BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

pubs.acs.org/biochemistry

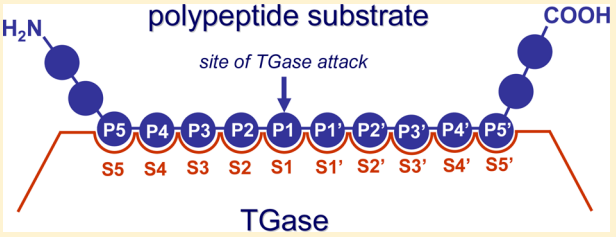
Local Unfolding Is Required for the Site-Specific Protein Modification by Transglutaminase

Barbara Spolaore,* Samanta Raboni,† Amparo Ramos Molina, Abhijeet Satwekar, Nunzio Damiano, and Angelo Fontana

CRIBI Biotechnology Centre, University of Padua, Viale G. Colombo 3, 35121 Padua, Italy

Supporting Information

ABSTRACT: The transglutaminase (TGase) from *Streptomyces mobaraensis* catalyzes transamidation reactions in a protein substrate leading to the modification of the side chains of Gln and Lys residues according to the A-CONH$_2$ + H$_2$N-B → A-CONH-B + NH$_3$ reaction, where both A and B can be a protein or a ligand. A noteworthy property of TGase is its susbstrate specificity, so that often only a few specific Gln or Lys residues can be modified in a globular protein. The molecular features of a globular protein dictating the site-specific reactions mediated by TGase are yet poorly understood. Here, we have analyzed the reactivity toward TGase of apomyoglobin (apoMb), α-lactalbumin (α-LA), and fragment 205−316 of thermolysin. These proteins are models of protein structure and folding that have been studied previously using the limited proteolysis technique to unravel regions of local unfolding in their amino acid sequences. The three proteins were modified by TGase at the level of Gln or Lys residues with dansylcadaverine or carbobenzoxy-L-glutaminylglycine, respectively. Despite these model proteins containing several Gln and Lys residues, the sites of TGase derivatization occur over restricted chain regions of the protein substrates. In particular, the TGase-mediated modifications occur in the "helix F" region in apoMb, in the β-domain in apo-α-LA in its molten globule state, and in the N-terminal region in fragment 205−316 of thermolysin. Interestingly, the sites of limited proteolysis are located in the same chain regions of these proteins, thus providing a clear-cut demonstration that chain flexibility or local unfolding overwhelmingly dictates the site-specific modification by both TGase and a protease.



Transglutaminases (TGases) are enzymes that catalyze the cross-linking between ε-amino groups of lysine (Lys) residues and γ-carboxamide groups of glutamine (Gln) residues of proteins through the formation of ε-(γ-glutamyl)lysine isopeptide bonds.[1] This intermolecular cross-linking reaction is a most relevant and physiologically important enzymatic reaction of TGase,[2] but this is also of biotechnological interest for the production of protein derivatives. Indeed, as outlined in Figure 1, TGases can catalyze the conjugation of alkylamine-containing chemical moieties to Gln residues of proteins.[3] The modification of proteins at the level of Lys residues is also feasible, even if less exploited, requiring a ligand containing a Gln residue or a Gln analogue.[4,5] Because mammalian TGases are not suitable for biotechnological applications because of their high cost of production and their availability in small quantities, a microbial TGase extracted from *Streptomyces mobaraensis* is preferred for these purposes.[6] Indeed, microbial TGase is available in large quantities and is easier to handle because at variance from mammalian TGases the microbial form is calcium-independent and has a molecular mass nearly half that of mammalian TGases (i.e., 37.9 kDa). The applications of microbial TGase include food treatment, tissue engineering, leather processing, and, in particular, protein derivatization.[6,7] In this respect, it has been used to conjugate a variety of chemical moieties, such as fluorescent probes,[8] radionuclide chelating agents for diagnostic and therapeutic applications,[9] functional molecules such as DNA[10] and lipids,[11] and polymer chains of pharmaceutical interest such as poly(ethylene glycol) (PEG),[12] to proteins as well as for immobilization of protein onto solid supports.[13]

The main advantages of the TGase-mediated protein conjugation compared to chemical reactions reside in the fact that the modification of a protein substrate occurs under physiological conditions and, more importantly, in the exceptional selectivity of the reaction at the level of specific Gln residues, thus leading to the production of homogeneous protein bioconjugates. In recent studies, it has been demonstrated that microbial TGase can be successfully used for the enzymatic PEGylation of proteins, sometimes leading to the conjugation of a PEG moiety at a unique Gln residue among the many Gln residues of the protein substrate.[12,14,15] The specificity determinants have been studied for the mammalian enzymes using model peptide and protein substrates, while less information is available for microbial TGase.[5,16−19] In general, these studies revealed that there is no obvious consensus sequence around the Gln residues modified

Received: July 26, 2012
Revised: October 8, 2012
Published: October 19, 2012

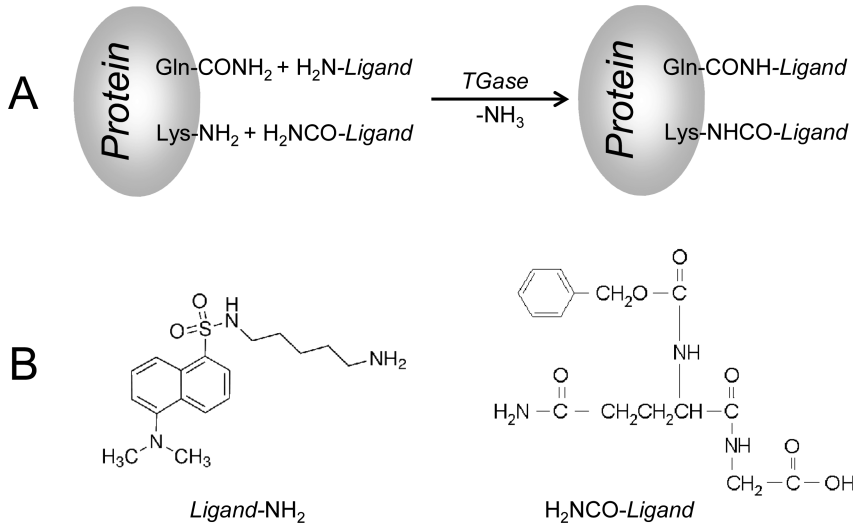


Figure 1. (A) Schematic drawing of the two possible reactions of TGase-mediated protein derivatization. Protein conjugation can be directed either at the level of Gln residues using an amine-containing ligand or at the level of Lys residues with a Gln-containing ligand. (B) Chemical structures of the model substrates used in the TGase-mediated reactions. The amine-containing ligand is dansylcadaverine (DC), while the Gln-containing ligand is carbobenzoxy-L-glutaminylglycine (ZQG).

by TGases and that both the amino acid sequence and the local conformation near Gln residues can contribute to enzyme specificity.[16,19,20] Some structural requirements for the Lys derivatization have been analyzed using human erythrocyte TGase, but an understanding of the structural features of the site of enzymatic attack is still vague.[21] Perhaps microbial and mammalian TGases can display some differences in the sites of modification with protein substrates,[5,12,22] considering that there are substantial structural differences between the human and microbial TGases.[23]

Here, we demonstrate that the selectivity of the derivatization reaction catalyzed by microbial TGase can be explained solely by the presence of Gln or Lys residues embedded in flexible or disordered regions of the globular protein substrate. We report the results of TGase-mediated labeling experiments at the level of both Gln and Lys residues conducted on the model proteins apomyoglobin [myoglobin devoid of heme (apoMb)], α-lactalbumin (α-LA), and fragment 205−316 of thermolysin (205−316Th). These proteins all contain several Gln and Lys residues that can be potentially modified by TGase (Figure S1 and Table S1 of the Supporting Information), but these proteins can be modified at very specific Gln or Lys residues. Of interest, the sites of TGase-mediated derivatization and the sites of hydrolysis by proteases in limited proteolysis experiments occur in the same chain regions of the protein substrates. In the past, in our laboratory we conducted systematic studies of the molecular aspects of the limited or site-specific proteolysis of globular proteins,[24−26] demonstrating that the specific chain fissions invariably occur in regions of the protein substrate characterized by local unfolding. Consequently, we conclude that chain flexibility or protein disorder is the overwhelming parameter dictating the selectivity of the reaction by both TGase and a protease. Finally, we advance the hypothesis that other enzymatic site-specific protein modifications can be explained in terms of a disordered protein substrate.

EXPERIMENTAL PROCEDURES

Materials. TGase from *S. mobaraensis* was purchased from Ajinomoto Co. (Tokyo, Japan), whereas bovine α-LA (type I, calcium-saturated), horse heart myoglobin (Mb), thermolysin (Th) from *Bacillus thermoproteolyticus*, and V8 protease from *Staphylococcus aureus* were from Sigma-Aldrich (Milwaukee, WI). Carbobenzoxy-L-glutaminylglycine (ZQG) and dansylcadaverine (DC) were purchased from Sigma, while porcine trypsin was from Promega (Madison, WI). All other chemicals where purchased from Sigma. ApoMb was obtained from horse heart myoglobin by removal of heme using the acetone extraction procedure.[27] 205−316Th was prepared by limited proteolysis of the protein as previously described.[28] Protein concentrations were determined on the basis of the absorbance at 280 nm.[29] Aliquots of the reaction mixtures were analyzed by RP-HPLC using an Agilent series 1100 HPLC system with an on-line UV detection from Agilent Technologies (Waldbroon, Germany). The separation conditions such as the gradients of acetonitrile (ACN) with 0.085% TFA and water with 0.1% TFA and the RP column used for these analyses were optimized for each protein (see below).

TGase-Mediated Conjugation of DC and ZQG to ApoMb. ApoMb was dissolved in 0.1 M phosphate buffer (pH 7.0) at a concentration of ~0.8 mg/mL. For the reactions with DC and ZQG, the apoMb/ligand molar ratio was 1/25 (the stock solution DC concentration was 20 mg/mL in methanol, while the ZQG concentration was 34 mg/mL in DMSO). TGase was added at an enzyme/substrate (E/S) ratio of 1/50 (w/w), and the reaction mixtures were incubated at 25 °C. Aliquots were collected after 1, 4, and 6 h, and reactions were stopped by lowering the pH with an aqueous solution of 1% TFA for the reaction with DC or by addition of a solution of iodoacetamide (final concentration of 100 μM) for the reaction with ZQG. Aliquots of the reaction mixtures were analyzed by RP-HPLC using a C18 Phenomenex column (150 mm × 4.6 mm) by applying a two-step gradient from 5 to 40% ACN over 5 min and from 40 to 47% ACN over 17 min. Fractions collected from the RP-HPLC analyses were lyophilized and analyzed by MS.

dx.doi.org/10.1021/bi301005z | *Biochemistry* 2012, 51, 8679−8689

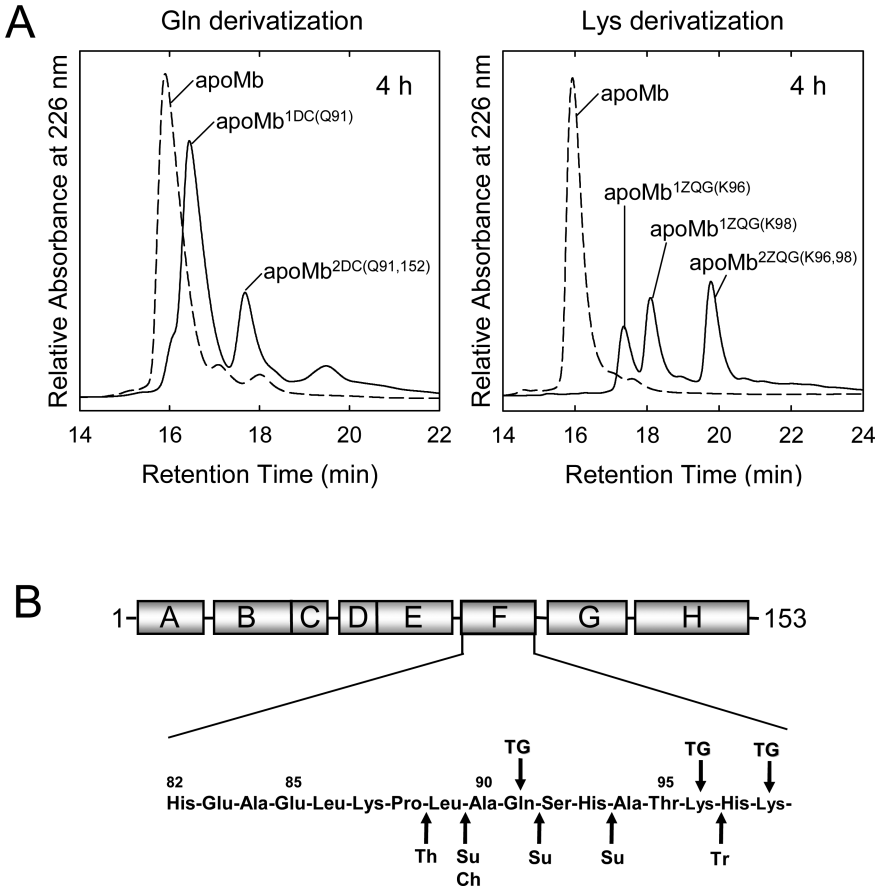


Figure 2. TGase-mediated conjugation of apoMb at Gln and Lys residues. (A) RP-HPLC analyses of the TGase-mediated derivatives of apoMb. A dashed line and a straight line indicate the chromatograms of native apoMb and the DC- and ZQG-conjugated species, respectively, after reaction for 4 h. (B) Scheme of the secondary structure of the 153-residue polypeptide chain of holoMb.[32] The eight helices (A–H) are indicated by boxes, and the amino acid sequence of helix F is explicitly shown. The sites of TGase-mediated modification of apoMb at Gln91, Lys96, and Lys98 are denoted with arrows. The sites of initial proteolytic cleavage of apoMb by thermolysin (Th), subtilisin (Su), trypsin (Tr), and chymotrypsin (Ch) are also shown.[33]

TGase-Mediated Conjugation of DC and ZQG to α-LA. α-LA was dissolved (~0.9 mg/mL) in 20 mM Tris and 5 mM EDTA (pH 7.5) for apo-α-LA and in 20 mM Tris-HCl and 8 mM $CaCl_2$ (pH 7.5) for the holo form of the protein. The apo- and holo-α-LA solutions were mixed with a solution 20 mg/mL DC (in DMSO) at an α-LA/DC molar ratio of 1/30, while TGase was added to the reaction mixtures at an E/S ratio of 1/50 (w/w). The reactions were allowed to proceed at 37 °C, and aliquots were collected after 0, 1, 5, 10, and 30 min for apo-α-LA and 0 min, 30 min, and 2 h for holo-α-LA. The TGase-mediated reaction was inhibited by addition of an equal volume of an aqueous solution of 1% TFA. Aliquots of the reaction mixtures were analyzed by RP-HPLC using a C4 Phenomenex column (150 mm × 4.6 mm) by applying a two-step gradient from 5 to 30% ACN over 5 min and from 30 to 50% ACN over 20 min. Fractions collected from the RP-HPLC analyses were lyophilized and analyzed by MS.

For the derivatization with ZQG, the same solutions of apo- and holo-α-LA described for the modification with DC were used. ZQG (34 mg/mL in DMSO) was added to the two protein solutions at a molar ratio of 1/50 (α-LA/ZQG), while TGase was added at an E/S ratio of 1/50 (w/w). The reaction mixtures were incubated at 37 °C and the reactions stopped by the addition of a solution of iodoacetamide (final concentration of 100 μM). Aliquots were removed after 0 min, 5 min, 15 min, 30 min, 1 h, and 2 h both for holo- and apo-α-LA and analyzed by RP-HPLC as described above for the reaction with DC.

TGase-Mediated Conjugation of DC and ZQG to 205–316Th. 205–316Th was dissolved in 0.1 M phosphate buffer (pH 7.0) at a concentration of ~0.8 mg/mL. For the reaction with DC, a protein/substrate molar ratio of 1/50 was used (stock solution of 20 mg/mL DC in methanol), while in the case of the modification with ZQG, the molar ratio was 1/10 [stock solution of 8 mg/mL ZQG in 0.1 M phosphate buffer (pH 7.0)]. In both reactions, TGase was added at an E/S ratio of 1/25 (w/w). Reaction mixtures were incubated at 25 °C, and aliquots were removed after 0, 1, 2, 4, 6, and 24 h. The enzymatic reaction was stopped by lowering the pH with an aqueous solution of 1% TFA for the reaction with DC or by addition of a solution of iodoacetamide for the reaction with ZQG. RP-HPLC analyses were performed on a C18 Phenomenex column (150 mm × 4.6 mm) by applying a two-step gradient from 5 to 40% ACN over 5 min and from 40 to 60% ACN over 25 min. Fractions collected from the RP-HPLC analyses were lyophilized and analyzed by MS.

Proteolytic Digestion of DC- and ZQG-Modified Proteins. DC- or ZQG-modified α-LA derivatives purified from the RP-HPLC analyses were dissolved in 50 mM NH_4HCO_3 (final protein concentration of ~0.35 mg/mL), and stock solutions containing V8 protease and trypsin were

dx.doi.org/10.1021/bi301005z | Biochemistry 2012, 51, 8679–8689

added to yield a final E/S ratio of 1/25 for V8 protease and 1/50 for trypsin (w/w). Proteolysis was allowed to proceed at 37 °C overnight, and then the mixture was acidified with an aqueous solution of 1% formic acid and 7 mM TCEP to yield a final TCEP concentration of 2.5 mM. Samples were then incubated for 3 h at 37 °C to allow the reduction of the disulfide bonds and then analyzed by ESI-MS and MS/MS. The DC derivatives of apoMb and DC- or ZQG-modified 205−316Th (~0.5 mg/mL) were digested with trypsin at an E/S ratio of 1/100 (w/w) in 0.1 M NH_4HCO_3 overnight at 25 °C, while the ZQG derivatives of apoMb were similarly digested using both trypsin and V8 protease. Because apoMb and 205−316Th do not contain disulfide bridges, for these samples the TCEP reduction step was omitted. Digestion mixtures were diluted with equal volumes of 0.1% formic acid in an ACN/water mixture (1/1) and then analyzed by ESI-MS and MS/MS.

Mass Spectrometry Analyses. Mass spectrometry-based analyses were performed with a Micromass (Manchester, U.K.) Q-Tof Micro mass spectrometer equipped with an electrospray source (ESI-MS). Samples were dissolved in 0.1% formic acid in an ACN/water mixture (1/1) and analyzed in MS and MS/MS mode. Measurements were taken at a capillary voltage of 3 kV and at cone and extractor voltages of 35 and 1 V, respectively (positive ion mode). Tandem MS (MS/MS) analyses of the modified peptides were conducted on the Q-Tof Micro mass spectrometer at variable collision energy values using argon as the collision gas. External calibration was performed using a solution of 0.1% phosphoric acid in 50% aqueous ACN for peptide mass determination, whereas for protein analysis, a solution of 10 μM horse heart apomyoglobin in 50% (v/v) aqueous acetonitrile with 0.1% (v/v) formic acid was used. Instrument control, data acquisition, and processing were achieved with Masslynx (Micromass).

RESULTS AND DISCUSSION

Apomyoglobin. Horse apomyoglobin (apoMb), myoglobin without the heme, is a small monomeric protein of 153 amino acid residues that contains 6 Gln and 19 Lys residues (Figure S1 and Table S1 of the Supporting Information). In this study, apoMb has been modified by TGase at the level of Gln residues using dansylcadaverine (DC) as the amino donor (Figure 1B) and, in a separate experiment, at the level of Lys residues using carbobenzoxy-L-glutaminylglycine (ZQG) as the Gln-containing substrate. In the case of Gln conjugation in the presence of DC, RP-HPLC analysis of the reaction mixture after incubation for 4 h reveals the complete derivatization of apoMb and the formation of two main products (Figure 2A). ESI-MS analysis allowed us to identify the products as mono- and di-DC conjugates (Table S2 of the Supporting Information), and fingerprinting analysis of these derivatives provided evidence that Gln91 is the major site of derivatization and Gln152 a very minor one (Table S3 of the Supporting Information). Indeed, no evidence of singly modified Gln152 is detected in the digestion mixture of apoMb^{1DC}, indicating that modification at this site is much slower than that at Gln91.

RP-HPLC analysis of the reaction mixture of apoMb with ZQG after incubation for 4 h reveals the disappearance of the chromatographic peak corresponding to native apoMb and formation of three new peaks with similar intensities (Figure 2A). The protein species eluting in these peaks were identified by ESI-MS analysis as two different mono-ZQG derivatives and one di-ZQG derivative of apoMb (Table S2 of the Supporting Information). ESI-MS analysis of the apoMb monoderivatives after digestion with trypsin and V8 protease allowed us to identify Lys96 and Lys98 as the two sites of modification. Indeed, in the apoMb derivative conjugated to two molecules of ZQG, both Lys96 and Lys98 are modified (Table S3 of the Supporting Information). Overall, the TGase-mediated reaction of apoMb with both DC and ZQG is very specific, despite the many Gln and Lys residues present in apoMb (see Figure S1 and Table S1 of the Supporting Information). Interestingly, the amino acid residues that are reactive toward TGase (Gln91, Lys96, and Lys98) are all located in the region of helix F of apoMb (Figure 2B), with the exception of Gln152 that is located at the C-terminal end of the 153-residue chain of the protein.

The heme-free apoMb has long been used by numerous investigators as a model protein for studies of protein folding and stability.[30,31] In the past, for the sake of simplicity, the structure of apoMb was often assumed to be similar to that of the heme-containing myoglobin (holoMb), which in the crystal state shows the classic globin fold constituted by eight helices (named A−H) (Figure 2B).[32] With a view of probing the structure and dynamics of apoMb through the limited proteolysis approach, in our laboratory we have conducted a series of proteolysis experiments utilizing subtilisin, thermolysin, chymotrypsin, and trypsin.[33] At neutral pH and 25 °C, these proteases initially cleave apoMb in a very restricted region of the 153-residue chain (see Figure 2B), which in the crystal state of holoMb encompasses helix F (residues 82−97). When reacted under identical experimental conditions, holoMb was instead fully resistant to proteolysis.[33] The results of limited proteolysis experiments, therefore, indicate that in apoMb the chain segment corresponding to helix F in holoMb is highly flexible or disordered.[24−26] The conclusions reached from proteolysis experiments with apoMb were in agreement with a number of additional experimental and theoretical studies of the structure and dynamics of apoMb.[30] In particular, the structural characterization of apoMb by NMR spectroscopy evidenced that the segment of residues 82−102 is disordered.[34]

The overall results of the TGase-mediated modification of apoMb at the level of Gln and Lys residues indicate that the enzymatic modification occurs at the level of the region encompassing helix F, shown previously to be the most flexible or unfolded region in apoMb.[34] Gln152, which is a secondary site of modification, is located at the flexible C-terminus of the protein chain, as given by NMR measurements.[34] The observation that the modification at Gln152 by TGase occurs slower than that at Gln91 can be explained also in light of its proximity to the negatively charged carboxylate of the C-terminus of the protein. Indeed, negative charges near Gln residues can slow the enzymatic modification because of the unfavorable interaction of the polypeptide substrate with the negatively charged active site of microbial TGase.[23,35]

An exception to the correlation between the flexibility of the protein substrate and the specificity of the TGase-mediated reaction appears to be Lys87, which is also in the region of helix F but is not modified. Lys87 is followed by Pro88, and this fact likely can account for the resistance to modification. Indeed, similar effects of the cyclic amino acid Pro are known for proteases, which typically do not hydrolyze X−Pro peptide bonds. For example, trypsin cleaves peptide bonds specifically at the C-terminal side of Lys and Arg residues, but when these residues have a Pro residue at their carboxylic side, they become highly resistant to cleavage.[36] Indeed, trypsin does not

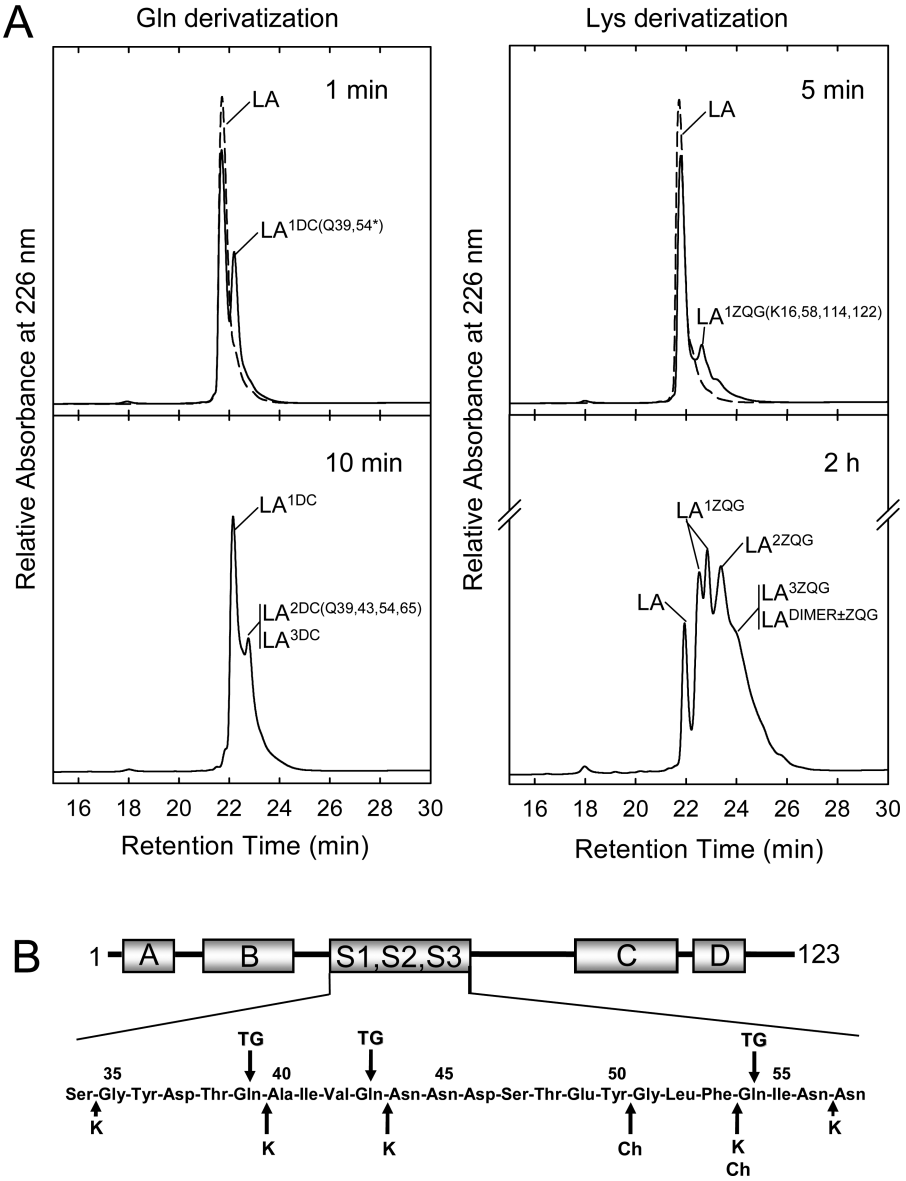


Figure 3. TGase-mediated conjugation of apo-α-LA at Gln and Lys residues. (A) RP-HPLC analyses of the TGase-mediated derivatives of apo-α-LA. A dashed line and a straight line indicate the chromatograms of native apo-α-LA and the DC-conjugated (left) and ZQG-conjugated (right) species at different reaction times. The initial and major reaction after 1 min occurs at Gln39, while a very minor derivatization of Gln54 (*) has been detected by ESI-MS. (B) Scheme of the secondary structure of α-LA. The four α-helices (A–D) and the three β-strands (S1–S3) along the 123-residue chain of the protein are indicated by boxes.[40] The amino acid sequence of the region of residues 34–57 encompassing the β-strands of the protein is explicitly shown, and the sites of TGase modification at Gln39, Gln43, and Gln54 are indicated. The initial sites of limited proteolysis of apo-α-LA by proteinase K (K) and chymotrypsin (Ch) are indicated by long arrows, while short arrows indicate secondary sites of cleavage.[42,43]

hydrolyze the Lys87–Pro88 peptide bond in limited proteolysis experiments conducted with apoMb.[33] The effect of Pro on the TGase reaction has not been studied so far for Lys residues, whereas it has been demonstrated that Gln residues located N-terminally to Pro residues react at a very slow rate[37] or their derivatization is completely inhibited, as in the case of the TGase-mediated modification of G-CSF.[38] Therefore, our results appear to indicate that Pro also has an inhibitory effect on Lys derivatization by TGase.

α-Lactalbumin. α-LA is a calcium binding protein[39] whose polypeptide sequence contains 6 Gln and 12 Lys residues that can be potentially derivatized by TGase (Table S1 of the Supporting Information). The three-dimensional (3D) structure of α-LA is characterized by two domains, a discontinuous helical domain with four α-helices denoted A–D, comprising regions of residues 1–39 and 81–123 and a β-sheet/coil domain (β-domain), which encompasses the rest of the polypeptide chain (Figure S1 of the Supporting Information).[40] For the past few decades, α-LA has been extensively studied as a model of protein folding, and in particular, the partly folded or molten globule state that it acquires at acidic pH (A-state) or upon removal of calcium at neutral pH has been extensively investigated.[41] Here, we analyze the reactivity of Gln and Lys residues of apo-α-LA in its molten globule state at neutral pH and moderately high temperatures. The derivatization reaction of apo-α-LA was indeed performed at 37 °C, while with the other two protein substrates, apoMb and fragment 205–316 of thermolysin (see below), the TGase reaction was conducted at 25 °C. These experimental conditions were those previously

dx.doi.org/10.1021/bi301005z | *Biochemistry* 2012, 51, 8679–8689

used in our laboratory to conduct limited proteolysis experiments with these proteins.[42,43]

The TGase-mediated derivatization of Gln residues of apo-α-LA with DC occurs very fast, as observed from the RP-HPLC analyses of the reaction mixture (Figure 3A). Indeed, after reaction for 1 min, the product is a monoderivative species eluting with a slightly higher retention time compared to that of the native protein (Table S4 of the Supporting Information). After reaction for 10 min, the chromatographic peak corresponding to the native protein has disappeared and two new peaks are detected in the chromatogram, corresponding to mono-, bi-, and triderivatives of the protein (Table S4 of the Supporting Information). To identify the sites of initial modification, the monoderivative species produced after reaction for 1 min was subjected to mass fingerprinting analysis. Gln39 was identified as the first site of TGase modification (Table S5 of the Supporting Information), while Gln54 is only partly derivatized. The same analysis was performed on the bi- and triderivatives produced after incubation for 10 min, and Gln43 and, to a minor extent, Gln65 were identified as secondary sites of modification (Table S5 of the Supporting Information). The results reported here differ somewhat from those of previous studies of the reactivity to TGase of Gln residues of apo-α-LA, where Gln54 or Gln39 and Gln43 were identified as preferential sites of modification.[44,45]

The TGase-mediated derivatization of apo-α-LA incubated in the presence of ZQG proceeds much more slowly than the derivatization of Gln residues with DC. Indeed, RP-HPLC analysis of the reaction mixture after incubation for 5 min shows the formation of a new chromatographic peak with a slightly higher retention time but a low intensity (Figure 3A). This peak was shown by ESI-MS analysis to contain a mono-ZQG derivative of α-LA (Table S4 of the Supporting Information) conjugated at the level of one Lys residue (Lys16, Lys58, Lys114, or Lys122), as determined by mass fingerprinting analysis (Table S5 of the Supporting Information). RP-HPLC of the reaction mixture after incubation for 2 h still showed the presence of the peak corresponding to the native protein, together with four additional poorly separated peaks, which by MS were shown to contain mono-, bi-, and tri-ZQG derivatives of α-LA (Table S4 of the Supporting Information). After incubation for 2 h, we also detected the presence of protein dimers and dimer species conjugated to up to three molecules of ZQG. Overall, the Lys modification of apo-α-LA by TGase is slow and not specific.

In previous studies,[40,42,43] we have shown that limited proteolysis of apo-α-LA in its molten globule state occurs at a rather short region of the 123-residue chain of the protein, from approximately residue 34 to 56, encompassing the three β-strands (β-domain) in the native protein (Figure 3B). Therefore, limited proteolysis experiments indicated that in the molten globule of apo-α-LA the chain region corresponding to the β-domain of the native protein is flexible or unfolded, whereas the helical domain remains folded, in agreement with the results of other physicochemical measurements.[41,46] Here, we show that the TGase-mediated derivatization of apo-α-LA at Gln residues is rather specific, the first site of DC conjugation being Gln39, followed by a slower reaction at Gln43, Gln54, and Gln65. Most interestingly, the modified Gln residues are all located in the disordered region encompassing the β-domain, shown to be flexible or disordered by limited proteolysis experiments and physicochemical measurements (Figure 3B).

While the TGase-mediated reaction at Gln residues occurs fast, the reactivity of Lys residues of apo-α-LA in the presence of the acyl donor ZQG is much slower (see Figure 3A). The reason for the observed lower reactivity, as well as specificity, of the TGase reaction at Lys residues of α-LA can be envisaged in the absence of Lys residues in the amino acid sequence encompassing the flexible β-domain region. Indeed, there are no Lys residues in the region of residues 17−57 (see Figure S1 of the Supporting Information). Nevertheless, in apo-α-LA there is a substantial weakening of native tertiary interactions, as indicated by the consistent reduction of the circular dichroism signal in the near-UV region and by NMR data.[43] TGase in apo-α-LA can thus derivatize, albeit with low efficiency, also Lys residues, even if outside the disordered β-domain (see above). In this study, we have conducted limited proteolysis experiments on apo-α-LA with trypsin (Figure S4 and Table S6 of the Supporting Information) to attempt to correlate the Lys reactivity in apo-α-LA observed with TGase with that of trypsin. Indeed, trypsin hydrolyzes a protein specifically at the level of Lys and Arg residues, and these residues are absent in the β-domain region of the protein (see Figure S1 of the Supporting Information). Like TGase, the initial sites of nicking of trypsin are also randomly distributed along the polypeptide chain of α-LA (see Figure S4 and Table S6 of the Supporting Information), while under identical conditions, proteases with broad substrate specificity (e.g., chymotrypsin and proteinase K) initially hydrolyze the protein selectively at the level of the β-domain (Figure 3B). Therefore, we see a parallel of Lys modification by TGase and trypsin in that both modify the apo-α-LA substrate slowly and with poor selectivity, because amino acid residues suitable for both enzymes are absent in the disordered β-domain region.

Here, we also have investigated the reactivity to TGase of the calcium-bound form of α-LA (holo-α-LA), which was shown before to be resistant to proteolytic digestion.[42,43] It was found that the holoprotein is fully resistant to the TGase-mediated modification at Gln with DC, as given by RP-HPLC and MS analysis of the reaction mixture (not shown). A simple interpretation of the lack of reactivity of Gln residues is that the folded and rigid structure of the native holoprotein prevents the enzymatic reaction of TGase. On the other hand, incubation of holo-α-LA with TGase in the presence of ZQG leads to the quantitative modification of the protein at Lys, as shown by the RP-HPLC analysis of the reaction mixture after incubation for 2 h (Figure S5A of the Supporting Information). ESI-MS and mass fingerprinting analysis identified the product of the reaction as a monoderivative and Lys122 as the site of modification (Tables S4 and S5 and Figure S3 of the Supporting Information). The reactivity of Lys122 can be explained by considering that the N- and C-terminal ends of the polypeptide chains of globular proteins usually are quite flexible. Moreover, the flexibility at the C-terminus of holo-α-LA results also by an examination of the temperature or *B* factor profile along the 123-residue chain of the protein (Figure S5B of the Supporting Information).[47,48] Indeed, the *B* factor provides a graphic image of the degree of mobility existing along a protein chain, and the highest values of this factor occur at the C-terminal end of the protein chain. Thus, the highly specific derivatization of Lys122 in holo-α-LA derives from the high flexibility of the substrate.

Fragment 205−316 of Thermolysin. The C-terminal fragment 205/206−316 is capable of folding into a nativelike structure in a manner independent of the rest of the

dx.doi.org/10.1021/bi301005z | *Biochemistry* 2012, 51, 8679−8689

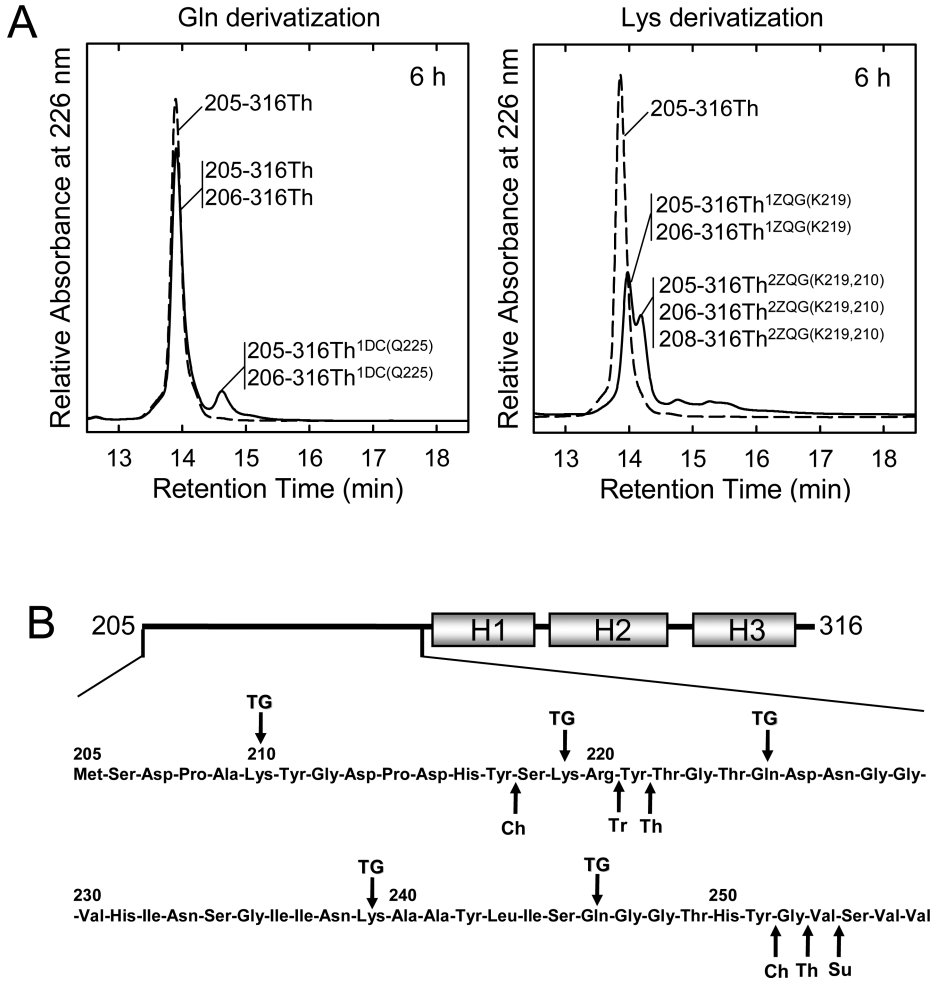


Figure 4. TGase-mediated conjugation of 205−316Th at Gln and Lys residues. (A) RP-HPLC analyses of the TGase-mediated derivatives of 205−316Th. A dashed line and a straight line indicate the chromatograms of the native thermolysin fragment and of the DC- and ZQG-conjugated species, respectively, after reaction for 6 h. (B) Scheme of the secondary structure of 205−316Th.[51,52] The three helices (H1−H3) are depicted as boxes. The amino acid sequence of the N-terminally disordered region of residues 205−256 is explicitly shown, and the sites of TGase-mediated modification at Gln225, Gln246, Lys210, Lys219, and Lys239 are denoted by arrows. The sites of initial proteolytic cleavage of the fragment by thermolysin (Th), subtilisin (Su), trypsin (Tr), and chymotrypsin (Ch) are also shown.[49]

polypeptide chain, thus possessing protein domain properties.[49−52] Here, we explored the TGase modification at both Gln and Lys residues of fragment 205−316, containing in its amino acid sequence seven Gln and seven Lys residues (Figure S1 and Table S1 of the Supporting Information). Analysis by RP-HPLC and ESI-MS of the reaction mixture with DC after reaction for 6 h showed that the N-terminal Met can be partly removed during the reaction (Figure 4A), as shown by the ESI-MS data (Table S7 of the Supporting Information). Because the removal of Met205 occurs when the fragment is incubated in the presence of TGase only and without DC or ZQG (data not shown), it seems that the TGase sample likely is contaminated by a protease. Indeed, the commercial preparation of TGase used in this study is produced by fermentation of a strain of *S. mobaraensis*, and it has already been suggested that it can contain a cosecreted protease.[53] The reaction with TGase and DC leads to a derivative that elutes at a retention time that is higher than that of the original fragment, corresponding to a monoderivative of fragment 205/206−316 with DC linked at Gln225 (Figure 4A and Tables S7 and S8 of the Supporting Information). Prolonged incubation of the reaction mixture for up to 24 h yields a doubly derivatized species of fragment 206−316, and Gln246 was identified as the second site of derivatization (Tables S7 and S8 of the Supporting Information). The conjugation reaction at Gln225 appears to be quite slow, probably because of the presence of nearby Asp226 (Figure 4B). The negatively charged side chain carboxylate of Asp226 can make difficult the interaction of the substrate at the negatively charged active site of TGase (see also above).[23,35]

The Lys conjugation of fragment 205−316 to ZQG occurs much faster than the Gln conjugation, because the RP-HPLC analysis of the reaction mixture after incubation for 6 h reveals the full disappearance of the chromatographic peak of the fragment and the appearance of two poorly separated peaks, identified by ESI-MS as mono- and biconjugated fragment derivatives (Figure 4A and Table S7 of the Supporting Information). Derivatives of fragments 206−316Th and 208−316Th were also detected, likely because of a proteolytic digestion during the incubation with the commercial sample of TGase (see above). Trypsin digestion of the ZQG derivatives followed by ESI-MS analysis demonstrated that the first site of modification is at Lys219, while the secondary site is at Lys210 (Table S8 of the Supporting Information). Further incubation

dx.doi.org/10.1021/bi301005z | *Biochemistry* 2012, 51, 8679−8689

of the reaction mixture for up to 24 h yields a third site of modification, identified as Lys239 (Table S8 of the Supporting Information). The TGase reaction at Lys appears to occur in a stepwise fashion, because the preferential site of modification is Lys219, followed by Lys210 and Lys239.

In previous studies, limited proteolysis experiments were conducted on the cyanogen bromide fragment of residues 206−316 of thermolysin with the aim of defining the minimal size of a fragment capable of an independent folding.[49] This fragment was digested with several proteases (Figure 4B), and it was found that proteolysis occurred only in the N-terminal region, leading to the rather short fragment of residues 255−316 that is quite resistant to further digestion with subtilisin, implying a tightly folded conformation.[26] NMR measurements on 205−316Th confirmed the disordered N-terminal region (approximately residues 205−261) and a folded and nativelike helical structure in the C-terminal region.[51,52] Overall, the results of the derivatization of Gln and Lys residues of 205−316Th with TGase correlate with those of limited proteolysis, because TGase and proteases attack the fragment substrate at its disordered N-terminal region (see Figure 4B).

CONCLUSIONS

Here, we have convincingly shown that the site-specific reactions mediated by microbial TGase occur in flexible or disordered regions of the polypeptide substrates. In this study, we have used as model proteins apoMb, apo-α-LA, and 205/206−316Th, previously subjected in our laboratory to systematic limited proteolysis experiments. The main finding is that both TGase and proteases can selectively attack these model proteins in their disordered chain regions, as outlined in Figure S6 of the Supporting Information. The fact that both TGase and proteases appear to recognize similar molecular features in their polypeptide substrates is in keeping with the notion that TGase can be considered a "reverse protease" synthesizing instead of hydrolyzing amide moieties (see Figure 1) and the fact that the active site of TGase is given by the catalytic Cys-Asp-His triad typical of Cys proteases such as papain.[23] Therefore, we are inclined to conclude that the molecular features that dictate a site-specific proteolytic event on globular proteins apply also to TGase-mediated selective reactions. Here, we report the experimental results obtained with microbial TGase, and therefore, one could argue that the results with the microbial enzyme do not necessarily apply to those that can be obtained with other mammalian TGases, considering their significant differences in terms of amino acid sequence, molecular mass, and 3D structure. Nevertheless, we have analyzed a number of published mammalian TGase-mediated modifications of proteins with known 3D structures and dynamics and concluded that chain flexibility or local unfolding can invariably account for the observed site-specific modifications (to be published).

Site-Specific Derivatization of Gln Residues. The rate-limiting step for the TGase-mediated catalysis involves the interaction of a Gln residue of a polypeptide substrate with the Cys64 residue of the enzyme's active site, forming a reactive thioacyl moiety. Once the reactive thioester intermediate is formed, it undergoes a reaction with a primary amine, leading to an amide bond, as schematically shown in Figure 5. The polypeptide substrate should bind at the catalytic Cys64 residue, which in microbial TGase is buried at the bottom of a 16 Å deep cleft (Figure 5A).[23] Moreover, alanine scanning mutagenesis studies on this enzyme indicated that an extended

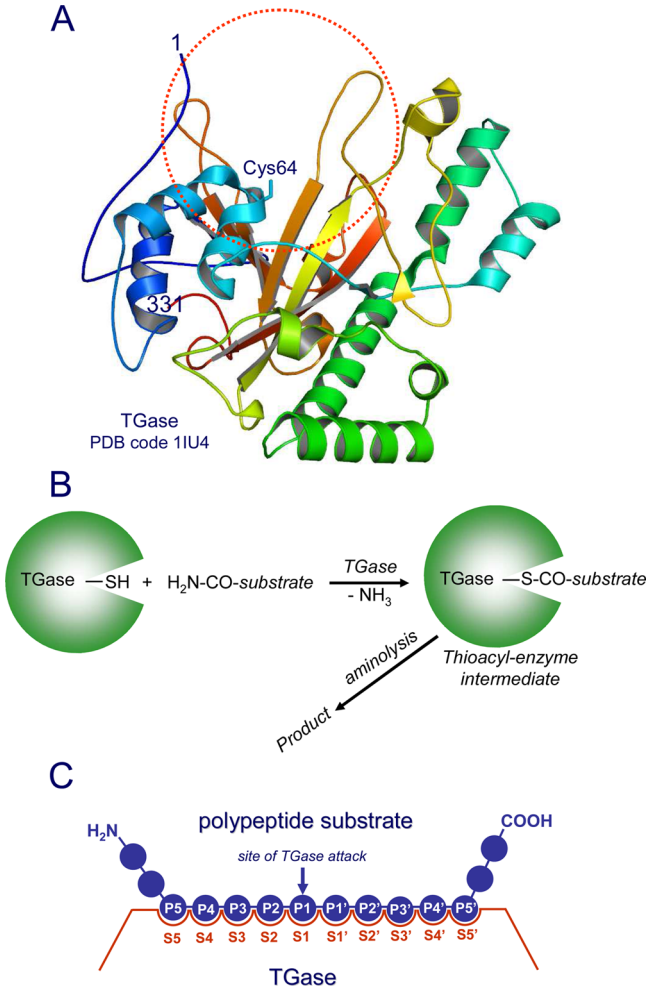


Figure 5. Interaction of a polypeptide substrate at the active site of TGase. (A) Three-dimensional structure of microbial TGase from *S. mobaraensis* (PDB entry 1IU4).[23] The numbers 1 and 331 indicate the N- and C-terminus, respectively, of the 331-residue chain of TGase. The active site area of the enzyme is circled, and the catalytic residue Cys64 is indicated. Clearly, Cys64 is quite buried in the protein interior. (B) Schematic drawing of the TGase-mediated derivatization of a protein substrate at the level of Gln residues. The key step in the enzymatic reaction involves formation of the thioacyl−enzyme intermediate at the level of Cys64. (C) Schematic representation of the binding of a polypeptide substrate at the active site of TGase. A 10−12-residue segment of a polypeptide chain interacts with its side chain residues (P) at a series of subsites (S) of TGase. The P1 side chain residue interacting with the S1 binding site of TGase is the carboxamido side chain group of a Gln residue. The scheme highlights the fact that the polypeptide substrate interacts at the enzyme's active site in an extended conformation (see the text). The figure and nomenclature are adapted from the representation of a similar binding of a polypeptide substrate at the protease's active site introduced by Schechter and Berger.[58]

surface along the active site cleft is involved in the recognition of a protein substrate, larger than the one involved in the interaction with a small molecule as ZQG.[54] In this respect, the interaction between TGase and a protein substrate is similar to that described for proteases.[14,55] In this case, the protease hydrolysis requires the interaction at the enzyme's active site of a stretch of up to 12 amino acid residues of the polypeptide substrate, as given by modeling studies.[56] Moreover, the analysis of many crystallographic structures of polypeptide

dx.doi.org/10.1021/bi301005z | *Biochemistry* 2012, 51, 8679−8689

substrates bound to proteases revealed that the efficient binding at the active site invariably involves a polypeptide substrate in an extended conformation.[57] Clearly, significant backbone flexibility is required for the productive binding of a polypeptide substrate at the protease's active site, thus explaining why folded globular proteins are rather stable to digestion while the unfolded ones are much more easily attacked by a protease. These data support the nomenclature used for describing the interaction of a polypeptide substrate at the protease's active site introduced by Schechter and Berger.[58] An at least 10−12-residue segment of the polypeptide substrate interacts with its side chain residues (P) at a series of subsites (S) of a protease. Considering the strong correlation between sites of TGase reactions and sites of proteolysis demonstrated herein, we propose that a similar representation can be used to describe the interaction of a polypeptide substrate at the TGase's active site (see Figure 5C). Therefore, the site-specific modification by TGase at Gln residues requires that this residue be embedded in a flexible or disordered region of a protein substrate, as firmly demonstrated by the results of our study.

Site-Specific Derivatization of Lys Residues. The ε-amino group in Lys residues is spaced from the polypeptide backbone by four carbon atoms, and this exposed alkylamino chain can in principle react easily with the thioacyl ester intermediate to form the amide bond (see Figure 5B).[3] Indeed, the reactivity of Lys residues toward TGase-mediated reactions is considered to be less site-specific than that for Gln residues. However, we demonstrate here that some backbone flexibility is also required for the TGase-mediated reaction at Lys. Indeed, in apoMb and fragment 205−316, the Lys modification occurs selectively only in the disordered regions of these proteins (see Figures 2B and 4B). Moreover, with the rigid and folded holo-α-LA, a strikingly selective modification occurs at the level of the Lys122 residue that is located at the flexible C-terminal end of the polypeptide chain, while many other Lys residues of the protein do not react with TGase. The requirement of chain flexibility also explains why only a Lys residue embedded in a large loop in an engineered alkaline phosphatase can be modified by TGase.[59] Similarly, the TGase-mediated modification of human growth hormone occurs specifically at Lys145, this residue being embedded in a disordered region of the protein (residues 134−149) prone to limited proteolysis.[60] Therefore, our results overall demonstrate that, besides Gln, Lys derivatization by TGase requires some flexibility of the protein substrate. The site-specific modification at Lys has not been exploited so far for preparing bioconjugates of proteins, but it seems that with some protein substrates Lys derivatization can provide the advantages already proven with the TGase modification at Gln.

Outlook. Previous studies of the determinants of TGase specificity identified amino acid sequence requirements for Gln modification.[5,35,61] Different rates of reaction were indeed measured upon variation of the charge and hydrophobicity of the amino acid sequence near the Gln residue. These observations can be related to analogous results obtained in studying the substrate specificities of proteolytic enzymes (see the ExPASy website at http://web.expasy.org/peptide_cutter/peptidecutter_enzymes.html), such as the inhibitory effect on tryptic cleavage of negatively charged amino acid residues near Lys and Arg residues. In the case of TGase, the preferred substrate sequences are expected to influence the modification reaction when short and flexible peptides are used as substrates, while flexibility or unfolding is the most critical parameter in dictating the enzymatic reaction in the case of globular protein substrates. In other studies, it was proposed that the surface accessibility of a reactive Gln (or Lys) could dictate the TGase-catalyzed modification.[16] However, surface exposure alone does not justify at all the fact that plenty of exposed Gln (and Lys) residues in native globular proteins with known 3D structures do not react with TGase.[14] Instead, our data indicate that a local unfolding of the chain region of a polypeptide substrate strongly favors a site-specific protein modification by TGase, provided that a Gln or a Lys residue is present in that region. Therefore, it is possible to predict the site(s) of TGase-mediated modification of a protein on the basis of its structure and dynamics and, consequently, the likely effects on its physicochemical and functional properties.

It is tempting to propose that the molecular features herein emphasized for a specific enzymatic reaction with TGase or a protease on a polypeptide substrate can also be extended to other site-specific enzymatic reactions occurring with proteins, such as phosphorylation, glycosylation, and other reactions. For example, it is intriguing to observe that the TGase-mediated PEGylation of G-CSF occurs at Gln134, while in vivo glycosylation occurs at the nearby Thr133.[38] Protein disorder, therefore, appears to be important in a variety of biomolecular recognition processes of proteins.[62,63] In summary, we hypothesize that the requirement of disordered regions in the protein substrate for an effective interaction at the enzyme's active site applies to TGases and proteases, and likely to other enzymes that in vivo perform site-specific post-translational modifications of proteins.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

Amino acid sequences of apoMb, α-LA, and 205−316Th (Figure S1), MS/MS spectrum of the triply charged ion at m/z 766.68 of peptide 32−49 of apo-α-LA modified at Gln39 with DC (Figure S2), MS/MS spectrum of the doubly charged ion at m/z 734.36 of peptide 115−123 of holo-α-LA modified at Lys122 with ZQG (Figure S3), RP-HPLC analysis of the reaction mixture of apo-α-LA reacted with trypsin after reaction for 15 min (Figure S4), TGase-mediated conjugation of ZQG to Lys122 in holo-α-LA (Figure S5), schematic representation of the 3D structures of myoglobin, α-lactalbumin, and thermolysin fragment 205/206−316 (Figure S6), the number of Gln and Lys residues in the amino acid sequences of the three model proteins being studied (Table S1), molecular masses of apoMb and its DC and ZQG derivatives (Table S2), molecular masses of the fragments obtained upon digestion of DC and ZQG derivatives of apoMb with only trypsin or with both trypsin and V8 protease (Table S3), molecular masses of the DC and ZQG derivatives of α-LA (Table S4), molecular masses of fragments obtained from the DC and ZQG derivatives of α-LA upon digestion with trypsin and V8 protease, followed by reduction of the reaction mixture with TCEP (Table S5), molecular masses of the nicked forms of apo-α-LA produced by limited proteolysis with trypsin, followed by reduction with TCEP (Table S6), molecular masses of the DC and ZQG derivatives of 205−316Th (Table S7), and molecular masses of the fragments obtained upon digestion with trypsin of the DC and ZQG derivatives of 205−316Th (Table S8). This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*CRIBI Biotechnology Centre, University of Padua, Viale G. Colombo 3, 35121 Padua, Italy. Telephone: +39-049-827 6155. Fax: +39-049-8276159. E-mail: barbara.spolaore@unipd.it.

Present Address

[†]Department of Biochemistry and Molecular Biology, University of Parma, Parma, Italy.

Author Contributions

B.S. and S.R. contributed equally to this work.

Funding

This work was supported by Research Project 2009 of the University of Padua. A.R.M. was a recipient of an Erasmus fellowship, while A.S. and N.D. were supported by a grant from the CARIPARO Foundation (Padua, Italy).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Marcello Zambonin for the preparation of thermolysin fragment 205−316 and for conducting some of the experiments reported here.

ABBREVIATIONS

ACN, acetonitrile; apoMb, apomyoglobin; holoMb, heme-containing myoglobin; α-LA, α-lactalbumin; apo-α-LA, Ca^{2+}-depleted α-LA; holo-α-LA, Ca^{2+}-bound α-LA; CD, circular dichroism; DC, dansylcadaverine; DMSO, dimethyl sulfoxide; EDTA, ethylenediaminetetraacetic acid; E/S, enzyme/substrate; ESI-MS, electrospray ionization mass spectrometry; HPLC, high-performance liquid chromatography; PDB, Protein Data Bank; PEG, poly(ethylene glycol); RP, reverse-phase; RT, retention time; TCEP, tris(carboxyethyl)phosphine; TFA, trifluoroacetic acid; TGase, transglutaminase; Th, thermolysin; ZQG, carbobenzoxy-L-glutaminylglycine.

REFERENCES

(1) Folk, J. E. (1983) Mechanism and basis for specificity of transglutaminase-catalyzed ε-(γ-glutamyl)lysine bond formation. *Adv. Enzymol. Relat. Areas Mol. Biol. 54*, 1−56.

(2) Griffin, R., Casadio, R., and Bergamini, C. M. (2002) Transglutaminases: Nature's biological glues. *Biochem. J. 368*, 377−396.

(3) Ohtsuka, T., Sawa, A., Kawabata, R., Nio, N., and Motoki, M. (2000) Substrate specificities of microbial transglutaminase for primary amines. *J. Agric. Food Chem. 48*, 230−233.

(4) Kulik, C., Heine, E., Weichold, O., and Möller, M. (2009) Synthetic substrates as amine donors and acceptors in microbial transglutaminase-catalysed reactions. *J. Mol. Catal. B: Enzym. 57*, 237−241.

(5) Ohtsuka, T., Ota, M., Nio, N., and Motoki, M. (2000) Comparison of substrate specificities of transglutaminases using synthetic peptides as acyl donors. *Biosci., Biotechnol., Biochem. 64*, 2608−2613.

(6) Yokoyama, K., Nio, N., and Kikuchi, Y. (2004) Properties and applications of microbial transglutaminase. *Appl. Microbiol. Biotechnol. 64*, 447−454.

(7) Zhu, Y., and Tramper, J. (2008) Novel applications for microbial transglutaminase beyond food processing. *Trends Biotechnol. 26*, 559−565.

(8) Kamiya, N., Abe, H., Goto, M., Tsuji, Y., and Jikuya, H. (2009) Fluorescent substrates for covalent protein labeling catalyzed by microbial transglutaminase. *Org. Biomol. Chem. 7*, 3407−3412.

(9) Jeger, S., Zimmermann, K., Blanc, A., Grünberg, J., Honer, M., Hunziker, P., Struthers, H., and Schibli, R. (2010) Site-specific and stoichiometric modification of antibodies by bacterial transglutaminase. *Angew. Chem., Int. Ed. 49*, 9995−9997.

(10) Tominaga, J., Kemori, Y., Tanaka, Y., Maruyama, T., Kamiya, N., and Goto, M. (2007) An enzymatic method for site-specific labeling of recombinant proteins with oligonucleotides. *Chem. Commun.,*, 401−403.

(11) Abe, H., Goto, M., and Kamiya, N. (2011) Protein lipidation catalyzed by microbial transglutaminase. *Chemistry 17*, 14004−14008.

(12) Sato, H. (2002) Enzymatic procedure for site-specific pegylation of proteins. *Adv. Drug Delivery Rev. 54*, 487−504.

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

(14) Fontana, A., Spolaore, B., Mero, A., and Veronese, F. M. (2008) Site-specific modification and PEGylation of pharmaceutical proteins mediated by transglutaminase. *Adv. Drug Delivery Rev. 60*, 13−28.

(15) Mero, A., Spolaore, B., Veronese, F. M., and Fontana, A. (2009) Transglutaminase-mediated PEGylation of proteins: Direct identification of the sites of protein modification by mass spectrometry using a novel monodisperse PEG. *Bioconjugate Chem. 20*, 384−389.

(16) Coussons, P. J., Price, N. C., Kelly, S. M., Smith, B., and Sawyer, L. (1992) Factors that govern the specificity of transglutaminase-catalysed modification of proteins and peptides. *Biochem. J. 282*, 929−930.

(17) Case, A., and Smith, R. L. (2003) Kinetic analysis of the action of tissue transglutaminase on peptide and protein substrates. *Biochemistry 42*, 9466−9481.

(18) Sugimura, Y., Hosono, M., Wada, F., Yoshimura, T., Maki, M., and 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.

(19) Facchiano, F., and Facchiano, A. (2005) Transglutaminases and their substrates. *Prog. Exp. Tumor Res. 38*, 37−57.

(20) Csosz, E., Bagossi, P., Nagy, Z., Dosztanyi, Z., Simon, I., and Fesus, L. (2008) Substrate preference of transglutaminase 2 revealed by logistic regression analysis and intrinsic disorder examination. *J. Mol. Biol. 383*, 390−402.

(21) Murthy, S. N., Lukas, T. J., Jardetzky, T. S., and Lorand, L. (2009) Selectivity in the post-translational, transglutaminase-dependent acylation of lysine residues. *Biochemistry 48*, 2654−2660.

(22) Shimba, N., Yokoyama, K., and Suzuki, E. (2002) NMR-based screening method for transglutaminases: Rapid analysis of their substrate specificities and reaction rates. *J. Agric. Food Chem. 50*, 1330−1334.

(23) Kashiwagi, T., Yokoyama, K., Ishikawa, K., Ono, K., Ejima, D., Matui, H., and Suzuki, E. (2002) Crystal structure of microbial transglutaminase from *Streptoverticillium mobaraense*. *J. Biol. Chem. 277*, 44252−44260.

(24) Fontana, A., Fassina, G., Vita, C., Dalzoppo, D., Zamai, M., and Zambonin, M. (1986) Correlation between sites of limited proteolysis and segmental mobility in thermolysin. *Biochemistry 25*, 1847−1851.

(25) Fontana, A., Polverino de Laureto, P., De Filippis, V., Scaramella, E., and Zambonin, M. (1997) Probing the partly folded states of proteins by limited proteolysis. *Folding Des. 2*, R17−R26.

(26) Fontana, A., Polverino de Laureto, P., Spolaore, B., Frare, E., Picotti, P., and Zambonin, M. (2004) Probing protein structure by limited proteolysis. *Acta Biochim. Pol. 51*, 299−321.

(27) Ascoli, F., Fanelli, M. R., and Antonini, E. (1981) Preparation and properties of apohemoglobin and reconstituted hemoglobins. *Methods Enzymol. 76*, 72−87.

(28) Fassina, G., Vita, C., Dalzoppo, D., Zamai, M., Zambonin, M., and Fontana, A. (1986) Autolysis of thermolysin: Isolation and characterization of a folded three-fragment complex. *Eur. J. Biochem. 156*, 221−228.

(29) Gill, S. C., and von Hippel, P. H. (1989) Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem. 182*, 319−326.
(30) Jennings, P. A., and Wright, P. E. (1993) Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin. *Science 262*, 892−896.
(31) Uzawa, T., Nishimura, C., Akiyama, S., Ishimori, K., Takahashi, S., Dyson, H. J., and Wright, P. E. (2008) Hierarchical folding mechanism of apomyoglobin revealed by ultra-fast H/D exchange coupled with 2D NMR. *Proc. Natl. Acad. Sci. U.S.A. 105*, 13859−13864.
(32) Evans, S. V., and Brayer, G. D. (1990) High-resolution study of the three-dimensional structure of horse heart metmyoglobin. *J. Mol. Biol. 213*, 885−897.
(33) Fontana, A., Zambonin, M., Polverino de Laureto, P., De Filippis, V., Clementi, A., and Scaramella, E. (1997) Probing the conformational state of apomyoglobin by limited proteolysis. *J. Mol. Biol. 266*, 223−230.
(34) Eliezer, D., and Wright, P. E. (1996) Is apomyoglobin a molten globule? Structural characterization by NMR. *J. Mol. Biol. 263*, 531−538.
(35) Sugimura, Y., Yokoyama, K., Nio, N., Maki, M., and Hitomi, K. (2008) Identification of preferred substrate sequences of microbial transglutaminase from *Streptomyces mobaraensis* using a phage-displayed peptide library. *Arch. Biochem. Biophys. 477*, 379−383.
(36) Halfon, S., Baird, T. T., and Craik, C. S. (2004) Trypsin. In *Handbook of Proteolytic Enzymes* (Barrett, A. J., Rawings, N. D., and Woesnner, J. F., Eds.) pp 1483−1488, Elsevier Academic Press, San Diego.
(37) Piersma, S. R., van de Pijpekamp, A., Wijngaards, G., Gruppen, H., and Boumans, H. (2002) Quantitation and localisation of (in vitro) transglutaminase-catalyzed glutamine hydroxylation using mass spectrometry. *Enzyme Microb. Technol. 30*, 266−272.
(38) Fontana, A., Spolaore, B., Mero, A., and Veronese, F. M. (2009) The site-specific TGase-mediated PEGylation of proteins occurs at flexible sites. In *PEGylated Protein Drugs: Basic Science and Clinical Applications* (Veronese, F. M., Ed.) pp 89−112, Birkhäuser Verlag, Basel, Switzerland.
(39) Permyakov, E. A., and Berliner, L. J. (2000) α-Lactalbumin: Structure and function. *FEBS Lett. 473*, 269−274.
(40) Pike, A. C., Brew, W. K., and Acharya, K. R. (1996) Crystal structures of guinea-pig, goat and bovine α-lactalbumin highlight the enhanced conformational flexibility of regions that are significant for its action in lactose synthase. *Structure 4*, 691−703.
(41) Kuwajima, K. (1996) The molten globule state of α-lactalbumin. *FASEB J. 10*, 102−109.
(42) Polverino de Laureto, P., Scaramella, E., Frigo, M., Wondrich, F. G., De Filippis, V., Zambonin, M., and Fontana, A. (1999) Limited proteolysis of bovine α-lactalbumin: Isolation and characterization of protein domains. *Protein Sci. 8*, 2290−2303.
(43) Polverino de Laureto, P., Frare, E., Gottardo, R., and Fontana, A. (2002) Molten globule of bovine α-lactalbumin at neutral pH induced by heat, trifluoroethanol and oleic acid: A comparative analysis by circular dichroism spectroscopy and limited proteolysis. *Proteins 49*, 385−397.
(44) Matsumura, Y., Chanyongvorakul, Y., Kumazawa, Y., Ohtsuka, T., and Mori, T. (1996) Enhanced susceptibility to transglutaminase reaction of α-lactalbumin in the molten globule state. *Biochim. Biophys. Acta 1292*, 69−76.
(45) Nieuwenhuisen, W. F., Dekker, H. L., de Koning, L. J., Gröneveld, T., de Koster, C. G., and de Jong, G. A. (2003) Modification of glutamine and lysine residues in holo and apo α-lactalbumin with microbial transglutaminase. *J. Agric. Food Chem. 51*, 7132−7139.
(46) Schulman, B. A., Kim, P. S., Dobson, C. M., and Redfield, C. (1997) A residue-specific NMR view of the non-cooperative unfolding of a molten globule. *Nat. Struct. Biol. 4*, 630−634.
(47) Frauenfelder, H., Petsko, G. A., and Tsernoglou, D. (1979) Temperature-dependent X-ray diffraction as a probe of protein structural dynamics. *Nature 280*, 558−563.
(48) Ringe, D., and Petsko, G. A. (1985) Mapping protein dynamics by X-ray diffraction. *Prog. Biophys. Mol. Biol. 45*, 197−235.
(49) Dalzoppo, D., Vita, C., and Fontana, A. (1985) Folding of thermolysin fragments: Identification of the minimum size of a carboxyl-terminal fragment that can fold into a stable native-like structure. *J. Mol. Biol. 182*, 331−340.
(50) Vita, C., and Fontana, A. (1982) Domain characteristics of the carboxyl-terminal fragment 206−316 of thermolysin: Unfolding thermodynamics. *Biochemistry 21*, 5196−5202.
(51) Conejero-Lara, F., González, C., Jiménez, M. A., Padmanabhan, S., Mateo, P. L., and Rico, M. (1997) NMR solution structure of the 205−316 C-terminal fragment of thermolysin. An example of dimerization coupled to partial unfolding. *Biochemistry 36*, 11975−11983.
(52) Rico, M., Jiménez, M. A., Gonzales, C., De Filippis, V., and Fontana, A. (1994) NMR solution structure of the C-terminal fragment 255−316 of thermolysin: A dimer formed by subunits having the native structure. *Biochemistry 33*, 14834−14847.
(53) Sommer, C., Hertel, T. C., Schmelzer, C. E., and Pietzsch, M. (2012) Investigations on the activation of recombinant microbial pro-transglutaminase: In contrast to proteinase K, Dispase removes the histidine tag. *Amino Acids 42*, 997−1006.
(54) Tagami, U., Shimba, N., Nakamura, M., Yokoyama, K., Suzuki, E., and Hirokawa, T. (2009) Substrate specificity of microbial transglutaminase as revealed by three-dimensional docking simulation and mutagenesis. *Protein Eng., Des. Sel. 22*, 747−752.
(55) Fontana, A., Polverino de Laureto, P., Spolaore, B., and Frare, E. (2012) Identifying disordered regions in proteins by limited proteolysis. *Methods Mol. Biol. 896*, 297−318.
(56) Hubbard, S. J. (1998) The structural aspects of limited proteolysis of native proteins. *Biochim. Biophys. Acta 1382*, 191−206.
(57) Tyndall, J. D. A., Nall, T., and Fairlie, D. P. (2005) Proteases universally recognize β strands in their active site. *Chem. Rev. 105*, 973−999.
(58) Schechter, I., and Berger, A. (1967) On the size of the active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun. 27*, 157−162.
(59) Mori, Y., Goto, M., and Kamiya, N. (2011) Transglutaminase-mediated internal protein labeling with a designed peptide loop. *Biochem. Biophys. Res. Commun. 410*, 829−833.
(60) Buchardt, J., Selvig, H., Nielsen, P. F., and Johansen, N. L. (2010) Transglutaminase-mediated methods for site-selective modification of human growth hormone. *Biopolymers 94*, 229−235.
(61) Tanaka, T., Kamiya, N., and Nagamune, T. (2004) Peptidyl linkers for protein heterodimerization catalyzed by microbial transglutaminase. *Bioconjugate Chem. 15*, 491−497.
(62) Dunker, A. K., Brown, C. J., Lawson, L. D., Iakoucheva, L. M., and Obradovic, Z. (2002) Intrinsic disorder and protein function. *Biochemistry 41*, 6573−6582.
(63) Uversky, V. N. (2002) Natively unfolded proteins: A point where biology waits for physics. *Protein Sci. 11*, 739−756.