
Glia- α 20	F	R	P	Q	P	Y	P	Q		4/20 [17]	✓♦
Gli-17	P	F	S	Q	Q	Q	P	V		^c	✓
DQ8- α -I	Q	G	S	F	Q	P	S	Q		-	
DQ8-glutenin	Q	G	Y	P	T	S	P	Q		-	
DQ2- ω 17mer	Q	F	Q	P	F	P	Q	P	F	5/14 [25]	
26mer	F	L	Q	P	Q	P	F	P	Q	P	
33mer	L	Q	L	P	F	P	Q	L	P	F	
	P	Q	P	F						14/14 [31], 13/13 [14]	

Glutamine residues targeted by TG2 are given in bold.

^a Peptide POPQLPYQPQLPY harboring both epitopes, DQ2- α -II and DQ2- α -III, was tested.

^b Peptide LQPQPFQPPQPPYQQPQ harboring both epitopes, DQ2- γ -II/DQ8- γ -I and DQ2- γ -VI, was tested.

^c Peptide QQPFPSQQQQPLPQ (previously called "Gli-17" epitope) was tested for T-cell response in CD patients: 3/20 [17]

♦ Identified TG2 peptide substrates harbor incomplete 9mer core binding regions.

Table 2: TG2 peptide substrates identified by nano-LC MS/MS

Peptide #	Peptide identified	Modifications	Identified protein (# of entries in data base)	T-cell epitope	Intact 9mer core epitope
1	<p> PQLPYPPQLPYPPQPF PQLPYPPQLPYPPQPF PQLPYPPQLPYPPQPF </p>	<p> 5BP: Q11 5BP: Q4 5BP: Q11, DA: Q4 </p>	<p> α-gliadin (3) </p>	<p> DQ2-α-II (2x), DQ2-α-III </p>	<p> Yes Yes Yes </p>
2	<p> LPYPQLPYPPQPF </p>	<p> 5BP: Q7 </p>	<p> α-gliadin (3) </p>	<p> DQ2-α-II, DQ2-α-III </p>	<p> Yes </p>
3	<p> QLQFFPQLPYPPQLPYPPQPF QLQFFPQLPYPPQLPYPPQPF </p>	<p> 5BP: Q9/Q16^a 5BP: Q9/Q16, DA: Q9/Q16^b </p>	<p> α-gliadin (3) </p>	<p> DQ2-α-I, DQ2-α-II (2x), DQ2-α-III </p>	<p> Yes Yes </p>
4	<p> QLQFFPQLPYPPQ </p>	<p> 5BP: Q9 </p>	<p> α-gliadin (5) </p>	<p> DQ2-α-I, DQ2-α-II </p>	<p> Yes </p>
5	<p> QLQFFPQLPYPPQPF </p>	<p> 5BP: Q9 </p>	<p> α-gliadin (4) </p>	<p> DQ2-α-I, DQ2-α-II </p>	<p> Yes </p>
6	<p> RQLPYPPQ </p>	<p> 5BP: Q3 </p>	<p> α-gliadin (9) </p>	<p> DQ2-α-20 </p>	<p> No </p>
7	<p> LQLQFFPQLPYPPQPF </p>	<p> 5BP: Q3/Q4^a </p>	<p> α-gliadin (3) </p>		
8	<p> LQPNFSQQLPYPPQVPL LQPNFSQQLPYPPQVPL </p>	<p> 5BP: Q14 5BP: Q14, DA: Q9 </p>	<p> α-gliadin (6) α-gliadin (6) </p>		
9	<p> VVFPQLQPNFSQQLPYPPQVPL VVFPQLQPNFSQQLPYPPQVPL </p>	<p> 5BP: Q19 5BP: Q19, DA: Q14 </p>	<p> α-gliadin (4) α-gliadin (4) </p>		

23	S <u>Q</u> Q <u>Q</u> PFV <u>X</u> PQQFS ^d	5BP: Q2/Q3/Q4 ^a	LMW glutenin (2/1)	DQ2-glutenin-17	No
24	PQQPFPSQQQPVLPQQSPFPQ ^d	5BP: Q10	LMW glutenin (5)	DQ2-glutenin-17	Yes
25	SHQQPPFPQQFPY <u>Q</u> QPYPS ^e SHQQPPFP <u>Q</u> YFPQQPYPS ^e	5BP: Q14 5BP: Q9	LMW glutenin (1)	Novel DQ2-glutenin epitope	Yes Yes
26	PQQPPFSQ <u>Q</u> QPx	5BP: Q9/Q10 ^a	LMW glutenin (3)	Cross-reactive to DQ2-glutenin-17	Yes
27	SQQQQPPFSQQQPPFSQQQ <u>Q</u> QL	5BP: Q20	LMW glutenin (1)	Cross-reactive to DQ2-glutenin-17	Yes
28	QQQPPFSQ <u>Q</u> QPx	5BP: Q9	LMW glutenin (11)		
29	S <u>Q</u> QLPPFSQQQSPF	5BP: Q2	LMW glutenin (1)		
30	SFPQPF <u>Q</u> QPPQPS	5BP: Q8	LMW glutenin (1), ω-gliadin (1)		
31	S <u>Q</u> QLFPQPS	5BP: Q2/Q3/Q4 ^a	LMW glutenin (1)		

Glutamine residues targeted by TG2 are given in bold.

DA, deamidation

The 9mer core region of T-cell epitopes are dashed underlined.

^a "x" indicates I or L

^b not possible to determine which Q residue is transamidated (shown in bold and italic)

^c not possible to determine whether Q2 or Q4 is targeted; however, this sequence has previously been shown to be targeted at Q4 [17].

^d identified in a sample incubated with TG2 for one minute

^e two possible 9mer binding registers to HLA-DQ2.5

Figure 1

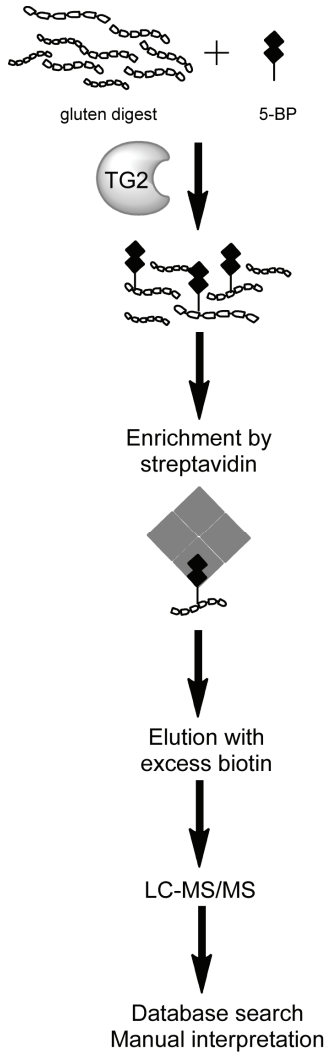


Figure 2

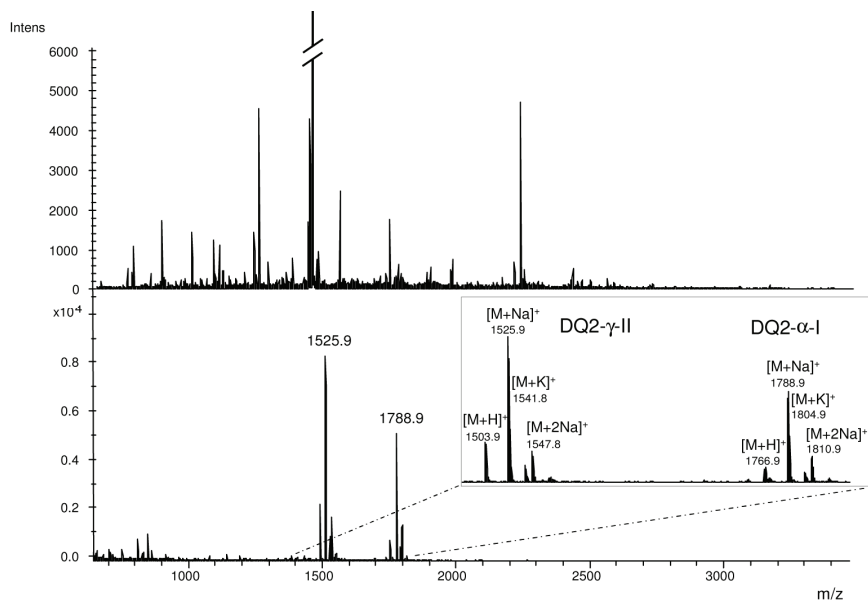


Figure 3

