Amino Acids (2009) 36:619–624
DOI 10.1007/s00726-008-0126-6

REVIEW ARTICLE

Preferred substrate sequences for transglutaminase 2: screening using a phage-displayed peptide library

Kiyotaka Hitomi · Miyako Kitamura · Yoshiaki Sugimura

Received: 3 April 2008 / Accepted: 10 May 2008 / Published online: 24 July 2008

Abstract A large number of substrate proteins for tissue transglutaminase (TGase 2) have been identified in vivo and in vitro. Preference in primary sequence or secondary structure around the reactive glutamine residues in the substrate governs the reactivity for TGase 2. We established a screening system to identify preferable sequence as a glutamine-donor substrate using a phage-displayed peptide library. The results showed that several peptide sequences have higher reactivity and specificity to TGase 2 than those of preferable sequences previously reported. By analysis of the most reactive 12-amino acid sequence, T26 (HQSYVDPWMLDH), residues crucial to the enzymatic reaction were investigated. The following review summarizes the screening system and also the preference in substrate sequences that were obtained by this method and those previously reported.

Keywords Transglutaminase · Substrate · Phage-displayed peptide library · Immobilization

Abbreviations

Bio-Cd	Biotinylated cadaverine
GST	Glutathione-*S*-transferase
MDC	Monodansylcadaverine
TGase 2	Tissue-specific transglutaminase

K. Hitomi (✉) · M. Kitamura · Y. Sugimura
Department of Applied Molecular Biosciences,
Graduate School of Bioagricultural Sciences,
Nagoya University, Chikusa, Nagoya 464-8601, Japan
e-mail: hitomi@agr.nagoya-u.ac.jp

Introduction

Transglutaminases (TGases; EC 2.3.2.13) are a family of enzymes that catalyze protein modification leading to formation of covalent isopeptide linkage (Chen and Mehta 1999; Griffin et al. 2002; Fesus and Piacentini 2002; Lorand and Graham 2003). The binding of Ca^{2+} is essential in the reaction, as this induces conformational changes of the enzyme to expose a cysteine residue in the active site. Then this cysteine residue reacts with a glutamine residue in the substrate, resulting in the formation of an acyl-enzyme intermediate. This complex reacts with a primary amine or lysine residue, after which the enzyme is released. Water can replace the "glutamine-acceptor" molecule, resulting in deamidation of the glutamine residue and formation of a glutamic acid residue. Because the formation of the acyl-intermediate as the first reaction is rate-limiting, TGase is selective for the glutamine residue in the substrates, while the second reaction toward lysine residue or primary amine is less selective.

So far, eight isozymes of TGase have been identified in humans: TGase 1–TGase 7 and Factor XIII. Most of them are characterized in terms of their enzymatic activity as active form or zymogen type. Each isozyme has unique substrates that are modified by transamidation or deamidation, which are involved in several biological events in various tissues (Esposito and Caputo 2005). The same enzymatic reactions have been observed in various organisms, from microorganisms to animals. Although all the TGases follow the same reaction step to form an acyl-enzyme intermediate, the glutamine residues in the substrates are different in the reactions. For example, in vitro catalytic reaction products of small proline rich proteins (SPR) or trichohyalin (both of which are structural proteins in the epidermis and hair follicle cells) by TGases 1, 2, and 3

showed different cross-linking patterns (Tarcsa et al. 1997; Steinert et al. 1999). Additionally, the products of enzymatic reactions with TGase 2 and Factor XIII are different in the case of cross-linking of β-casein-derived peptide or fibrin subunits as substrates (Gorman and Folk 1984; Hettasch et al. 1997). Thus, each TGase recognizes a different cross-linking site in their substrate protein, in which the primary sequence and/or tertiary structure around the reactive glutamine residue govern the interaction between enzymes and their substrates.

Therefore, the preferable sequences or structures in substrates should be characterized in order to investigate the physiological roles of each TGase. Furthermore, dissection of the interaction of TGases and their substrates would provide a clue for development of efficient regulatory factors for the enzymatic reaction in vitro and in vivo. Useful database for TGase substrate proteins have produced information regarding sequence preference. TRANSDAB (http://genomics.dote.hu/wiki/index.php/Main_Page) provides a list of the identified substrate previously reported, while TRASIT (http://bioinformatica.isa.cnr.it/TRANSIT/) is available for evaluation of potent primary sequences as possible substrates (Facchiano et al. 2003). Both databases have been developed and used for searches of possible substrate for TGases on web sites. Additionally, analysis of the reactive glutamine residues in substrates by functional proteomics has been carried out (Ruoppolo et al. 2003). However, there has not been investigation of such substrate preference using a random peptide library. Therefore, we have established a screening system using unbiased phage-displayed peptide library (Sugimura et al. 2006). As a first target among TGase isozymes, we chose TGase 2 because many substrates and reactive glutamine residues have been already identified, providing information on sequence preference.

In this review, preferable sequences and structures for TGase 2 are discussed on the basis of our identified peptides and also previously reported information.

Screening procedure and results

As a peptide library for screening, we used genetically modified M13 phages that have a 12-mer random peptide on their coat protein. The phage-displayed peptide library has been established as a powerful tool for epitope mapping and for identification of antibody-like molecules and peptide mimics of non-peptide ligands (Clackson and Lowman 2004). Each variant peptide is displayed on the surface of the phage, while the displayed peptide can be simply identified by sequencing the encoding DNA.

In the screening, we took an advantage of the displayed peptides as glutamine-donor substrates (Fig. 1). In the TGase catalytic reaction, phages that displayed an efficient "glutamine-donor" sequence incorporated biotin-labeled cadaverine (Bio-Cd) onto their peptide-fused coat proteins. In the following procedure, the "biotinylated" phages could be purified by avidin-affinity chromatography and were normally infectious to bacteria for amplification. The selected phages were subjected to additional enzymatic reactions and panning. After 5 rounds, several phages were analyzed for their DNA sequences encoding for the displayed peptides.

Among the obtained sequences, based on the alignment of possible reactive glutamine residue, several consensus amino acid residues were observed (Sugimura et al. 2006). Sequences of the representative 11 clones are shown in Fig. 2a. Most of the clones contained Q-x-P, where x represents a non-conserved amino acid. The sequences were divided into three groups: Q-x-P-ψ-D-P (where ψ represents a hydrophobic amino acid), Q-x-P-ψ, and Q-x-x-ψ-D-P. In most sequences, bulky amino acid was observed at the -1 position relative to the glutamine. Each peptide sequence in these 11 clones was further investigated for its reactivity as a fusion protein with glutathione-*S*-transferase (GST) in bacteria (Fig. 3a). By measuring the amount of monodansyl cadaverine (MDC) incorporated into the fusion proteins, each sequence was evaluated for its reactivity and specificity to TGase 2. Significant differences were found in reactivity among the obtained sequences, suggesting preference around the possible reactive glutamine residue (Fig. 2a). It was found that the preferred substrate peptides with high reactivity contained the Q-x-(P, Y)-ψ-D-(P) sequence. Among the obtained sequences, T26 (HQSYVDPWMLDH: shown at the top of Fig. 2a), was most reactive and highly specific to TGase 2. This sequence appeared to be a favorable substrate in the form of a synthesized peptide that was enzymatically incorporated into casein, a glutamine-acceptor substrate.

Preferred substrate peptides that have been identified as partial sequences from prominent substrate proteins have been described in the previous reports. By using this approach, several sequences including NQEQVSPLTLLK (α2PI; as a Factor XIII substrate), PGGQQIV (fibronectin), QQIV (fibronectin), TVQQEL (crystalline), and RPKPQQFFGLM (substance P) have been found as favored substrate peptides (Ichinose et al. 1983; Lorand et al. 1991; Parameswaran et al. 1990; Groenen et al. 1992; Pastor et al. 1999). The consensus sequence (GQDPVK) of trappin family, which is a protein family that includes anti-leukoproteinase and sodium–potassium ATP inhibitor-2 in skin, has also been characterized as a good substrate (Trappin is an acronym for *tr*ansglutaminase substrate and whey *a*cidic *p*rotein motif-containing *prote*i*n*) (Zeeuwen et al. 1997; Schalkwijk et al. 1999). Several peptides have also been identified in other studies on sequence preference

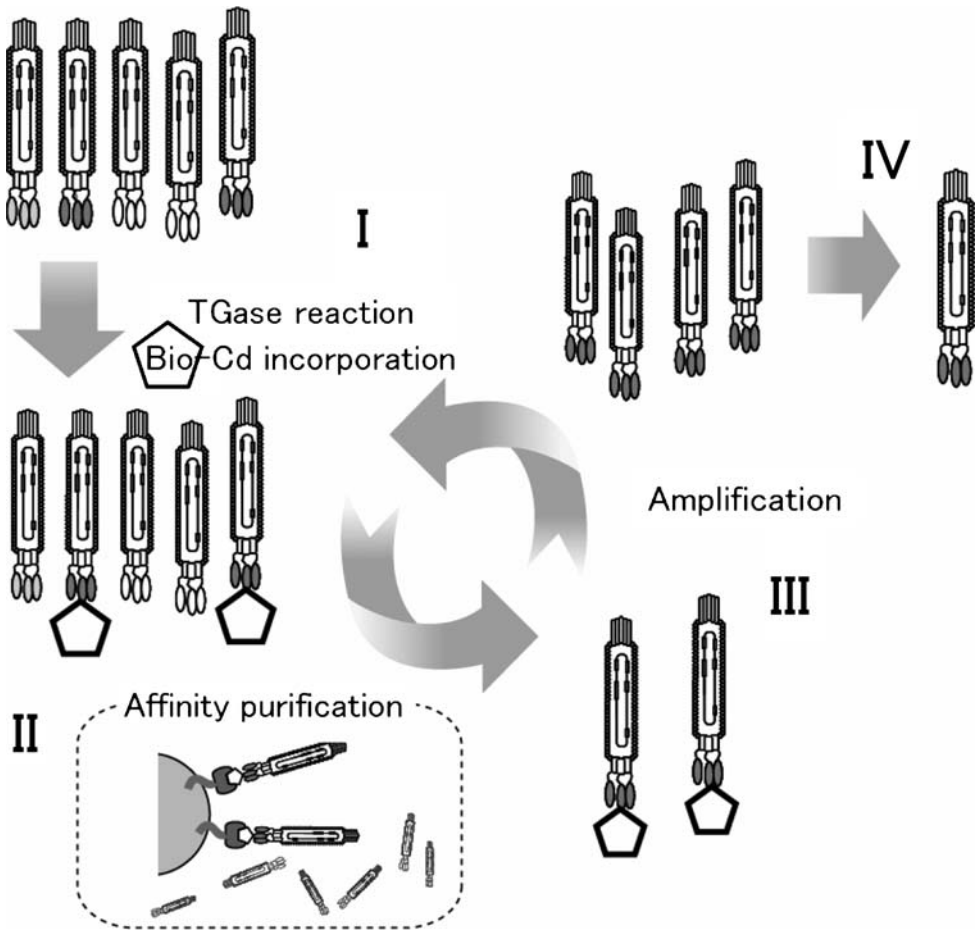

Fig. 1 Screening procedure for substrate sequences preferred by TGase using a phage displayed random peptide library. Panning procedure (I–III) was performed five rounds: *I* incubation of 1–2 × 10^{11} phage clones displaying a 12-mer peptide sequence with biotinylated cadaverine (Bio-Cd) and TGase in an appropriate reaction buffer, *II* affinity purification of phage clones that incorporated Bio-Cd using mono-avidin column chromatography, *III* amplification of selected phage clones in host bacteria, *IV* sequence analysis of DNA isolated from each phage clone

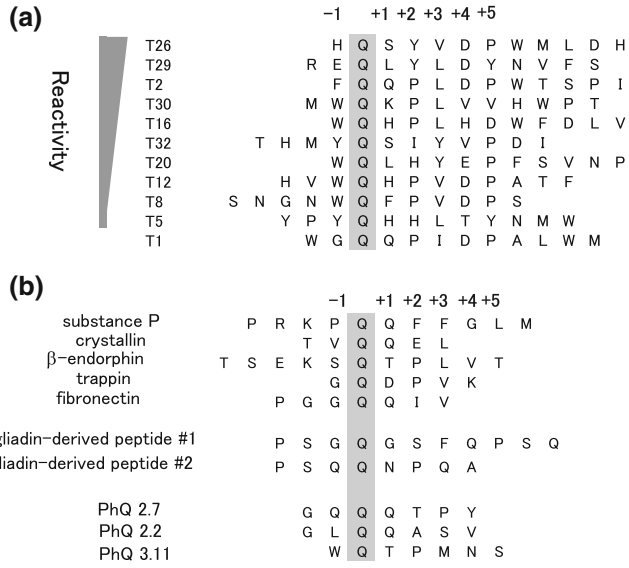

Fig. 2 Alignments of the preferred substrate sequences for the selected phage clones (**a**) and previously reported sequences described in the text (**b**). In **a**, the sequences aligned based on reactivity as GST-fusion proteins (see legend to Fig. 3). The gliadin-derived peptides #1 and #2 indicate the sequences P205-Q215 and P213-A220 in HLA-DQ8-restricted, T-cell stimulatory peptide, respectively. PhQ2.7, PhQ2.2, and PhQ3.11 are phage clone names in the reference (Keresztessy et al. 2006)

(as described later). Figure 2b shows these sequences by alignment based on the reactive glutamine residue.

Our obtained favorable sequences did not completely coincide with those previously identified sequences. Therefore, we compared the reactivity of T26 with previously reported sequences among good "glutamine-donor" peptide substrates (Fig. 3b). Comparison of the reactivity with those for the previously identified peptides, which were also produced as fusion proteins with GST, revealed that T26 is the most prominent substrate.

Sequence preference by substitution of the T26 peptide sequence

In order to evaluate the significance of the amino acid residues in the T26 peptide sequence, GST fusion proteins in which each amino acid residue was replaced to alanine were produced. By incorporation assays of MDC, the reactivities of the fusion proteins were compared (Fig. 4). The results showed that the reactivity was significantly decreased by substitution of tyrosine (+2 position relative to glutamine residue), valine (+3) and aspartic acid (+4), suggesting that these residues are important for access of

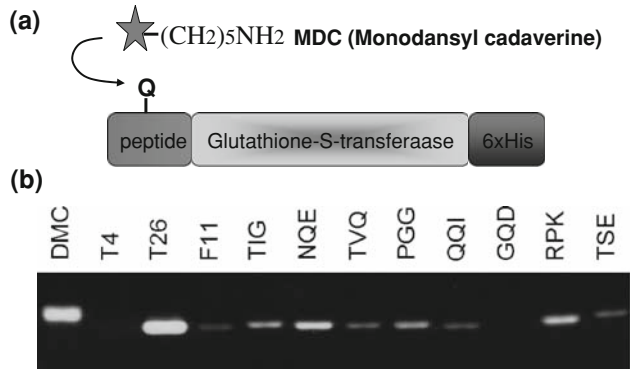

Fig. 3 Reactivities of preferred substrate sequences for TGase 2. **a** Recombinant fusion proteins to evaluate the peptide by genetic fusion with glutathione-*S*-transferase. 6xHis indicates hexahistidine to purify the recombinant proteins. The proteins were reacted with monodansylcadaverine (MDC) in the presence of the enzyme. The reaction products were separated on SDS-PAGE, and a fluorograph of the gel was obtained by UV illumination to detect incorporated MDC. Fluorescence intensity was evaluated as reactivity. **b** Comparison with 10-min reaction of GST-fusion proteins with several partial sequences of TGase substrates: TIG (fibrin γ-chain; TIGEGQQHHLGGA), NQE (α2 plasmin inhibitor, NQEQVSPLTLLK), TVQ (αB-crystalline, TVQQEL), PGG (fibronectin, PGGQQIV), QQI (fibronectin, QQIV) and GQD (trappin, GQDPVK), RPK (substance P, RPKPQQFFGLM), and TSE (β-endorphin, TSEKSQTPLVT). The reaction products of GST-fusion proteins with T4 (WGHTIYHLHPTI, as a negative control), T26 (HQSYVDPWMLDH) and F11 (DQMMLPWPAVKL), and DMC (dimethyl casein, a positive control) were paralleled

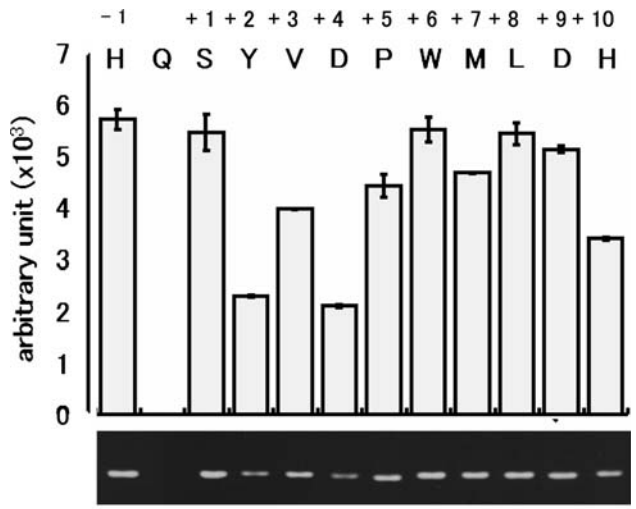

Fig. 4 Influence of substitution of each amino acid residue into alanine residue in T26 on reactivity for TGase 2. GST-fusion proteins with a modified peptide sequence from T26, in which each amino acid was replaced to alanine, were produced in a bacterial expression system and purified to homogeneity. These genetically modified T26–GST fusion proteins were reacted with MDC in the presence of TGase 2, and then the products were analyzed as described in the legend to Fig.3. The column represents the arbitrary reactivity based on the intensity of the fluorescence

the enzyme. Consistently, these amino acid residues are frequently observed among our obtained preferred sequences (Sugimura et al. 2006). Additionally, replacement of +4 (aspartic acid) and +7 (methionine) apparently influenced the preference. Interestingly, +9 position (histidine) was also significantly effective in spite of distant site from the glutamine.

Sequence preference as TGase substrate

For these years, several characteristics in primary sequence as preferable substrates for TGase 2 have been investigated. Among them, much information has been contributed by a series of studies on celiac disease, which is a chronic inflammatory disorder and developed by intolerance to ingested wheat gluten (Sollid 2002). In the small intestine of the patients with celiac disease, peptides from the digested gluten become a substrate for TGase 2, resulting in stimulation of the reactive T cells. In this process, through TGase 2-mediated deamidation of several glutamine residues, epitopes within peptides are produced. Studies on the peptide sequences around the deamidation site have provided information on the favorable primary structure that TGase 2 would access (Vader et al. 2002; Fleckenstein et al. 2002; Qiao et al. 2005) (Fig. 2b).

Vader et al. investigated the influence of amino acid residues around the deamidation site using a gliadin-derived peptide library and mass spectrometry. They systematically analyzed the deamidation in each amino acid substitution analog, PSGQGSFQPSQQNPQA (underlined residues are deamidation sites.) and provided information for several favorable and discouraging amino acid residues. In their studies, they demonstrated that +1 or +3 proline residue relative to the reactive glutamine residue decreased the reactivity, while +2 proline showed a higher rate of deamidation. Substitution to hydrophobic residue at the +3 position also allowed favorable reactivity.

To understand how TGase 2 is involved in formation of the epitopes, Fleckenstein et al. also investigated the effect of amino acid position on preference for deamidation by modifying the peptide TSEKSQTPLVT, which is a partial sequence of β-endorphin, known as good substrate (Pucci et al. 1988). They also showed that amino acid residues in positions −1, +1, +2, and +3 influence TGase reaction. Proline in position +2 had a particularly favorable effect, but glycine and proline at +1 or +3 had a negative influence. This tendency is observed in our obtained preferred peptide sequences. They also revealed that the ratio of deamidation to transamidation reaction increased when pH was lowered below neutral pH.

Keresztessy et al. successfully established a screening system for preference using a phage-displayed peptide

library based on a principle different from that of our system (Keresztessy et al. 2006). They used a phage library displaying a heptapeptide random peptide and performed screening by direct binding between the active site of TGase 2 and displayed peptide. This screening revealed several possible substrate sequences and identified highly preferred sequences: GQQQTPY (PhQ 2.7), GLQQASV (PhQ 2.2), and WQTPMNS (PhQ 3.11) (Fig. 2b). According to their data regarding identified peptide sequences, they also showed that –1, +1, +2, +3 positions relative to the targeted glutamine residue have characteristics to affect the preference: –1, where most residues are polar; +2, where serine, threonine and proline are frequently observed; and +3, where aliphatic amino acids are predominant.

The above-described data and our data indicate that the effective amino acid residues are mostly –1 (polar or bulky), +1 (any amino acid except for proline and glycine), +2 (proline or tyrosine), +3 (hydrophobic, mainly leucine) relative to the reactive glutamine residue. Positions at +4 (aspartic acid) and +5 (proline) weakly contribute to the favorable reaction.

Application of the obtained preferred substrate sequences

In principle, our screening method would be applicable for any TGase. In parallel to the screening of preferable sequences of TGase 2, we were successful in identification of the sequences in the case of Factor XIII (Sugimura et al. 2006). The obtained peptide (F11, DQMMLPWPAVKL) also showed higher specificity and reactivity than partial sequences previously reported, such as α2 plasmin inhibitor (α2PI) and fibrinogen. Accumulation of data on the preferable peptide sequences for the eight isozymes would contribute to the development of isozyme-specific inhibitors or regulator molecules. Furthermore, analyses of co-crystallization with the obtained peptide and enzyme would provide much information for the precise reaction mechanism.

As a useful application, attachment of the preferred substrate peptide enables functional proteins to act as prominent substrates for TGase. Taking advantage of the catalytic activity for formation of covalent cross-linking, this reaction is applicable to immobilization of functional proteins. We recently succeeded in TGase 2-mediated immobilization of single-chain fragment antibody (scFv) and glutathione-*S*-transferase (GST), both of which were genetically modified by fusion with a T26 peptide-tag (Sugimura et al. 2007). In this technique, T26 sequence works as a "glue flap" in a genetically fused protein, in which TGase 2 acts as "glue". This genetic modification

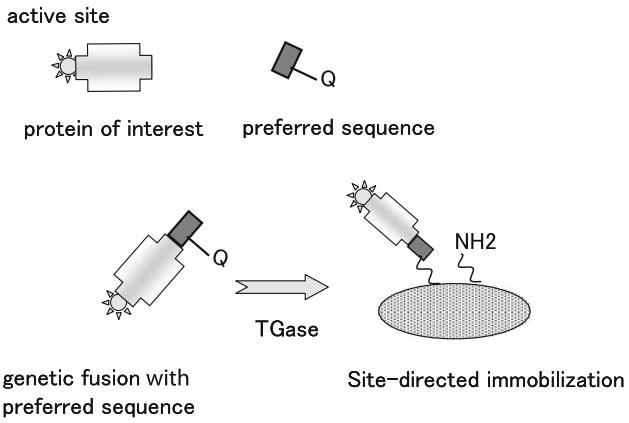

Fig. 5 Procedure for possible application of the preferred substrate sequence for immobilization of functional proteins. The preferred sequences are fused to the protein of interest, as a recombinant protein, at a distant site from its active site or essential region for function. The fusion proteins become recognized by TGase 2 as favorable substrates and possible for cross-linking at the amino-terminated solid phase. In the procedure, binding is designed in a site-directed manner in order to avoid steric hindrance by immobilization

and TGase enzymatic reaction enable functional proteins to attach to amino-terminated gel in a site-directed immobilization (Fig. 5). In fact, using microbial TGase, several methodologies to immobilize fluorescent proteins, enzymes, and food protein as targets have also been developed (Kamiya et al. 2005; Valdivia et al. 2006; Tanaka et al. 2007). These methodologies can be applied for mammalian TGase 2 and T26 sequences. Conversely, identification of the preferred substrate sequences for microbial TGase will contribute to more efficient cross-linking immobilization.

Information on the preferable sequences for lysine-donor substrates would be valuable to clarify the reaction mechanism and to provide more utility for application of the enzyme. Although favorable peptides from partial sequences of substrates have been reported, consensus sequences have not been obtained, probably because of less selectivity toward the lysine residue of the acyl-enzyme intermediate (Groenen et al. 1992; Grootjans et al. 1995; Ruoppolo et al. 2003). Screening using the phage-displayed random library would also be effective to identify more preferable sequences.

Conclusion

We established a screening system to identify preferred peptide sequences for TGase 2 by a simple and time-saving procedure. The obtained peptides will provide structural and sequence information for the development of specific inhibitors and also specific tags for cross-linking of functional proteins. Furthermore, results of analysis for the

influence of each amino acid residue will be valuable for investigation of the recognition mechanism for each TGase isozyme to unique substrates. This screening system can be used for any TGase reaction, and investigation for other TGases is in progress.

Acknowledgment This work was supported by Grant-in-Aid for Scientific Research (C) 19580103 (to K.·H.).

References

Chen JSK, Mehta K (1999) Tissue transglutaminase: an enzyme with a split personality. Int J Biochem Cell Biol 31:817–836

Clackson T, Lowman HB (2004) Phage display. Oxford University Press, UK

Esposito C, Caputo I (2005) Mammalian transglutaminases: identification of substrates as a key to physiological function and physiological relevance. FEBS J 272:615–631

Facchiano AM, Facchiano A, Facchiano F (2003) Active sequences collection (ASC) database: a new tool to assign functions to protein sequences. Nucleic Acids Res 31:379–382

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

Fleckenstein B, Molberg Ø, Qiao SW, Schmid DG, von der Mulbe F, Elgstoen K, Jung G, Sollid LM (2002) Gliadin T cell epitope selection by tissue transglutaminase in celiac disease. J Biol Chem 277:34109–34116

Gorman JJ, Folk JE (1984) Structural features of glutamine substrates for transglutaminases: role of extended interactions in the specificity of human plasma factor XIIIa and of the guinea pig liver enzyme. J Biol Chem 259:9007–9010

Griffin M, Casadio R, Bergamini CM (2002) Transglutaminases: nature's biological glues. Biochem J 368:377–396

Grootjans JJ, Groenen PJ, de Jong WW (1995) Substrate requirements for transglutaminases. J Biol Chem 270:22855–22858

Groenen PJ, Bloemendal H, de Jong WW (1992) The carboxy-terminal lysine αB-crystalline is an amine donor substrate for tissue transglutaminase. Eur J Biochem 205:671–674

Hettasch JM, Peoples KA, Greenberg CS (1997) Analysis of factor XIII substrate specificity using recombinant human factor XIII and tissue transglutaminase chimeras. J Biol Chem 272:25149–25156

Ichinose A, Tamaki T, Aoki N (1983) Factor XIII-mediated cross-linking of NH2-terminal peptide of α2-plasmin inhibitor to fibrin. FEBS lett 153:369–371

Kamiya N, Doi S, Tominaga J, Ichinose H, Goto M (2005) Transglutaminase-mediated protein immobilization to casein nanolayers created on a plastic surface. Biomacromolecules 6:35–38

Keresztessy Z, Csosz E, Harsfalvi J, Csomos K, Gray J, Lightowlers RN, Lakey JH, Balajthy Z, Fesus L (2006) Phage display selection of efficient glutamine-donor substrate peptides for transglutaminase 2. Protein Sci 9:2466–2480

Lorand L, Parameswaran KN, Velasco PT (1991) Sorting-out of acceptor-donor relationship in the transglutaminase catalyzed cross-linking of crystallines by the enzyme-directed labeling of potential sites. Proc Natl Acad Sci USA 88:82–83

Lorand L, Graham RM (2003) Transglutaminases: crosslinking enzymes with pleiotropic functions. Nat Rev Mol Cell Biol 4:140–156

Parameswaran KN, Velasco PT, Wilson J, Lorand L (1990) Labeling of ε-lysine cross-linking sites in proteins with peptide substrate of factor XIIIa and transglutaminase. Proc Natl Acad Sci USA 87:8472–8475

Pastor MT, Diez A, Peretz-Paya E, Abad C (1999) Addressing substrate glutamine requirements for tissue transglutaminase using substance P analogue. FEBS Lett 451:231–234

Pucci P, Malorni A, Marino G, Metafora S, Esposito C, Porta R (1988) β-endorphin modification by transglutaminase in vitro: identification by FAB/MS of glutamine-11 and lysine-29 as acyl donor and acceptor sites. Biochem Biophys Res Commun 154:735–740

Qiao SW, Bergseng E, Molberg Ø, Jung G, Fleckenstein B, Sollid LM (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

Ruoppolo M, Orru S, D'Amato A, Francese S, Rovero P, Marino G, Esposito C (2003) Analysis of transglutaminase protein substrates by functional proteomics. Protein Sci 12:1290–1297

Schalkwijk J, Wiedow O, Hirose S (1999) The trappin gene family: proteins defined by an N-terminal transglutaminase substrate domain and a C-terminal four-disulphide core. Biochem J 340:569–577

Sollid LM (2002) Coeliac disease: dissecting a complex inflammatory disorder. Nat Rev Immunol 2:647–655

Steinert PM, Candi E, Tarcsa E, Marekov LN, Sette M, Paci M, Ciani B, Guerrieri P, Melino G (1999) Transglutaminase crosslinking and structural studies of the human small proline rich 3 protein. Cell Death Differ 6:916–930

Sugimura Y, Hosono M, Wada F, Yoshimura T, Maki M, Hitomi K (2006) 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 281:17699–17706

Sugimura Y, Ueda H, Maki M, Hitomi K (2007) Novel site-specific immobilization of a functional protein using a preferred substrate sequence for transglutaminase 2. J Biotechnol 131:121–127

Tarcsa E, Marekov LN, Andreoli J, Idler WW, Candi E, Chung S-I, Steinert PM (1997) The fate of trichohyalin: sequential post-translational modifications by peptidyl-arginine deiminase and transglutaminase. J Biol Chem 272:27893–27901

Tanaka Y, Tsuruda Y, Nishi M, Kamiya N, Goto M (2007) Exploring enzymatic catalysis at a solid surface: a case study with transglutaminase-mediated protein immobilization. Org Biomol Chem 5:1764–1770

Vader LW, Ru A, van der Wal Y, Kooy YMC, Benckhuijsen W, Mearin ML, Drijfhout JW, van Veelen P, Koning F (2002) Specificity of tissue transglutaminase explains cereal toxicity in celiac disease. J Exp Med 195:643–649

Valdivia A, Villalonga R, Di Pierro P, Perez Y, Mariniello L, Gomez L, Porta R (2006) Transglutaminase-catalyzed site-specific glycosidation of catalase with aminated dextran. J Biotechnol 122:326–333

Zeeuwen PL, Hendriks W, de Jong WW, Schalkwijk J (1997) Identification and sequence analysis of two new members of the SKALP/elafin and SPAI-2 gene family. J Biol Chem 272:20471–20478