UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Università degli Studi di Padova

Centro di Ricerca Interdipartimentale per le Biotecnologie Innovative

– CRIBI –

SCUOLA DI DOTTORATO DI RICERCA IN BIOCHIMICA E BIOTECNOLOGIE

INDIRIZZO DI BIOTECNOLOGIE

XXIII CICLO

SITE-SPECIFIC MODIFICATION OF PROTEINS BY TRANSGLUTAMINASE

Direttore della scuola: Ch.mo Prof. Giuseppe ZANOTTI

Coordinatore d'indirizzio: Ch.mo Prof. Giorgio VALLE

Supervisore: Ch.mo Prof. Angelo FONTANA

Dottorando: Abhijeet Ajit SATWEKAR

Dedicated to

My Parents…

Acknowledgements

My thesis work was carried at the Proteomics facility of the CRIBI Biotechnology Center at the University of Padova, Italy, under the supervision of Prof. Angelo FONTANA.

First of all, I would like to thank the organizers of the International PhD program, Fondazione "Cassa di Risparmio di Padova e Rovigo (CARIPARO)" for covering my PhD fellowship. I wish to express my deepest appreciation to Prof. Angelo Fontana, for giving me the possibility to join his group. I thank him for involving me in this project. Especially, I would like to thank him for the scientific guidance and the encouragement. I have always been inspired from his scientific approaches and suggestions. I am also grateful to him, for providing me the opportunities to attend the congresses and summer schools.

I am also deeply grateful to the researchers, Barbara Spolaore and Erica Frare for their excellent cooperation in providing me the scientific input and advice during the experiments, and also for the proof reading of my thesis. Moreover, I wish to extend my thanks to the whole group members, Patrizia Polverino de Laureto, Giorgia De-franceschi, Marcello Zambonin, Odra, Serena, Micaela, Samanta, Chiara, Nunzio, and all the visiting thesis students, for their cooperation and constant support during my stay at the CRIBI Biotechnology Center, Italy.

My special thanks to the thesis evaluation committee of my doctoral school, Prof. Guiseppe Zannoti, (Director) and Prof. Giorgio Valle (Co-coordinator).
Furthermore, I wish to acknowledge Anna Mero, from the Department of Pharmaceutical Sciences, University of Padova, Italy, for her guidance during some experiments.

I wish to thank all my friends, staying in the ESU student residenza (Morgani, Goito, Copernico, Galileo Galilee) for making my life enjoyable and comfortable during the three years of my stay in Italy.

Finally, I wish to thank my parents, relatives and all my friends from India, for their constant family support.

TABLE OF CONTENTS

ABBREVATIONS

AcCN	acetonitrile
apoMb	apomyoglobin
CAM	carbamidomethylation
CM	carboxymethylation
DC	dansyl cadaverine
E:S	enzyme to substrate ratio
ESI	electrospray ionization
Gdn·HCl	guanidine hydrochloride
HA	hydroxyl amine
IA	iodoacetic acid
LC-MS/MS	liquid chromatography tandem mass spectrometry
LYS	lysozyme
$LYS^{CM6, 127}$	3-disulphide derivative of lysozyme
m/z	mass to charge ratio
MALDI	matrix-assisted laser desorption ionization
mPEG	monodisperse, amine-Poly(ethylene) glycol (mass of 556 Da)
MS	mass spectrometry
MS/MS	tandem mass spectrometry
MW	molecular weight
PAGE	polyacrylamide gel electrophoresis
PEG20k	amine-Poly(ethylene) glycol (mass of 20 kDa approx.)
PEG5k	polydisperse, amine-Poly(ethylene) glycol (mass of 5 kDa)
PEGB	biotin-PEG-amine
Q-Tof	quadrupole-time of flight
RNase	ribonuclease A (1-124)
RNase del-Th1	ribonuclease A (1-34/37-124)
RNase Th1	ribonuclease A (1-34/35-124)
RP-HPLC	reverse phase high pressure liquid chromatography
RT	Retention time
SDS	sodium dodecyl sulphate
TCEP	tris(2-carboxyethyl)fosfine
TEMED	tetra(methylethylene)diamine
TFA	trifluoroacetic acid
TFE	trifluoro ethanol
Tris·HCl	tris(hydroxymethyl)aminomethane
UV	Ultraviolet
ZQG	Carbobenzoxy-Glutamyl-Glycine (CBZ-Gln-Gly)

Abbrevations of amino acids

Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamine	Gln	Q
Glutamic acid	Glu	E
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

ABSTRACT

Site-specific modification of proteins by transglutaminase

Transglutaminase (TGase; EC 2.3.2.13) catalyzes the reaction between the γ-amido group of a protein-bound Gln residue (–$CONH_2$, the acceptor) and an amino group (–NH_2, the donor) of an alkyl-amine (protein-$CONH_2$ + H_2N-ligand → protein-CONH-ligand + NH_3). In the case of protein substrates, TGase causes an intra- and inter-molecular crosslinking of proteins by formation of an isopeptide bond involving the side chains of Gln and Lys residues. The acyl donor can be also a small amido-ligand mimicking the Gln residue, so that TGase allows a useful and interesting variability of substrates, thus leading to the modification of proteins at the level of Gln or Lys residues using appropriate substrate reagents.

Recently, the microbial TGase from *S. mobaraensis* has attracted a strong interest for protein modification, considering its stability, high reactivity and small size. The X-ray structure of this TGase has been solved and shown to contain an active site given by a triad Cys-His-Asp in analogy to a protease. A striking result of recent studies is that reactions mediated by TGase can be site-specific with some proteins, sometimes leading to the modification of only one Gln residue among the many Gln residues of a protein substrate. On the other hand, there is only a moderate specificity for Lys residues. With the view to shed light into the molecular features dictating the site-specific reaction(s) of TGase, in this Thesis a number of TGase-mediated reactions have been studied using proteins of known structure and dynamics, as apomyoglobin (apoMb), egg-white lysozyme (LYS) and bovine pancreatic ribonuclease A (RNase).

Amino- as well amido-ligands have been used in the TGase-mediated reactions, so that it was possible to analyse the specificity of modification of both Gln and Lys in the examined proteins. We have found an almost strict specificity of TGase-mediated reactions at the level Gln91 of apoMb, a residue embedded in the highly flexible or unfolded helix-F of the holo protein, as given by previous NMR measurements and limited proteolysis data. Also a Gln-mimicking ligand can be covalently linket by TGase at a Lys residue of the same chain region. Thus, we concluded that local enhanced flexibility or even fully local unfolding dictates the site-specific reaction with TGase.

While RNase can be selectively modified by using an amido-ligand at the level of the ε-amino group of Lys1, a similar Lys1 of LYS was instead fully unreactive. It was possible to relate this finding to the flexibility and rigidity of the N-terminal region of RNase and LYS, respectively, on the basis of the crystallographically determined *B*-factor values (a measure of

chain flexibility) of these two proteins. A nicked species of RNase with the single peptide bond Asn34-Leu35 cleaved (RNase Th1) and a LYS derivative with a single disulfide bridge reduced among the four of native LYS ($LYS^{CM6,\ 127}$) were shown to be much more reactive in the TGase-mediated reactions than the parent intact proteins, in agreement with their enhanced flexibility or partial unfolding. Moreover, we could demonstrate that the sites or regions susceptible to TGase reactions are also prone to limited proteolysis phenomena, implying that both TGase and a protease require some local unfolding for a site-specific enzymatic reaction. Indeed, this in keeping with view that the biorecognition phenomenon is similar for both enzymes, considering also the fact that TGase acts as a reverse protease (amide synthesis instead of hydrolysis).

An interesting outcome of these studies resides in the fact that we can envisage a novel enzymatic method of covalent coupling of an amino-polymer as poly(ethylene)glycol (PEG) to specific Gln residue(s) of proteins of pharmaceutical interest. Indeed, using TGase and an amino derivative of PEG ($PEG\text{-}NH_2$), it was possible to prepare homogeneous PEGylated derivatives of apoMb, human growth hormone (hGH) and granulocyte colony-stimulating factor (G-CSF). Overall, we have interpreted our findings as indicating that the selective TGase-mediated reactions require a flexible or unfolded polypeptide substrate. Therefore, it is possible to predict the sites of TGase attack on a protein substrate, provided that its structure and dynamics are known. Considering the increasing relevance of PEGylated protein drugs and the high regulatory demands for their approval, it can be anticipated that the innovative methods for the site-specific PEGylation of proteins using TGase will be considered a useful advance in the methodologies used for protein modification.

RIASSUNTO

Modifica sito-specifica di proteine con transglutaminasi

La transglutaminasi (TGasi; EC 2.3.2.13) catalizza la reazione tra il gruppo γ-ammidico di un residuo di Gln (-$CONH_2$, accettore) ed un gruppo amminico (-NH_2, donatore) di una alchil-ammina (proteina-$CONH_2$ + H_2N-ligando → proteina-CONH-ligando + NH_3). Con substrati proteici la TGasi determina una reticolazione intra- e inter-molecolare di proteine mediante la formazione di un legame isopeptidico tra le catene laterali dei residui di Gln e Lys. Il donatore acilico può essere anche un ammido-ligando in grado di mimare la catena laterale di Gln e, pertanto, la TGase consente una modifica di proteine a livello dei residui di Gln o Lys utilizzando opportuni reagenti.

Recentemente, la TGase microbica da *S. mobaraensis* ha suscitato un notevole interesse per la modifica enzimatica di proteine, considerando la sua stabilità, elevata reattività e piccole dimensioni. La struttura ai raggi-X di questa TGase ha rivelato la presenza di un sito attivo costituito da una triade Cys-His-Asp in analogia al sito attivo di una proteasi. Recenti studi hanno dimostrato che le reazioni mediate da TGase possono essere sito-specifiche, portando talvolta alla modifica di un solo residuo Gln tra i molti residui Gln di un substrato proteico. D'altra parte, è stata accertata solo una moderata specificità per i residui di Lys. Con lo scopo di chiarire i motivi strutturali che determinano la specificità di azione della TGase, in questa Tesi sono state studiate una serie di reazioni TGase-mediate utilizzando proteine di struttura e dinamica note, come apomioglobina (apoMb), lisozima da bianco d'uovo (LYS) e ribonucleasi pancreatica bovina (RNase).

Sono stati utilizzati sia ammino- che ammido-ligandi nelle reazioni con TGase ed in tal modo è stato possibile analizzare la specificità di modifica sia a livello di Gln che di Lys con le proteine esaminate. E' stata accertata una specificità rigorosa delle reazioni TGase-mediate a livello di Gln91 di apoMb, un residuo localizzato nel segmento corrispondente all'elica F della proteina nativa e risultato molto flessibile o *unfolded* in apoMb da precedenti misure NMR e da dati di proteolisi limitata. Anche un ligando in grado di mimare il residuo di Gln può essere legato da TGase a livello di una Lys incorporata nello stesso segmento flessibile. Pertanto, è stato concluso che una elevata flessibilità locale o *unfolding* determina la reazione sito-specifica di TGase.

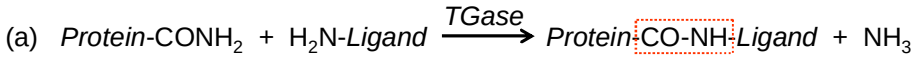
Mentre con la RNase si può ottenere con TGase ed un ammido-ligando la specifica modifica a livello del gruppo ε-aminico di Lys1, la simile Lys1 del LYS non reagisce affatto in simili condizioni. E 'stato possibile mettere in relazione questo fatto con la flessibilità e la rigidità della regione N-terminale di RNase e LYS, rispettivamente, sulla base dei valori del

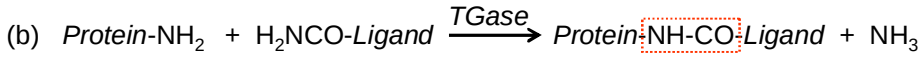
fattore-*B* (correlato alla flessibilità della catena polipeptidica) di queste due proteine ottenuti da dati cristallografici. Un derivato di RNase con il singolo legame peptidico Asn34-Leu35 idrolizzato (RNase Th1) ed un derivato di LYS con un ponte disolfuro ridotto ($LYS^{CM6,\ 127}$) sono risultati molto più reattivi nelle reazioni con TGase rispetto alle corrispondenti proteine native, in accordo con la loro maggiore flessibilità o parziale denaturazione. Inoltre, abbiamo potuto dimostrare che i siti o regioni suscettibili di reazioni con TGase subiscono anche fenomeni di proteolisi limitata, significando che sia la TGase che una proteasi necessitano di un *local unfolding* per determinare una reazione enzimatica sito-specifica. In effetti, questo è linea con il fatto che il bioriconoscimento del substrato è simile per entrambi gli enzimi, considerando il fatto che la TGase è una proteasi inversa (sintesi invece che idrolisi di una ammide).

Un risultato interessante di questi studi risiede nel fatto che si apre la strada ad un nuovo metodo enzimatico di *binding* covalente di poli(etilene)glicole (PEG) a livello di specifici residui di Gln in proteine di interesse farmaceutico. Infatti, utilizzando TGase ed un ammino-derivato di PEG (PEG-NH_2), è stato possibile preparare derivati peghilati omogenei di apoMb, ormone della crescita umano (hGH) e *granulocyte colony-stimulating factor* (G-CSF). In generale, i risultati ottenuti hanno indicato chiaramente che le reazioni selettive di TGase richiedono un substrato flessibile o *unfolded.* Pertanto, è possibile prevedere i siti di attacco di TGase su un substrato proteico, se sono note la sua struttura e dinamica. Considerando la crescente importanza dei farmaci proteici peghilati e le speciali richieste di omogeneità dettate dalle autorità regolatorie, si può prevedere che la peghilazione sito-specifica con TGase di proteine di interesse farmaceutico troverà utili applicazioni.

EXTENDED ABSTRACT

Background. Transglutaminase (TGase; EC 2.3.2.13) catalyzes an acyl transfer reaction in which the γ-carboxyamide groups of peptide- or protein-bound Gln residues act as the acyl donors. The most common acyl acceptors of TGase are the ε-amino groups of lysine residues within proteins, thus leading to protein crosslinking (*1-4*). Alternatively, the amino donor in the TGase reaction can be an aliphatic amine resembling the side-chain of Lys residue or, in general, an amino-containing ligand. Therefore, TGase allows permutation of substrates, so that it can be used to modify proteins at the level of both Gln or Lys residues using a ligand containing an amino or amido group mimicking the side-chain of a Lys or Gln residue, respectively.





The crystal structure of a microbial TGase from *Streptoverticillium mobaraensis* (*5*) has been solved and shown to contain a catalytic site given by a catalytic Asp-His-Cys triad (*6*). The cysteine (Cys) residue in position 64 of the 331-residue chain of the enzyme is essential for the catalytic activity and it is located at the bottom of a deep cleft in the globular protein. Several molecular and functional characteristics of microbial TGase, including calcium-independence, protein stability, higher reaction rate and small molecular size, are advantageous for its practical and industrial applications. Actually, microbial TGase is nowadays used to improve the physical properties of food proteins, including albumin, actins, myosin, fibrin, milk casein, α-lactalbumin, β-lactoglobulin and other proteins. Because of the usefulness and generality of the enzymatic reactions mediated by TGase (see above), interesting and novel applications of this enzyme in protein research and technology have been already demonstrated and additional ones are expected to be further explored (*7*).

TGase was shown to display a high degree of specificity towards the acyl donor, *i.e.* Gln residues, and a lower specificity towards the acyl acceptor, *i.e* Lys residues (*4*). Several studies were conducted in the past in order to unravel the mechanism of action of TGase and, in particular, in order to explain in molecular terms the reason why TGase reacts with some preference or even strict specificity with protein-bound Gln residue(s) (*8*, and ref.es cited therein). The main aim of this PhD work was to examine the TGase reactions using few model proteins of known structure and dynamics in order to unravel the molecular features of the protein substrate dictating the selective enzymatic reaction. To this aim, experiments were conducted on apomyoglobin (apoMb), lysozyme (LYS) and ribonuclease A (RNase). The results thus obtained on one side helped to clarify the protein substrate requirements for the TGase selectivity and on the other to develop suitable and novel procedures for the TGase-mediated modification of important protein drugs.

Apomyoglobin. The TGase-mediated reactions on the 153-residue chain of apomyoglobin (apoMb) were studied using dansyl-cadaverine and other primary

amines as acyl acceptors, as well as N-carbobenzoxy-Gln-Gly-OH (ZQG) as an acyl donor for the TGase-mediated reactions. Also derivatives of ZQG containing a fluorescent moiety at the C-terminus were prepared and used in the TGase reactions. The sites of protein modification were determined by fingerprinting and ESI mass spectrometry. Myoglobin in its holo form (with heme bound to the protein) is not susceptible to TGase reactions due to its rigid conformation, while the apo form (apoMb) was modified by TGase at the level of Gln or Lys residues embedded in helix F (chain segment 82–99). NMR studies on apoMb have earlier demonstrated increased flexibility or even unfolding of helix F (*9, 10*) and, moreover, several proteases cleave the 153-residue chain of apoMb at the level of helix F only (*11, 12*). Therefore, the chain region attacked by both TGase and proteases is flexible or unfolded.

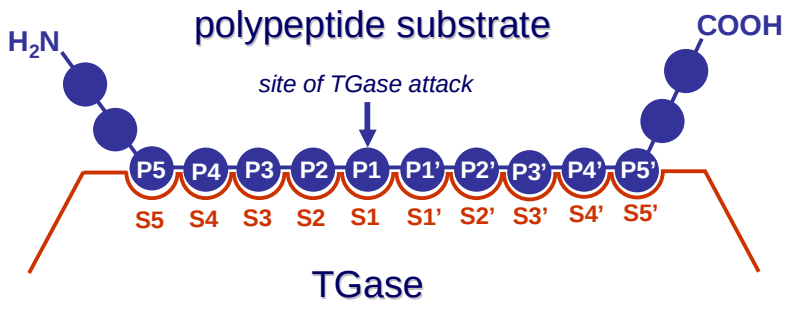
Lysozyme. Lysozyme (LYS) is a 129-residue protein chain that was shown to be resistant to the TGase action using both amino- or amido-containing ligands, implying that there are no Gln or Lys residues available for the TGase reactions. This protein is also resistant to proteolysis, highly thermostable and rigid, mostly due to the stabilising role of the four intramolecular disulfide bridges. In order to induce some flexibility in this protein, a procedure was developed to selectively cleave the disulfide bridge Cys6-Cys127, thus preparing and isolating a three disulfide containing protein ($LYS^{CM6,\ 127}$). This derivative possesses a 3D structure similar to that of native LYS, while some perturbation of structure occurs at the N- and C-termini, as deduced from spectroscopic studies (*13*). Moreover, $LYS^{CM6,\ 127}$ was shown to be more labile to heat treatment, implying a more flexible protein structure. It was shown that $LYS^{CM6,\ 127}$can be attacked by both TGase and chymotrypsin at its terminal ends, thus allowing us to conclude that some protein chain flexibility is required for the enzymatic reaction with TGase or a protease.

Ribonuclease. The modification of bovine pancreatic ribonuclease A (RNase) with TGase using N-carbobenzoxy-Gln-Gly-OH (ZQG) occurs at Lys1 at the N-terminal end of the 124-residue chain of the protein. In order to induce some flexibility in the protein chain, we produced a nicked version of RNase by cleaving the protein with thermolysin at the level of the Asn34-Leu35 peptide bond (*14*). The nicked protein RNase Th1 was shown to retain the overall 3D features of the intact protein, but has a slightly distorted tertiary structure confined at the site of thermolysin cleavage (*14*). Comparative TGase-mediated reactions performed on nicked and intact protein revealed an enhanced susceptibility of the nicked species to the TGase attack at the level of Lys residues using ZQG as substrate. In particular, Lys1, Lys37 and Lys91 were shown to be the sites of attack in RNase Th1, in agreement with structural and dynamic data regarding the nicked protein (*14*). Of interest, we may note that both intact RNase and LYS contain an N-terminal Lys residue, but that only RNase reacts with TGase (see above). This finding strikingly correlates with the fact that the *B*-factor (*15*) values of corresponding protein chains reveal a very flexible and rigid N-terminal region in RNase and LYS, respectively. These results provide a strong evidence for the requirement of flexibility or disorder for the TGase selectivity at the level of Lys-residues.

PEGylation of proteins. Protein drugs may possess several shortcomings that can limit their usefulness in therapy, including susceptibility to degradation by proteases, rapid kidney clearance and propensity to generate neutralizing antibodies (*16, 17*). Among the techniques so far explored for the development of safer and more useful protein drugs, undeniably the protein surface modification by covalent attachment of poly(ethylene glycol) (PEG) became an extremely valuable technique for producing protein drugs more water-soluble, non-aggregating, non-immunogenic and more stable to proteolytic digestion (*18-20*). The most used chemical methods for the PEGylation of proteins involve the covalent conjugation of PEG at the level of the ε-amino group of lysine residues by using acylating PEG derivatives. A drawback of these procedures resides in multiple sites of conjugation and thus in the substantial heterogeneity of the PEGylated proteins (*20*). During this PhD work, it was shown that TGase can be used to covalently link a PEG polymer to a protein-bound Gln residue(s) by using an amino-derivative of PEG (PEG-NH_2). The PEGylation reaction was studied in more detail with apomyoglobin (apoMb), showing that PEGylation occurs selectively at Gln91 and to a very minor extent to Gln151 at the C-terminus of the 153-residue chain of the protein. The results obtained indicated that it is possible to achieve a selective protein modification and that the PEGylation reaction occurs at the level of a disordered site of the protein substrate, in analogy to the TGase reactions occurring with other protein systems (cf. above).

Other proteins. During this PhD work I collaborated also to studies going on in the Laboratory of Protein Chemistry of CRIBI regarding TGase-mediated reactions on other proteins. The TGase-mediated reactions were investigated with human growth hormone (hGH) and granulocyte-colony stimulating factor (G-CSF), with special emphasis to the PEGylation reaction (see below). Overall, the results obtained clearly established that TGase can attack both hGH and G-CSF at their flexible or locally unfolded sites, in agreement with the results obtained with the model proteins listed above. Of particular interest was the fact that it was possible to prepare mono-conjugated derivatives of these important protein drugs.

Conclusion. The results obtained during this PhD work using several protein systems indicate that the sites of TGase reactions on a globular protein substrate are characterised by enhanced chain flexibility or even protein disorder. Indeed, it is demonstrated that there is a correlation between sites of TGase attack and sites of enhanced chain flexibility, this last deduced from the crystallographically determined *B*-factor (*15*). Moreover, we have established that the sites or chain regions that are amenable to TGase reactions are also prone to limited proteolysis phenomena with a variety of proteases, thus indicating that both TGase and a protease require an unfolded polypeptide substrate for their selective enzymatic attack (*21-23*). This is in keeping with a view that the enzymatic reaction involves the binding and adaptation of a 10-12 residue polypeptide substrate at the enzyme's active site (*24*).



Therefore, it is possible to predict the site(s) of TGase-mediated modification of a protein on the basis of its 3D structure and dynamics and, consequently, the likely effects on its physicochemical and functional properties. An interesting outcome of these studies is the possibility to develop novel and mild procedures of PEGylation of proteins of pharmaceutical interest (*25*, *26*). By using an amino-derivative of PEG (PEG-NH_2) it is possible to link the polymer chain at the level of a specific Gln residue, allowing the preparation of homogeneous PEGylated protein derivatives and thus overcoming the over-labelling and heterogeneity of similar derivatives prepared by using chemical methods.

References

1. Folk, J. E. (1969) Mechanism of action of guinea pig liver transglutaminases: Order of substrate addition, *J. Biol. Chem. 244*, 3707–3713.

2. Folk, J. E. (1980) Transglutaminases, *Annu. Rev. Biochem. 49*, 517–531.

3. Lorand, L., and Conrad, S. M. (1984) Transglutaminases, *Mol. Cell. Biochem. 58*, 9–35.

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

5. Ando, H., Adachi, M., Umeda, K., Matsuura, A., Nonaka, M., Uchio, R., Tanaka, H., and Motoki, M. (1989) Purification and characterization of a novel transglutaminase derived from microorganisms, *Agric. Biol. Chem. 53*, 2613–2617.

6. Kashiwagi, T., Yokoyama, K., Ishikawa, K., Ono, K., Ejima, D., Matsui, H., and Suzuki, E. (2002) Crystal structure of microbial transglutaminase from *Streptoverticillium mobaraensis*, *J. Biol. Chem. 277*, 44252–44260.

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

8. 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.

9. Eliezer, D., and Wright, P. E. (1996) Is apomyoglobin a molten globule? Structural characterization by NMR, *J. Mol. Biol. 263,* 531–538.

10. Eliezer, D., Yao, J., Dyson, H. J., and Wright, P. E. (1998) Structural and dynamic characterization of partially folded states of apomyoglobin and implications for protein folding, *Nature Struct. Biol. 5*, 148–155.

11. 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.

12. Picotti, P., Marabotti, A., Negro, A., Musi, V., Spolaore, B., Zambonin, M., and Fontana, A. (2004) Modulation of the structural integrity of helix F in apomyoglobin by single amino acid replacements, *Protein Sci. 13*, 1572–1585.

13. Radford, S. E., Woolfson, D. N., Martin, S. R., Lowe, G., and Dobson, C. M. (1991) A three-disulphide derivative of hen lysozyme: Structure, dynamics and stability, *Biochem. J. 273*, 211–217.

14. Polverino de Laureto, P., Scaramella, E., De Filippis, V., Bruix, M., Rico, M., and Fontana, A. (1997) Limited proteolysis of ribonuclease A with thermolysin in trifluoroethanol, *Protein Sci. 6*, 860–872.

15. Ringe, D., and Petsko, G. A. (1985) Mapping protein dynamics by X-ray diffraction, *Prog. Biophys. Mol. Biol. 45*, 197–235.

16. Leader, B., Baca, Q. J., and Golan, D. E. (2008) Protein therapeutics: A summary and pharmacological classification, *Nature Rev. Drug Discov. 7*, 21–39.

17. Pavlou, A., and Reichert, J. (2004) Recombinant protein therapeutics: Success rates, market trends and values to 2010, *Nature Biotechnol. 22*, 1513–1519.

18. Roberts, M. J., Bentley, M. D., and Harris, J. M. (2002) Chemistry for peptide and protein PEGylation, *Adv. Drug Deliv. Rev. 54*, 459–476.

19. Harris, J. M., and Chess, R. (2003) Effect of PEGylation on pharmaceuticals, *Nature Rev. Drug Discovery 2*, 214–221.

20. Veronese, F. M., and Pasut, G. (2005) PEGylation, successful approach to drug delivery, *Drug Discovery Today 10,* 1451–1458.

21. 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.

22. 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.

23. 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.

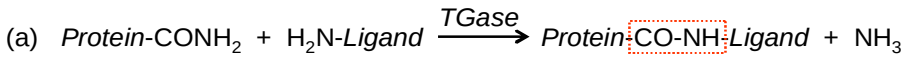
24. Schechter, I., and Berger, A. (1967) On the size of the active site in proteases. I. Papain, *Biochem. Biophys. Res. Commun. 27*, 157–162.

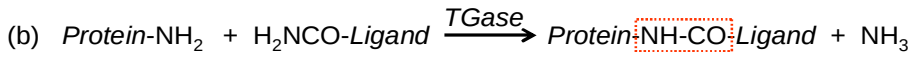
25. Sato, H. (2002) Enzymatic procedure for site-specific PEGylation of proteins, *Adv. Drug Deliv. Rev. 54*, 487–504.

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

RIASSUNTO ESTESO

Generalità. La transglutaminasi (TGase; EC 2.3.2.13) catalizza la reazione di condensazione tra il gruppo γ-carbossammidico della glutammina (Gln) di una proteina o peptide ed un componente amminico, con liberazione di ammoniaca (*1*, *2*). I più comuni accettori acilici in questa reazione sono i gruppi ε-amminici della lisina (Lys), determinando quindi la reticolazione (*crosslinking*) intra- ed inter-molecolare delle proteine (*3*, *4*). In alternativa, il donatore amminico nella reazione con TGase può essere una alchil-ammina o, in generale, un ligando contenente un gruppo amminico simile alla catena laterale di Lys. Inoltre, un ligando con un gruppo ammidico simile a quello della Gln può svolgere la funzione di donatore acilico per le catene laterali ε-amminiche di Lys di una proteina. Pertanto, la TGase permette una interessante permutazione dei suoi substrati e conseguentemente è possibile modificare le proteine a livello sia di residui Gln che Lys, utilizzando un ligando contenente un gruppo amminico o ammidico, mimando rispettivamente la catena laterale di un residuo di Lys o Gln (cf. Schema).





La struttura cristallina della TGase microbica da *Streptoverticillium mobaraensis* (*5*) è stata determinata di recente, evidenziando la presenza di un sito attivo dell'enzima formato dalla triade catalitica Asp–Cys–His, in analogia al sito attivo di alcune proteasi (*6*). Il residuo di cisteina in posizione 64 (Cys64) della catena di 331-residui della TGase è essenziale per l'attività catalitica e si trova nascosto (*buried*) all'interno di una tasca idrofobica dell'enzima. Varie caratteristiche molecolari e funzionali della TGase microbica, tra cui l'attività non Ca^{2+}-dipendente, la stabilità proteica, la maggiore velocità di reazione e le piccole dimensioni molecolari, costituiscono importanti vantaggi pratici per varie applicazioni di questo enzima. Infatti, la TGase microbica è attualmente utilizzata per migliorare le proprietà fisiche delle proteine alimentari, tra cui albumina, actina, miosina, fibrina, caseina del latte, α-lattalbumina, β-lattoglobulina ed altre proteine. Le reazioni enzimatiche mediate da TGase (cf. sopra) per la modifica di proteine hanno già trovato interessanti ed utili applicazioni ed in un prossimo futuro si attendono ulteriori sviluppi metodologici ed applicazioni in un ambito di ricerca sia di base che biotecnologica (*7*).

E' stato dimostrato che la TGase presenta un elevato grado di specificità nei confronti del donatore acilico (residui di Gln), mentre si è osservata una bassa specificità per l'accettore acilico (residui di Lys) (*4*). Diversi studi sono stati condotti in passato al fine di chiarire il meccanismo d'azione della TGase e per comprendere in termini molecolari la ragione per cui la TGase reagisce con una certa preferenza o addirittura stretta specificità con residui di Gln in proteine globulari. I risultati ottenuti con questi studi non hanno però portato ad una chiara spiegazione della selettività di azione della TGase con substrati proteici (*8*). Lo scopo principale delle ricerche condotte in questa Tesi di dottorato è stato quello di esaminare le reazioni con TGase

utilizzando alcune proteine modello di struttura tridimensionale nota al fine di chiarire le caratteristiche molecolari del substrato proteico che sono in grado di determinare la selettività della reazione enzimatica. Per questo scopo, gli esperimenti sono stati condotti su apomioglobina (apoMb), lisozima (LYS) e ribonucleasi A (RNase). I risultati ottenuti da un lato hanno portato a proporre una chiara spiegazione molecolare della selettività della TGase ed inoltre nuove ed innovative metodologie per la modifica TGase-mediata di farmaci proteici.

Apomioglobina. Le reazioni mediate da TGase sulla catena polipeptidica di 153 residui di apomioglobina (apoMb) sono state studiate utilizzando come accettori acilici ammine primarie (ad es., dansil-cadaverina), nonchè N-carbobenzoxy-Gln-Gly-OH (ZQG) come agente acilante per residui di Lys in presenza di TGase. Inoltre, sono stati preparati derivati di ZQG con una molecola fluorescente covalentemente legata al C-terminale di questo substrato peptidico e successivamente utilizzati per la modifica di apoMb a livello di Lys in presenza di TGase. I siti di modifica delle proteine sono stati determinati mediante tecniche di *fingerprinting* (proteolisi, seguita da analisi dei frammenti peptidici) e spettrometria di massa ESI. Mentre la mioglobina nella sua forma olo, cioè con il suo groppo eme, non reagisce in presenza di TGase, la forma apo è stata modificata con TGase a livello dei residui Gln e Lys dell'elica F (segmento 82–99). In precedenza è stato dimostrato mediante misure NMR che il segmento corrispondente all'elica F in apoMb è caratterizzato da una elevata flessibilità (*9, 10*) e che varie proteasi sono in grado di idrolizzare la catena di 153 residui di apoMb solo a livello dell'elica F (*11, 12*). Pertanto, è stato concluso che le reazioni selettive sia di TGase che di proteasi avvengono in una regione del substrato proteico molto flessibile o disordinato (*unfolded*).

Lisozima. Il lisozima (LYS) è una proteina di 129 residui amminoacidicidel tutto resistente all'azione di TGase utilizzando sia ammino- che ammido-ligandi (cf. sopra), il che implica che non vi sono residui Gln o Lys accessibili alle reazioni con TGase. Questa proteina è anche resistente alla proteolisi, altamente termostabile e rigida, principalmente a causa del ruolo stabilizzante dei suoi quattro ponti disolfuro intramolecolari. Al fine di indurre una certa flessibilità in questa proteina, è stata studiato un metodo per ridurre selettivamente uno dei quattro ponti disolfuro di LYS ed in particolare è stato possibile produrre ed isolare allo stato omogeneo un derivato con il ponte disolfuro Cys6-Cys127 ridotto, una proteina cioè con tre ponti disolfuro ($LYS^{CM6,\ 127}$). Studi spettroscopici hanno accertato che questo derivato possiede una struttura 3D simile a quella del LYS nativo, con variazioni strutturali solo a livello delle zone N- e C-terminali (*13*). Inoltre, $LYS^{CM6,\ 127}$ è risultato meno termostabile della proteina nativa, implicando quindi una struttura proteica più flessibile. E 'stato dimostrato che il campione di $LYS^{CM6,\ 127}$viene modificato da TGase in presenza di ammino- ed ammido-ligandi a livello delle zone terminali. Inoltre, anche la chimotripsina è in grado di idrolizzare la catena polipeptidica di 129 residui di $LYS^{CM6,\ 127}$ alle estremità della catena polipeptidica. Pertanto, è stato possibile concludere che una certa flessibilità (*unfolding*) del substrato proteico è necessaria per poter reagire sia con la TGase che con una proteasi.

Ribonucleasi. La modifica della ribonucleasi A bovina pancreatica (RNase) con TGase utilizzando N-carbobenzoxy-Gln-Gly-OH (ZQG) si verifica selettivamente

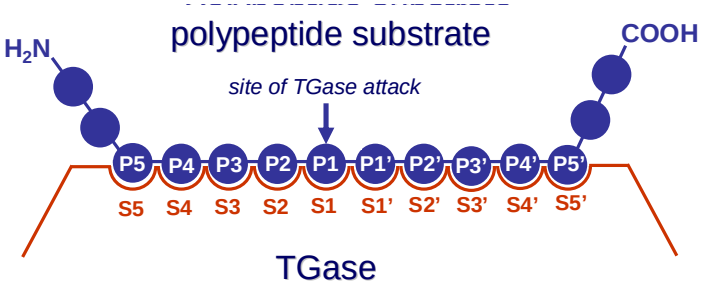
a livello di Lys1, residuo N-terminale della catena di 124 residui della proteina. Al fine di indurre una certa flessibilità nella catena proteica, è stato prodotto un derivato di RNase mediante l'azione proteolitica di termolisina, che determina l'idrolisi del solo legame peptidico Asn34-Leu35 (*14*). Questo derivato (RNase-Th1, forma *nicked*) mantiene la struttura 3D della proteina intatta ed anche una parziale attività enzimatica, mentre una distorsione strutturale è localizzata a livello del legame peptidico idrolizzato dalla termolisina (*14*). Reazioni comparative con TGase condotte sulla proteina intatta e *nicked* hanno chiaramente rivelato un aumento significativo della suscettibilità della specie *nickeld* all'attacco di TGase a livello dei residui di Lys utilizzando ZQG come substrato (cf. sopra). In particolare, è stato dimostrato che i siti di attacco di ZQG in RNase-Th1 avvengono a livello di Lys1, Lys37 e Lys91, in accordo con i dati spettroscopici riguardanti la struttura e la dinamica della proteina *nicked*. E' interessante osservare che le specie proteiche intatte sia di RNase che LYS contengono entrambe un residuo N-terminale di Lys, ma che solo RNase reagisce con TGase (cf. sopra). Questo fatto è ottimamente correlato al fatto che il fattore-*B* (che evidenzia flessibilità) (*15*) è elevato e molto basso nella zona N-terminale di RNase e LYS, rispettivamente, dimostrando quindi che la zona N-terminale è flessibile in RNase e rigida in LYS. Pertanto, questi risultati forniscono una chiara dimostrazione che la flessibilità o *unfolding* determina la reazione della TGase anche a livello di residui di Lys.

Peghilazione delle proteine. Farmaci proteici possono avere varie proprietà che possono limitare la loro utilità in terapia, compresa la suscettibilità alla degradazione con proteasi, la rapida *clearance* renale e la propensione a generare anticorpi neutralizzanti (*16, 17*). Tra le tecniche finora esplorate per lo sviluppo di farmaci proteici più sicuri ed utili, la modifica della superficie proteica mediante il *binding* covalente di poli(etilenglicole) (PEG) rappresenta una tecnica molto utile per ottenere farmaci proteici più solubili in acqua, non aggreganti, non immunogenici e più stabili alla digestione proteolitica (*18-20*). Attualmente i metodi chimici più usati per la peghilazione di proteine comportano la coniugazione covalente di PEG a livello dei gruppi ε-amminici di residui di Lys utilizzando quali agenti acilanti derivati di PEG contenenti un gruppo carbossilico C-terminale reattivo (*18-20*). Un inconveniente di questi metodi chimici risiede nel fatto che si hanno più siti di peghilazione e quindi notevole eterogeneità delle proteine peghilate (*20*). Durante questa Tesi di dottorato è stato dimostrato che la TGase può essere utilizzata per legare covalentemente un polimero PEG ad un residuo di Gln in un substrato proteico utilizzando un ammino-derivato del PEG (PEG-NH_2). La reazione di peghilazione è stata studiata in dettaglio con apomioglobina (apoMb), dimostrando che la peghilazione avviene selettivamente a livello di Gln91 e, in misura molto minore, di Gln151 localizzata nella zona C-terminale della catena di 153 residui amminoacidici di apoMb. I risultati ottenuti hanno indicato quindi che è possibile ottenere una modifica della proteina molto selettiva e che la peghilazione si verifica a livello di un sito *unfolded* del substrato proteico, in analogia alle reazioni di TGase con altri sistemi proteici (cf. sopra).

Altre proteine. Durante questa Tesi di dottorato sono state condotte anche ricerche in collaborazione con altri ricercatori del Laboratorio di Chimica delle Proteine del CRIBI rigurdanti reazioni TGase-mediate su altre proteine. Le reazioni

TGase-mediate sono state studiate con l'ormone della crescita umano (hGH) ed il fattore stimolante i granulociti (G-CSF), con particolare attenzione alla peghilazione. Complessivamente i risultati ottenuti hanno chiaramente stabilito che la TGase può attaccare sia hGH che G-CSF a livello di siti flessibili o localmente denaturati (*unfolded*), in accordo con i risultati ottenuti con le proteine modello di cui sopra. Di particolare interesse è il fatto che è stato possibile preparare derivati mono-peghilati di questi importanti farmaci proteici.

Conclusione. I risultati ottenuti durante questa Tesi di dottorato hanno chiaramente dimostrato che i siti di reazione della TGase su un substrato proteico sono caratterizzati da una maggiore flessibilità della catena polipeptidica o addirittura disordine (*unfolding*) della proteina. Infatti, è stato dimostrato che esiste una correlazione tra i siti di attacco TGase e dei siti di maggiore flessibilità della catena polipeptidica. Questa particolare flessibilità è stata dedotta dall'analisi dei valori di fattore-*B* lungo la catena polipeptidica ottenuti con metodi cristallografici (*15*). Inoltre, è stato dimostrato che i siti o le regioni di una catena polipeptidica che subiscono reazioni con TGase sono anche suscettibili di fenomeni di proteolisi limitata con una varietà di proteasi, indicando in tal modo che sia la TGase che una proteasi richiedono un substrato polipeptidico disordinato (*unfolded*) per poter determinare un loro attacco enzimatico selettivo (*21-23*). Questa conclusione è in linea con il fatto che una reazione enzimatica implica il *binding* e l'adattamento del substrato polipeptidico al sito attivo dell'enzima e questo *binding* coinvolge un segmento di 10-12 residui amminoacidici (*24*).



Pertanto, è possibile prevedere il sito di modifica TGase-mediata di una proteina sulla base della sua struttura 3D e della sua dinamica e, conseguentemente, i probabili effetti sulle sue proprietà chimico-fisiche e funzionali. Un risultato interessante ed una ricaduta importante di questi studi è la possibilità di sviluppare nuove e blande procedure di peghilazione di proteine di interesse farmaceutico (*25, 26*). Utilizzando ammino-derivati di PEG (PEG-NH_2) è possibile coniugare una catena polimerica a livello di un residuo di Gln specifico, consentendo la preparazione di campioni omogenei di proteine peghilate, ovviando quindi all'eccessiva eterogeneità delle proteine modificate con metodi chimici.

Bibliografia

1. Folk, J. E. (1969) Mechanism of action of guinea pig liver transglutaminases: Order of substrate addition, *J. Biol. Chem. 244*, 3707–3713.

2. Folk, J. E. (1980) Transglutaminases, *Annu. Rev. Biochem. 49*, 517–531.

3. Lorand, L., and Conrad, S. M. (1984) Transglutaminases, *Mol. Cell. Biochem. 58*, 9–35.

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

5. Ando, H., Adachi, M., Umeda, K., Matsuura, A., Nonaka, M., Uchio, R., Tanaka, H., and Motoki, M. (1989) Purification and characterization of a novel transglutaminase derived from microorganisms, *Agric. Biol. Chem. 53*, 2613–2617.

6. Kashiwagi, T., Yokoyama, K., Ishikawa, K., Ono, K., Ejima, D., Matsui, H., and Suzuki, E. (2002) Crystal structure of microbial transglutaminase from *Streptoverticillium mobaraensis, J. Biol. Chem. 277*, 44252–44260.

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

8. 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.

9. Eliezer, D., and Wright, P. E. (1996) Is apomyoglobin a molten globule? Structural characterization by NMR, *J. Mol. Biol. 263*, 531–538.

10. Eliezer, D., Yao, J., Dyson, H. J., and Wright, P. E. (1998) Structural and dynamic characterization of partially folded states of apomyoglobin and implications for protein folding, *Nature Struct. Biol. 5*, 148–155.

11. 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.

12. Picotti, P., Marabotti, A., Negro, A., Musi, V., Spolaore, B., Zambonin, M., and Fontana, A. (2004) Modulation of the structural integrity of helix F in apomyoglobin by single amino acid replacements, *Protein Sci. 13*, 1572–1585.

13. Radford, S. E., Woolfson, D. N., Martin, S. R., Lowe, G., and Dobson, C. M. (1991) A three-disulphide derivative of hen lysozyme: Structure, dynamics and stability, *Biochem. J. 273*, 211–217.

14. Polverino de Laureto, P., Scaramella, E., De Filippis, V., Bruix, M., Rico, M., and Fontana, A. (1997) Limited proteolysis of ribonuclease A with thermolysin in trifluoroethanol, *Protein Sci. 6*, 860–872.

15. Ringe, D., and Petsko, G. A. (1985) Mapping protein dynamics by X-ray diffraction, *Prog. Biophys. Mol. Biol. 45*, 197–235.

16. Leader, B., Baca, Q. J., and Golan, D. E. (2008) Protein therapeutics: A summary and pharmacological classification, *Nature Rev. Drug Discov. 7*, 21–39.

17. Pavlou, A., and Reichert, J. (2004) Recombinant protein therapeutics: Success rates, market trends and values to 2010, *Nature Biotechnol. 22*, 1513–1519.

18. Roberts, M. J., Bentley, M. D., and Harris, J. M. (2002) Chemistry for peptide and protein PEGylation, *Adv. Drug Deliv. Rev. 54*, 459–476.

19. Harris, J. M., and Chess, R. (2003) Effect of PEGylation on pharmaceuticals, *Nature Rev. Drug Discovery 2*, 214–221.

20. Veronese, F. M., and Pasut, G. (2005) PEGylation, successful approach to drug delivery, *Drug Discovery Today 10*, 1451–1458.

21. 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.

22. 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.

23. 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.

24. Schechter, I., and Berger, A. (1967) On the size of the active site in proteases. I. Papain, *Biochem. Biophys. Res. Commun. 27*, 157–162.

25. Sato, H. (2002) Enzymatic procedure for site-specific PEGylation of proteins, *Adv. Drug Deliv. Rev. 54*, 487–504.

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

I. INTRODUCTION

1. Transglutaminase

Transglutaminases (TGase) are enzymes that catalyze the isopeptide bond formation between glutamine and lysine residues in proteins. The term "transglutaminase" was first introduced by Clarke *et al* in 1957 (*1*) for the transamidating activity detected in the liver of guinea-pig. Further, Pisano *et al.* 1968 (*2*) performed studies on the stabilization of fibrin monomers during blood clotting and demonstrated that the TGase mediated cross-linking occurs through an acyl transfer reaction between the γ-carboxamide group of a protein/peptide bound glutamine and the ε–amino group of a protein/peptide bound lysine, the product of the reaction being a ε–(γ-glutamyl) lysine isopeptide bond. Since then, the blood coagulation factor XIII (TGase) has been extensively studied and characterized (*3-8*).

TGases are widely distributed in nature, both in prokaryotes and eukaryotes. They occur inside the cell as well as in the extracellular milieu and consist of a large family of enzymes. Various forms of TGase are present in animals which are Ca^{2+} dependent and are distributed in different tissues such as liver, lung, intestine, epidermis, placenta, blood etc. (*9-11*). Up to eight members of the mammalian TGase family (Table 1) were identified from the genomic data (*12*), of which six isozymes have been purified and characterized as calcium dependent thiol enzymes. All the members from the mammalian TGase family are structurally homologous and arise from different genes, due to the rearrangement and duplication mechanisms in cell. However, these TGases differ in their molecular weight and biochemical properties. Apart from the mammals, TGase activity has been found in fish (*13*), lobster (*14*), horseshoe crab (*15*), grasshopper (*16*), yeast *Candida albicans (17)*, filarial nematode (*18*) and in plant *Medicago sativa* L. (alfalfa) (*19*). In plants, TGases have been reported to be present in chloroplasts, mitochondria, cytoplasm and cell walls and they are primarily involved in functions related to plant growth, stress, differentiation and programmed cell death (*20*). TGase activity has also been reported in bacteria, and in particular the *Bacillus subtilis (21)*, *Streptoverticillium* sp. (*22*), and *Streptomyces* sp (*23*).

Table 1: Members of the mammalian TGase family

TGase	**Synonym**	**Role**	**Localization**
Factor XIII	Fibrin stabilizing factor	Blood clotting and wound healing	Cytosol, extracellular
TG1	Keratinocyte transglutaminase	Cornified envelope assembly in surface epithelia	Cytosol, membrane
TG2	Tissue transglutaminase	Matrix assembly, adhesion, cell death/differentiation	Cytosol, nucleus, membrane, cell surface, extracellular
TG3	Epidermal transglutaminase	Cornified envelope assembly in surface epithelia	Cytosol
TG4	Prostate transglutamianse	Semen coagulation in rodents	Unknown
TG5	Transglutaminase X	Epidermal differentiation	Nuclear matrix, cytoskeleton
TG6	Transglutaminase Y	Unknown	Unknown
TG7	Transglutaminase Z	Unknown	Unknown

2. Microbial TGases

TGase has also been purified, characterized and cloned from the sporulating gram positive bacteria *Bacillus subtilis* (*21*) and Streptoverticillium sp. TGase has been reported to be present extracellularly in Streptoverticillium sp., whereas intracellularly in Bacillus sp. The sequence of the two TGases showed similarity for the presence of a cysteine residue at the active site of the enzyme. In the non-sporulating gram negative bacteria *E. coli*, TGase activity was also measured. The sequence of the *E. coli* enzyme was determined and it showed the presence of a

cysteine residue at the active site (*24*). The low homology with mammalian suggests that TGases present in prokaryotes may not have the same conserved active site region as that of the mammalian counterpart, but they maintain a cysteine residue within the active site. Indeed, both mammalian and microbial TGases are inhibited by sulphahydryl group blocking entities such as parachloromercuribenzoic acid, iodo acetate and N-ethyl maleimide (*22*).

TGase identification from the bacterial source involved the initial screening of 5000 micro organisms by Ando (*22*) which was based on the detection of hydroxamate formation (*22*). TGase activity was reported in many microbes, however strong activity was only found in the actinomycete strain belonging to the genus *Streptoverticillum*. TGase was purified from the species of *Streptoverticillium (25-26)* and *Streptomyces (23)*. Both these organisms belong to the family Streptomycetacae, order Actinomycetales and are characterized by the formation of arthropores on the aerial mycelium (*27*). TGase activity increases with the mycelium growth and bacterial morphological differentiation. However, the exact role of TGase in these organisms is unknown. Purification of microbial TGase was first achieved from a variant of *Streptoverticillum mobaraense*, which is *Streptoverticillium* S-8112 (*22*). TGase is synthesized as a precursor protein consisting of 406 amino acid residues, which is an inactive form of the enzyme. During the extracellular migration of the inactive TGase from the bacterial cell, a proteolytic cleavage occurs which generates the active TGase consisting of 331 amino acid residues.

The enzyme studied in this Thesis is microbial TGase from *Streptoverticillum mobaraense*. Microbial TGase is a monomeric protein with the molecular weight of 37863 Da, which is half as that of mammalian TGase (85000 Da) and bears little homology with factor XIII like TGases (*28*). The hydropathic profile of the active site region in microbial TGase bears resemblance with the mammalian TGase, which suggests a similar secondary structural environment at the active site region. Importantly, the activity of the microbial TGase does not require the presence of Ca^{2+} ions and can act well in either the presence or absence of Ca^{2+}. However, Zn^{2+} and Cu^{2+} were shown to inhibit the activity, while other metals had no effect (*22*). Moreover, the iso-electric point of microbial TGase is 8.9, which is also different from that of mammalian TGase (4.5). NMR studies have demonstrated a high reaction

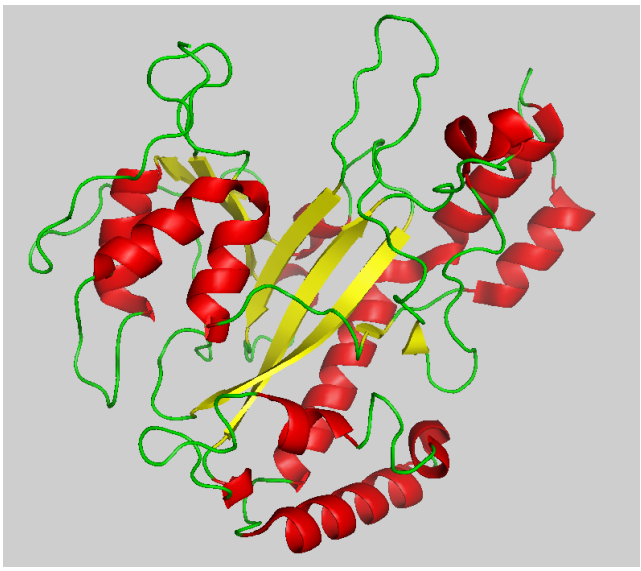
rate and lower acyl donor substrate specificity for microbial TGase as compared with the other TGases (*29*). Also, microbial TGase has low deamidation rate as compared with the other TGases (*30*). Optimal temperature for the activity of TGase from *Streptoverticillium mobaraense* is 55° C but the stability of the enzyme decreases at high temperatures (*22*).

The discovery of microbial TGase (MTGase) has opened the possibility to an industrial production of the enzyme and to its use in biotechnological applications. Indeed, along with the cost effective production of TGase from *Streptoverticillium* sp. by fermentation the microbe secretes the enzyme extracellularly in the culture medium, which reduces the purification steps and eventually the cost of production. All these features favour the use of MTGase in industrial applications. These opportunities were indeed limited in the case of mammalian TGases due to their high cost of productiom and their non-availability in high quantities.

3. Structural features of microbial TGase

Crystal structure of microbial TGase isolated from *Streptoverticillium* sp. S-8112 revealed the structural features of the enzyme, which consists of 331 amino acid residues, arranged in a compact domain with overall dimensions of 65 × 59 × 41 A° (*31*) (Figure 1). A deep cleft of 16 A° within the enzyme constitutes the catalytic cleft. The catalytic site consists of the Cys-His-Asp triad typical of cysteine proteases, implying a similar mechanism of catalysis (Figure 2). The triad consists of Cys64, Asp255 and His274 residues in the microbial TGase, which almost superimposes with the triad of factor XIII-like TGase. The cleft of the microbial TGase active site is composed of acidic residues (Asp1, Asp3, Asp4, Glu249, Asp255 and Glu300), whereas the surface around the cleft consists of aromatic residues (Trp59, Tyr62, Trp69, Tyr75, Tyr278, Tyr291 and Tyr302) (Figure 3). The characteristic arrangements of the acidic and aromatic residues appear to have an effect on the substrate specificity of TGase (*31*).

A



B

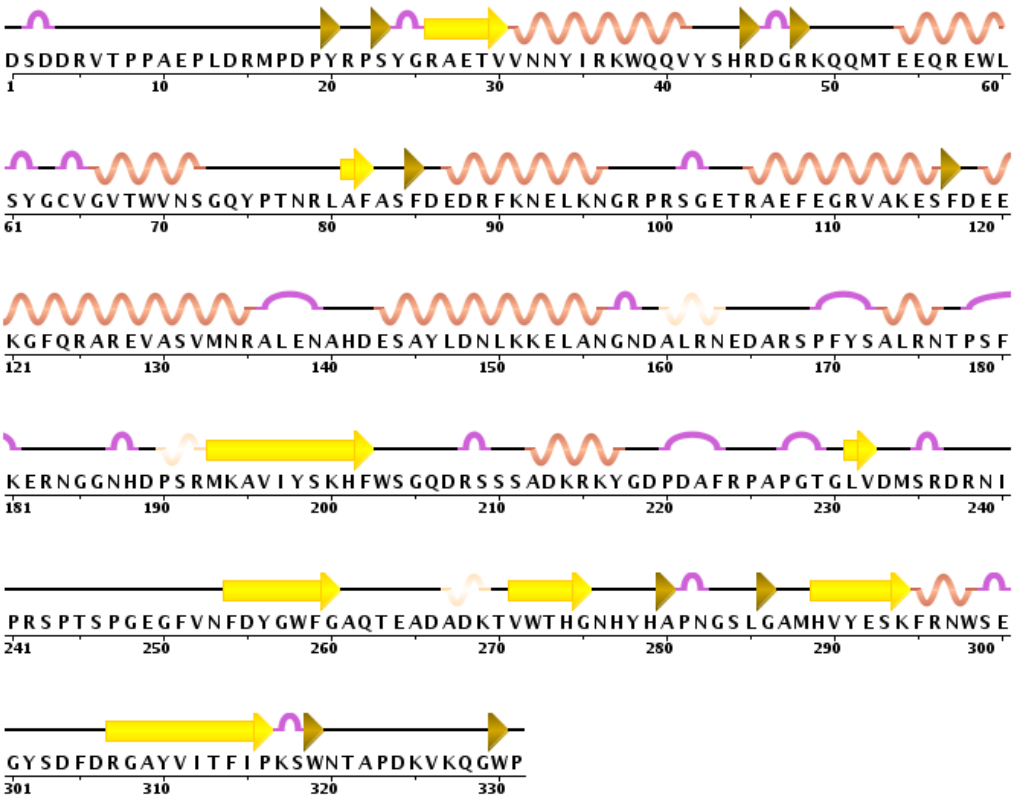


Figure 1. Microbial TGase. A) 3D structure of microbial TGase (PDB file 1IU4). B) Amino acid sequence and secondary structure of microbial TGase. Helical segments along the protein chain are indicated above the amino acid sequence by squiggled red lines and beta strands by yellow arrows.

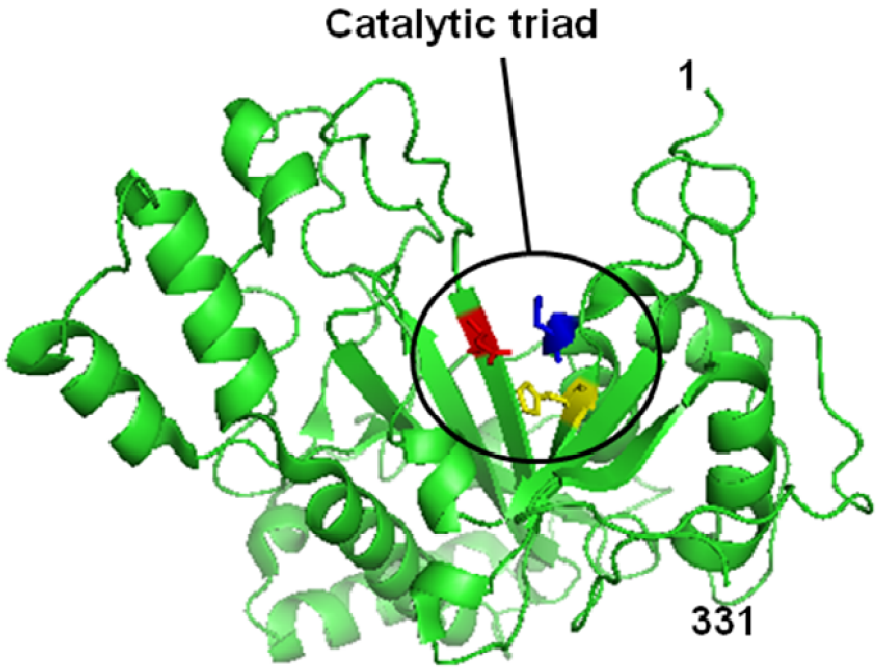


Figure 2. Three-dimensional structure of microbial TGase from *Streptomyces mobaraense*. The protein model was prepared from the X-ray structure of the enzyme (PDB file 1IU4) using the software PyMOL [47]. In the structure, numbers indicate the N- and C-terminus of the protein. The catalytic triad of the enzyme is circled and residues of the catalytic triad Cys64, Asp255 and His254 are indicated in blue, red and yellow colors, respectively.

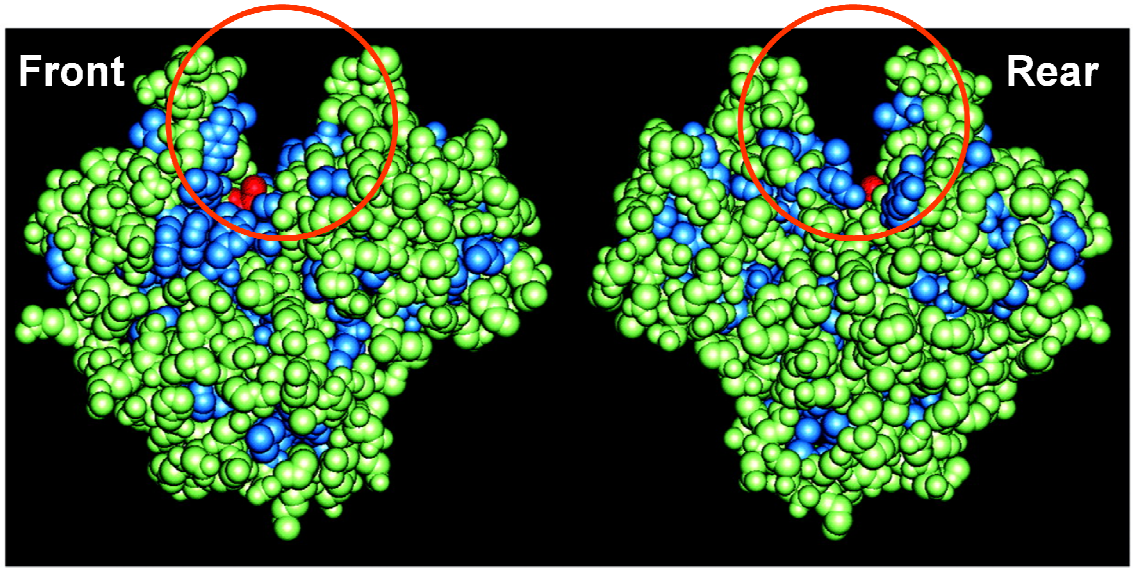


Figure 3. Structure of microbial TGase from *Streptomyces mobaraense*. The protein model was prepared from the X-ray structure of the enzyme available in PDB (code 1IU4) using the software PyMOL. The front and rear views of TGase are shown. The catalytic site of the enzyme is circled and residue Cys64 is shown in red color. Hydrophobic residues are indicated in blue.

4. The catalytic mechanism of TGases

TGases catalyse a ping pong reaction mechanism (Figure 4) as postulated by Folk (*32*) in which: a) the glutamine containing substrate binds to the cysteine at the active site of the enzyme to form a thioester. The formation of an acyl enzyme covalent intermediate occurs with the liberation of ammonia. b) The acyl enzyme intermediate reacts with a second substrate which is a protein bound lysine or a primary amine. This acyl acceptor forms a ε–(γ-glutamyl) lysine isopeptide bond with the glutamine residue. Through this mechanism, TGase catalyses the crosslinking between protein-bound lysine and glutamine residues forming ε-(γ-glutamyl)lysyl bridges, which is the transamiding activity. Moreover, TGase can catalyse also the deamidation of glutamine residues in the absence of a nucleophilic amine, in which water acts as a second substrate. Both the transamidation and deamidation reactions proceed through a common intermediate, which is the thioester between the carbonyl group of a glutamine residue and the cysteine residue of the enzyme.

TGase belongs to the cysteine protease family with a phylogenetic relation to papain. However, TGase differs from them, by no water access at its catalytic site. Water exclusion and attraction of nucleophilic uncharged moieties at the catalytic site favours the acyl transfer reaction onto the nucleophile. The glutamine (acyl or donor) substrate engages with the nucleophile (primary amine or acceptor) substrate in the water deprived catalytic site of TGase, leading to the formation of an isopeptide bond (*33*). Hence, two types of substrates are required for the TGase mediated bioconjugation reaction:

a) Proteins containing an acyl donor (glutamine residue) and

b) Proteins containing an acyl acceptor (lysine residue).

a) Proteins containing an acyl donor (glutamine residue) TGase shows strict specificity in recognition of glutamine protein substrates. The reactivity of the glutamine residue depends upon the type of TGase and the environment surrounding the glutamine residue (*29*). Several studies have been conducted to elucidate the pattern of amino acids near the reactive glutamine residues for unravelling the specificity of TGase (*34-38*). Sequence arrangement of QXP favours the TGase-

mediated conjugation while QXXP is not favourable for conjugation (*39*). Presence of bulky or hydrophobic amino acids near the N-terminus of the glutamine residue accelerated the reaction for Microbial TGase, whereas the presence of hydrophobic amino acids at the C terminus accelerated the reaction with fish (red sea bream) liver (FTGase) and guinea pig liver (GTGase) (*37*).

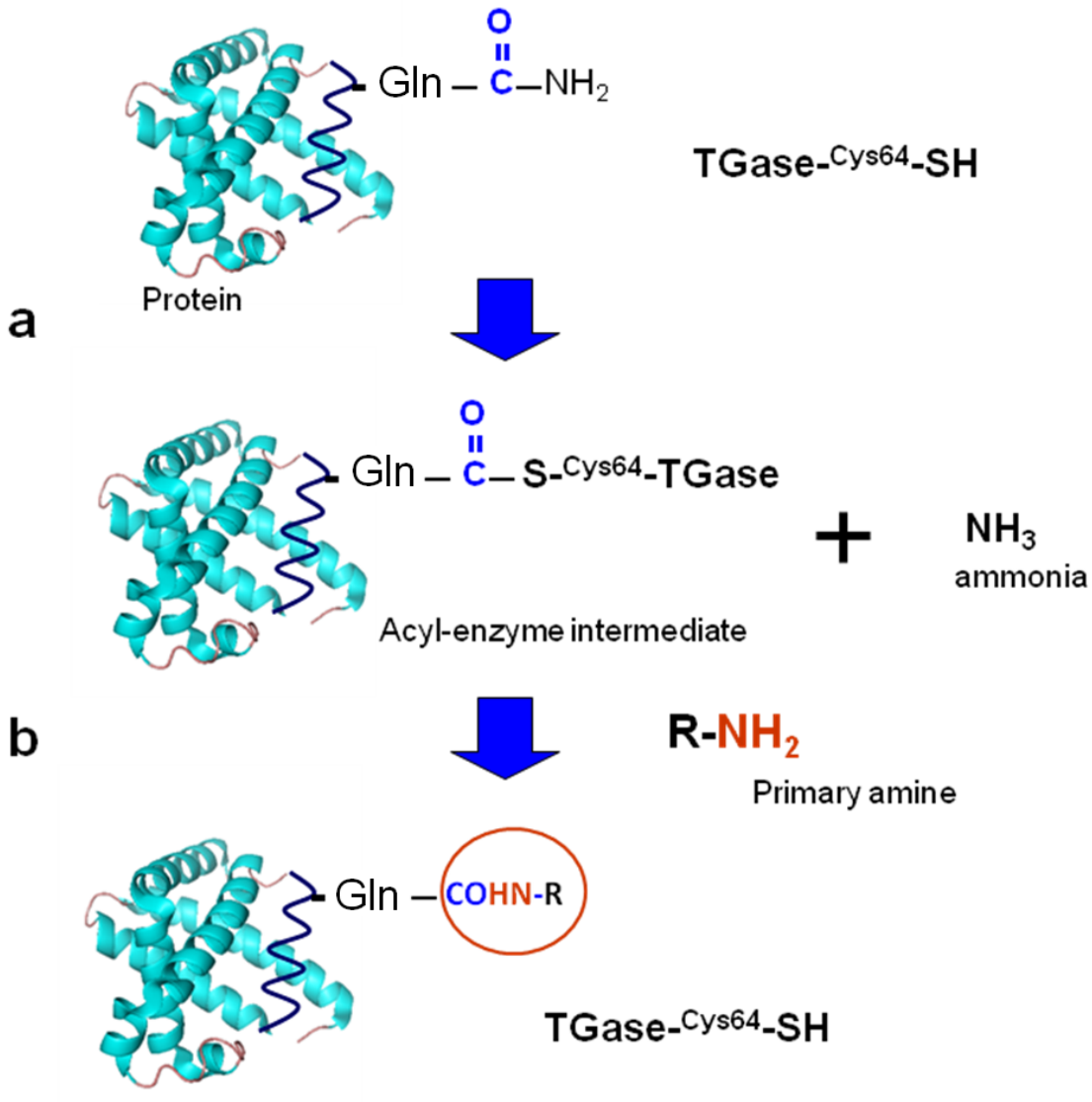


Figure 4. Schematic representation of the TGase mechanism of catalysis. a) Protein bound glutamine (Gln) forms a thioester bond with the cysteine (Cys64) in active site of TGase, leading to the formation of ammonia. b) The acyl enzyme intermediate undergoes an acyl transfer to a primary amine or protein bound lysine resulting in the formation of a covalent iso-peptide bond (indicated by the circle) with release of the free enzyme.

However, unfolded or partially unfolded globular proteins show an increased susceptibility to TGase, suggesting that consensus sequences are not enough to predict the specificity of TGase. Moreover, if two reactive glutamines are present in the protein, the modification at one glutamine with the formation of a crosslink with a lysine can change the behaviour of the substrate in terms of specificity. Therefore, presently it is not possible to predict the reactivity of specific glutamines in the protein substrate (*29*).

b) Proteins containing an acyl acceptor (lysine residue) The TGase catalysed reaction is not only specific towards glutamine residues, but also towards the amine substrates. The TGase mediated conjugation of primary amines to proteins occurs through formation of a γ- glutamyl linkage. As indicated in Figure 4, the amine containing substrate reacts with TGase after formation of the acyl-enzyme intermediate leading to the release of the free enzyme. Therefore, the site of interaction for the amine derivative is present in the acyl-enzyme intermediate. Different functional groups can be incorporated at the level of selective glutamine residues in the protein. However, in the chemical structure of the substrate these groups need to be separated from the primary amine by a spacer. Indeed, the chemical structure of the molecule carrying the primary amine plays a critical role for the TGase selectivity. Studies on the length of spacers between a chemical moiety and an amine indicated that aliphatic chains up to 6 carbon atoms are well tolerated by microbial TGase as well as GTGase (*40*).

It has also been demonstrated that the efficiency of the interaction of the acyl-enzyme intermediate with lysine containing peptides is dependent on the side chains of the amino acid residues adjacent to the TGase reactive lysine. Indeed, studies by Folk. (*41*) on the reactivity of peptides carrying a lysine residue reported that the presence of leucine before the TGase reactive lysine accelerates the reaction, whereas its presence after the lysine has a negative effect on the rate of the reaction. Moreover, occurrence of leucine in the positions not directly adjacent to the TGase reactive lysine has little or no effect on the conjugation reaction. It was proposed that a leucine residue positioned at the N-terminal of the TGase reactive lysine favours the reaction because the side chain of leucine forms hydrophobic interactions with the enzyme. On

the other end, the reduction in reactivity for Lys-Leu containg peptides can be attributed to the fact that the interaction of the hydrophobic side chain of leucine dictates the orientation of the substrate on the surface of the acyl-enzyme intermediate (*41*). Further studies have indicated that the presence of residues with basic or polar side chains preceding the lysine residue has also a positive effect for the proper orientation of the peptide at the TGase active site (*42*). On the other end, the presence of proline or glycine at the C-terminal of the TGase reactive lysine has a negative effect on the enzyme reactivity as compared to a preceding leucine, serine, alanine, or arginine residue (*42*).

5. Biotechnological applications of TGase

The TGase-mediated conjugation reaction has a wide range of applications in basic sciences as well as in the biotechnology industry. Glutamine and lysine residues in proteins can be conjugated to lysine mimicking ligands or with glutamine donor peptides, respectively (Figure 5). Glutamine containing peptides or primary amine containing substrates can carry several types of molecules as fluorescent or radioactive labels, biotin, a drug or a polymer. Among the numerous applications of the TGase reaction, the derivatization of protein drugs with poly(ethyleneglycol) (PEG) to improve their pharmacodynamic and pharmacokinetic properties will be discussed in the following paragraph along with other applications.

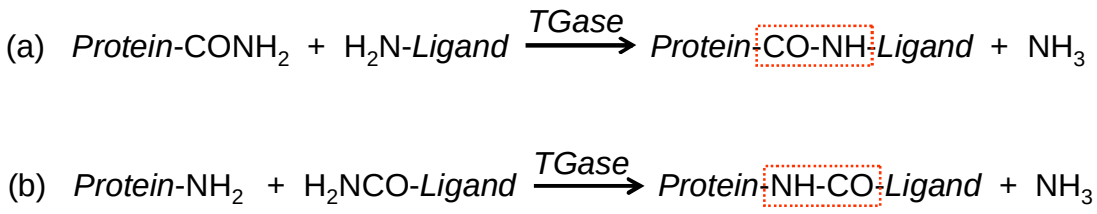


Figure 5. Schematic representation of the reaction catalysed by TGase. a) Protein bound glutamine can be conjugated with a primary amine containing ligand. Similarly, b) protein bound lysine can be conjugated with a ligand containing the glutamine donor. The ligand can be a fluorescent label, biotin, polymer, a drug entity or a chelating agent.

5.1 TGase-mediated PEGylation of protein drugs

PEGylation of protein drugs. Protein drugs when administered *in vivo* display a low bioavailability, which is mainly due to their degradation by proteolytic enzymes, their rapid kidney clearance, and their propensity to generate an immunogenic response. Several approaches have been developed to improve the clinical properties of protein drugs, which involve the alteration of the amino-acid sequences of proteins to reduce their degradation by proteases and the immunogenic response, the fusion to immunoglobulin or albumin and the incorporation of the protein into liposomes (*43-47*).

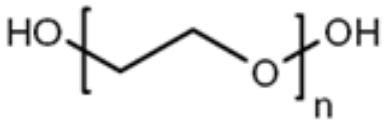


Figure 6. Chemical structure of poly(ethylene glycol) (PEG). PEG is a linear polymer given by repeating ethylene oxide units (44 Da). PEG is produced by polymerization of ethylene oxide (–CH_2-CH_2-O-) in the presence of water.

However, these approaches also show several limitations as the reduction of the catalytic efficiency or the receptor binding affinity of the protein drug. Liposomes also activate the complement cascade thus causing pseudo-allergic reactions damaging heart and liver cells (*48*). An advantageous alternative to these approaches is the PEGylation of protein drugs, which is the modification of the surface of proteins with PEG (Figure 6). In 1976, glucose-6-phosphate was first PEGylated to alter the phase partitioning constant of the enzyme in a two phase bioreactor (*49*). Later, Abuchowski (*50*) demonstrated the reduced immunogenicity of PEGylated bovine serum albumin as compared to its native form. Since then an extensive work has been done to improve protein properties by PEGylation. Several PEGylated proteins drugs have been developed (Table 2) mainly with a aim to decrease the immunogenicity and increase the biological half life of the protein drug (*51*).

Table 2. Commercially available PEGylated protein drugs

PEG-conjugated protein (Commercial name) Company	PEGylation method	Disease	Date of approval
PEG ademase bovine (Adagen®) Enzon	Amine chemistry with random, multiple linear 5kDa PEG	Adenosine deaminase deficiency associated with SCID	1990 (US)
PEG-L-asparaginase (Oncaspar®) Enzon	Amine chemistry with random, multiple linear 5 kDa PEG	Acute lymphoblastic leukemia	1994 (US, EU)
PEG interferon alfa-2b (Pegintron®) Schering-Plough	Amine chemistry with random linear 12 kDa PEG	Hepatitis C, cancer, multiple sclerosis, AIDS	2000 (EU) 2001 (US)
PEG interferon alfa-2a (Pegasys®) Hoffmanm-La Roche	Amine chemistry with random, branched 40kDa PEG (two 20 kDa linear)	Hepatitis C	2002 (EU, US)
PEG filgrastim (Neulasta®) Amgen	N terminal selective amine chemistry with Linear 20kDa PEG	Induced neutropenia	2002 (US) 2003 (EU)
Pegvisomant (Somavert®) Pharmacia & Upjohn	Amine chemistry with random 4-6 linear 5kDa PEG	Acromegaly	2002 (US) 2003 (EU)
Pegaptanib sodium (Macugen®) OSI/Pfizer	Selective amine chemistry with branched 40kDa PEG (two 20 kDa linear)	Age related macular degeneration	2004 (US) 2006 (EU)
mPEG-epoetin beta CERA (Mircera®) Roche	Amine chemistry with random linear 30kDa PEG	Anemia associated with chronic renal failure	2007 (US, EU)
Certolizumab pegol (Cimzia®) UCB	Thiol chemistry with selective branched 40 kDa	Crohn's disease	2008 (US)

PEGylation of protein drugs leads to an improvement of the efficacy and safety of the drug, mainly by three reasons (Figure 7):

a) Reduced proteolytic degradation and immunogenicity. The presence of PEG prevents the binding of the protein to proteases and antibodies, thus increasing the proteolytic stability and reducing the immunogenicity of these drugs. These effects eventually provide a longer bioavailability (*51*). Even hypersensitivity reactions have been demonstrated to be decreased with the use of PEG-protein conjugates (*52*). Indeed, due to an increase in the half life of the drug, the protein is administered at low dosages reducing the side effects.

b) Longer half life and improved therapeutic efficacy. The clearance of the protein *in vivo* is due to passive glomerular filtration, which is dependent on the size of the molecule. The increased molecular size and larger hydrodynamic radius of the PEGylated protein drug provides slower clearance of the protein-PEG conjugate leading to a longer biological half life. However, in the preparation of PEG-protein conjugates, a threshold of 60kDa and a hydrodynamic radius of 45 Å has to be considered. Indeed, over this limit the PEG-conjugate remains in circulation for a very long period and is accumulated in the liver. Systematically administered PEG-protein conjugates have also demonstrated a decrease in their cellular clearance i.e. by RES or specific cell receptor-protein interaction. Lactoferrin is cleared by its interaction with a receptor on the hepatocytes. Studies on lactoferrin conjugated with PEG-2000 and PEG-4000 by Beauchamp (*53*) clearly showed the increase in the half life of the protein by 5 fold and 20 fold, respectively (*53*).

c) Improved physico-chemical properties. The PEG chain is hydrophilic in nature and forms a hydrated water shell, thus increasing the solubility of the PEG-protein conjugate. This characteristic is quite advantageous during the formulation and administration of protein drugs at physiological pH. IL-2 (interleukin-2) precipitates at pH 7, whereas PEG-IL-2 is readily soluble even at 20 mg/ml (*51*). Conjugation to PEG also reduces the tendency of aggregation that leads to the formation of aggregates at the administration sites causing phlebitis. Importantly, conformational analysis of proteins after conjugation to PEG indicates that the polymer does not cause major changes at the structural level (*54-55*).

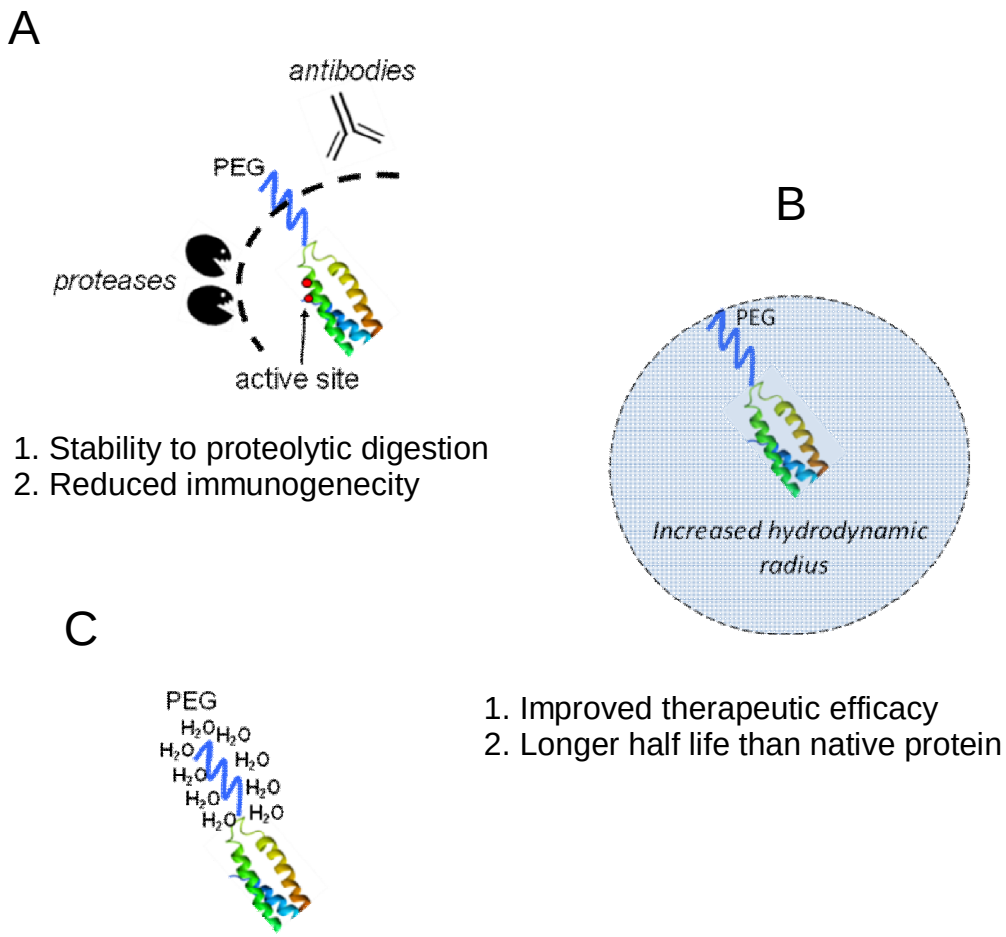


Figure 7. Advantages of PEGylation of the protein drugs. A) Reduced proteolytic degradation and immunogenicity, B) improved therapeutic efficacy and longer half life and C) improved physical properties.

Chemical methods used for the PEGylation of proteins. Several methods of PEGylation of proteins have been reported in the literature (*56-61*). They can be divided into methods that lead to the addition of several chains of PEG to the protein and methods that allow the production of mono-conjugated PEG derivatives. Clearly, the production of a heterogeneous mixture of PEGylated protein derivatives leads to a number of issues such as a non consistent bioactivity and variations in the physical and pharmacological properties in different production batches. Moreover, the purification of the PEGylated protein isoforms is tedious and economically unfeasible, which eventually makes the characterization of the site(s) of conjugation and further scale-up of the production process quite difficult. Even the drug approval authorities demand a detailed characterization of the PEGylated protein with the description of the polydispersity index of PEG, degree of PEGylation and exact site(s) of modification (*62-63*). Therefore, much of the recent research has been focused on the development of site specific PEGylation approaches to generate a homogenous product. An overview of the chemical methods of protein PEGylation is given below, with particular attention to the approaches that allow a site-specific derivatization of the protein.

a)Conjugation at the level of amine groups. In the first generation of the chemical methods of PEG conjugation, the amine selective chemistry was widely utilized for the generation of PEGylated proteins. In this approach, the PEG reagents were conjugated on the primary amine groups in the target protein. The chemistry of the acylating PEG reagents was based on carbonate and succinimidyl esters with the formation of amide bonds. PEG carbonate esters display a low reactivity towards amines, which allows selectivity in the reaction e.g. PEG benzotriazolyl carbonate (*64*), whereas PEG succinimidyl succinate esters have high reactivity towards amines providing a low selectivity e.g. PEG succinimidyl succinate (*65*). However, this amine selective chemistry approach had the drawback of side reactions with the side chain hydroxyl groups of serine or tyrosine.

The amine-chemistry based PEGylation approach was improved by the replacement of the terminal hydroxyl group of monomethoxy-PEG with a carbonyl group (*66*), followed by its esterification with N hydroxysuccinimide (NHS) to

generate an activated ester, which readily reacts with amines in the protein. The newly formed amide bond is resistant towards hydrolysis. However, the PEG-NHS reagent is highly reactive and undergoes hydrolysis, creating an obstacle to its use. The stability problem was overcome by increasing the length of the alkyl chain e.g. butanoic NHS ester (*59*).

A further improvement in the selectivity of the amine derivatization chemistry was the development of PEG-aldehydes (Figure 8A). In this reaction, the PEG-aldehyde reacts with the N-terminal amino group of the protein to form a reversible Schiffs base, which is reduced by sodium cyanoborohydride to form a stable alkyl amine bond. The reaction is specific towards the N-terminal amino group of the protein due to its lower pKa value (pKa 7.6-8.0) as compared with the ε- amine groups of lysine residues (pKa 10.0–10.2) (*67-68*). The derivatization of proteins with PEG-aldehyde greatly reduces the formation of heterogeneous and multi-PEGylated derivatives. A successful example of this approach is PEG-filgrastim (G-CSF) which has been commercialized under the brand Neulasta® by Amgen. The PEGylation chemistry involves the conjugation of PEG-aldehyde (20 kDa, linear) on the N-terminal methionine residue of G-CSF (*67, 69*). However, the implementation of this approach on other proteins is difficult because the conditions for the N-terminal PEGylation reaction are different for each protein. Indeed, for every protein it is important to determine the optimal pH to achieve a site-specific reaction at the amine terminus, since small differences in the pKa along the protein sequence can lead to multiply conjugated protein derivatives (*70*).

b) Conjugation at the level of the thiol group. In this approach, free thiol groups in the target protein are conjugated with PEG reagents. PEG vinyl sulphone and PEG maleimide reacts through a Michael addition reaction (*71*). The reaction with PEG vinyl sulphone proceeds rapidly, but also results in side reactions with the primary amine groups of lysines, whereas the reaction with PEG maleimide is quite unstable in water. PEG-iodoacetamide reacts through a nucleophilic substitution reaction (*58*) which is quite slow and requires the use of an the PEG reagent. PEG o-pyridyl disulphide reacts through the disulphide exchange reaction (*67*), which requires the presence of a free cysteine group in the protein.

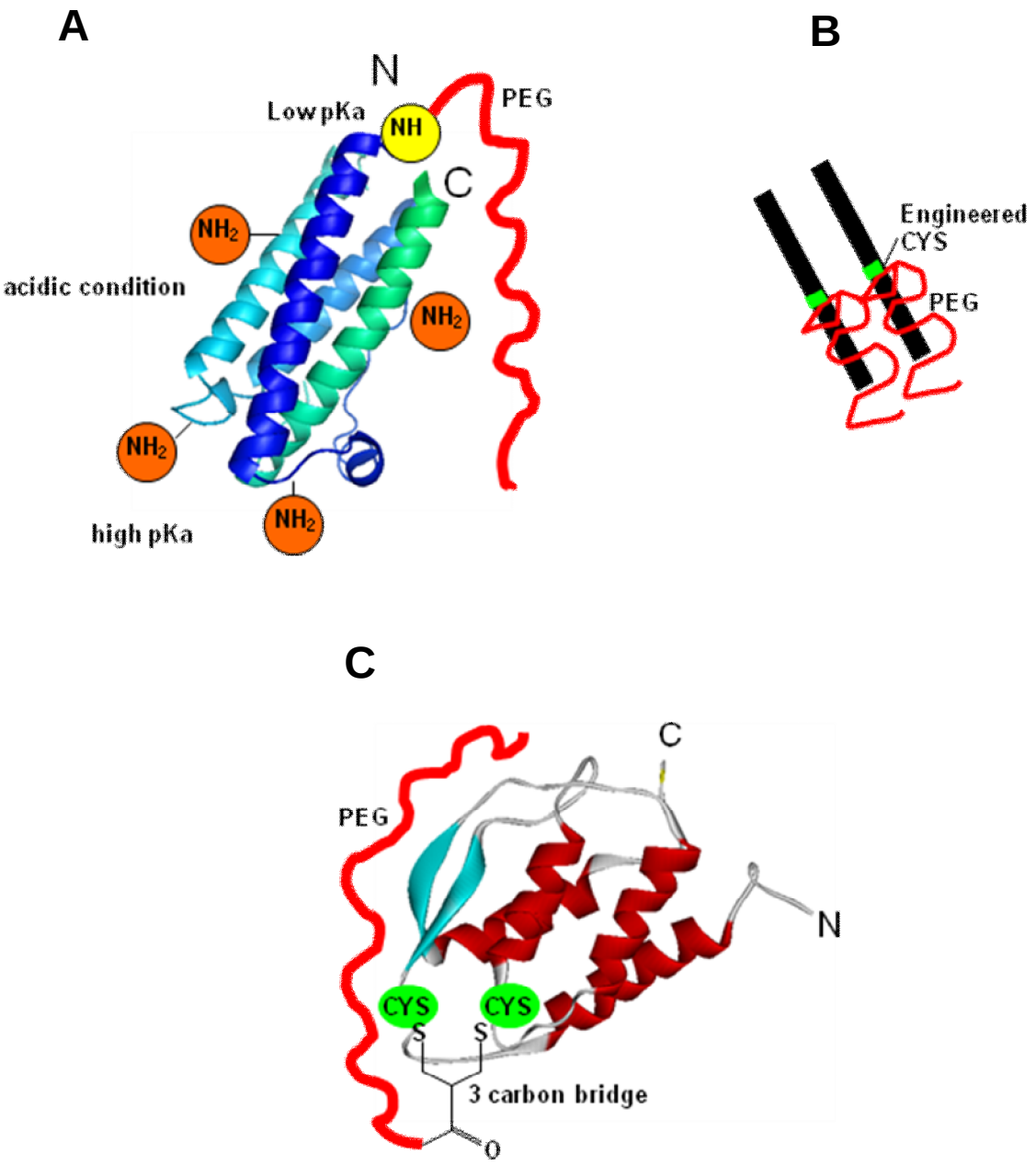


Figure 8. Schematic representation of site-specific chemical methods of PEGylation. A) N-terminal site specific PEGylation. The N-terminal amino group (the site of the modification) is shown in yellow whereas the other amino groups of the proteins in orange. The PEG chain is coloured in red. B) Site specific PEGylation at the level of cysteine residues. The engineered cysteine residue is shown in green on the Fab' fragment of the anti-TNF α mAb (see text). The attached PEG chain is coloured in red. C) Disulphide bridge mediated site-specific PEGylation of proteins. Native cysteine residues in the protein are shown in green. The thiol groups of these cysteines are linked by a 3 carbon bridge derivatised with the PEG chain.

At present, the coupling at the free cysteine using the thiol reactive PEG reagent is a highly selective and rapid method for the PEGylation of proteins. However, the occurrence of a free cysteine residue is very rare in protein sequences as compared to the amino group containing amino acid residues and most of the cysteines residues form disulphide bridges within the protein. By site-specific mutagenesis a free cysteine can be inserted at any desired position in the protein and this approach has been successfully implemented in the commercialized anti TNF-α monoclonal antibody (mAb) (Figure 8B) (*72*). Anti-TNF-α-mAb-PEG (Certolizumab pegol) was approved in 2008 and it is marketed under the brand CIMZIA® by UCB (*73*). It is a PEGylated antigen-binding fragment (Fab') devoid of the Fc region. A free cysteine is engineered at the hinge region of Fab' which is site specifically conjugated with a 40 kDa PEG. The PEG chain is located at the C-terminus of the mAb, hence it does not interfere with the active Fab' region. However, the implementation of this approach on other complex proteins is challenging and expensive. Even though this method provides high selectivity and yield of the PEGylation reaction, the insertion of a free cysteine residue can lead to structural changes in the protein with increased immunogenicity and aggregation, and a reduction in its activity. Moreover, the free cysteine has a thiol group, which is highly reactive and can form intramolecular crosslinked dimers.

Another strategy based on the thiol chemistry is the disulphide bridging based site specific PEGylation. It involves the conjugation of PEG through a three carbon bridge, which replaces a native disulphide bridge. Protein consists of accessible and buried disulphide bridges. The accessible disulphide bridges can be easily reduced under mild reducing conditions. These reduced thiols are further reacted with the bis (thiol) specific reagent containing PEG, which leads to the bridging of the cysteine thiols maintaining the native structure of the protein (Figure 8C) (*56, 70*). The approach has proved to be quite efficient with L- asparaginase. The disulphide bridge based PEGylated L-asparaginase showed no loss of activity as compared to the commercially available PEGylated form PEG-L-asparaginase (Oncaspar®) marketed by Enzon. The three carbon disulphide bridged PEGylated Interferon α-2b (IFN) showed antiviral activity, which was approximately 8% of the native IFN but similar to the activity of the commercially available PEGylated IFN α-2b (Pegasys®)

marketed by Hoffmanm-La Roche (~ 7 % of native IFN) (*56, 74-75*). Interestingly, the decrease in the biological activity of the disulphide bridging based PEGylated proteins was seen to be independent of the size of PEG. Clearly, the use of this approach is restricted to few proteins since they have to contain an accessible disulphide bridge having the same length as that of the three carbon based linker.

c) Other conjugation chemistries. Other approaches for chemical PEGylation involve the use of PEG-isocyanate for the conjugation at the hydroxyl groups of a protein (*76*). PEG-hydrazide specifically PEGylates the carboxylic groups at low pH conditions (4.5-5.0). (*77*). PEG-1,3-dioxocompounds are used for the PEGylation at the level of arginine residues but the conjugation reaction lacks specificity i.e. the conjugation reaction may also occur on lysine and histidine residues (*78-79*).

TGase mediated PEGylation. The PEGylation methods described above suffer from the limitations of generating an heterogeneous product with a decreased bioactivity. Therefore, the development of alternative approaches for the site-specific PEGylation of proteins is essential for the generation of more useful and safer protein drugs which are economically feasible. The use of TGase offers a valid alternative for the production of PEGylated protein drugs. Indeed, recent studies have demonstrated that TGase catalyses the conjugation of PEG-amine at the level of specific glutamine residues, even in proteins that contain numerous glutamines in their amino acid sequence (*80*). In the literature, examples of the TGase-mediated PEGylation using PEG-amines have already been published. They include a mono-PEGylated interleukin-2 derivative, in which out of 6 glutamine residues the conjugation was observed only at the level of Gln74 (*75-76*). Granulocyte- colony stimulating factor (G-CSF) was PEGylated at Gln134 among the 17 glutamines in the G-CSF amino acid sequence (*81*). In the case of human growth hormone (hGH), PEGylation by TGase occurred at the level of Gln40 and Gln141, in spite of the presence of 13 glutamines in the amino acid sequence (*80-81*). In PEGylated erythropoietin (EPO), the TGase mediated conjugation with dansyl cadaverine occurred at the level of two glutamine residues out of the 7 glutamines present in the sequence of the protein (*82*).

5.2 TGase applications in the food industry

Microbial TGase is widely used in the food industry to improve the physical properties of the food products, such as soy, muscle, myosin, globulin, casein, whey etc (*83-85*). Indeed, TGase catalyses the formation of a covalent bond between glutamine and lysine residues resulting into cross-linking of proteins. Such modification changes the properties of these protein as viscosity, size, conformation, stability and gelation, which eventually improves the texture of food (*80-81*). For example, TGase treatment of wheat flour improves the texture and distribution of pores resulting in higher loaves of bread and pastry (*86*). Better mechanical properties were demonstrated for surmi gels derived by the TGase treatment of silver carp (*87*). TGase treatment of milk increased the firmness of yogurt due to the formation of iso-peptide bonds (*88*).

5.3 TGase applications in protein labelling

The TGase mediated conjugation reaction has been extensively used for research purposes to obtain the labelling of proteins. For example, fluorescein and its derivatives offer advantages in being relatively high absorptive and in providing excellent fluorescent quantum yields and high water solubilities. Generally, chemical methods of modification involve the incorporation of the fluorescent label by amine-reactive (fluorescein-NHS ester) (*89*) and thiol reactive (fluorescein-maleimide) reagents (*90-91*). In these strategies the number of reactive amines or sulfhydryl groups has a major impact on the degree of labelling, which leads to an heterogeneous labelling of proteins. The enzymatic modification of proteins by TGase for the incorporation of fluorescent labels is a promising alternative with the potential to obtain a homogenous site-selective labelling (*92*).

6. AIMS OF THE THESIS

The reaction mechanism of transglutaminase (TGase) involves the intermediate formation of a thioester at the level of the Cys residue of the active site of the enzyme (TGase-SH + H_2NCO-protein → TGase-SCO-protein + NH_3) and then the reactive thioester reacts with a nucleophile, which can be the ε-amino group of a lysine (Lys) residue or a variety of other nucleophiles, including water. In this last case, the original enzyme is formed and the protein substrate suffers hydrolysis of a protein-bound Gln residue to a glutamic acid (Glu) residue. Clearly, the formation of the reactive thioester intermediate involves the binding and adaptation of a Gln-containing peptide segment of the protein substrate at the TGase's active site, so that it can be anticipated that this intermediate would be difficult to be made if Gln is embedded in a globular, well-folded protein substrate. On the other hand, an easy formation of the intermediate is expected with a small, flexible Gln-peptide or, in general, a small ligand bearing an amido moiety mimicking a Gln residue. This general reasoning would explain the fact that in a number of cases it has been found that TGase does not react at all with some globular proteins, while rather unstructured and flexible proteins instead are easily attacked by TGase. In few cases, even a very selective reaction of TGase was shown to occur at the level of just one Gln residue among the many present in a protein. While data reported in the literature have documented some specificity for Gln residue (s), protein-bound Lys residues instead do appear to display a poor selectivity in TGase-mediated reactions.

Several studies were conducted in the past to unravel the TGase mechanism of action and, in particular, to explain in molecular terms the reason why TGase reacts with some preference or even strict specificity with protein-bound Gln residue(s). In general, the idea was expressed that surface exposure of a protein-bound Gln would dictate the reactivity towards TGase. This proposal was based on the fact that several globular proteins do not react in their native state, but only if denatured. However, it was also clear that several Gln residues are exposed to solvent in a native protein and, therefore, exposure or protrusion of Gln appears to be not at all sufficient to explain the observed selectivity of reaction.

The aim of this PhD work was to examine the TGase-mediated reactions using few model proteins of known structure and dynamics in order to unravel molecular features of the protein substrate dictating the enzymatic reactions at the level of Gln residue(s). Studies were performed on three proteins, namely apomyoglobin (apoMb), ribonuclease (RNase) and lysozyme (LYS), extensively investigated in the past and used as model proteins for investigating fundamental aspects of protein structure, folding, stability and function of proteins. The TGase-mediated reactions at the level of Gln on these protein substrates were studied by using several amino-derivatives as acyl acceptors, as for example dansyl-cadaverine. The modified proteins were isolated and characterised by fingerprinting techniques combined with mass spectrometry MS) in order to precisely identify the reacted Gln residues along the protein chains. In order to analyse the reactivity of Lys residues we used the Gln-containing reagent N-carbobenzoxy-Gln-Gly-OH (ZQG) as acyl donor for the TGase reactions and similarly we identified the reacted Lys in a protein. Moreover, the protein substrates were used in their native, as well as in a perturbed state by cleaving a single peptide bond (RNase) or just one disulphide bridge (LYS).

To provide an explanation of the reacted Gln or Lys, an analysis of the structure and dynamics of the investigated proteins was conducted by using their known 3D structures, concluding that protein flexibility or even unfolding at the site of TGase attack strongly dictates the observed site-specific TGase-mediated reactions at the level of Gln residues. Conversely, we have shown that there is a poor selectivity for Lys residues in TGase reactions, but that Lys residues embedded in a flexible or unfolded chain segment of the protein substrate react faster in a TGase reaction. It seems possible to explain the reactivity of Lys in respect to that of Gln by considering that (i) the reactive thioester intermediate can react with a variety of non-specific nucleophiles (see above) and the fact that (ii) the long $-(CH_2)_4-NH_2$ alkylamino side chain of a Lys residue can allow both exposure and flexibility of the reactive amino group, thus acting as an arm that facilitates the aminolysis of the acyl-donor thioester.

A main conclusion deriving from this work is that the enzymatic action of TGase can be related to that of a protease. In particular, it will be shown that the sites

or chain regions in a protein substrate prone to a TGase attack can also be selectively attacked by a protease. Of interest, flexibility or segmental mobility has been previously used to explain limited proteolysis phenomena and, therefore, we conclude that the same dynamic features, *i.e.* enhanced chain flexibility or local unfolding of the protein substrate, are dictating site-specific reactions with both TGase and a protease. Of interest, in this respect we can consider that TGase acts as a reverse protease that allow the synthesis instead of hydrolysis of an amide bond and that the active site of TGase involves a catalytic triad Cys-His-Asp in analogy to a protease.

An additional and practical aim of this PhD work was to possibly use TGase for the modification of proteins of pharmaceutical interest using an amino-derivative of poly(ethylene)glycol (PEG-NH_2) as an acyl-acceptor in a TGase-mediated reaction, resulting in a protein PEGylated at the level of Gln residue(s). Initially, enzymatic PEGylation was investigated with the model protein apoMb and then extended to other protein drugs. Of interest is the fact that TGase can mediate a site-specific PEGylation, thus allowing to produce homogeneous protein bioconjugates. This fact offers a great advantage over the chemical methods in current use to produce more soluble, non-aggregating e non-immunogenic protein drugs, since these methods usually suffer from the drawbacks of leading to heterogeneous and over-labelled protein derivatives.

7. References

1. Clarke, D.D., Mycek, M.J., Neidle, A., and Waelsch, H. (1959) The incorporation of amines into protein, *Archives of Biochemistry and Biophysics 79*, 338–354.

2. Pisano, J.J., Finlayson, J.S., and Peyton, M.P. (1968) Cross-link in Fibrin Polymerized by Factor XIII: ε-(γ-Glutamyl)lysine, *Science 160*, 892–893.

3. Gladner, J.A., and Nossal, R. (1983) Effects of crosslinking on the rigidity and proteolytic susceptibility of human fibrin clots, *Thromb Res 30*, 273–288.

4. Ichinose, A., McMullen, B.A., Fujikawa, K., and Davie, E.W. (1986) Amino acid sequence of the b subunit of human factor XIII, a protein composed of ten repetitive segments, *Biochemistry 25*, 4633–4638.

5. Lorand, J.B., Pilkington, T.R., and Lorand, L. (1966) Inhibitors of fibrin cross-linking: relevance for thrombolysis, *Nature 210*, 1273–1274.

6. Bishop, P.D., Teller, D.C., Smith, R.A., Lasser, G.W., Gilbert, T., and Seale, R.L. (1990) Expression, purification, and characterization of human factor XIII in Saccharomyces cerevisiae, *Biochemistry 29*, 1861–1869.

7. Hornyak, T.J., Bishop, P.D., and Shafer, J.A. (1989) Alpha-thrombin-catalyzed activation of human platelet factor XIII: relationship between proteolysis and factor XIIIa activity, *Biochemistry 28*, 7326–7332.

8. Hornyak, T.J., and Shafer, J.A. (1991) Role of calcium ion in the generation of factor XIII activity, *Biochemistry 30*, 6175–6182.

9. Folk, J.E., and Cole, P.W. (1965) Structural Requirements of Specific Substrates for Guinea Pig Liver Transglutaminase, *J. Biol. Chem. 240*, 2951–2960.

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

11. Noguchi, K., Ishikawa, K., Yokoyama, K., Ohtsuka, T., Nio, N., and Suzuki, E. (2001) Crystal structure of red sea bream transglutaminase, *J. Biol. Chem. 276*, 12055–12059.

12. Beninati, S., and Piacentini, M. (2004) The transglutaminase family: an overview: minireview article, *Amino Acids 26*, 367–372.

13. Araki, H., and Seki, N. (1993) Comparison of reactivity of transglutaminase to various fish actomyosins. , *Nippon Suisan Gakkaish 59* 711–716.

14. Myhrman, R., and Bruner-Lorand, J. (1970) Lobster muscle transpeptidase, In *Methods Enzymol.* (Gertrude E. Perlmann, L. L., Ed.), pp 765–770, Academic Press.

15. Tokunaga, F., Yamada, M., Miyata, T., Ding, Y.L., Hiranaga-Kawabata, M., Muta, T., Iwanaga, S., Ichinose, A., and Davie, E.W. (1993) Limulus hemocyte transglutaminase. Its purification and characterization, and identification of the intracellular substrates, *Journal of Biological Chemistry 268*, 252–261.

16. Singer, M.A., Hortsch, M., Goodman, C.S., and Bentley, D. (1992) Annulin, a protein expressed at limb segment boundaries in the grasshopper embryo, is homologous to protein cross-linking transglutaminases, *Dev. Biol. 154*, 143–159.

17. Ruiz-Herrera, J., Iranzo, M., Elorza, M., Sentandreu, R., and Mormeneo, S. (1995) Involvement of transglutaminase in the formation of covalent cross-links in the cell wall of <i>Candida albicans</i&gt, *Archives of Microbiology 164*, 186–193.

18. Singh, R.N., and Mehta, K. (1994) Purification and characterization of a novel transglutaminase from filarial nematode Brugia malayi, *European Journal of Biochemistry 225*, 625–634.

19. Margosiak, S.A., Dharma, A., Bruce-Carver, M.R., Gonzales, A.P., Louie, D., and Kuehn, G.D. (1990) Identification of the Large Subunit of Ribulose 1,5-Bisphosphate Carboxylase/Oxygenase as a Substrate for Transglutaminase in Medicago sativa L. (Alfalfa), *Plant Physiol. 92*, 88–96.

20. Della Mea, M., Caparros-Ruiz, D., Claparols, I., Serafini-Fracassini, D., and Rigau, J. (2004) AtPng1p. The First Plant Transglutaminase, pp 2046–2054.

21. Ramanujam, M.V., and Hageman, J.H. (1990) Intracellular transglutaminase (EC 2.3.2.13) in a prokaryote: evidence from vegetative and sporulating cells of Bacillus subtilis., *FASEB J. 4*.

22. Ando, H., Adachi, M., Umeda, K., Matsuura, A., Nonaka, M., Uchio, R., Tanaka, H., and Motoki, M. (1989) Purification and characteristics of a novel transglutaminase derived from microorganisms, *Agric. Biol. Chem. 53*, 2613.

23. Færgemand, M., Otte, J., and Qvist, K.B. (1997) Enzymatic cross-linking of whey proteins by a Ca2+-independent microbial transglutaminase from Streptomyces lydicus, *Food Hydrocolloids 11*, 19–25.

24. Schmidt, G., Selzer, J., Lerm, M., and Aktories, K. (1998) The Rho-deamidating cytotoxic necrotizing factor 1 from Escherichia coli possesses transglutaminase activity. Cysteine 866 and histidine 881 are essential for enzyme activity, *J. Biol. Chem. 273*, 13669–13674.

25. Duran, R., Junqua, M., Schmitter, J.M., Gancet, C., and Goulas, P. (1998) Purification, characterisation, and gene cloning of transglutaminase from Streptoverticillium cinnamoneum CBS 683.68, *Biochimie. 80*, 313–319.

26. Tsai, G.-J., Lin, S.-M., and Jiang, S.-T. (1996) Transglutaminase from Streptoverticillium ladakanum and application to minced fish product, *Journal of Food Science 61*, 1234–1238.

27. Locci, R. (1989) Streptomyces and related Genera., In *Bergey's Manual of Systematic Bacteriology* (Krieg, N. R., and Holt, J. G., Eds.), pp 2451–2492, Williams and Wilkins, Baltimore, MD.

28. Pasternack, R., Dorsch, S., Otterbach, J.T., Robenek, I.R., Wolf, S., and Fuchsbauer, H.L. (1998) Bacterial pro-transglutaminase from Streptoverticillium mobaraense--purification, characterisation and sequence of the zymogen, *Eur. J. Biochem. 257*, 570–576.

29. 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.

30. Ohtsuka, T., Umezawa, Y., Nio, N., and Kubota, K. (2001) Comparison of Deamidation Activity of Transglutaminases, *Journal of Food Science 66*, 25–29.

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

32. Folk, J.E. (1969) Mechanism of action of guinea pig liver transglutaminase. VI. Order of substrate addition, *J. Biol. Chem. 244*, 3707–3713.

33. Nemes, Z., Csosz, E., Petrovski, G., and Fesus, L. (2005) Structure-Function Relationship of Transglutaminse- A Contemporary view., In *Transglutaminase. Prog. Exp. Tum. Res.* (Mehta, K., and Eckert, R., Eds.), pp 19–36, Karger, Basel.

34. 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 (Pt 3)*, 929–930.

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

36. 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.

37. 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.

38. 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.

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

40. 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*, 6230–6233.

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

42. Grootjans, J.J., Groenen, P.J., and de Jong, W.W. (1995) Substrate requirements for transglutaminases. Influence of the amino acid residue preceding the amine donor lysine in a native protein, *J. Biol. Chem. 270*, 22855–22858.

43. Harris, J.M., and Chess, R.B. (2003) Effect of pegylation on pharmaceuticals, *Nat Rev. Drug Discov. 2*, 214–221.

44. Lyczak, J.B., and Morrison, S.L. (1994) Biological and pharmacokinetic properties of a novel immunoglobulin-CD4 fusion protein, *Arch. Virol. 139*, 189–196.

45. Mateo, C., Lombardero, J., Moreno, E., Morales, A., Bombino, G., Coloma, J., Wims, L., Morrison, S.L., and Perez, R. (2000) Removal of amphipathic epitopes from genetically engineered antibodies: production of modified immunoglobulins with reduced immunogenicity, *Hybridoma 19*, 463–471.

46. Syed, S., Schuyler, P.D., Kulczycky, M., and Sheffield, W.P. (1997) Potent antithrombin activity and delayed clearance from the circulation characterize recombinant hirudin genetically fused to albumin, *Blood 89*, 3243–3252.

47. Allen, T.M. (1997) Liposomes. Opportunities in drug delivery, *Drugs 54 Suppl 4*, 8–14.

48. Szebeni, J. (1998) The interaction of liposomes with the complement system, *Crit Rev. Ther. Drug Carrier Syst. 15*, 57–88.

49. Pollak, A., and Whitesides, G.M. (1976) Organic synthesis using enzymes in two-phase aqueous ternary polymer systems, *Journal of the American Chemical Society 98*, 289–291.

50. Abuchowski, A., van Es, T., Palczuk, N.C., and Davis, F.F. (1977) Alteration of immunological properties of bovine serum albumin by covalent attachment of polyethylene glycol, *J. Biol. Chem. 252*, 3578–3581.

51. Delgado, C., Francis, G.E., and Fisher, D. (1992) The uses and properties of PEG-linked proteins, *Crit. Rev. Ther. Drug Carrier Syst. 9*, 249–304.

52. Fuertges, F., and Abuchowski, A. (1990) The clinical efficacy of poly(ethylene glycol)-modified proteins, *Journal of Controlled Release 11*, 139–148.

53. Beauchamp, C.O., Gonias, S.L., Menapace, D.P., and Pizzo, S.V. (1983) A new procedure for the synthesis of polyethylene glycol-protein adducts; Effects on function, receptor recognition, and clearance of superoxide dismutase, lactoferrin, and [alpha]2-macroglobulin, *Anal. Biochem. 131*, 25–33.

54. Suzuki, T., Kanbara, N., Tomono, T., Hayashi, N., and Shinohara, I. (1984) Physicochemical and biological properties of poly(ethylene glycol)-coupled immunoglobulin G, *Biochim. Biophys. Acta 788*, 248–255.

55. Veronese, F.M., Sartore, L., Schiavon, O., and Caliceti, P. (1990) A comparative study of enzymatic, structural, and pharmacokinetic properties of superoxide dismutase isolated from two sources and modified by monomethoxypolyethylene glycol using different methods of coupling, *Ann N Y Acad. Sci. 613*, 468–474.

56. Balan, S., Choi, J.W., Godwin, A., Teo, I., Laborde, C.M., Heidelberger, S., Zloh, M., Shaunak, S., and Brocchini, S. (2007) Site-specific PEGylation of protein disulfide bonds using a three-carbon bridge, *Bioconjug. Chem. 18*, 61–76.

57. DeFrees, S., Wang, Z.G., Xing, R., Scott, A.E., Wang, J., Zopf, D., Gouty, D.L., Sjoberg, E.R., Panneerselvam, K., Brinkman-Van der Linden, E.C., Bayer, R.J., Tarp, M.A., and Clausen, H. (2006) GlycoPEGylation of recombinant therapeutic proteins produced in Escherichia coli, *Glycobiology 16*, 833–843.

58. Hinds, K.D. (2005) Protein conjugation, cross-linking, and PEGylation, In *Biomaterials for Delivery and Targeting of Proteins and Nucleic Acids* (Mahato, R. I., Ed.), pp 119–185, CRC Press LLC, Boca Raton, Fla.

59. Roberts, M.J., Bentley, M.D., and Harris, J.M. (2002) Chemistry for peptide and protein PEGylation, *Adv. Drug Deliv. Rev. 54*, 459–476.

60. Sato, H. (2002) Enzymatic procedure for site-specific pegylation of proteins, *Adv. Drug Deliv. Rev. 54*, 487–504.

61. Veronese, F.M., and Pasut, G. (2006) PEGylation of Proteins as Tailored Chemistry for Optimized Bioconjugates, In *Polymer Therapeutics I Polymers as Drugs, Conjugates and Gene Delivery Systems* (Satchi-Fainaro, R., and Duncan, R., Eds.), pp 95–134, Springer.

62. Fontana, A., Spolaore, B., Mero, A., and Veronese, F. (2009) The site-specific TGase-mediated PEGylation of proteins occurs at flexible sites., In *PEGylated Protein Drugs: Basic Science and Clinical Applications* (Veronese, F., Ed.), pp 89–112, Birkhäuser.

63. Reichert, J.M. (2003) Trends in development and approval times for new therapeutics in the United States, *Nat. Rev. Drug Discov. 2*, 695–702.

64. Dolence, E.K., Hu C., Tsang R., Sanders, G.;, C., and S., O. (1997) Electrophilic polyethylene oxides for the modification of polysaccharides, polypeptides (proteins) and polymer surfaces. , (Surface Engineering Technologies, D. o. I., Inc., USA, Ed.).

65. Kogan, T.P. (1992) The Synthesis of Substituted Methoxy-Poly(Ethyleneglycol) Derivatives Suitable for Selective Protein Modification, *Synthetic Communications: An International Journal for Rapid Communication of Synthetic Organic Chemistry 22*, 2417–2424.

66. Zalipsky, S., Seltzer, R., and Menon-Rudolph, S. (1992) Evaluation of a new reagent for covalent attachment of polyethylene glycol to proteins, *Biotechnol. Appl. Biochem. 15*, 100–114.

67. Kinstler, O., Molineux, G., Treuheit, M., Ladd, D., and Gegg, C. (2002) Mono-N-terminal poly(ethylene glycol)-protein conjugates, *Adv. Drug Deliv. Rev. 54*, 477–485.

68. Lee, H., Jang, I.H., Ryu, S.H., and Park, T.G. (2003) N-terminal site-specific mono-PEGylation of epidermal growth factor, *Pharm. Res. 20*, 818–825.

69. Kinstler, O.B., Brems, D.N., Lauren, S.L., Paige, A.G., Hamburger, J.B., and Treuheit, M.J. (1996) Characterization and stability of N-terminally PEGylated rhG-CSF, *Pharm. Res. 13*, 996–1002.

70. Choi, J., Godwin, A., Balan, S., Bryant, P., Cong, Y., Pawlisz, E., Porssa, M., Rumpf, N., Singh, R., Powell, K., and Brocchini, S. (2009) Rebridging disulphides: site-specific PEGylation by sequential bis-alkylation, In *PEGylated Protein Drugs: Basic Science and Clinical Applications* (Veronese, F., Ed.), pp 47–74, Birkhäuser.

71. Morpurgo, M., Veronese, F.M., Kachensky, D., and Harris, J.M. (1996) Preparation of characterization of poly(ethylene glycol) vinyl sulfone, *Bioconjug. Chem. 7*, 363–368.

72. Bradley, J.R. (2008) TNF-mediated inflammatory disease, *J. Pathol. 214*, 149–160.

73. Nesbitt, A.M., Stephens, S., and Chartash, E.K. (2009) Certolizumab pegol: a PEGylated anti-tumour necrosis factor alpha biological agent., In *PEGylated Protein Drugs: Basic Science and Clinical Applications* (Veronese, F., Ed.), pp 229–254, Birkhäuser.

74. Grace, M., Youngster, S., Gitlin, G., Sydor, W., Xie, L., Westreich, L., Jacobs, S., Brassard, D., Bausch, J., and Bordens, R. (2001) Structural and biologic characterization of pegylated recombinant IFN-alpha2b, *J Interferon Cytokine Res. 21*, 1103–1115.

75. Youngster, S., Wang, Y.S., Grace, M., Bausch, J., Bordens, R., and Wyss, D.F. (2002) Structure, biology, and therapeutic implications of pegylated interferon alpha-2b, *Curr. Pharm. Des. 8*, 2139–2157.

76. Greenwald, R.B. (1997) Drug delivery systems: anticancer prodrugs and their polymeric conjugates, *Expert Opinion on Therapeutic Patents 7*, 601–609.

77. Zalipski, S., and Menon-Rudolph, S. (1997) Hydrazide derivatives of poly(ethylene glycol) and their bioconjugates. , In *Poly(ethylene glycol) chemistry and biological applications.*, pp 318–341, ACS Symp Ser. .

78. Bonora, G.M., and Drioli, S. (2009) Reactive PEGs for protein conjugation. , In *PEGylated Protein Drugs: Basic Science and Clinical Applications* (Veronese, F., Ed.), pp 33–46, Birkhäuser.

79. Yankeelov Jr, J.A. (1972) Modification of arginine by diketones, In *Methods Enzymol.* (C. H. W. Hirs, S. N. T., Ed.), pp 566–579, Academic Press.

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

81. 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, *Bioconjug. Chem. 20*, 384–389.

82. Pool, C.P., PA, US). (2006) Formation of novel erythropoietin conjugates using transglutaminase, United States.

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

84. Mariniello, L., and Porta, R. (2005) Transglutaminases as Biotechnological Tools. In *Transglutaminases. Prog. Exp. Tum. Res.* (Mehta, K., and Eckert, R., Eds.), pp 174–191, Karger, Basel.

85. Yokoyama, K., Nio, N., and Kikuchi, Y. (2004) Properties and applications of microbial transglutaminase, *Applied Microbiology and Biotechnology 64*, 447–454.

86. Hazová, B., Jančovičová, J., Dodok, L., Buchtová, V., and Staruch, L. (2002) Use of transglutaminase for improvement of pastry produced by frozen-dough technology. , *Czech J. Food Sci. 20*, 215–222.

87. Ramírez, J.A., Santos, I.A., Morales, O.G., Morrissey, M.T., and Vázquez, M. (2000) Application of microbial transglutaminase to improve mechanical properties of surimi from Silver carp, *Ciencia y Tecnologia Alimentaria 3*, 21–28.

88. Bönisch, M.P., Huss, M., Weitl, K., and Kulozik, U. (2007) Transglutaminase cross-linking of milk proteins and impact on yoghurt gel properties, *International Dairy Journal 17*, 1360–1371.

89. Holmes, K.L., and Lantz, L.M. (2001) Protein labeling with fluorescent probes, *Methods Cell Biol. 63,* 185–204.

90. Sippel, T.O. (1981) New fluorochromes for thiols: maleimide and iodoacetamide derivatives of a 3-phenylcoumarin fluorophore, *J Histochem. Cytochem. 29,* 314–316.

91. Wolff, C., and Lai, C.S. (1989) Fluorescence energy transfer detects changes in fibronectin structure upon surface binding, *Arch. Biochem. Biophys. 268,* 536–545.

92. Taki, M., Shiota, M., and Taira, K. (2004) Transglutaminase-mediated N- and C-terminal fluorescein labeling of a protein can support the native activity of the modified protein, *Protein Eng. Des. Sel. 17,* 119–126.

II. MATERIALS AND METHODS

1. Materials

Carbobenzoxy-L-glutamyl-glycine (ZQG), L-glutamic acid γ-mono hydroxamate, tryptamine-HCl, hydroxylamine-HCl, glucosamine, N, N′-dicyclohexylcarbodiimide (DCCI), sodium cynoborohydride, 2,2,2- trifluroethanol (TFE), avidin (Monomeric) HC agarose (50 % suspension), Sephadex G-25 SF resin, thermolysin, endoproteinase Glu-C from *Staphylococcus aureus* (V8 protease) and sequencing grade TLCK treated α- chymotrypsin from bovine pancreases were purchased from Sigma-Aldrich (Milwaukee, WI, USA). Trypsin was obtained from Promega (Madison, WI, USA). PepClean™ C-18 Spin Columns and dialysis cassettes were supplied by Pierce (Rockford, Il, USA). Dansylcadaverine, iodoacetamide, iodoacetic acid, tris (2-Carboxyethyl) phosphine-HCl (TCEP), 1-hydroxybenzotriazole hydrate (HoBt), trifluoroacetic acid (TFA) and the other high purity salts used for spectroscopic analyses were obtained from Fluka (Buchs, Switzerland). The acetonitrile used for reverse phase HPLC chromatography is from Carlo Erba Reagenti (Italy). Boc-PEG-NH_2 (556.36 Da) was purchased from LCC Engineering GmbH (Egerkingen, Switzerland), PEG-NH_2 (5 kDa) and PEG-CHO (20 kDa) from Iris Biotech (Marktredwitz, Germany). Resource S column was purchased from Pharmacia LKB Biotechnology (Uppsala, Sweden). Biotin-PEG-amine was obtained from Thermo Fisher Scientific (Waltham, USA). Reagents and solvents for electrophoresis were purchased from Bio-Rad (Richmond, IL, USA). All other chemicals were of analytical reagent grade and were obtained from Sigma or Fluka. The composition of the PBS buffer is 10 mM Na_2HPO_4, 1.5 mM KH_2PO_4, 2.7 mM KCl and 0.137 M NaCl, pH 7.4.

2. Instrumentation

a) Mass spectrometry (MS)

Mass spectrometry based analyses were performed with a Micromass mass spectrometer Q-Tof Micro (Manchester, UK) equipped with an electrospray source (ESI-MS). Samples were dissolved in 0.1% formic acid in ACN: water (1:1) and analyzed in MS and MS/MS mode. The measurements were conducted at a capillary voltage of 3 kV and at a 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 a collision energy of 40 V and using argon as collision gas. External calibration was performed using a solution of 0.1% (v/v) phosphoric acid in 50% (v/v) aqueous acetonitrile for peptide mass determination, whereas for protein analysis a solution of 10 μM horse heart apomyoglobin in 50% (v/v) aqueous acetonitrile, 0.1% (v/v) formic acid was used. Instrument control, data acquisition and processing were achieved with masslynx software (Micromass).

Protein fingerprinting was performed by dissolving the protein in 0.1 M ammonium carbonate at a concentration of about 1 mg/ml. Trypsin or the V8 protease were added at a enzyme/protein ratio of 1:100, by weight. The reactions were incubated overnight at 37°C and then analysed directly by MS.

b) High performance liquid chromatography system (HPLC)

HPLC system used for the RP-HPLC analysis was an Agilent series 1100 HPLC with an online U.V. detection from Agilent Technologies (Waldbroon, Germany). The system operation and data analysis was controlled with the software Agilent Chemstation. RP-HPLC analysis was performed with the gradient of acetonitrile (AcCN) containing 0.085 % TFA against water containing 0.1 % TFA.

c) Fast protein liquid chromatography system (FPLC)

The FPLC system was ane AKTA (Amersham Biosciences, Sweden) for the liquid chromatography. The system was equipped with an Monitor UPC-900 for online monitoring of U.V. absorption, pH and conductivity..

3. Microbial Transglutaminase (TGase)

TGase (1000 units/mg protein) isolated from *Streptoverticillium mobaraense* was purchased from Ajinomoto Co. (Tokyo, Japan) as TGase MP. The activity of the solutions of enzyme was determined by hydroxamate formation in the TGase catalysed reaction between the substrate carbobenzoxy-Gln-Gly (ZQG) and hydroxylamine. Prior to the assay, the concentration of TGase was determined from its absorbance at 280 nm according to Gill and von Hippel (*1*) (extinction coefficient of 1.89 mg/ml). The assay was performed in a total 0.5 ml of an enzyme-substrate

solution containing 0.1 M hydroxylamine, 0.1 ml of enzyme and 30 mM ZQG in 0.1 M phosphate buffer pH 7.0. Reaction was kept at 37 °C for 10 minutes and stopped using an equal volume of a solution obtained upon mixing a solution of 5% $FeCl_3$ in 0.1 N HCl and a solution of 12 % trichloroacetic acid in H_2O at a ratio of 1:3 by volume, respectively. Optical density was recorded with a Perkin Elmer spectrofometer at 525 nm, which measures the formation of the colored Fe^{3+}-hydroxamate complex. A standard plot was prepared using different known concentrations of L-glutamic acid γ-monohydroxamate. One unit of TGase was defined as the formation of 1 micromole of hydroxamate per min at 37 °C.

4. Apomyoglobin

Horse heart myoglobin was purchased commercially from Sigma-Aldrich (Milwaukee, WI, USA). For the removal of heme group, 25 mg of myoglobin were dissolved in 10% acetic acid and kept on ice. 2M HCl in acetone was added drop wise until precipitation of apoMb occurs. The tube was kept on ice for 30 min and then centrifuged at 3000 rpm for 10 min. The pellet was collected, washed three times with 2M HCl in acetone and then dissolved in 3 ml of 5 % acetic acid. The concentration of this solution apoMb was determined from its absorbance at 280 nm according to Gill and von Hippel (*1*) (extinction coefficient of 0.825 mg/ml). Aliquots of 1 mg of the solution of apoMb were prepared by lyophilization. For experimental use, the aliquots were first dissolved in 50 µl of 0.085 % TFA and then reaction buffer was added to obtain the desired concentration.

4.1. TGase mediated conjugation of apoMb with primary amines

a) Tryptamine and dansyl cadaverine

A solution of ApoMb at a concentration of 0.75 mg/ml in 0.1 M phosphate buffer pH 7.0 was mixed with few microliters of solutions of tryptamine or dansyl cadaverine (stock solution about 8 mg/ml in methanol) at different molar ratios (1:10, 1:25, and 1:50 protein: tryptamine or dansyl cadaverine). The reaction mixture was incubated with TGase at an enzyme:substrate ratio of 1:50, by weight. The reaction was allowed to proceed at 25° C. Sample aliquots of 70 µg were removed at 1.5, 3, 6 and 24 hr of reaction and the reaction was stopped by addition of an equal volume of

an aqueous solution of 1% TFA. Aliquots were further analyzed by RP-HPLC using a C18 Phenomenex (150 x 4.6 mm) column. The separation was performed with a linear gradient of AcCN containing 0.85 % TFA from 40 to 50 % in 25 min. Fractions collected from the RP-HPLC analysis were collected, lyophilized and analyzed by MS. The site(s) of conjugation in the products of the reation were determined by protein fingerprinting (see above, *2.a. Mass spectrometer*).

b) Biotin-PEG-amine

A solution of ApoMb (0.75 mg/ml): biotin-PEG-amine in the molar ratios of 1:25 was prepared in 0.1 M phosphate buffer pH 7.0 and incubated with TGase at an enzyme:protein ratio of 1:50 by weight at 25° C. Sample aliquots were removed at different time intervals and stopped upon addition of an equal volume of 1% TFA. These aliquots were analyzed by RP-HPLC using the conditions described for the reaction with dansylcadaverine and tryptamine. Fractions were collected, lyophilized and then analyzed with the ESI-Q-TOF mass spectrometer (MS). The modified protein was digested with trypsin in 0.1 M ammonium bicarbonate buffer pH 8.9. Purification of biotinylated peptides was performed on a monomeric avidin agarose matrix (50% slurry, Sigma) by modification of the manufacturer's instructions. Monomeric avidin agarose matrices were chosen because they allow the elution of bound biotin-derivatives under mildly solvent conditions in respect to columns of tetrameric avidin. 0.6 ml of slurry was used to prepare the column in 2ml centrifuge tubes. The columns were equilibrated with 2X PBS buffer and then with 2 mM biotin in PBS buffer pH 7.2 to block sites of avidin that can bind biotin irreversibly. The column was then washed with 0.1 M glycine pH 2.8 (2.5 ml) and re-equilibrated with 2X PBS (2 ml). Protein digests (~30 µg) were loaded on the column and incubated at room temperature for 30 minutes. The column was washed with 2X PBS (5 ml) and 50 mM ammonium bicarbonate (5 ml). Biotinylated peptides were recovered by elution with 0.3 % HCOOH in 30 % acetonitrile (100 µl,) and analyzed by mass spectrometry.

c) Hydroxylamine and glucosamine

Solutions of ApoMb (0.75 mg/ml) and hydroxylamine or glucosamine in the molar ratios of 1:25 were prepared in 0.1 M phosphate buffer pH 7.0 and incubated with TGase at a protein: enzyme ratio of 1:50 by weight at 25° C. Sample aliquots were removed and stopped with an equal volume of 1% TFA at different time intervals. These aliquots were analyzed as described in the previous paragraphs on the derivatization of apoMb by TGase. Fractions were collected, lyophilized and then analyzed with the ESI-QTOF mass spectrometer (MS). Modified proteins were digested with trypsin and analyzed by MS. Identification of the site of conjugation in the modified peptide was performed by MS/MS analysis.

4.2. TGase mediated conjugation of apoMb with CBZ-Gln-Gly and its derivatives

a) Preparation of CBZ-Gln-Gly derivatives

33.7 mg of CBZ-Gln-Gly (Carbobenzoxy-glutamine-glycine, ZQG) were dissolved in 5 ml of DMSO and then 15 ml of chloroform were added. Coupling reagents were added according to the molar ratio of ZQG: HoBt: DCCI of 1: 2.5: 2.5. The reaction mixture was incubated for 1 hour at room temperature. 0.3 molar equivalents of primary amines [amine-PEG (556 Da) or tryptamine] were added to the reaction mixture and incubated overnight on a shaker. Three times washing with basic water (pH 9.0) were performed using a separating funnel. The lower fractions were lyophilized and dissolved in acetonitrile. The derivatives were purified by RP-HPLC using the conditions described above for the apoMb conjugation reactions. Fractions were collected, lyophilized and analyzed by MS.

b) Quantitation of CBZ-Gln-Gly derivatives

An assay procedure to estimate the yield in ZQG derivatives was developed by modification of the TGase activity assay (see above, 3. Microbial Transglutaminase). A standard plot was made using 0.5 ml of a solution containing 0.1 M hydroxylamine, 0.1 M phosphate buffer pH 7.0, enzyme (0.01 mg/ml) and differing concentrations of ZQG (1-10 mM). The estimation of the amount of purified ZQG derivatives was performed in a total 0.5 ml of an enzyme-substrate solution-1 containing 0.1 M hydroxylamine, and 0.1 ml of enzyme (0.01 mg/ml), ZQG derivative in 0.1 M

phosphate buffer pH 7.0. Reaction was kept at 37 °C and aliquots were removed at 10, 20 and 30 minutes to confirm the complete derivatization of the substrate and stopped using an equal volume of a solution obtained upon mixing a solution of 5% FeCl3 in 0.1 N HCl and a solution of 12 % trichloroacetic acid in H2O at a ratio of 1:3 by volume, respectively. Optical density was recorded with a Perkin Elmer spectrometer at 525 nm. One mole of ZQG from the standard plot was defined to be equivalent to 1 mole of ZQG derivative.

c) TGase mediated conjugation of apoMb using ZQG derivatives

Solutions of apoMb (0.75 mg/ml) and ZQG, ZQG-tryptamine or ZQG-mPEG in the molar ratios of 1:25 were prepared in 0.1 M phosphate buffer pH 7.0 and incubated with TGase at an enzyme:protein ratio of 1:50 by weight at 25° C. Sample aliquots were removed and stopped with equal volumes of 1% TFA at different time intervals. These aliquots were analyzed by RP-HPLC using the conditions described above. Fractions were collected, lyophilized and then analyzed by mass spectrometry. Modified proteins were digested with both trypsin and V8 protease at an enzyme:protein ratio of 1:100 by weight in 0.1 M ammonium bicarbonate buffer pH 8.9 and analyzed by MS. Tandem mass spectrometry (MS/MS) was performed to identify the site of conjugation in the modified peptide.

4.3. Chemical and TGase mediated methods of PEGylation of proteins

a) N-terminal site-specific PEGylation of apoMb

A solution of ApoMb (0.75 mg/ml) and PEG20k-CHO at a molar ratio of 1:10 was prepared in 20 mM sodium acetate buffer pH 5.0 and incubated at 37° C with NaCNBH3 at a molar ratio with apoMb of 1:10 (ApoMb:NaCNBH3). Sample aliquots were removed and stopped with equal volumes of 1% TFA at different time intervals. These aliquots were analyzed by RP-HPLC using the Phenomenex C4 HPLC column (150 × 4.6 mm). The separation was performed with a linear gradient of AcCN containing 0.085 % TFA from 40 to 50 % in 25 min. Fractions were collected, lyophilized and then analyzed by SDS-PAGE using a 15 % gel (2). Reaction conditions were optimized by changing molar ratios in order to increase the yield of the mono-conjugated derivative. Best conditions were found to be the

following molar ratios: apoMb: PEG20k-CHO 1:5 and apoMb: NaCNBH3 1:5. For the identification of the site of conjugation, the protein derivatives collected in the RP-HPLC fractions were lyophilized and dissolved in 0.1 M ammonium bicarbonate buffer pH 8.9 in the presence of trypsin at a enzyme:protein ratio of 1:100, by weight. The reaction was allowed to proceed at 37° C and then protein digests were analyzed by RP-HPLC using a Phenomenex C18 HPLC column (150 × 4.6 mm). The separation was performed with a linear gradient of AcCN containing 0.085 % TFA from 25 to 60 % in 30 min. Fractions were collected, lyophilized and analyzed by mass spectrometry.

b) TGase mediated derivatization of apoMb with PEG-amines

Solutions of ApoMb (0.75 mg/ml) and monodisperse PEG-amine (556 Da, mPEG) or polydisperse PEG-amine (5000 Da, PEG5k) in the molar ratios of 1:100 were prepared in 0.1 M phosphate buffer pH 7.0 and incubated at 25° C with TGase at an enzyme:protein ratio of 1:50, by weight. Sample aliquots were removed and stopped with equal volumes of 1% TFA at different time intervals. These aliquots were analyzed using the conditions described above (4.1.a). Fractions were collected, lyophilized and then analyzed by ESI-Q-TOF mass spectrometry.

5. Lysozyme

Chicken egg white lysozyme (LYS) was purchased from Sigma-Aldrich (Milwaukee, WI, USA). The concentration of the protein was determined using the extinction coefficient □=2.69 (mg/ml) according to Gill and von Hippel (*1*).

5.1. Preparation of three disulphide derivative of lysozyme ($LYS^{CM6,127}$)

$LYS^{CM6,127}$ was prepared by a modification of the method of Radford *et al* 1991 (*4*). Selective reduction of the Cys^{6}-Cys^{127} disulphide bridge was performed by using TCEP and the reduced cysteines were blocked by iodoacetic acid. The selective reduction and carboxymethylation was performed in a single step, and time dependent formation of $LYS^{CM6,127}$ derivative was studied in different experimental conditions to generate an optimum method of preparation.

LYS (1 mM) was dissolved in 0.1 M Tris·HCl buffer, pH 7.8, and incubated with TCEP (2 mM) and iodoacetic acid (5 mM) at 25^{o} C in the dark. Sample aliquots were removed after 30, 60 and 90 min of reaction. The precipitate formed was removed by centrifugation at 5000 rpm for 2 min. The soluble fraction was immediately desalted against water on a Sephadex G-25 column connected to an AKTA-FPLC system (Amersham Biosciences, Sweden). The protein fractions were collected and lyophilized with a Speed-vac concentrator. The carboxymethylated (CM) derivatives of lysozyme were purified by ion-exchange chromatography on a Resource S column connected to the AKTA-FPLC system. The column was equilibrated with 20 mM sodium phosphate buffer, pH 6.8. The elution buffer (B) contained 0.4 M NaCl along with 20 mM sodium phosphate buffer, pH 6.8. A linear gradient from 0 to 100 % of buffer B in 60 min was used for the elution of lysozyme and its carboxymethylated derivatives. Fractions collected were desalted against water on a Sephadex G-25 column and loaded onto a Phenomenex C4 HPLC column (150 × 4.6 mm) connected to the HPLC system (Agilent Technologies, Waldbronn, Germany). The elution of the protein species was obtained with a linear gradient of acetonitrile containing 0.085 % TFA from 22 to 52 % in 25 min. The effluent from the column was monitored by measuring the absorbance at 226 nm. The protein material contained in the RP-HPLC fractions was analyzed on a ESI-Q-Tof mass spectrometer (see above for details).

5.2. Characterization of LYS^{CM6,127}

a) Mass spectrometry

LYS^{CM6,127} was digested with trypsin (1:100 w/w) in 0.1 M ammonium bicarbonate buffer containing 1mM of TCEP for 24 hrs at 37^{o} C. The tryptic digest of LYS^{CM6,127} was loaded in a ESI-Q-Tof mass spectrometer (Waters, Milford, MA) and the analysis was carried out as mentioned above in *2.a. Mass spectrometry*

b) Circular dichroism

Circular dichroism (CD) spectra were obtained with a Jasco J-710 (Tokyo, Japan) dichrograph, equipped with a thermostated cell holder. The instrument was calibrated with *d*-(+) 10-camphorsulfonic acid. Quartz cells with a 1-mm and 5-mm

light pathlength were used for measurements in the far-UV and near-UV region, respectively. The spectra were recorded at 8° C from 240 to 190 nm in the far-UV region and from 320 to 240 nm for the near-UV region. The concentration of the protein (LYS and $LYS^{CM6,\ 127}$) was in the range 0.05-0.1 mg/ml in 10 mM phosphate buffer, pH 7.0, containing 50 mM NaCl. The mean residual ellipticity [θ] in $deg \cdot cm^2 \cdot dmol^{-1}$ was calculated using the formula $[\theta] = (\theta_{obs}/10)\ (MRW/lc)$, where θ_{obs} is the observed ellipticity in deg, MRW the mean residue molecular weight, *l* the path length in cm and *c* is the protein concentration in mg/ml. The thermal unfolding was monitored by recording the decrease of the CD signal at 222 nm as a function of the sample temperature (heating rate 50°C/h). Both CD signal and temperature data were recorded simultaneously by a computer program provided by Jasco. The melting temperature (T_m) was determined from the derivative curve of the CD signal at 222 nm *versus* temperature. The reversibility of the thermal unfolding process was determined by measuring the recovery of the CD signal upon cooling to 8 °C.

5.3. TGase-mediated conjugation of $LYS^{CM6,127}$ with dansyl-cadaverine

Solutions of $LYS^{CM6,127}$ (0.7 mg/ml) and dansyl cadaverine at 1:50 molar ratio was prepared in 0.1 M Tris·HCl buffer, pH 7.8. Therefore, the reaction mixture was incubated with a TGase:substrate ratio of 1:20 by weight at 25° C. Sample aliquots after 1.5, 3, 6 and 24 hr of reaction time were removed and the reaction was stopped by the addition of equal amount of 1% TFA. Aliquots were further analyzed by RP-HPLC on a C18 column with a linear gradient of acetonitrile containing 0.085 % TFA from 22 to 52 % in 30 min. The fractions were collected, lyophilized and analyzed by mass spectrometry. The site(s) of conjugation were determined with the mass spectrometry based protein fingerprinting technique as described above.

5.4. TGase-mediated conjugation of $LYS^{CM6,127}$ with CBZ-Gln-Gly

Solutions of $LYS^{CM6,127}$ (0.5 mg/ml) and carbobenzoxy (CBZ)-glutamine-glycine (ZQG) at the molar ratio of 1:500 was prepared in 0.1 M Tris·HCl buffer, pH 7.8. The reaction mixture was then incubated with a TGase: substrate ratio of 1:20 by weight at 25° C. Sample aliquots were removed after 0, 6 and 24 hrs of reaction and the reaction was stopped by the addition of an equal volume amount of 1% TFA.

Aliquots were further analyzed by RP-HPLC using a C18 column with a linear gradient of acetonitrile containing 0.085 % TFA from 22 to 52% in 30 min. Fractions were collected and further analyzed by mass spectrometry. The site(s) of conjugation of the obtained derivatives were determined with the mass spectrometry based protein fingerprinting technique (see above for details).

5.5. Limited proteolysis of $LYS^{CM6,127}$

$LYS^{CM6,127}$ was incubated at a 0.7 mg/ml concentration with chymotrypsin (enzyme:substrate ratio of 1:400, by weight) at 30° C. Sample aliquots were removed after 0, 10, 30, 60 and 90 min of reaction. These aliquots were analyzed with a tricine-SDS PAGE system (*5*) and RP-HPLC using a C18 column with a linear gradient of acetonitrile containing 0.085 % TFA from 25 to 60 % in 30 min. Fractions from RP-HPLC were collected and analyzed by mass spectrometry (see above for details).

6. Ribonuclease (RNase)

Bovine pancreatic ribonuclease A was purchased from Sigma-Aldrich (Milwaukee, WI, USA). Concentrations of the protein solutions were determined by measurement of the absorbance at 280 nm and using an extinction coefficient of 0.690 mg/ml, as determined according to Gill and von Hippel (*1*).

6.1. Preparation of RNase Th1

a) Proteolysis of RNase with thermolysin

RNase Th1 was prepared by a modification of the method reported by Polverino de Laureto (3). A solution of RNase (2.5 mg/ ml) was **incubated** with thermolysin (3:100 enzyme: substrate ratio, by weight) in 50 mM Tris-HCl buffer pH 7.0, containing 5 mM CaCl2 and 50 % TFE. The reaction mixture was incubated at 42° C and sample aliquots were taken at 0, 30, 60 and 90 minutes of reaction. The proteolysis reaction was stopped by acidification and analyzed by ion-exchange chromatography (see below).

b) Purification of RNase Th1

The main product of the proteolysis reaction RNase Th1 was purified by ion-exchange chromatography on a Resource S column connected to the AKTA-FPLC system (Amersham Biosciences). The column was equilibrated with 20 mM Sodium phosphate buffer pH 7.0. Elution buffer (B) contained 0.8 M NaCl along with 20 mM sodium phosphate buffer pH 7.0. A liner gradient from 0 to 60 % B in 20 min was used for the elution of RNase and RNase Th1. Fractions were collected, desalted against water using a Sephadex G 25 column and then loaded onto a Phenomenex C18 HPLC column (150 × 4.6 mm) connected to the HPLC system (Agilent technologies). Elution was performed using a linear gradient from 22 to 40 % AcCN containing 0.085 % TFA over a period of 34 min and by measuring the absorbance at 226 nm. Fractions from the RP-HPLC were analyzed using the ESI-Q-TOF mass spectrometer to identify the eluting protein material.

6.2. TGase mediated conjugation studies on RNase

Solutions of RNase (0.8 mg/ml) and dansyl cadaverine or ZQG at molar ratios of 1:50 were prepared in 0.1 M tris-HCl buffer pH 7.5 and incubated at 25° C with TGase at an enzyme:protein ratio of 1:20, by weight. Sample aliquots were removed at different time intervals and stopped with equal volumes of 1% TFA. These aliquots were analyzed by RP-HPLC using a Phenomenex C18 HPLC column (150 × 4.6 mm) with a linear gradient from 22 to 40 % of AcCN containing 0.085 % TFA in 34 min. Fractions were collected, lyophilized and then analyzed using the ESI-Q-TOF mass spectrometer (MS). Modified proteins were digested with trypsin in 0.1 M ammonium bicarbonate buffer pH 8.9 and analyzed by MS. Characterization of the site of conjugation on the modified peptide was performed by MS/MS analysis.

6.3. TGase mediated conjugation studies on RNase Th1

Solutions of RNase Th1 (0.8 mg/ml) and dansyl cadaverine or ZQG at molar ratios of 1:50 were prepared in 0.1 M tris-HCl buffer pH 7.5 and incubated at 25° C with TGase at an enzyme:protein ratio of 1:20, by weight. Sample aliquots were removed and stopped at different time intervals upon addition of equal volumes of 1% TFA. These aliquots were analyzed by RP-HPLC using the conditions described in the previous paragraph. Fractions were collected, lyophilized and then analyzed by

MS. Fractions containing the modified protein were digested with trypsin in 0.1 M ammonium bicarbonate buffer pH 8.9 and analyzed by MS, followed by the identification of site of conjugation in the modified peptide by MS/MS analysis.

6.4. References

1. Gill, S. C., and von Hippel, P. H. (1989) Calculation of protein extinction coefficients from amino acid sequence data, *Anal. Biochem. 182*, 319–326.

2. Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4, *Nature 227*, 680–685.

3. Polverino de Laureto, P., Scaramella, E., De Filippis, V., Bruix, M., Rico, M., and Fontana, A. (1997) Limited proteolysis of ribonuclease A with thermolysin in trifluoroethanol, *Protein Sci. 6*, 860–872.

4. Radford, S.E., Woolfson, D.N., Martin, S.R., Lowe, G., and Dobson, C.M. (1991) A three-disulphide derivative of hen lysozyme. Structure, dynamics and stability, *Biochem. J. 273(Pt 1)*, 211–217.

5. Schagger, H., and von Jagow, G. (1987) Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa, *Anal. Biochem. 166*, 368–379.

III. RESULTS

1. APOMYOGLOBIN (apoMb)

Myoglobin is a protein involved in the function of transport and storage of oxygen. From a structural point of view, the protein has a polypeptide chain of 153 amino acid residues that coordinates an heme group and folds into a secondary structure composed of eight α-helices, denoted as A-H (Figure 1.1 and 1.2) arranged into a globular tertiary structure. Apomyoglobin (apoMb), which is myoglobin without heme, has been widely used as a model protein in various studies on protein structure, folding and stability (*1-2*). Structurally, at neutral pH in solution, apoMb maintains a secondary structure similar to that of the holo form (*3-4*) with the exception of the region encompassing F helix (residues His82–His97) which shows a slightly reduced helical content on the basis of spectroscopic studies and in particular by NMR measurements (*1*). Limited proteolysis experiments conducted on apoMb have also confirmed that the removal of the heme group makes the peptide bonds located in the sequence of helix F more susceptible to the proteolytic attack, thus demonstrating that in apoMb the region of helix F is flexible/unfolded (*5-6*).

Earlier studies have reported the TGase-mediated site-specific conjugation of apoMb with PEG-amine to generate a PEGylated protein (*7*). In that paper, it was demonstrated that apoMb is modified selectively at the level of a Gln residue located in the region of helix F, leading to a mono-conjugated derivative. In this Thesis, the study of the reactivity of apoMb with TGase was extended. For the conjugation of glutamine residues, different primary amines were tested (Fig. 1.3), whereas for the derivatization of lysine residues, we performed the reaction using N-carbobenzoxy-Gln-Gly-OH (ZQG) as acyl donor. The aim of the present study is to unravel the effect of the chemical entity containing the primary amine or the glutamine residue on the selectivity and rate of the TGase conjugation reaction.

1. TGase-mediated conjugation of primary amines to apoMb

a) *Tryptamine*

TGase mediated conjugation studies were performed on apoMb with tryptamine (TP, Figure 1.3.A) as an acyl acceptor to obtain a site-specific conjugation at the level of glutamine residues. Different apoMb:tryptamine molar ratios were used

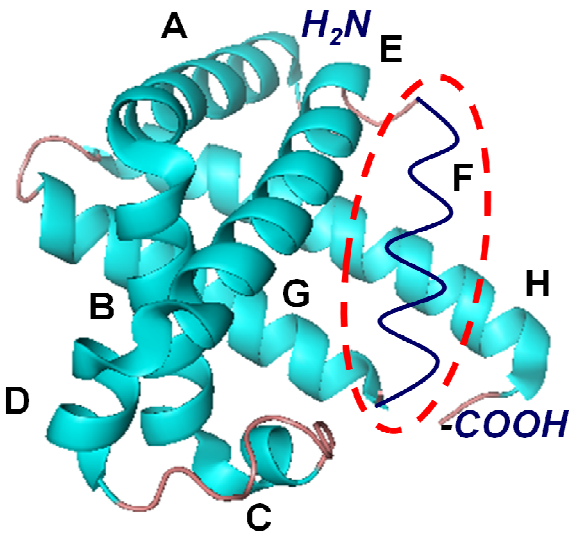


Figure 1.1. Apomyoglobin structure as obtained from the PDB file 1YMB. PyMOL generated apoMb structure showing the structural arrangement of the 153 residue chain. Helical segments (A to H) are shown. Chain segment (residues 82–97) corresponding to the disordered helix F as given by the NMR studies, is indicated by the red dotted circle.

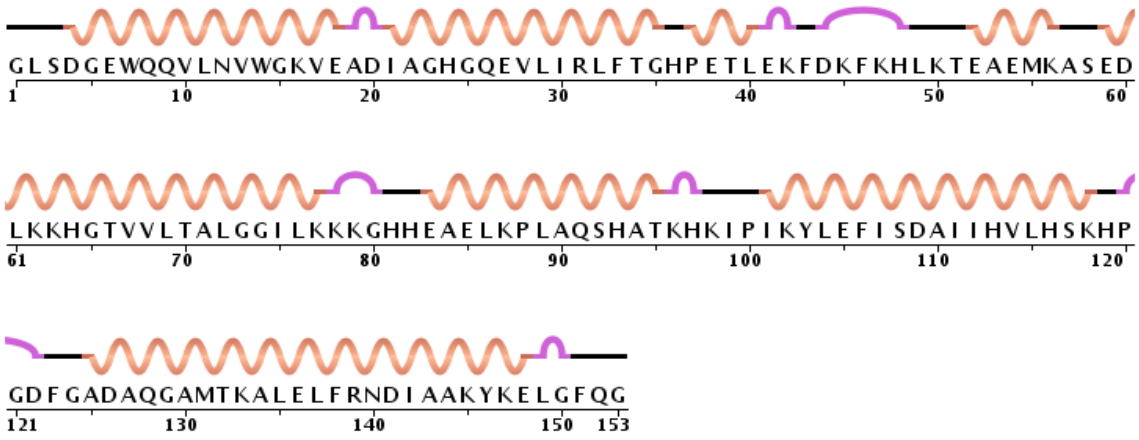


Figure 1.2. ApoMb sequence (PDB file 1YMB). Amino acid sequence and secondary structure of apoMb is shown. Helical segments along the protein chain are indicated by squiggled red lines above the amino acid sequence.

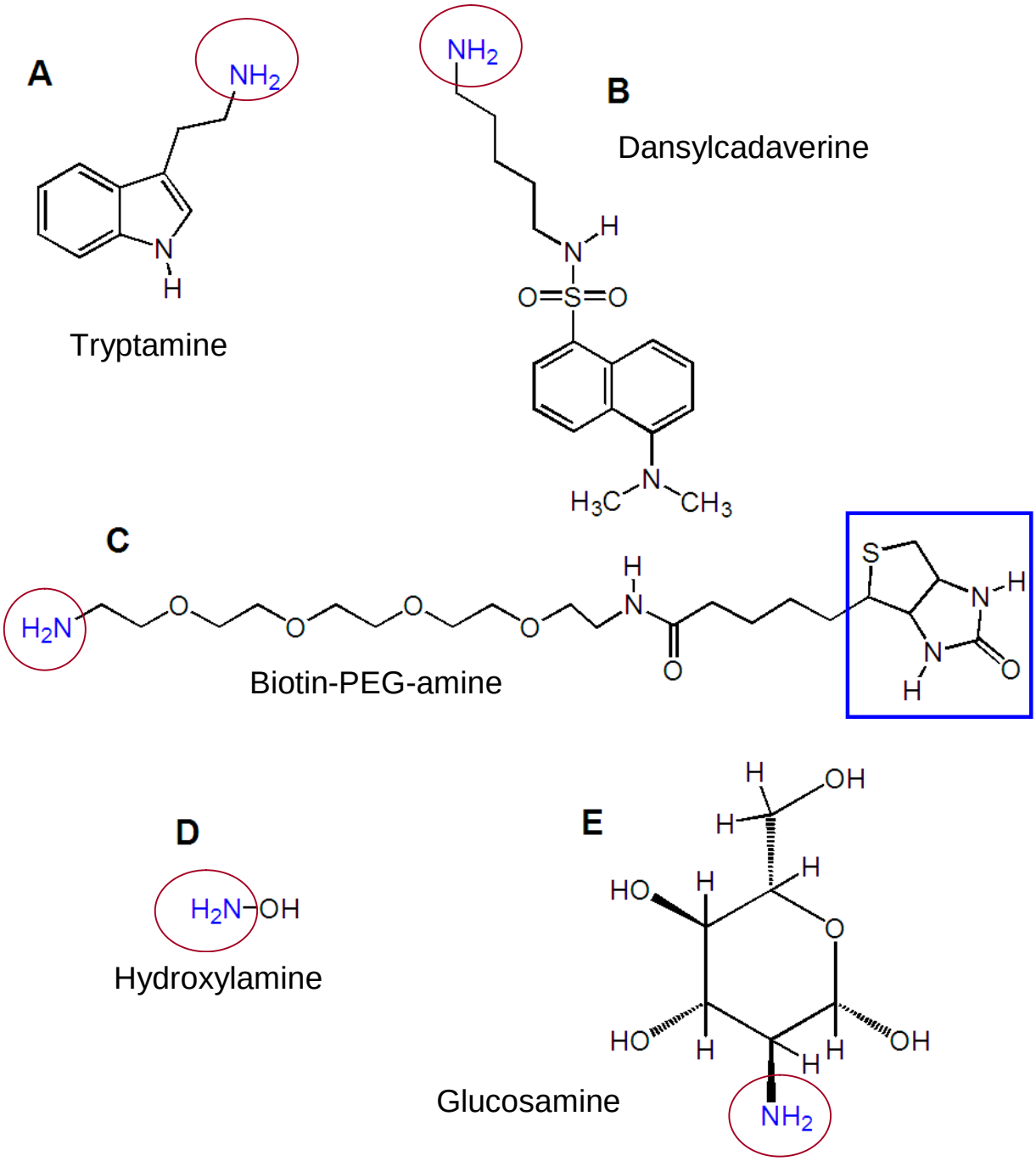


Figure 1.3. Chemical structures of primary amines. TGase accessible primary amine group is indicated by a dark red circle for A) tryptamine, B) dansyl cadaverine, C) Biotin-PEG-amine (the biotin group is indicated with a blue rectangle), D) hydroxylamine and E) glucosamine.

in order to optimize the reaction conditions to obtain the site-specific derivative. Figure 1.4 shows the RP-HPLC chromatogram of the TGase reaction performed at an apoMb:tryptamine, molar ratio of 1:25 after 5 hours of incubation. The chromatogram shows the presence of three separated peaks that elute at a retention time longer than the wild type protein, thus suggesting that they correspond to the protein derivatised with the hydrophobic chemical moiety of tryptamine (TP). Mass spectrometric measurements conducted on the protein material collected from the RP-HPLC analysis identified the three peaks as apoMb-tryptamine ($apoMb^{1TP}$, the first peak), apoMb conjugated to two tryptamine molecules ($apoMb^{2TP}$; the middle peak) and the dimer of apoMb crosslinked by TGase and derivatised with one molecule of tryptamine ($apoMb_2{}^{1TP}$; the last eluting peak) (Table 1.1). The mass values clearly demonstrated the incorporation of tryptamine in apoMb at two locations.

Identification of the sites of conjugation was performed by protein fingerprinting of the apoMb-tryptamine derivatives. The protein conjugates were digested with trypsin and analyzed by MS (Table 1.1). The modified peptides corresponding to the regions Lys79–Lys96 and Gly80–Lys96 of the apoMb sequence were observed in the tryptic digest of mono-conjugated derivative by the MS analysis. MS/MS analysis of peptide Gly80–Lys96 confirmed the site of conjugation to be at the level of Gln91 (Figure 1.5). In the digest of the $apoMb^{2TP}$ derivative, modified peptides were observed encompassing amino acids Lys79–Lys96, Gly80–Lys96 and Glu148–Gly153 of the apoMb sequence. Only Gln152 is located between the residues Glu148–Gly153, thus indicating that in the bis-derivative the sites of conjugation are located at the level of Gln91 and Gln152. Interestingly, these two Gln residues are both located in regions of apoMb that are characterized by an enhanced chain flexibility (see Comments).

In order to optimize the reaction conditions, time dependent kinetic studies (Figure 1.6) were performed at different molar ratios of apoMb:tryptamine (1:10, 1:25 and 1:50, respectively). RP-HPLC analysis of aliquots of the reaction mixtures quenched after different time of incubation were used to estimate the relative yield of each derivative for the different reaction conditions. The mono-derivative showed the highest yield (~ 80%) at 1:25 apoMb: tryptamine molar ratio after 2 hours of reaction. At 1:10 and 1:50 molar ratios, other derivatives are favored in respect to $apoMb^{TP}$.

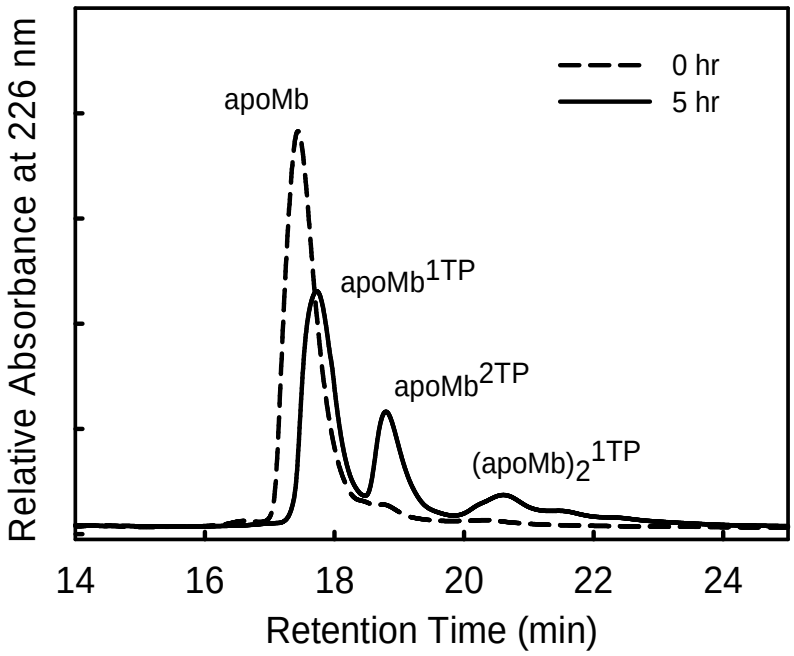


Figure 1.4. TGase mediated conjugation of apoMb with tryptamine (TP). RP-HPLC analysis of the reaction mixture after 0 hour (dotted line) and 5 hours (solid line) of incubation. Analyses were performed on a C18 Phenomenex column using a linear gradient of AcCN containing 0.085 % TFA from 40 to 50 % in 25 min. The absorbance of the effluent from the column was monitored at 226 nm and identities of the eluting protein material were confirmed by MS analysis.

Table 1.1. Molecular masses of apoMb derivatives conjugated by TGase to tryptamine (TP). The observed and calculated molecular masses of the protein material eluted in the RP-HPLC peaks (See figure 1.4) and of the modified peptides identified in the tryptic digests of the apoMb derivatives are reported. The site of conjugation is also indicated.

Protein Species	Conjugation Site(s)	Molecular Mass (Da)	
		Observed	Calculated
Tryptamine			160.22
apoMb		16951.88	16951.51
$apoMb^{1TP}$	Q91	17094.24	17094.51
$apoMb^{1TP}$ [79–96]		2124.32	2124.05
$apoMb^{1TP}$ [80–96]		1996.23	1995.95
$apoMb^{2TP}$	Q91, Q152	17238.29	17237.50
$apoMb^{1TP}$ [79–96]		2124.48	2124.05
$apoMb^{1TP}$ [80–96]		1996.11	1995.95
$apoMb^{1TP}$ [148–153]		791.98	792.31
$(apoMb)_2{}^{1TP}$		34030.54	34029.32

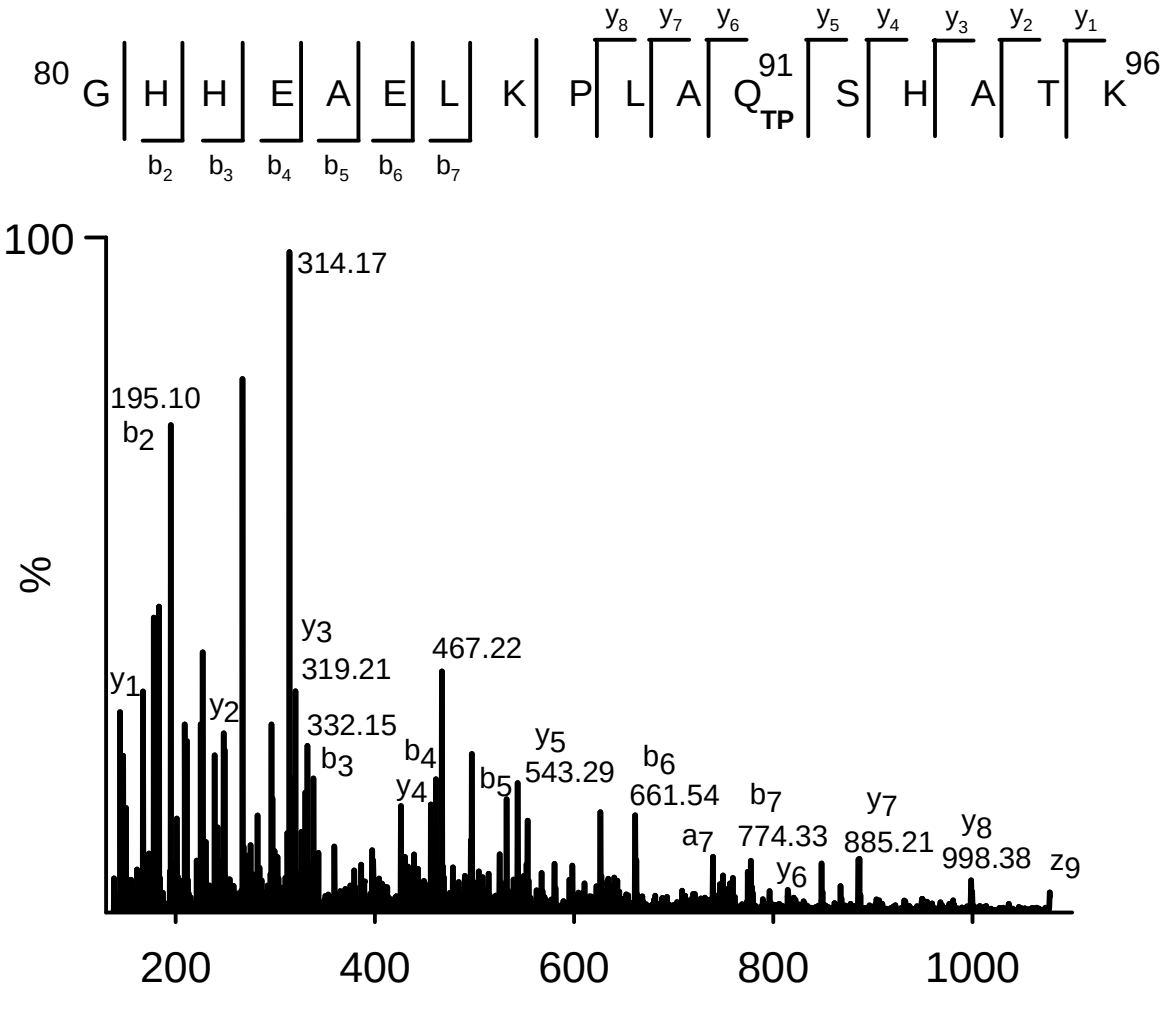


Figure 1.5. Electrospray MS/MS mass spectrum of the ion at 666.21 m/z of peptide 80–96 $[M+3H]^{3+}$ of $apoMb^{1TP}$. (Top) Fragments of the series *b* and *y* that were identified in the MS/MS spectrum are indicated on the sequence of the peptide. Modified Gln91 is indicated as Q_{TP}. (Bottom) In the MS/MS spectrum, ions assigned to fragments of the series *b* and *y* are indicated.

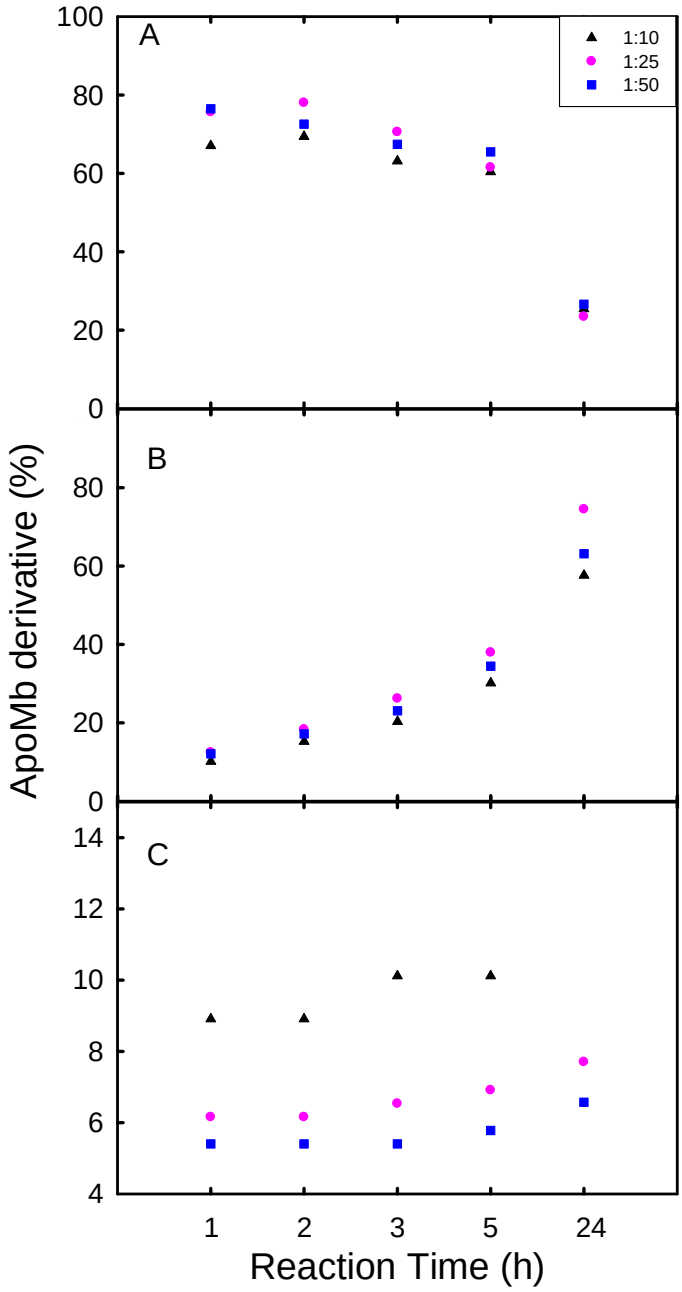


Figure 1.6. Kinetics of TGase-mediated derivatization of apoMb with tryptamine. Time dependent kinetic studies for the formation of, (A) single site specific derivative, (B) double site specific derivative and (C) $(apoMb)_2^{1TP}$ derivatives are indicated by symbols corresponding to different apoMb: trytamine molar ratios, (▲) 1:10, (●) 1:25, and (■) 1:50. The percent recovery of apoMb derivatives during the conjugation reaction is reported VS the reaction time.

Indeed, at a 1:50 molar ratio the maximum yield of the single site derivative was obtained after 1 hour of reaction (~ 80%), but it gradually decreased in time due to the formation of the double site specific derivative. The dimerization reaction also competes for the formation of the mono-conjugated apoMb derivative. The formation of $apoMb_2$ and $apoMb_2^{1TP}$ derivatives are favored at a molar ratio of 1:10 (~ 9 %), whereas at molar ratios of 1:50 and 1:25 the yield was much lower (~ 5 % and ~ 6%, respectively). Therefore, it can be concluded that the most favorable conditions to produce selectivity a single site-specific derivative of apoMb is to use a 1:25 apoMb:tryptamine molar ratio.

b) Dansyl cadaverine

To further study the TGase-mediated conjugation of apoMb using primary amines, dansylcadaverine (DC, Figure 1.3.B) was used as an acyl acceptor for the conjugation at the level of glutamine residues in apoMb. Different apoMb:dansyl cadaverine molar ratios were used in order to achieve the optimized conditions for the production of a site-specific derivative. The RP-HPLC chromatographic profile of the reaction mixture at 1:25 molar ratio after 24 hour of incubation (Figure 1.7) showed the presence of four peaks, the first eluting at the same retention time of apoMb, while the others at longer retention times. The identities of the eluting protein material were determined by MS and they were found to be apoMb, apoMb conjugated to one molecule of DC ($apoMb^{1DC}$) and to two molecules of DC ($apoMb^{2DC}$), and an apoMb crosslinked dimer ($apoMb_2$) (Table 1.2). The sites of conjugation of the apoMb derivatives were determined by trypsin digestion, followed by identification of the modified peptides by mass spectrometry (Table 1.2). The modified peptides corresponding to the regions Lys79–Lys96 and Gly80–Lys96 of the apoMb sequence were observed in the tryptic digest of the mono-conjugated derivative. MS/MS analysis of the peptide Gly80–Lys96 confirmed the site of conjugation at the level of Gln91 (Figure 1.8). In the digest of the $apoMb^{2DC}$ derivative, the modified peptides were observed for the regions Lys79–Lys96, Gly80–Lys96 and Glu148–Gly153 of the apoMb sequence thus confirming that dansyl cadaverine was selectively linked at the positions Gln91 and Gln152 of the apoMb sequence. Gln91 is the initial site of conjugation followed by Gln152 as seen in the conjugation studies using tryptamine.

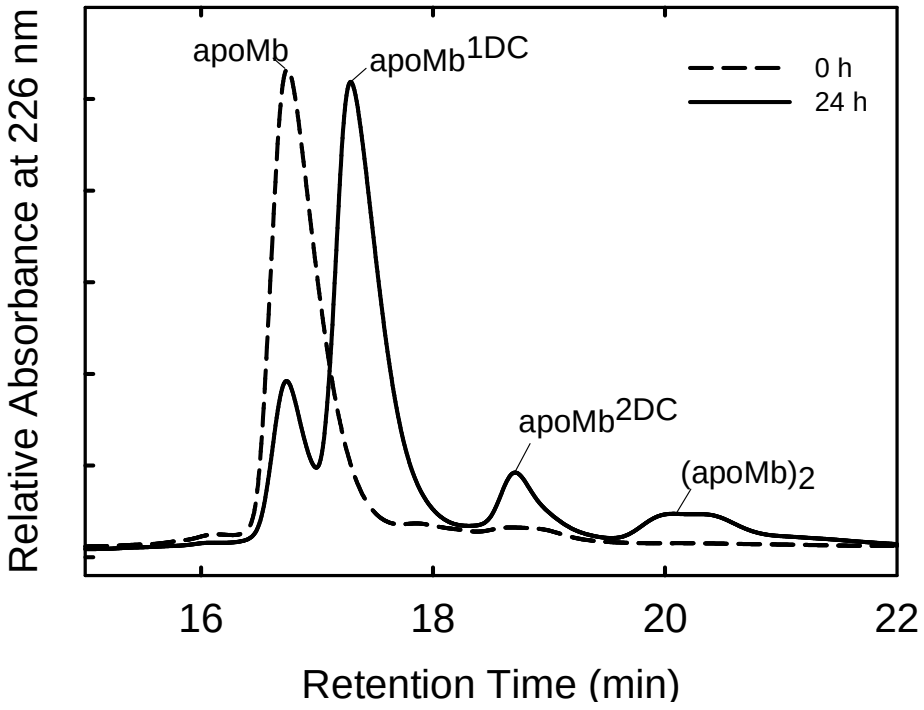


Figure 1.7. TGase mediated conjugation of apoMb with dansyl cadaverine (DC). RP-HPLC analysis of the reaction mixture after 0 hour (dotted line) and 5 hours (solid line) of incubation. Analyses were performed on a C 18 Phenomenex column using a linear gradient of AcCN containing 0.085 % TFA from 40 to 50 % in 25 min. The absorbance of the effluent from the column was monitored at 226 nm and identities of the proteins were confirmed by MS analysis.

Table 1.2. Molecular masses of the TGase mediated conjugation reaction of apoMb with dansylcadaverine as observed from the MS analysis of the RP-HPLC peaks. Modified peptides as identified by protein fingerprinting are reported.

Protein Species	Conjugation Site(s)	Molecular Mass (Da) Observed	Calculated
Dansyl cadaverine (DC)			335.47
apoMb		16952.88	
		16951.51	
$apoMb^{1DC}$	Q91	17270.94	17269.97
$apoMb^{1DC}$- [80–96]		2171.51	
		2171.42	
$apoMb^{2DC}$	Q91, Q152	17589.45	
$apoMb^{1DC}$- [80–96]		17588.47	
		2171.28	
		2171.42	
$apoMb^{1DC}$- [148–153]		968.56	968.78
$(apoMb)_2$		33886.43	33886.02

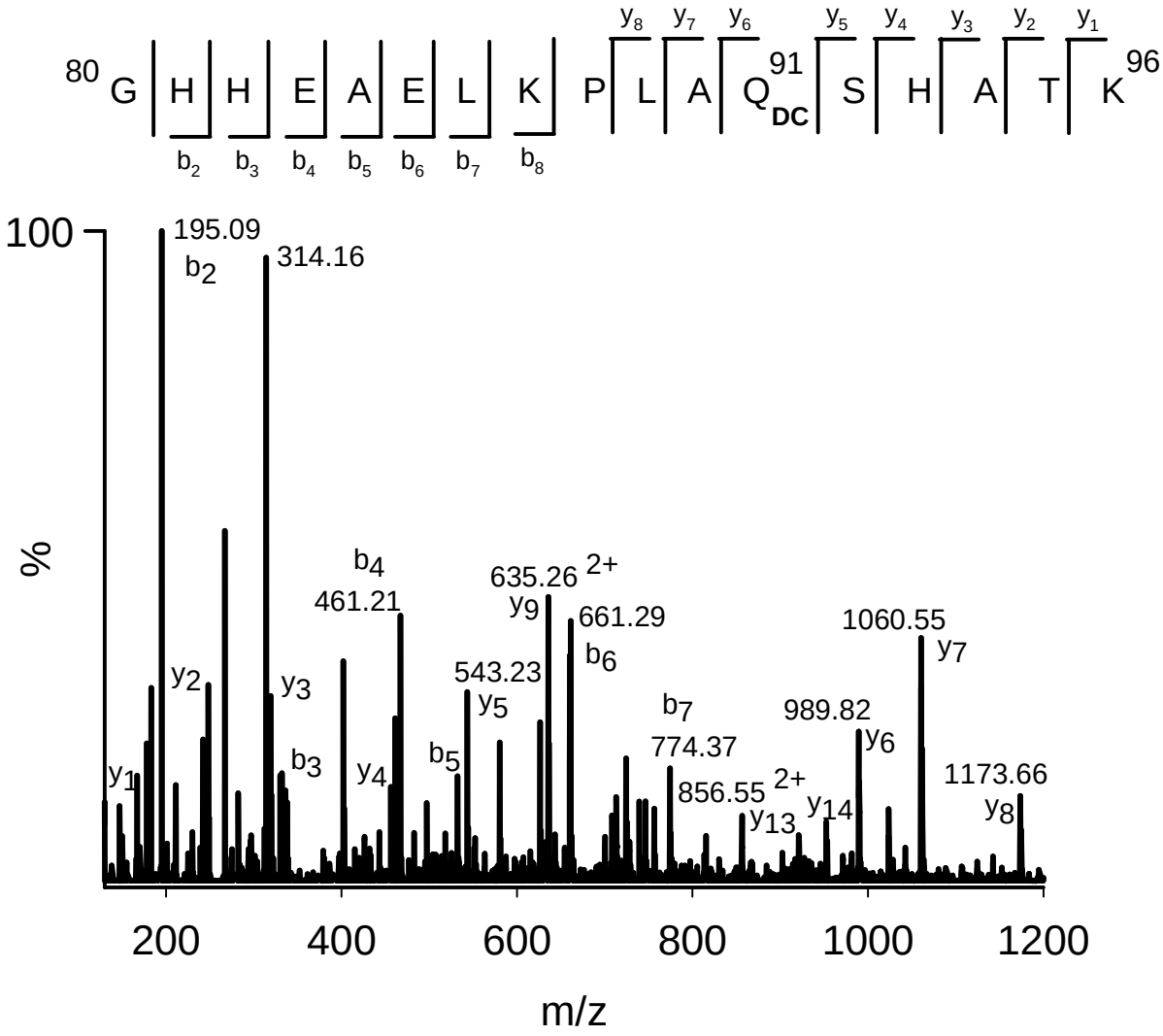


Figure 1.8. Electrospray MS/MS mass spectrum of the ion at 724.81 m/z of peptide 80–96 $[M+3H]^{3+}$ of $apoMb^{1DC}$. (Top) Fragments of the series *b* and *y* that were identified in the MS/MS spectrum are indicated on the sequence of the peptide. Modified Gln91 is indicated as Q_{DC}. (Bottom) In the MS/MS spectrum, ions assigned to fragments of the series *b* and *y* ions are indicated.

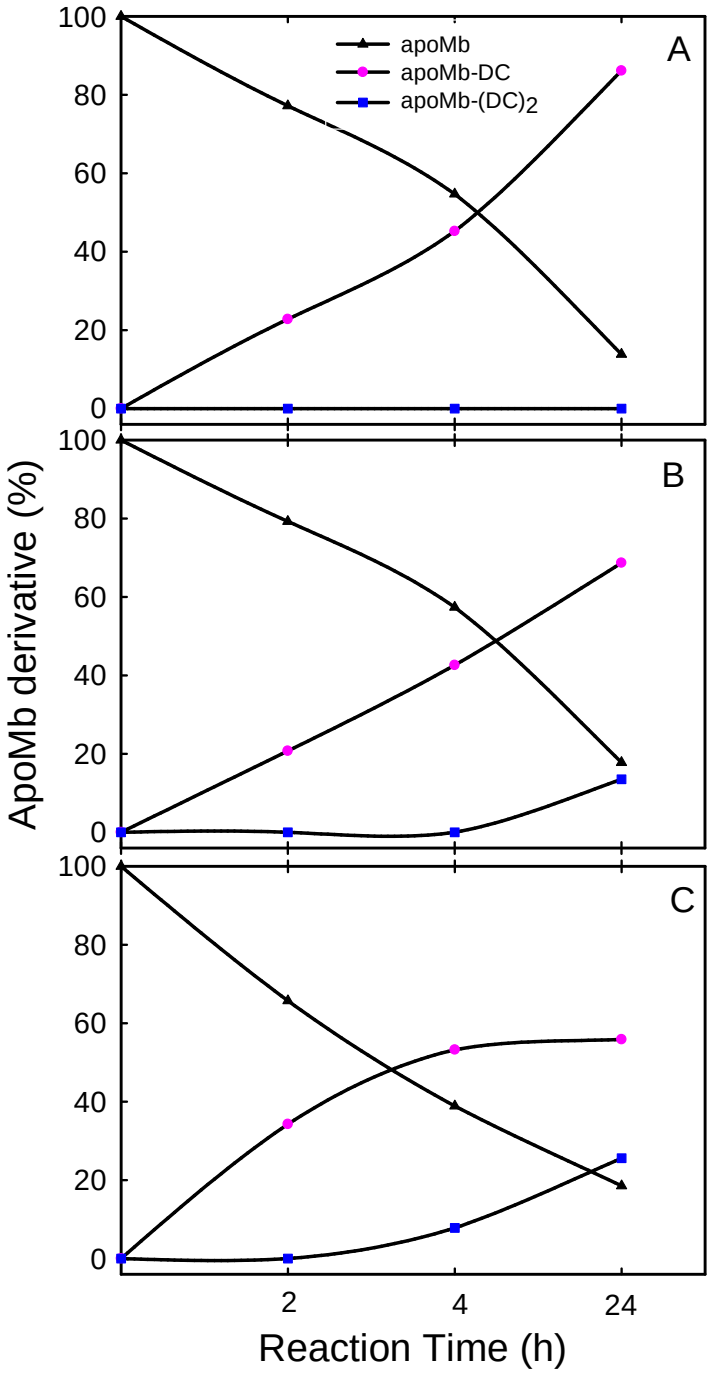


Figure 1.9. Kinetics of TGase-mediated derivatization of apoMb with dansyl cadaverine (DC). Time dependent kinetic studies for the formation of the reaction products at apoMb:DC molar ratios of (A) 1:10, (B) 1:25 and (C) 1:50. The percent recovery of the apoMb derivatives is reported VS the reaction time. ApoMb and its conjugated derivatives are indicated by symbols corresponding to, (▲) apoMb, (●) apoMb^{1DC}, and (■) apoMb^{2DC}.

Kinetic studies of the TGase conjugation reaction at different molar ratios of apoMb:DC showed that after 24 hours of reaction for a 1:10 apoMb:DC molar ratio (Figure 1.9.A), the yield of the mono conjugated derivative was maximum and it was almost quantitative (~80%) with the absence of the bis-conjugated derivative. For the molar ratio 1:25 (apoMb:DC, Figure 1.9.B), after 24 h of reaction, the mono-conjugated derivative was observed (~ 70 %) along with the formation of the bis-conjugated derivative (~15%). Further, for the ratio 1:50 (Figure 1.9.C), the reaction proceeded rapidly towards the formation of the bis-conjugated derivative. Formation of the mono-conjugated derivative was observed to be stationary (~55%) after the 4 h of reaction, whereas the formation of the bis-conjugated derivative was observed to increase exponentially after 2 h of reaction. These results indicate that 1:10 or 1:25 molar ratios of apoMb:dansylcadaverine are the optimal reaction conditions to produce a mono-conjugated apoMb derivative with a reduced formation of dimer and bis-conjugated derivative.

c) Biotin-PEG-amine (PEGB)

TGase mediated conjugation studies of apoMb were performed also using biotin-PEG-amine (PEGB) (Figure 1.3.D) as the acyl acceptor. PEGB has the advantage that the modified peptides can be enriched using an avidin affinity column for the rapid identification of the sites of derivatization. The RP-HPLC chromatographic profile (Figure 1.10) of the apoMb:PEGB reaction mixture (1:100, molar ratio) after 2 hours of incubation, and MS analysis of the eluting protein material showed the formation of an apoMb-PEGB derivative eluting at the same retention time of apoMb. The presence of an apoMb dimer was also observed along with mono-conjugated dimer derivative (Table 1.3). Time dependent MS analysis (Figure 1.11) of the RP-HPLC peak eluting at the same retention time of apoMb showed the gradual formation of a mono-conjugated apoMb derivative along with the decrease of native apoMb.

The protein fingerprinting technique was employed for the characterization of the TGase conjugation site of the apoMb-PEGB derivative. Derivatized apoMb was digested with trypsin and the modified peptides were enriched by purification of the protein digest on an avidin column. MS analysis of the eluted fraction by the ESI

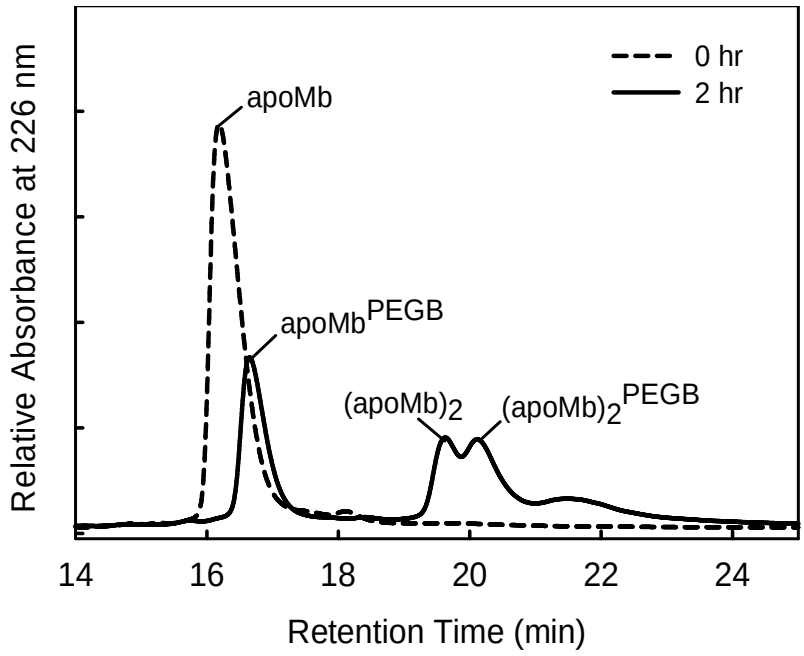


Figure 1.10. TGase-mediated conjugation of apoMb with PEGB. RP-HPLC analysis of the reaction mixture after 0 hour (dotted line) and 2 hours (solid line) of incubation. Analyses were performed on a C18 Phenomenex column using a linear gradient of AcCN containing 0.085 % TFA from 40 to 50 % in 25 min. The identities of the eluting protein material were confirmed by MS analysis.

Table 1.3. Molecular masses as observed from the MS analysis of the RP-HPLC peaks for the TGase mediated conjugation of apoMb to PEGB.

Protein Species	Molecular Mass (Da)	
	Observed	Calculated
PEGB	447.11	447.30
apoMb	16951.11	16951.51
apoMb1^{1PEGB}	17381.67	17381.51
$(apoMb)_2$	33887.55	33886.01
$(apoMb)_2^{1PEGB}$	34313.78	34316.01

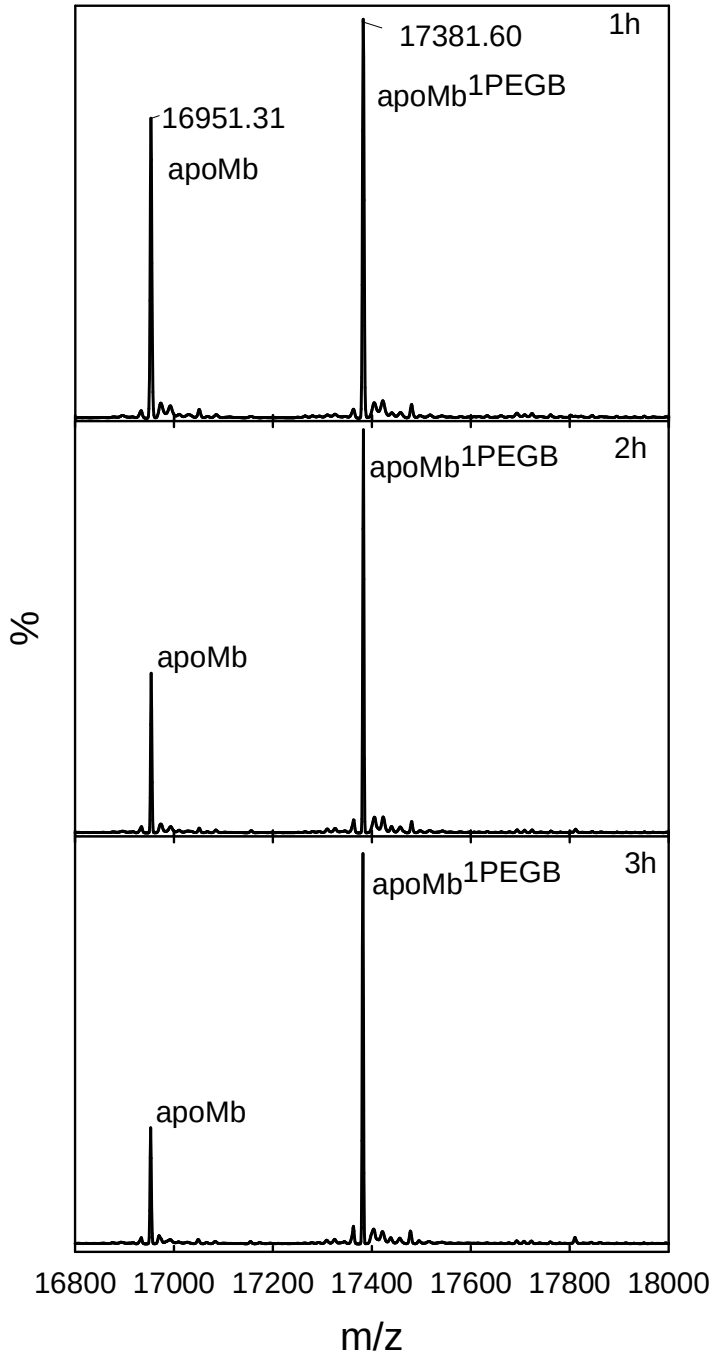


Figure 1.11. Deconvoluted MS spectra for the RP-HPLC fraction corresponding to the reaction time 1, 2 and 3 h of the TGase mediated conjugation of apoMb with PEGB. Observed molecular masses are assigned to the peaks along with its identity.

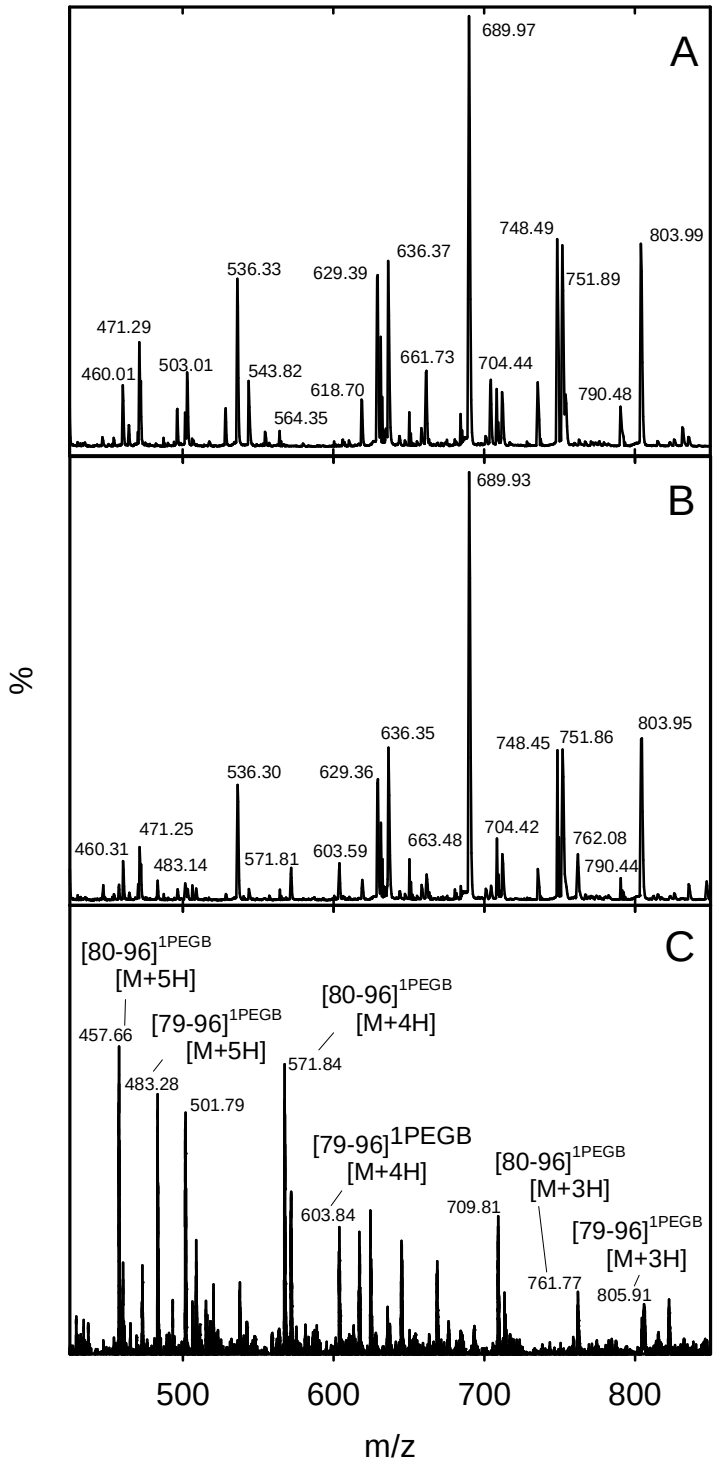


Figure 1.12. Identification of biotin-PEG-amine conjugation site(s) in apoMb. The region of the mass spectrum between 440 and 840 *m/z* is reported in order to show the derivatized tryptic peptides of apoMb. (A) Trypsin digest of non conjugated apoMb. (B) Trypsin digest of monoderivatised apoMb. (C) Trypsin digest of monoderivatised apoMb after avidin column purification. M/z signals of tryptic peptides conjugated to PEGB are indicated.

Table 1.4. Molecular masses as observed from the MS analysis of the apoMb-PEGB tryptic digest.

Fragment Species	Observed (m/z) [Charge state]	Molecular Mass (Da) Found	Calculated
apoMb1PEGB [80–96]	457.46 [M+5H]	2282.30	2283.25
apoMb^{1PEGB} [79–96]	483.09 [M+5H]	2410.45	2411.35
apoMb^{1PEGB} [80–96]	571.84 [M+4H]	2283.6	2283.25
apoMb^{1PEGB} [79–96]	603.86 [M+4H]	2411.44	2411.35
apoMb^{1PEGB} [80–96]	762.11 [M+3H]	2283.33	2283.25
apoMb^{1PEGB} [79–96]	804.83 [M+3H]	2411.49	2411.35

-QTOF mass spectrometer showed that modified peptides can be observed with a higher intensity signal in the MS spectrum in respect to the spectrum of the digest, thus allowing an easier identification of the TGase conjugation site(s) (Figure 1.12). Modified peptides corresponding to the region Gly80–Lys96 and Lys79–Lys96 of the apoMb sequence were observed (Table 1.4), which showed the exact mass increment of one biotin-PEG-amine molecule. These data confirm that the site of conjugation is again at the level of Gln91, since only one glutamine residue is present between the residues Lys79–Lys80 of the apoMb sequence. These results also clearly demonstrate that modification of proteins by TGase using biotin-PEG-amines allows an easy enrichment of the modified peptides on avidin columns and rapid identification of the conjugation sites by MS analysis.

d) Hydroxylamine

Hydroxylamine (HA) (Figure 1.3.D) was used as an acyl acceptor in the TGase-mediated conjugation studies on apoMb. In the RP-HPLC elution profile of the conjugation reaction after 30 minutes and 1 hour of incubation, only one peak was observed with the same retention time of apoMb (data not shown). MS analysis of these RP-HPLC peaks clearly indicated that apoMb and modified apoMb co-eluted in the same peak (Figure 1.13, Table 1.5). At longer reaction time, from the MS spectra it is evident that the amount of the mono-conjugated derivative increases along with a decrease of unmodified apoMb. The conjugation site of the apoMb-HA derivative was determined by trypsin digestion, followed by the identification of modified peptides by mass spectrometry. Modified peptides corresponding to the region Lys79–Lys96 and Gly80–Lys96 of the apoMb sequence were observed (Table 1.6), with a mass increment corresponding to the addition of one molecule of hydroxylamine by TGase. MS/MS analysis of peptide Gly80–Lys96 confirmed the conjugation at position Gln91 (Figure 1.14), the same glutamine residue derivatised in the reactions with tryptamine, dansylcadaverine and biotin-PEG-amine.

Table 1.5. Molecular masses as observed from the MS analysis of the RP-HPLC peaks for the TGase-mediated reaction of apoMb with HA.

Protein Species	Molecular Mass (Da)	
	Observed	Calculated
Hydroxylamine (HA)	-	33.03
apoMb	16951.11	16951.51
apoMb-hydroxylamine	16967.87	16967.54

Table 1.6. Molecular masses of the modified peptides as observed from the MS analysis of the apoMb^{1HA} tryptic digest.

Fragment Species	Molecular Mass (Da)	
	Found	Calculated
apoMb^{1HA} [80–96]	1869.12	1868.95
apoMb^{1HA} [79–96]	1997.24	1997.05

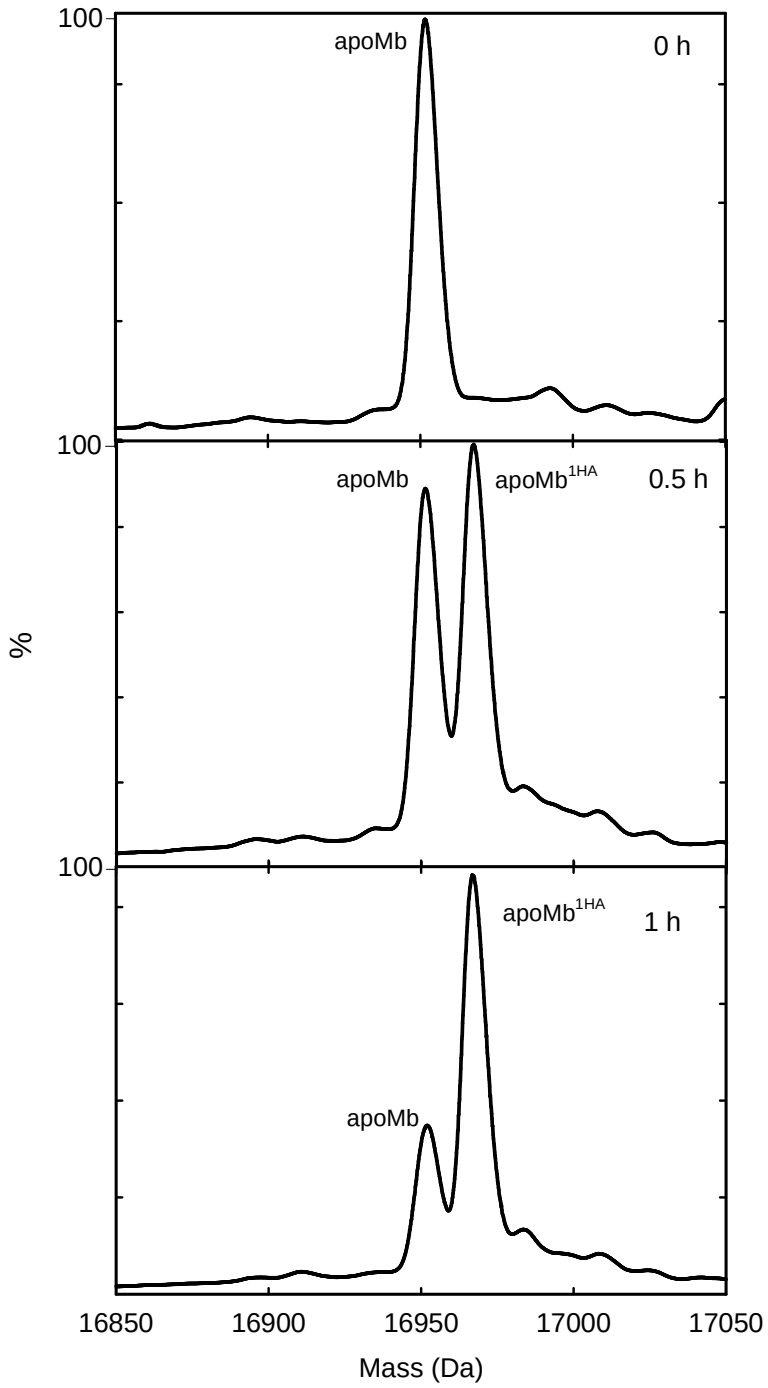


Figure 1.13. Deconvoluted MS spectra of the conjugation products obtained in the reaction between apoMb and hydroxylamine (HA) catalyzed by TGase. The reaction mixtures stopped at 0, 0.5 and 1 h were analysed by RP-HPLC and the fractions eluting at a retention time corresponding to that of apoMb were collected and analysed by MS. In the MS spectra, the observed molecular masses and the identities of the species are indicated.

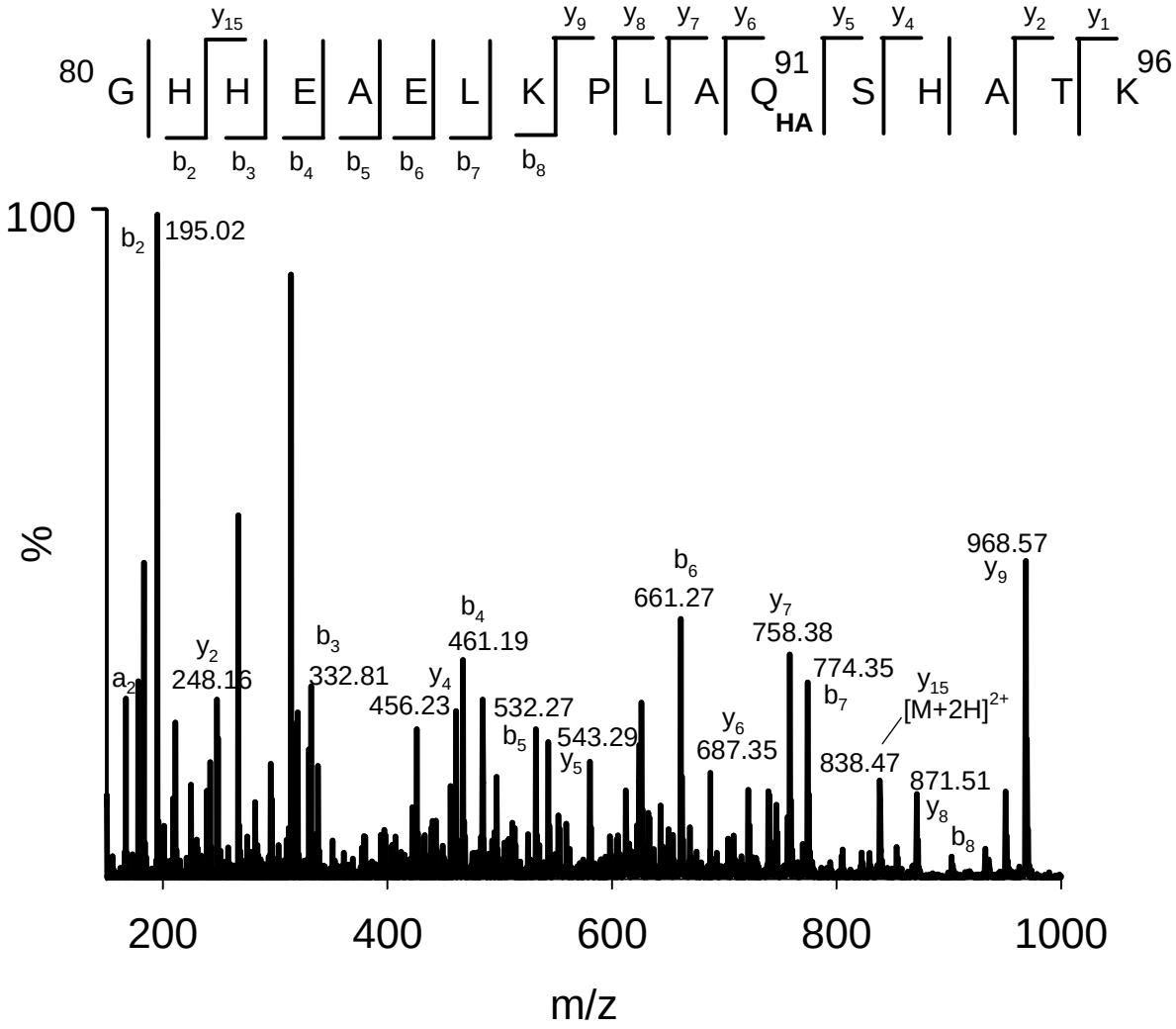


Figure 1.14. Electrospray MS/MS mass spectrum of the ion at 468.28 m/z of peptide 80–96 $[M+4H]^{4+}$ derivatized with HA. (Top) Fragments of the series *b* and *y* that were identified in the MS/MS spectrum are indicated on the sequence of the peptide. Modified Gln91 is indicated as Q_{HA}. (Bottom) In the MS/MS spectrum, ions assigned to fragments of the series *b* and *y* ions are indicated.

e) Glucosamine

TGase mediated conjugation studies on apoMb using glucosamine (Figure 1.3.E) as the acyl acceptor were also performed in order to investigate the reactivity of a primary amine bound to a sacaridic moiety. Indeed, glucosamine differs from the previous used amine derivatives for the absence of a aliphatic carbon chain (spacer) between the primary amine and the bulky group. The RP-HPLC analysis (Figure 1.15) of the reaction mixture after 1 hour of incubation and MS analysis of the collected peaks (Table 1.7) showed that the conjugation of apoMb with glucosamine did not occur and only the formation of the crosslinked dimer of apoMb was observed.

Table 1.7. Molecular masses as observed from the MS analysis of the RP-HPLC peaks of the reaction mixture containing apoMb, glucosamine and TGase.

Protein Species	Molecular Mass (Da)	
	Observed	Calculated
apoMb	16951.38	16951.51
$(apoMb)_2$	33886.55	33886.02

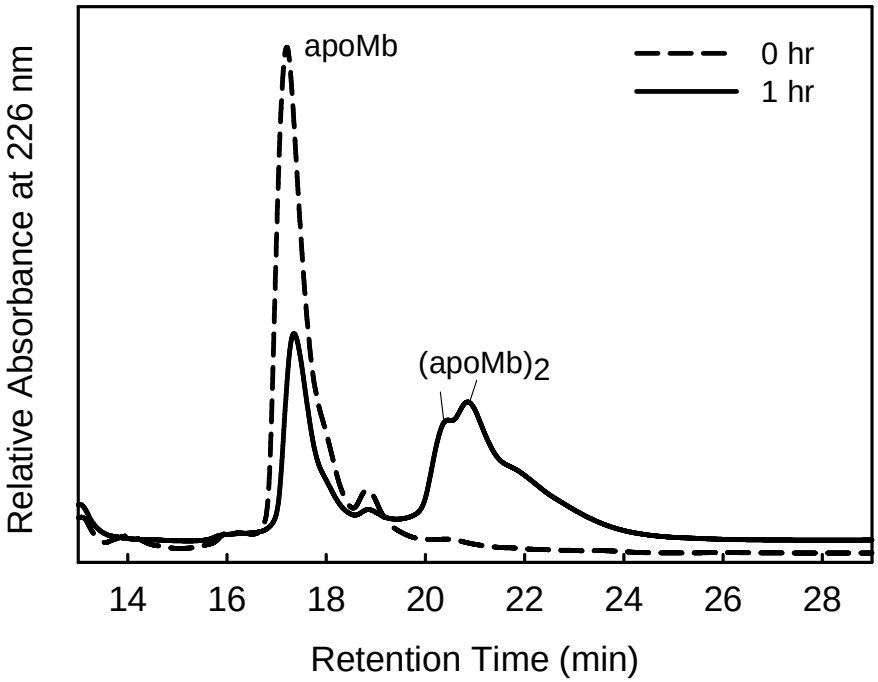


Figure 1.15. TGase mediated conjugation of apoMb with glucosamine. RP-HPLC analysis of the reaction on a C 18 Phenomenex column. The separation was performed with a linear gradient of AcCN containing 0.085 % TFA from 40 to 50 % in 25 min.

2. TGase-mediated conjugation of Z-Gln-Gly and its derivatives to apoMb

a) Z-Gln-Gly

The TGase-mediated conjugation of apoMb was performed using a tri-peptide, carbobenzoxy-glutamine-glycine (ZQG) (Figure 1.16.A) to study the selectivity of the conjugation reaction at the level of lysines. The RP-HPLC elution profile (Figure 1.17) of the reaction mixture after 5 hours of incubation showed the presence of three peaks (indicated as 1-3) eluting at a longer retention time than apoMb. MS analysis of these RP-HPLC peaks confirmed the formation of two mono-conjugated apoMb derivatives (peaks 2 and 3), and of a double ZQG conjugated derivative (peak 4) (Table 1.8).

Further characterization of the sites of conjugation was performed by digestion of the three protein derivatives collected from the RP-HPLC analysis followed by mass spectrometric analysis of the digests. The enzymatic hydrolysis was performed using trypsin and V8 protease in order to increase the sequence coverage of the protein. Indeed, the derivatization with ZQG at the level of lysine residues of apoMb hamper the trypsin hydrolysis at same lysine, thus leading to the formation of longer peptides that cannot be analysed by MS/MS. The use of V8 protease that hydrolyses the protein at the level of glutamic and aspartic acid residues allows the shortening of these long peptides for their further analysis by MS/MS.

Characterization of the conjugation site(s) by the protein fingerprinting technique revealed the presence of modified peptides corresponding to the region Leu86–Lys98 and His97–Lys102 of the apoMb sequence in peaks 2 and 3, respectively (Table 1.8). Peptide Leu86–Lys98 was generated by the V8-tryptic digestion. MS/MS analysis of the modified peptide Leu86–Lys98 (peak 2) confirmed the site of conjugation at the level of Lys96 (Figure 1.18). In the case of peak 3, the modified peptide His97–Lys102 contains only one lysine residue that can be derivatised by TGase (Lys98) indicating that this is the site of conjugation. In the case of peak 4, MS analysis of the tryptic digest of the bis-ZQG conjugated derivative of apoMb showed the presence of a double-modified peptide Leu86–Lys102 (Table 1.8), indicating that the derivatizations occurred at the level of both Lys96 and Lys98.

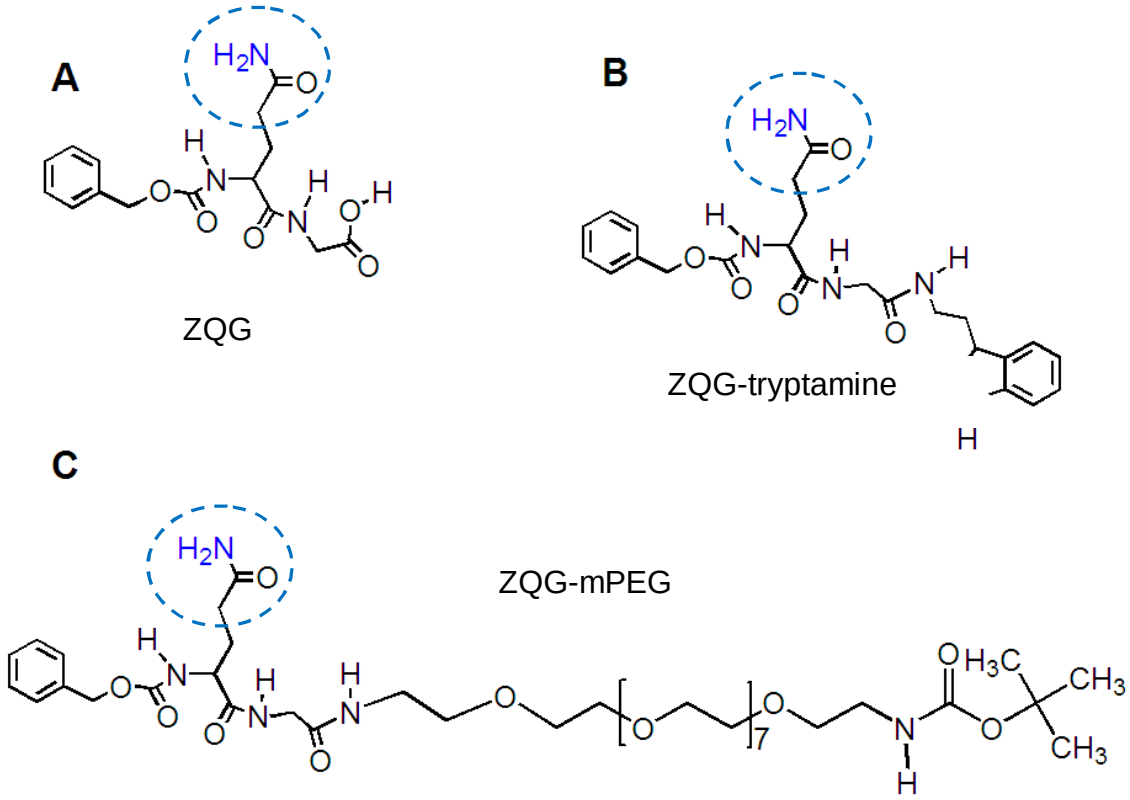


Figure 1.16. Chemical structures of ZQG and its derivatives. A) ZQG, B) ZQG-tryptamine and C) ZQG-mPEG. The TGase accessible γ-carboximide group of Gln is indicated by a dotted blue circle.

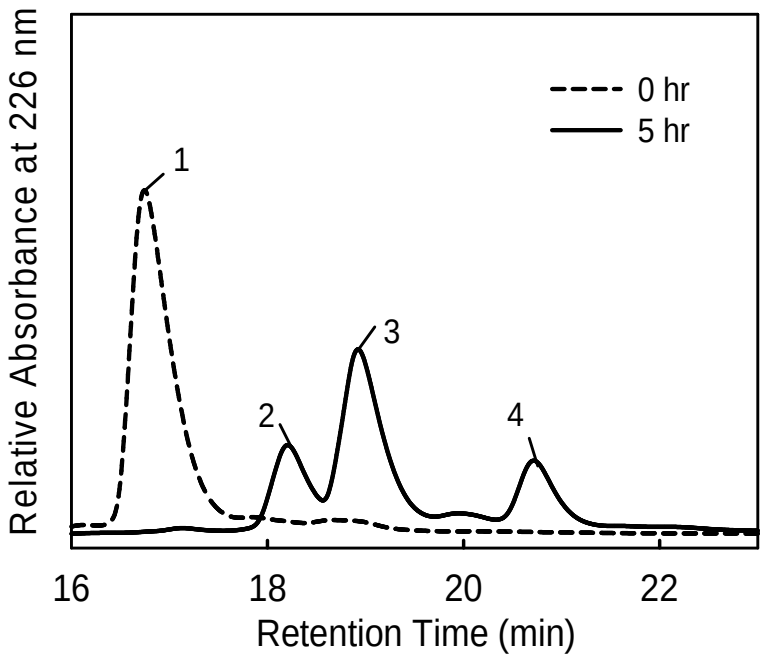


Figure 1.17. TGase mediated conjugation of apoMb with ZQG. RP-HPLC analysis of the reaction mixture after 0 hour (dotted line) and 5 hours (solid line) of incubation. Analyses were performed on a C 18 Phenomenex column using a linear gradient of AcCN containing 0.085 % TFA from 40 to 50 % in 25 min.

Table 1.8. Molecular masses of apoMb conjugated by TGase to ZQG. The observed and calculated molecular masses of the protein material eluted in the RP-HPLC peaks (See figure 1.17) and of the modified peptides identified in the tryptic digests of the apoMb derivatives are reported. The site of conjugation as determined from the modified peptides is also indicated.

RP-HPLC peak	Protein/fragment Species	Conjugation site(s)	Molecular Mass (Da)	
			Found	Calculated
	ZQG		-	337.33
1	apoMb		16951.56	16951.51
2	apoMb^{1ZQG}	K_{96}	17271.63	17271.56
	apoMb^{1ZQG} [86–98]		1778.83	1778.72
3	apoMb^{1ZQG}	K_{98}	17271.63	17271.56
	apoMb^{1ZQG} [97–102]		1054.48	1054.8
4	apoMb^{2ZQG}	K_{96}, K_{98}	17591.86	17591.67
	apoMb^{2ZQG} [86–102]		2550.45	2550.33

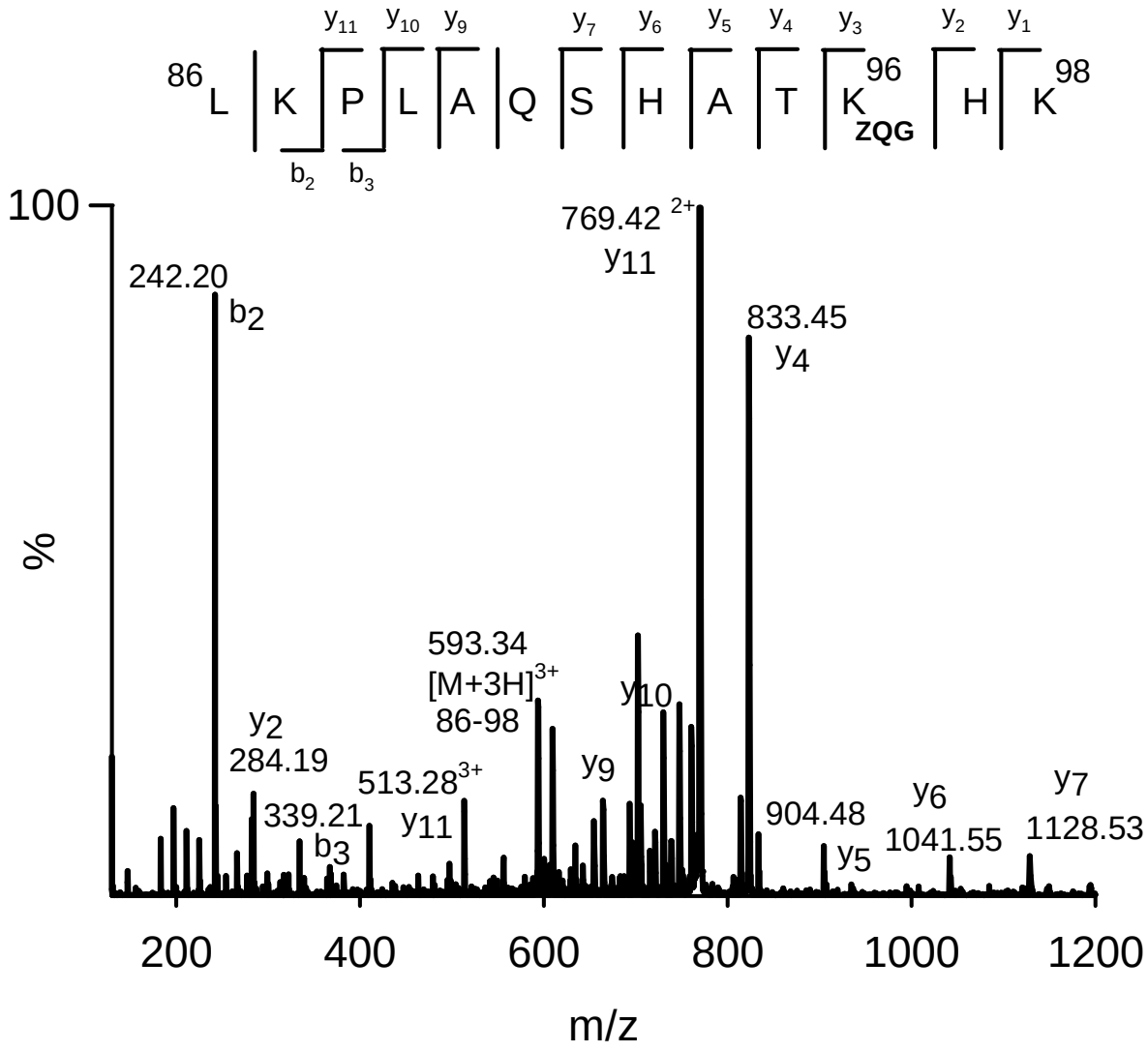


Figure 1.18. Electrospray MS/MS mass spectrum of the ion at 593.81 m/z of peptide 86–98 $[M+3H]^{3+}$ of apoMb modified with ZQG. (Top) Fragments of the series *b* and *y* that were identified in the MS/MS spectrum are indicated on the sequence of the peptide. Modified Lys96 is indicated as K_{ZQG}. (Bottom) In the MS/MS spectrum, ions assigned to fragments of the series *b* and *y* are indicated.

indeed, trypsin doesn't hydrolyze the protein at the level of Lys96 or Lys98 due to the presence of the ZQG moiety. Interestingly, Lys87 is not a site of derivatization by TGase even though it lies in the same flexible/disordered region His82–His97 as that of the TGase reactive Lys96 and Lys98This observation can be explained in analogy to proteases. Indeed, Lys87 is followed in the sequence by a proline residue (Figure 1.2). Considering the fact that trypsin cannot cleave the Lys–Pro or Arg–Pro peptide bonds (http://www.expasy.ch/tools/peptidecutter), it can be anticipated that the TGase mode of action is similar to that of proteases.

In the conjugation reaction of apoMb with ZQG, the derivatization was quantitative and formation of dimer was not observed. Moreover, the TGase mediated conjugation reaction at the level of lysines was seen to be very selective. Interestingly, both the conjugation sites lie in the flexible region His82–His97 of the apoMb sequence suggesting that the TGase mediated conjugation of proteins at the level of Lys residues also occurs in regions of enhanced chain flexibility.

b) Z-Gln-Gly-tryptamine and ZQG-mPEG

For the implementation in diagnostics (labeling) and therapy (PEGylation), it is expected that derivatives of ZQG containing a fluorescent label or a PEG chain can be conjugated to the proteins of interest by TGase at the level of only one or few lysines residues. In this Thesis, these derivatives of ZQG were synthesized and their use in the TGase conjugation reaction was tested on apoMb. Derivative of ZQG containing tryptamine as a fluorescent label (Figure 1.16.B) was prepared by the conventional HoBt/DCCI method. ZQG was chemically conjugated to tryptamine by the DCCI (dicyclohexylcarbodiimide)-HoBt (1-hydroxybenzotriazole) standard protocol. The primary amine group of tryptamine was conjugated at the –COOH group of glycine with the liberation of NH_3. The ZQG-tryptamine (ZQG-TP) derivative was purified by RP-HPLC. The elution profile of the reaction mixture after 24 hours of incubation consisted of three major peaks corresponding to tryptamine, ZQG and ZQG-TP (Figure 1.19). The identity of the peaks was confirmed by the mass spectrometric analysis (Table 1.9).

TGase-mediated conjugation studies were performed on apoMb using ZQG-TP as amine acceptor in order to obtain the selective conjugation at the level of

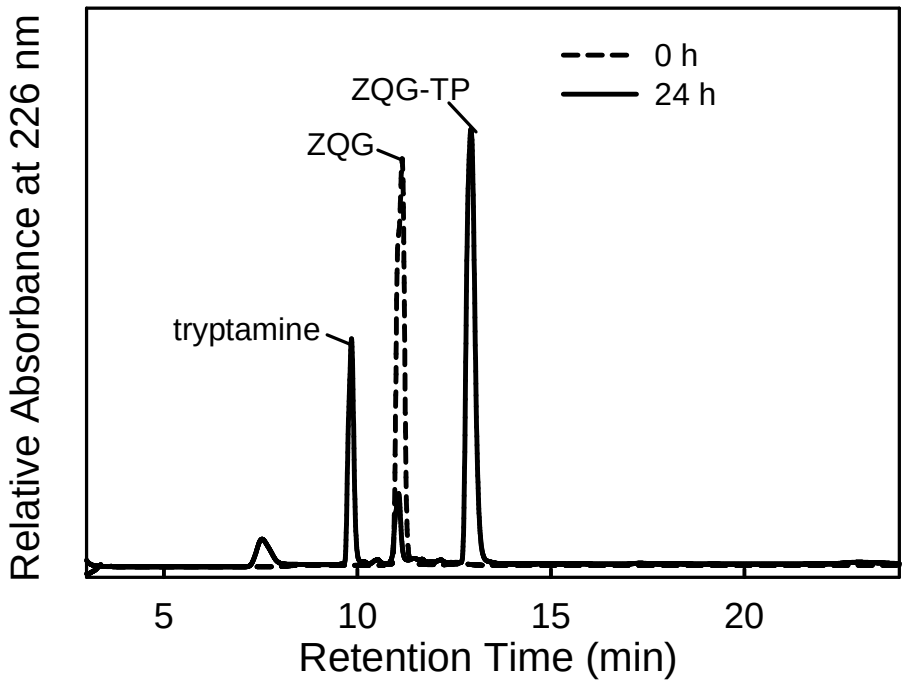


Figure 1.19. Preparation of ZQG-tryptamine derivative. RP-HPLC elution profile for the purification of the ZQG-TP derivative after 24 h of reaction (solid line). Analyses were performed using a C18 Phenomenex RP-HPLC column, at the flow rate of 0.8 ml/min, with a gradient from 40 to 50% of AcCN in 25 min.

Table 1.9. Molecular masses as observed from the MS analysis of the RP-HPLC peaks for the purification of the ZQG-TP derivative.

Identity	Molecular Mass (Da)	
	Found	Calculated
Tryptamine	-	160.22
ZQG	-	337.33
ZQG-TP	480.29	480.55

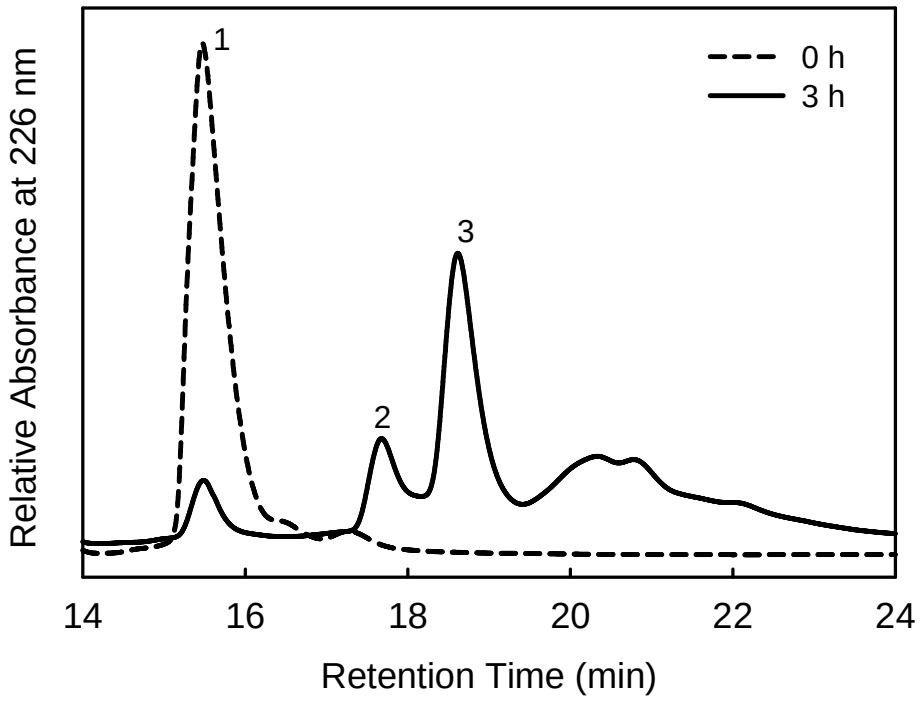


Figure 1.20. TGase-mediated conjugation of apoMb with ZQG-TP. RP-HPLC analysis of the reaction mixture after 0 hour (dotted line) and 3 hours (solid line) of incubation. Analyses were performed on a C 18 Phenomenex column using a linear gradient of AcCN containing 0.085 % TFA from 40 to 50 % in 25 min.

Table 1.10. Molecular masses of apoMb conjugated by TGase to ZQG-TP. The observed and calculated molecular masses of the protein material eluted in the RP-HPLC peaks (See figure 1.20) and of the modified peptides identified in the tryptic digests of the apoMb derivatives are reported. The site of conjugation as determined from the modified peptides is also indicated.

Protein/fragment Species	Conjugation site(s)	Molecular Mass (Da)	
		Found	Calculated
ZQG-TP		480.29	480.55
apoMb		16951.78	16951.51
$apoMb^{1ZQG\text{-}TP}$	K_{96}	17414.25	17414.50
$apoMb^{1ZQG\text{-}TP}$ [86–98]		1921.84	1921.71
$apoMb^{1ZQG\text{-}TP}$	K_{98}	17414.60	17414.50
$apoMb^{1ZQG\text{-}TP}$ [97–102]		1197.63	1197.48

lysines. The RP-HPLC elution profile of the reaction mixture after 3 hours of incubation (Figure 1.20) is similar to that obtained for the reaction in the presence of ZQG, since two main products are observed (peak 2 and 3) and a broad peak at a longer retention time. MS analysis of the protein material eluting in the RP-HPLC peaks 2 and 3 showed the presence of mono-conjugated derivatives (Table 1.10). Characterization of the conjugation sites by protein fingerprinting allowed the identification of modified peptides corresponding to the region Leu86–Lys98 and His97–Lys102 of the apoMb sequence for the RP-HPLC peaks 2 and 3, respectively. Therefore, the conjugation sites for the peak 2 and peak 3 are at the level of Lys96 and Lys98, in analogy to the results obtained for the derivatization with ZQG.

Derivative of ZQG containing a mono disperse amine mPEG moiety at the carboxyl terminus of glycine (Figure 1.16.C) was prepared by the conventional HoBt/DCCI method. ZQG was chemically conjugated to mPEG amine by the DCCI-HoBt standard protocol. The primary amine group of mPEG was conjugated at the –COOH group of glycine with the liberation of NH_3. ZQG-mPEG derivative was purified by RP-HPLC and the elution profile of the reaction mixture after 24 hours of incubation consisted of three major peaks corresponding to mPEG, ZQG and ZQG-mPEG (Figure 1.21). The identity of the peaks was confirmed by mass spectrometric analysis (Table 1.11).

TGase mediated conjugation studies were performed on apoMb using ZQG-mPEG as amine acceptor in order to obtain the selective conjugation at the level of lysines. The RP-HPLC elution profile (Figure 1.22) of the reaction mixture after 1 hour of incubation was similar to that of the reaction mixtures containing ZQG and ZQG-TP. MS analysis of the protein material eluting in the RP-HPLC peaks showed the presence of mono-conjugated apoMb derivatives for peak 2 and 3 and of a double conjugated derivative for peak 4 (Table 1.12). Characterization of the conjugation sites by the protein fingerprinting technique indicated that the conjugation sites are at the level of Lys96 and Lys98 for the peak 2 and peak 3, respectively. For the doubly conjugated derivative (peak 4), the conjugation is at the level of Lys96 and Lys98. These results indicate that the TGase mediated derivatization of apoMb with ZQG, ZQG-TP and ZQG-mPEG occur at the same lysine residues.

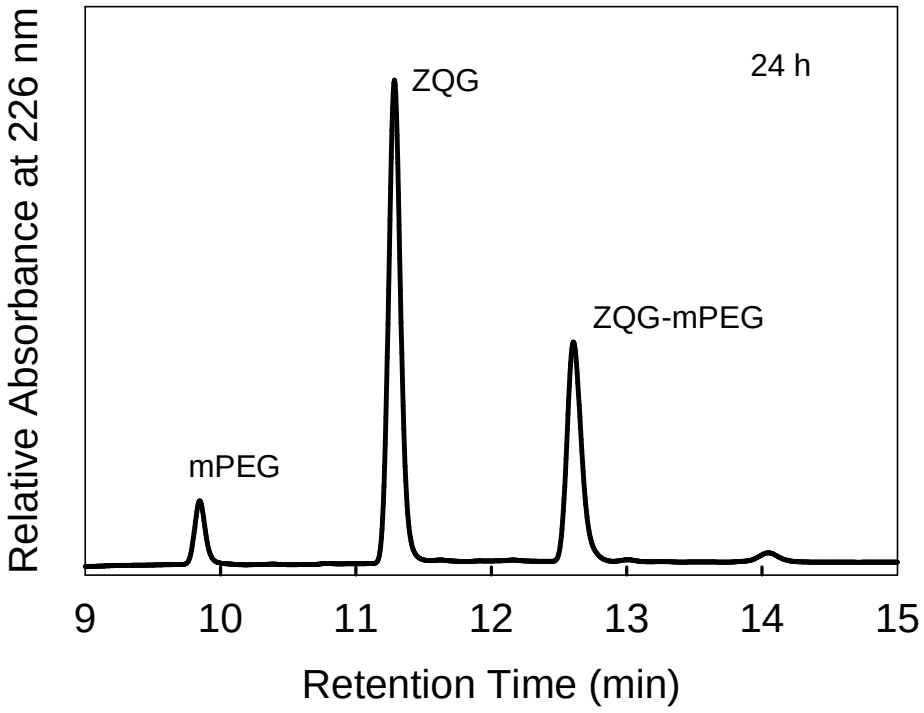


Figure 1.21. Preparation of ZQG-mPEG derivative. RP-HPLC elution profile for the purification of ZQG-mPEG derivative after 24 h of reaction. Analyses was performed using a C_{18} (Phenomenex) RP-HPLC column, at the flow rate of 0.8 ml/min, with a gradient from 40 to 50% of AcCN in 25 min.

Table 1.11. Molecular masses as observed from the MS analysis of the RP-HPLC peaks for the purification of the ZQG-TP derivative.

Identity	Molecular Mass (Da)	
	Found	Calculated
mPEG	-	556.35
ZQG	-	337.33
ZQG-mPEG	876.41	876.68

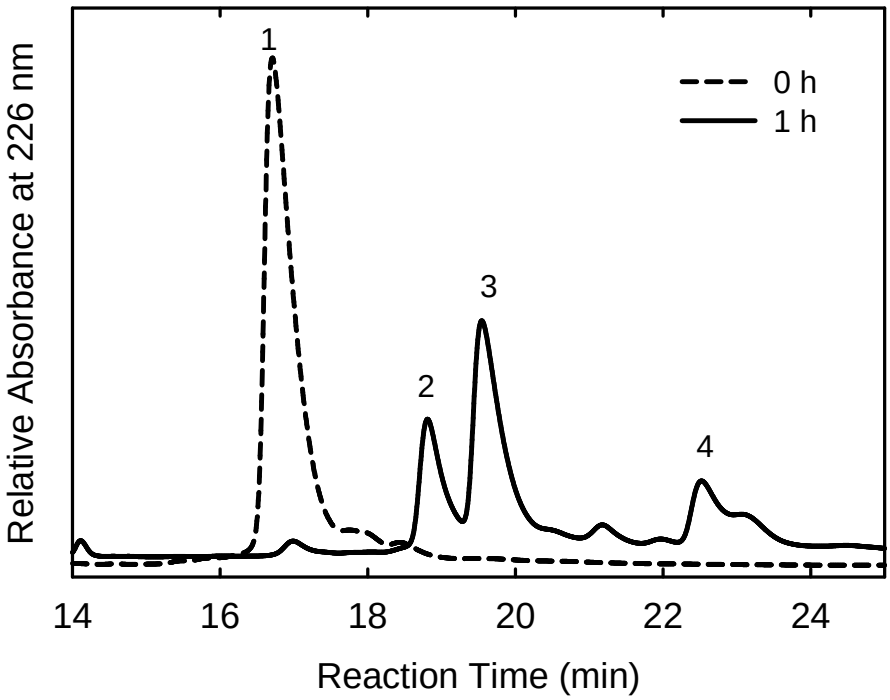


Figure 1.22. TGase-mediated conjugation of ZQG-mPEG with apoMb. RP-HPLC elution profile on a C 18 Phenomenex column showing the 0 hr (dotted line) and 1 hr (solid line) of reaction. The separation was performed with a linear gradient of AcCN containing 0.085 % TFA from 40 to 50 % in 25 min.

Table 1.12. Molecular masses of the products of the TGase-mediated reaction between apoMb and ZQG-mPEG (See figure 1.22). Molecular masses of the modified peptides detected by MS in the tryptic digests of the protein material eluting in the RP-HPLC peaks are reported. The sites of conjugation as determined in the modified peptides are also reported.

Peak	Protein/fragment Species	Conjugation site(s)	Molecular Mass (Da)	
			Found	Calculated
	ZQG- mPEG		876.41	876.68
1	apoMb		16951.78	16951.51
2	$apoMb^{1ZQG\text{-}mPEG}$	K_{96}	17810.63	17810.51
	$apoMb^{1ZQG\text{-}mPEG}$ [86–98]		2318.13	2317.72
3	$apoMb^{1ZQG\text{-}mPEG}$	K_{98}	17810.78	17810.51
	$apoMb^{1ZQG\text{-}mPEG}$ [97–102]		1593.63	1593.94
4	$apoMb^{2ZQG\text{-}mPEG}$	K_{96} , K_{98}	18669.95	18669.51

3. PEGylation of apoMb by chemical methods and by TGase

Protein drugs can be derivatised with poly(ethylene glycol) (PEG) to overcome the limitations of short half life and immunogenicity (see Introduction). PEGylation of protein indeed increases their bio-availability by reducing immune reactions, decreasing renal clearance and increasing the stability and resistance towards proteases (*7*). Recent research has been focused on the development of site-specific PEGylation approaches to generate homogenous products. Indeed, at present different chemical PEG conjugation strategies are available (*8-9*), which produce heterogeneously PEGylated proteins with a non-uniform chemical and pharmaceutical profile. The TGase mediated site-specific conjugation of PEG-amines can be used to generate homogenously PEGylated protein drugs (*7*). To demonstrate the efficiency and selectivity of the TGase mediated site-specific PEGylation approach, apoMb was PEGylated using two different strategies, the N terminal site-specific PEGylation strategy and with PEG-amines using TGase.

a) N- terminal site-specific PEGylation of apoMb

This approach involves a pH driven nucleophilic attack of the N-terminal amino acid to the electrophile PEG-aldehyde. The selectivity of the approach relies on the pKa difference between the N terminal amino group (pKa 7.6-8.0) and the ε-amine groups of lysines residues (pKa 10.0 – 10.2). In the first step, a labile Schiff base is formed and then by reduction, the formation of a stable alkyl amine bond occurs. PEG_{20k}-CHO was conjugated to apoMb in the presence of $NaCNBH_3$ in sodium acetate buffer pH. 5.0. The RP-HPLC elution profile of the reaction mixture after 24 hours showed the formation of more than one isoforms of PEGylated apoMb (Figure 1.23.A). Reaction conditions were optimized to reduce the heterogeneity of the products, by changing the ratio of the reactants from 1:10:10 to 1:5:5 (Figure 1.23.B). SDS-PAGE analysis of the protein material eluted in the RP-HPLC peaks (Figure 1.23.B) clearly indicated the elution of $apoMb^{PEG20k}$ conjugates in peaks 2, 3, and 4. SDS-PAGE lanes corresponding to peak 2 and 3 indicated the formation of mono-conjugated apoMb derivatives, whereas analysis of peak 4 indicated the conjugation of one or more PEG_{20k}-CHO to apoMb. The exact molecular weight

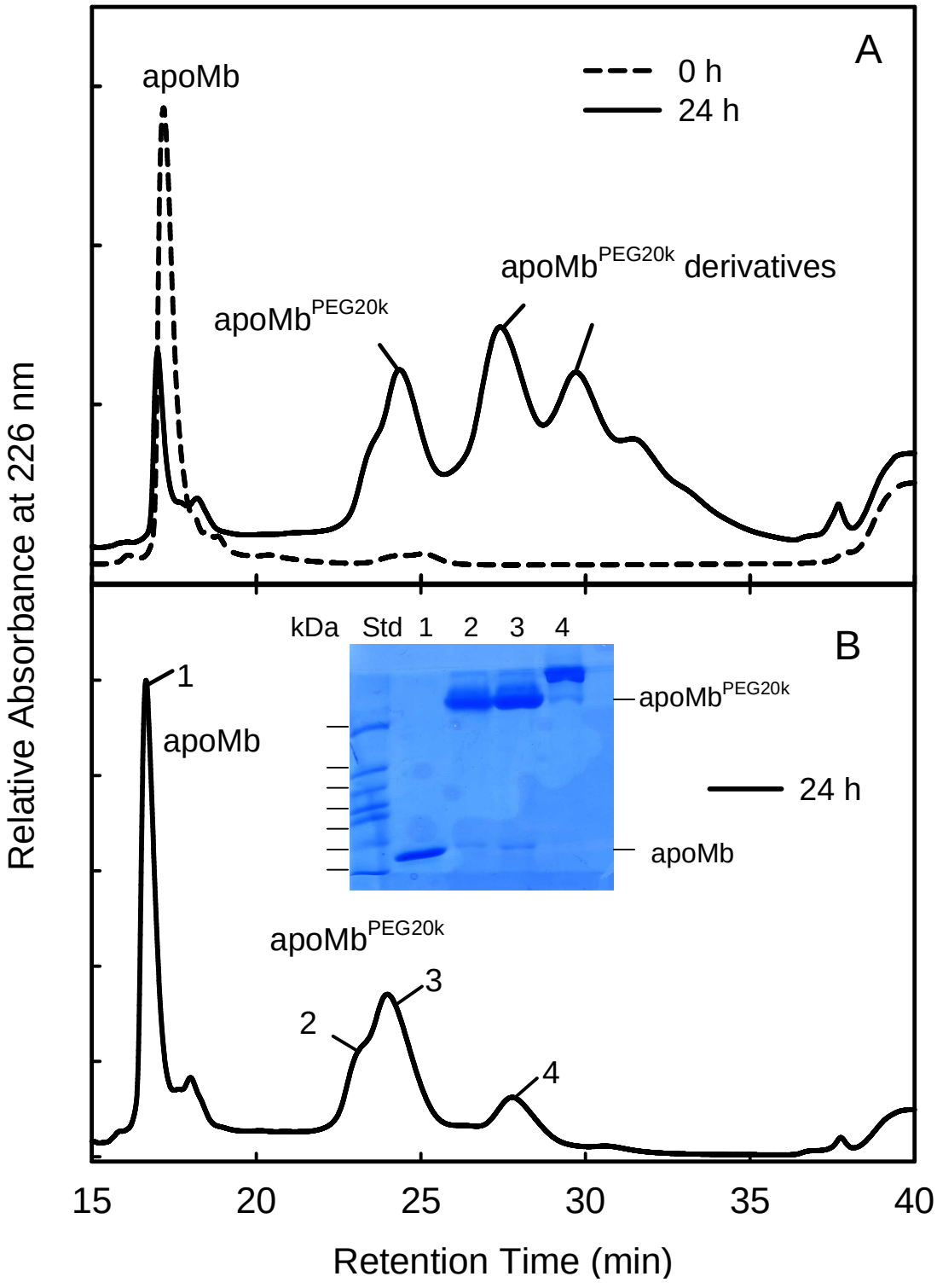


Figure 1.23. N-terminal PEGylation of apoMb. RP-HPLC elution profile comparison between the A) initial reaction and B) optimized reaction conditions after 24 hours of incubation. The separation was performed on a C4 phenomenex column with a linear gradient of AcCN containing 0.085 % TFA from 40 to 50 % in 25 min.The SDS-PAGE analysis of the peaks is also reported.

determination of the derivatives was not obtained by SDS-PAGE analysis due to the interaction of PEG with SDS, which causes the migration of the conjugates different than their actual molecular weight (*10*). Optimized reaction conditions did not lead to a quantitative production of apoMb^{PEG20k}. Indeed, the amount of unmodified apoMb (peak 1) was higher than that of the apoMb^{PEG20k} conjugate (peak 2) as evident from the respective heights and areas of the peaks (Figure 1.23). The presence of a shoulder in peak 2 also indicates that the apoMb^{PEG20k} derivative is heterogeneous.

Characterization of the sites of conjugation in the protein material eluting in peak 2 was performed by the protein fingerprinting technique. Since the measurement of the molecular mass of the 20kDa PEG conjugates by ESI-TOF mass spectrometry is hampered, the modified peptides were identified on the basis of their chromatographic shift in the RP-HPLC chromatogram. Conjugated and non-conjugated apoMb was digested with trypsin and analysed by RP-HPLC separately. Comparison of the two elution profiles (Figure 1.24) clearly showed the decrease in the peak 1 and the appearance of peak 2 for the tryptic digest of the conjugate. MS analysis of the eluting peptides revealed that peak 1 corresponds to the region Gly1–Lys16 of the apoMb sequence and peak 2 corresponds to the PEGylated peptide. Poly-dispersity of PEG hindered the proper characterization of the conjugation site by ESI-MS, which is due to the fact that this ionization technique produces multiply charged ions of the PEGylated peptide hampering the interpretation of mass spectra (*11-12*). Nonetheless, it is evident from the above results that at least one conjugation site occurred at the N-terminus of the apoMb amino acid sequence.

b) TGase-mediated PEGylation of apoMb using PEG amines

TGase mediated PEGylation of apoMb was achieved using PEG-amines as acyl acceptors. Mono-disperse PEG-amine (mPEG) and a poly-disperse PEG-amine of 5 kDa (PEG5k) were used in two separate studies to demonstrate that the length of the PEG chain does not affect the rate of the TGase reaction and its specificity. The RP-HPLC profile of the TGase-mediated conjugation with PEG-amines followed by the MS analysis of the chromatographic peaks showed the quantitative formation of mono-PEGylated derivatives for both the PEG-amines (1.25). Formation with a low yield of a bis-conjugated derivative was observed only in the TGase reaction with

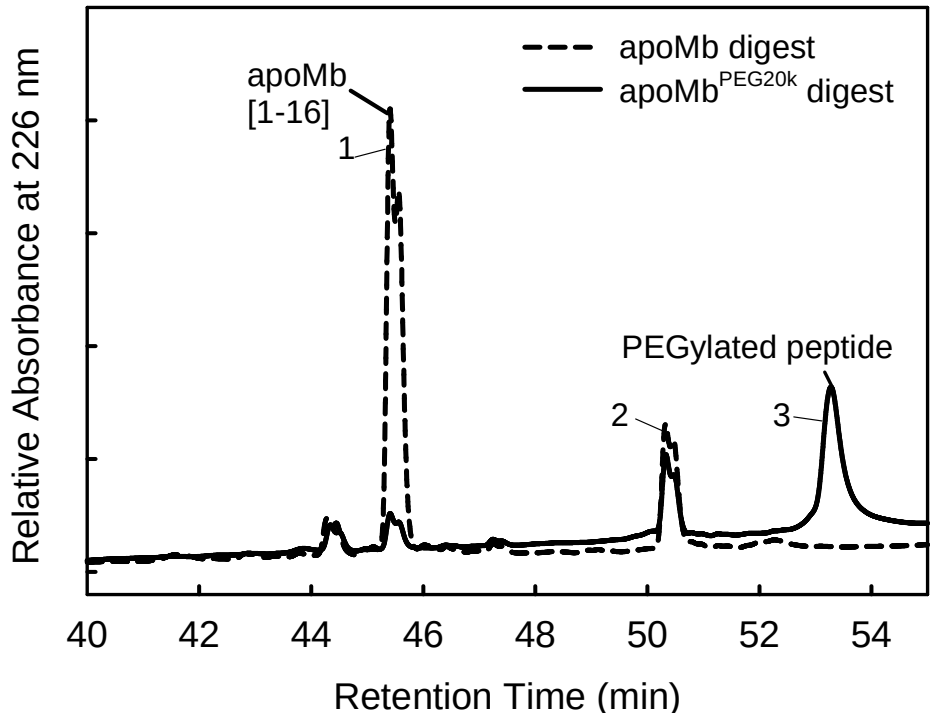


Figure.1.24. Characterization of the sites of derivatization of apoMb subjected to the N-terminal site specific conjugation reaction. RP-HPLC elution profile comparison for the tryptic digests of apoMb and apoMb^{PEG20k}. The identification of peptide material eluted in the RP-HPLC peaks obtained by ESI-QTOF MS analysis is shown.

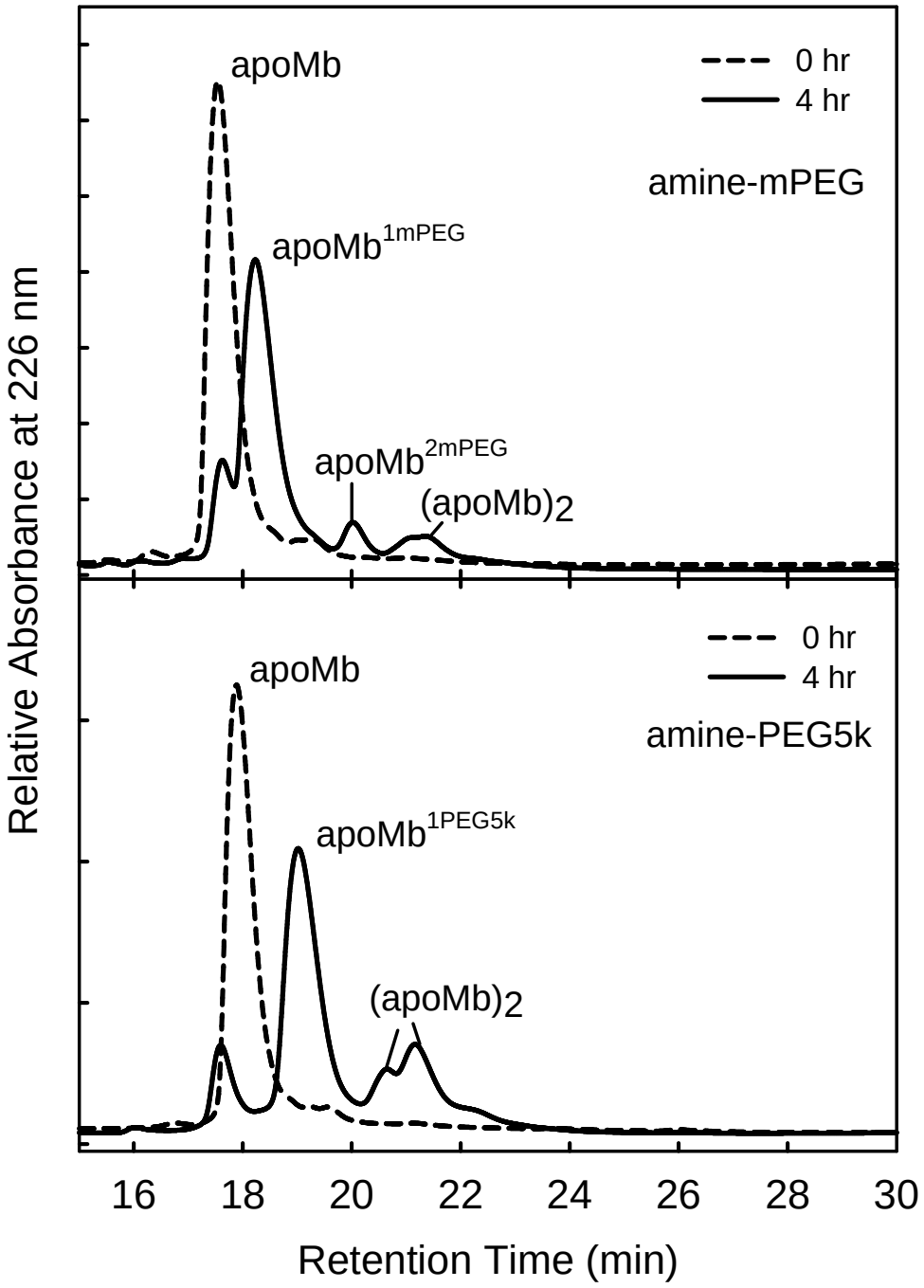


Figure 1.25. TGase mediated conjugation of amine-mPEG and polydisperse amine-PEG5k with apoMb. RP-HPLC elution profiles for the 4 hours TGase mediated reaction mixtures of apoMb with (A) amine-mPEG and (B) amine-PEG5k are shown. The analyses were performed using a C4 Phenomenex RP-HPLC column at the flow rate of 0.8 ml/min, with a gradient from 40 to 50% of AcCN containing 0.085 % TFA in 25 min.

mPEG, while small amounts of dimer of apoMb were observed in both the reactions. Molecular masses of the products of these reactions measured with the ESI-QTOF mass spectrometer showed the addition of one mPEG or PEG5k molecule to apoMb (Table 1.13).

Table 1.13. Molecular masses as observed from the MS analysis of the RP-HPLC peaks (Figure 1.25) for the TGase mediated conjugation of apoMb with PEG amines.

Protein Species	Molecular Mass (Da)	
	Found	Calculated
mPEG	556.01	556.31
PEG5k	-	5000
apoMb	16951.11	16951.51
$apoMb^{1mPEG}$	17491.13	17491.82
$apoMb^{2mPEG}$	18030.50	18030.23
$(apoMb)_2$	33887.55	33886.01

Comparison of the efficiency of the PEGylation of apoMb by N-terminal derivatization and TGase-mediated conjugation indicated that only the TGase-catalysed reaction leads in a short time (4 h) to generate a homogenous mono-conjugated derivative. The length of the PEG chain did not affect the specificity of the TGase reaction or its rate. This fact is in good agreement with the assumption that the TGase mediated conjugation is mainly dictated by the structural features of the protein substrate containing the glutamine residue and not by the size of the PEG chain (*7*). Overall, these results clearly demonstrated that TGase-mediated PEGylation is more efficient in producing homogenously PEGylated proteins derivatives than the chemical N-terminal PEGylation.

4. Comments

Derivatizatization of apoMb with TGase using primary amines. TGase-mediated conjugation studies on apoMb using primary amines demonstrated the efficiency of the conjugation reaction for tryptamine, dansylcadaverine, biotin-PEG-amine and hydroxyl amine. No conjugation reaction was observed with glucosamine,

which can be attributed to the fact that the lack of a spacer between the primary amine and the bulky functional group creates a steric hindrance to the interaction with TGase. Indeed, the chemical structures of tryptamine, dansylcadaverine and biotin-PEG-amine all consist of an aliphatic carbon chain spacer separating the primary amine and a bulky functional group (Figure 1.3). Therefore, these results suggest that at least a two carbon atom spacer is required between the primary amine and the bulky functional group for the interaction with TGase and the conjugation at the level of glutamine residues. As a consequence, if a spacer is present then different bulky functional groups can be successfully incorporated in apoMb by the TGase reaction. Hydroxylamine was also conjugated to apoMb, and this fact indicates that primary amines without a spacer and bulky functional groups can also act as substrates for the TGase reaction.

The identification by MS of the sites of TGase conjugation on apoMb indicated that independently on the primary amine used in the reactions, the conjugation sites are located at Gln91 and Gln152. Indeed, even though there are 6 glutamines in the apoMb sequence, only these two Gln residues which are located in flexible/disordered regions were conjugated. Therefore, the selectivity of the TGase reaction is governed by the flexibility of the polypeptide chain encompassing the Gln residues and not by the molecular features of the bulky functional groups of the primary amines herewith used as substrates.

The TGase conjugation studies with biotin-PEG-amine were also used to develop a very rapid procedure to identify the sites of TGase conjugation. In the pharmaceutical industry, proper characterization of protein drugs is essential. The derivatization of proteins with the poly-disperse high molecular weight PEG limits the use of MS for the characterization of the sites of modification (*11-12*). It has already been demonstrated that the molecular features of the bulky functional groups carrying the primary amines does not affect the specificity of the TGase reaction and that at an analytical level mono-disperse PEG-amines of small molecular weights can be used to identify the sites of modification by TGase (*12*). However, the MS based analysis by the protein fingerprinting technique is time consuming due to the presence in the mass spectra of a background of signals due to non-modified peptides. Therefore, by using

the biotin-avidin high affinity interaction it was demonstrated that the enrichment of the modified peptides is possible by affinity chromatography, thus allowing a more rapid identification by MS of the sites of conjugation.

Derivatizatization of apoMb with TGase using glutamine containing peptides. Conjugation studies on apoMb with ZQG and its derivatives by TGase demonstrated that out of 19 lysines, only two (Lys96 and Lys98) are selectively modified. Incorporation of PEG or a fluorescent group (tryptamine) at the C-terminus of glycine in ZQG did not affect the specificity and efficiency of the TGase reaction. Hence, it can be anticipated that ZQG can be used efficiently as a carrier for the TGase mediated incorporation of functional groups at the level of lysines.

*Correlation between sites of limited proteolysis and sites of conjugation by TGase in proteins***.** Proteolytic probes are used to analyse the structural and dynamic features of globular proteins. (*6, 13-15*). Indeed, it has been demonstrated that sites of limited proteolysis are located in regions of the polypeptide chain characterized by an enhanced chain flexibility (*14*). A clear correlation exists between the sites of limited proteolysis and the regions of high segmental mobility as given from the B factor values obtained from X ray crystallography.

Limited proteolysis studies on myoglobin have reported that the holo form is resistant towards proteolysis, whereas apoMb is quickly hydrolyzed at the level of the polypeptide chain encompassing helix F, indicating that this region is rather flexible/disordered (*5-6*). At room temperature and neutral pH, different proteases such as subtilisin, thermolysin, chymotrypsin and trypsin cleave apoMb at a single peptide bond in the restricted region His82–His97 of the apoMb sequence (Figure 1.26.A). Indeed, thermolysin hydrolyses the polypeptide chain between residues Pro88–Leu89, chymotrypsin at the Leu89–Ala90 peptide bond , subtilisin at Gln91–Ser92, His93–Ala94 and trypsin between Lys96–His97 (*6, 14*). These data clearly demonstrate that the chain segment His82–His97 of the apoMb sequence is highly flexible and thus it is subjected to selective proteolysis. NMR studies on apoMb have confirmed that the chain segment corresponding to the helix F (His82–His97) is unfolded (*1, 16*). In the present study, TGase conjugation sites were seen to occur in

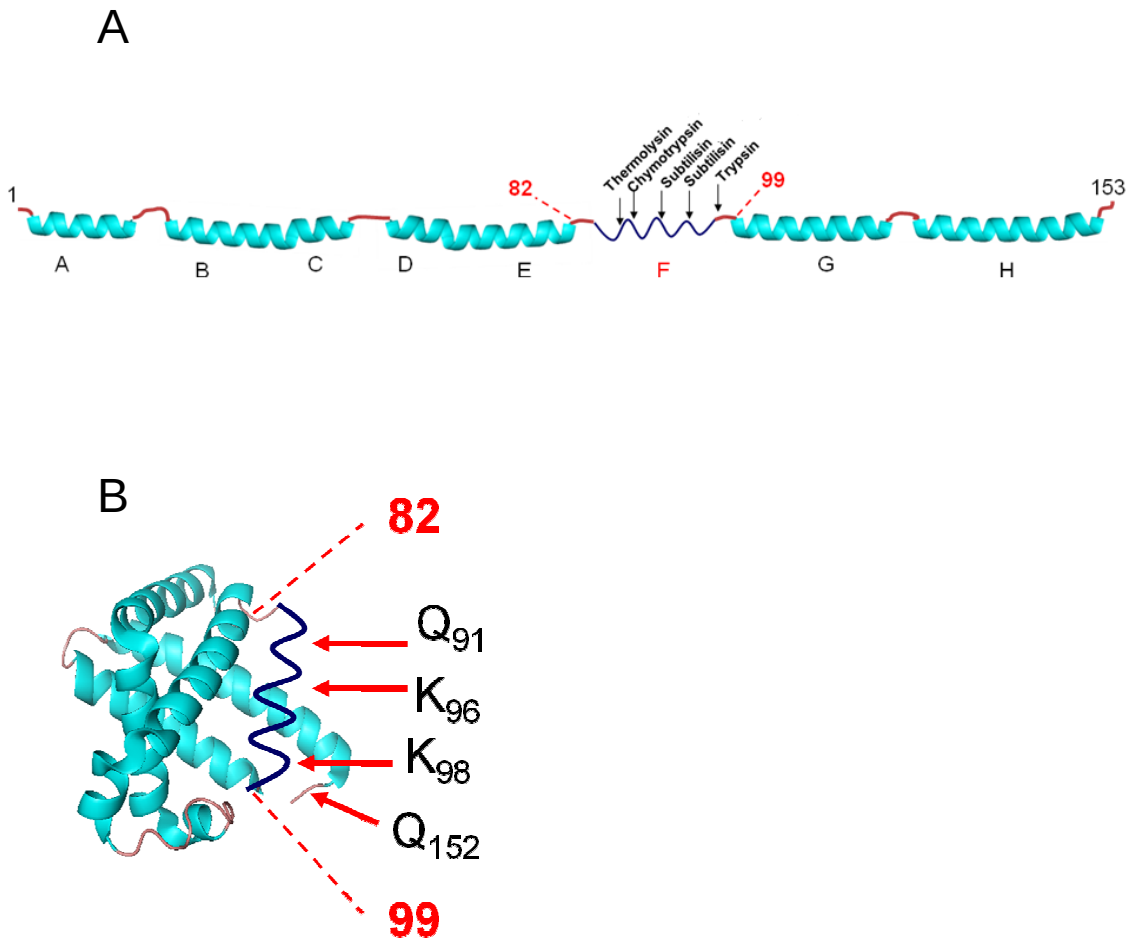


Figure 1.26. Correlation between sites of TGase conjugation and of limited proteolysis on apoMb. A) Secondary structure of apoMb is indicated with squiggled lines for the helical segments (A-H) along the protein chain. The disordered chain segment of helix F is indicated in dark blue. The sites of limited proteolysis are indicated by arrows for the respective proteases. B) 3D structure of apoMb as derived from the X-ray structure of the holo protein (PDB file 1YMB). The disordered helix F as given from NMR studies is shown in dark blue color. TGase-mediated conjugation sites at Q91, K96, K98 and Q152 are indicated by red arrows.

the regions His82–His97 and Phe151–Gly153 of the apoMb sequence (Figure 1.26.B). Indeed, Gln91 and Lys96 are localised in the region of helix F of apoMb, Lys98 is located in the flexible chain loop connecting helix F and G and Gln152 is located at the C-terminal of the sequence of apoMb. These regions show enhanced chain flexibility as given from the results of limited proteolysis experiments and NMR measurements. Out of 6 glutamines and 19 lysines in apoMb, only two Gln and two Lys residues located in flexible/disordered regions were conjugated. This observation provides clear indication that TGase acts similarly to a protease in selecting the site(s) of derivatization, which have to be located in flexible regions of the polypeptide chain.

Moreover, in this Thesis it is demonstrated that TGase can not derivatise Gln or Lys residues located at the N-terminus of a Pro residue. Similarly, trypsin cannot hydrolyse the peptide bonds in between Lys–Pro or Arg–Pro.

Use of TGase for the production of PEGylated protein drugs. In this Thesis, the specificity and efficacy of the TGase mediated conjugation in producing PEGylated proteins was compared to those of the N-terminal PEGylation chemistry. This reaction allows the conjugation of PEG-aldehyde to the N terminal residue of a protein, due to the difference in pKa between the N-terminal amine and the other amine groups in the protein (*17-18*). However, small differences in the pKa along the protein sequence often provide multiply conjugated proteins, as in the case of apoMb. Efforts to determine the optimal pH to achieve a site-specific reaction at the amine terminus are complicated and the conditions are specific for each protein (*19*). The studies here reported on the TGase-mediated PEGylation of apoMb with PEG amines demonstrated the easy production of homogenous PEGylated derivatives. The PEGylation reaction was also found to be almost quantitative and reproducible, thus allowing an easy scale up at an industrial level. Considering that TGase is available commercially at a low price and that the reaction conditions are mild (i.e. neutral pH and room temperature), these results demonstrate the several advantages of the use of TGase in producing PEGylated proteins..

5. References

1. Eliezer, D., Yao, J., Dyson, H.J., and Wright, P.E. (1998) Structural and dynamic characterization of partially folded states of apomyoglobin and implications for protein folding, *Nat. Struct. Biol. 5*, 148–155.

2. 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.

3. Hughson, F.M., Wright, P.E., and Baldwin, R.L. (1990) Structural characterization of a partly folded apomyoglobin intermediate, *Science 249*, 1544–1548.

4. Griko Iu, V., Privalov, P.L., Ven'iaminov, S., and Kutyshenko, V.P. (1988) Thermodynamic study of the structure of apomyoglobin, *Biofizika 33*, 18–26.

5. Picotti, P., Marabotti, A., Negro, A., Musi, V., Spolaore, B., Zambonin, M., and Fontana, A. (2004) Modulation of the structural integrity of helix F in apomyoglobin by single amino acid replacements, *Protein Sci. 13*, 1572–1585.

6. 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.

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

8. Roberts, M.J., Bentley, M.D., and Harris, J.M. (2002) Chemistry for peptide and protein PEGylation, *Adv. Drug Deliv. Rev. 54*, 459–476.

9. Veronese, F.M., and Pasut, G. (2006) PEGylation of Proteins as Tailored Chemistry for Optimized Bioconjugates, In *Polymer Therapeutics I Polymers as Drugs, Conjugates and Gene Delivery Systems* (Satchi-Fainaro, R., and Duncan, R., Eds.), pp 95–134, Springer.

10. Zheng, C., Ma, G., and Su, Z. (2007) Native PAGE eliminates the problem of PEG-SDS interaction in SDS-PAGE and provides an alternative to HPLC in characterization of protein PEGylation, *Electrophoresis 28*, 2801–2807.

11. Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F., and Whitehouse, C.M. (1989) Electrospray ionization for mass spectrometry of large biomolecules, *Science 246*, 64–71.

12. 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, *Bioconjug. Chem. 20*, 384–389.

13.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.

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

15.Fontana, A., Polverino de Laureto, P., De Filippis, V., Scaramella, E., and Zambonin, M. (1999) Limited proteolysis in the study of protein conformation, In *Proteolytic Enzymes: Tools and Targets* (Sterchi, E. E., and Stöcker, W., Eds.), pp 257–284, Springer Verlag, Heidelberg.

16.Eliezer, D., and Wright, P.E. (1996) Is apomyoglobin a molten globule? Structural characterization by NMR, *J. Mol. Biol. 263*, 531–538.

17.Kinstler, O., Molineux, G., Treuheit, M., Ladd, D., and Gegg, C. (2002) Mono-N-terminal poly(ethylene glycol)-protein conjugates, *Adv. Drug Deliv. Rev. 54*, 477–485.

18.Kinstler, O.B., Brems, D.N., Lauren, S.L., Paige, A.G., Hamburger, J.B., and Treuheit, M.J. (1996) Characterization and stability of N-terminally PEGylated rhG-CSF, *Pharm. Res. 13*, 996–1002.

19.Choi, J., Godwin, A., Balan, S., Bryant, P., Cong, Y., Pawlisz, E., Porssa, M., Rumpf, N., Singh, R., Powell, K., and Brocchini, S. (2009) Rebridging disulphides: site-specific PEGylation by sequential bis-alkylation, In *PEGylated Protein Drugs: Basic Science and Clinical Applications* (Veronese, F., Ed.), pp 47-74, Birkhäuser.

2. LYSOZYME (LYS)

1. Preparation of a three disulphide derivative of lysozyme ($LYS^{CM6,127}$)

Lysozyme (LYS) is an enzyme belonging to the family of glycoside hydrolases (EC. 3.2.1.17). It damages the bacterial cells by the hydrolysis of 1, 4-beta-linkages. It is abundantly present in secretions, such as tears, saliva, milk and mucus. LYS is also present in significant amount in the hen egg white. LYS has been widely used in structural studies for many years due to the easy conditions of purification and crystallization. The X-ray structure of LYS was solved in 1965 with a resolution of 2 Å (*1*) . LYS is constituted by 129 amino acid residues with an overall molecular weight of 14305.14 Da (*2*) and has an alpha/beta structure (*3*) (Figures 2.1 and 2.14). The alpha domain consists of four helices, whereas the beta domain consists of a triple stranded anti parallel beta-sheet and a small double stranded anti-parallel beta sheet near the N-terminal (*3*) (Figure 2.14). The LYS structure is held intact by four disulphide bridges located at the positions Cys6-Cys127, Cys30-Cys115, Cys76-Cys94 and Cys64-Cys80 (Figure 2.1 and 2.14). The B-factor profile of LYS indicates a very rigid conformation (Figure 2.2, bottom). This rigidity is likely to render LYS highly resistant to proteases (*4*), as well as to TGase-mediated conjugation of lysine residues (*5*).

Radford and co-authors (*6*) have reported the preparation of a three disulphide derivative of LYS by reduction and carboxymethylation of Cys6 and Cys127, which was denoted as $LYS^{CM6,127}$. The conformational features of $LYS^{CM6,127}$ have been analyzed by NMR and were proved to be very similar to those of native LYS, with differences located only near the site(s) of carboxymethylation (Figure 2.2, top). These chemical shift movements occur in both upside and downside directions, but do not correspond to typical random coil chemical shifts.

Herewith, an attempt to introduce a moderate disorder/flexibility in LYS was performed by the reduction and alkylation of the disulphide bridge Cys6-Cys127 of LYS. Indeed, a specific aim was to verify if such induced flexibility in the protein could favor TGase-mediated conjugation and proteolysis reactions.

$LYS^{CM6,\ 127}$ was prepared by a modification of the method of Radford *et al. (6)*. Selective reduction and carboxymethylation (CM) of cysteine residues was

performed in a single step, and the time-dependent formation of $LYS^{CM6,127}$ was studied under different experimental conditions to optimize the method of preparation (see Methods 5.1 for details). The ion exchange elution profile relative to the purification of $LYS^{CM6,127}$ after 30, 60 and 90 min of reaction is reported in Figure 2.3. The eluted fractions were manually collected and analyzed by mass spectrometry. The results show that the first eluting chromatographic peak at all reaction times (RT 30 min) corresponds to a three disulphide derivative (LYS^{2CM}), followed by a peak (LYS^{1CM}) (RT 30.8 min) containing a LYS derivative with only one CM group and a free cysteine residue (Figure 2.3 and Table 2.1). The last eluting peak corresponds instead to native lysozyme (LYS) (RT 40.9 min).

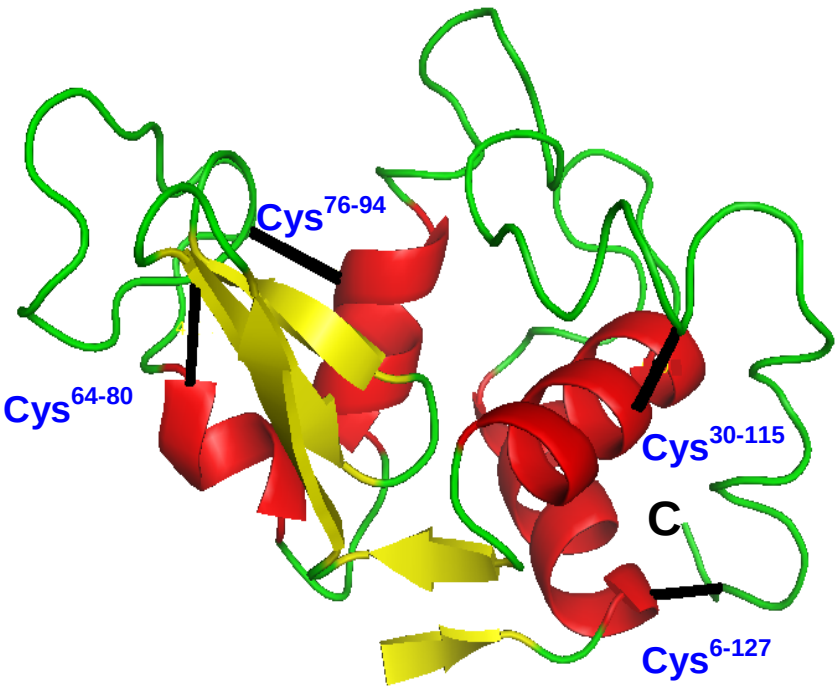


Figure 2.1. Schematic model of LYS (PDB code 2VB1) prepared from the X-ray structure of the protein using the PyMol molecular graphics system (www.pymol.org) (*7*).

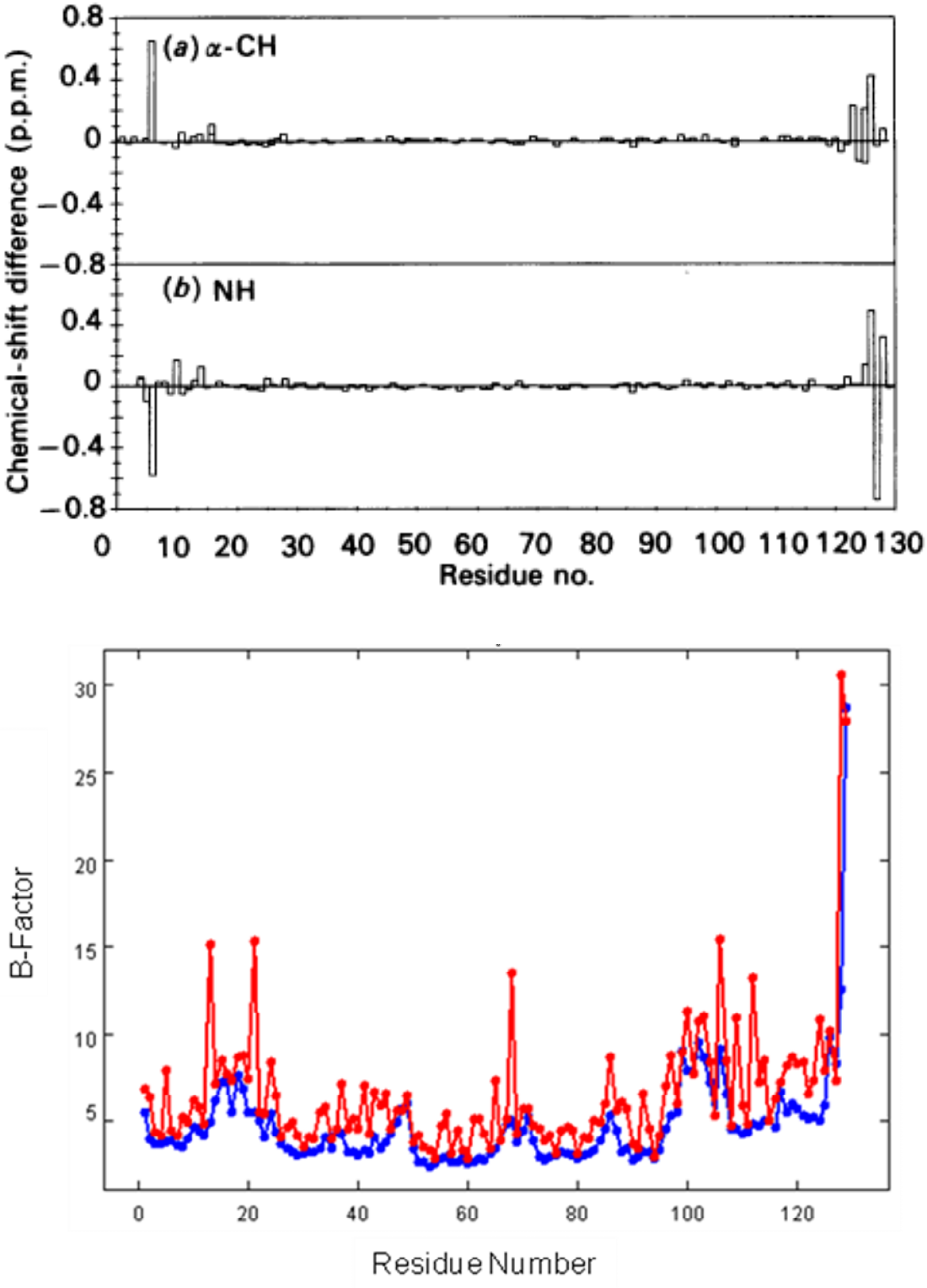


Figure 2.2. (Top) Plot of the chemical-shift differences between the amino acid sequence of native LYS and that of $LYS^{CM6,127}$ [adapted from (*6*)]. (Bottom) Plot of the *B*-Factor values along the polypeptide chain of LYS (129 residues) with intact disulphide bridges (PDB code 2VB1).

2. Characterization of $LYS^{CM6,127}$

In order to characterize the site of carboxymethylation (CM) of $LYS^{CM6,127}$, the fractions isolated by ion exchange chromatography were analyzed by protein fingerprinting coupled with mass spectrometry analysis of the obtained peptide fragments. Tryptic peptides 6–13 and 126–129 containing Cys6 and Cys127 showed a mass increase of 58 Da, which corresponds to the addition of a CM group (Table 2.2). On the contrary, the remaining cysteines were not modified. Hence, only a single three disulphide species was generated with consistent reproducibility, leading to the $LYS^{CM6,127}$ derivative.

Circular dichroism (CD) measurements have been carried out to unravel the differences in the conformational features of LYS and $LYS^{CM6,127}$. The far- and near-UV CD spectra of $LYS^{CM6,127}$ registered in 10 mM phosphate buffer, pH 7.0 containing 50 mM NaCl, resemble closely those of native LYS, both in terms of shape and intensity, suggesting a native-like secondary structure for $lysozyme^{CM6,127}$ (Figure 2.4). The thermal denaturation process of LYS and $LYS^{CM\ 6,\ 127}$ was also followed at 222 nm (Figure 2.5). The results of the thermal melting point of LYS (74.7 °C) and $LYS^{CM6,\ 127}$ (49.5 °C) are in good agreement with those already reported (*6*). Therefore, $LYS^{CM6,127}$ is dramatically destabilized if compared to LYS. The transition for both LYS and $LYS^{CM6,\ 127}$ is a highly co-operative and reversible event at pH 7.0 (95 and 92 %, respectively).

3. TGase-mediated conjugation of $LYS^{CM6,127}$ with dansyl cadaverine

TGase-mediated conjugation studies were performed on $LYS^{CM6,127}$, in order to verify if a selective conjugation at the level of glutamine residues occurred. Dansyl cadaverine (DC) was used as the acyl acceptor. The RP-HPLC elution profile (Figure 2.6) shows the presence of two main peaks corresponding to non-conjugated $LYS^{CM6,\ 127}$ (RT 21.4 min) and a $LYS^{CM6,127}$–DC conjugate (RT 22.7). The identity of the peaks was confirmed by the mass spectrometry spectra, which showed an observed mass of 14741.75 Da for conjugated $LYS^{CM6,127}$, in accord with the mass of LYS with a mass increment corresponding to one dansyl cadaverine moiety (14741.63 Da).

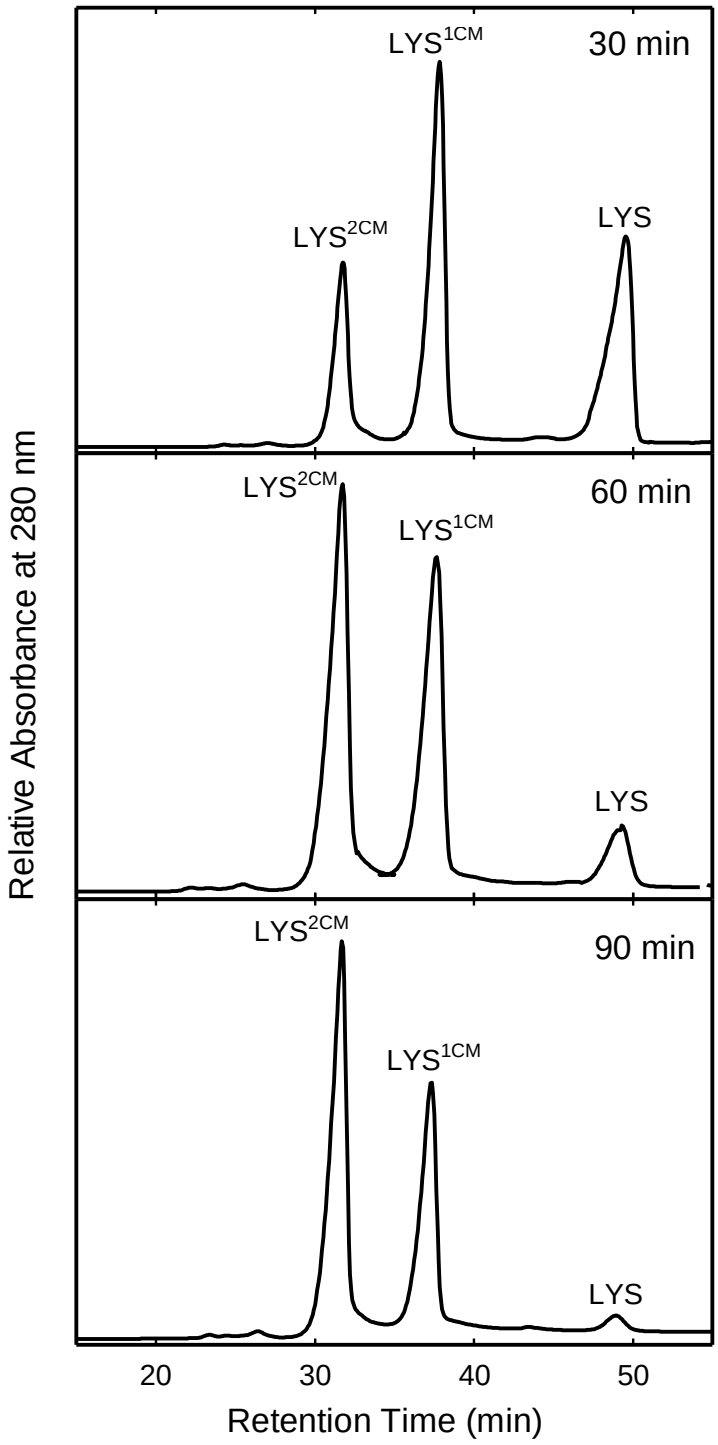


Figure 2.3. Isolation of $LYS^{CM6,127}$ (LYS^{2CM}) by ion-exchange chromatography. The chromatograms corresponding to 30, 60 and 90 min of the reduction reaction of LYS with TCEP and iodoacetic acid are reported (see Methods 5.1 for details). LYS species with one or two reduced and carboximetylated (CM) cysteines have been identified by mass spectrometry (see Table 2.1) and are indicated as LYS^{1CM} and LYS^{2CM}, respectively, on the top of the chromatographic peaks.

Table 2.1. Molecular masses of the protein species eluted in the ion-exchange chromatography fractions relative to the reduction reaction of LYS with TCEP and iodoacetic acid (see Figure 2.3), as determined by mass spectrometry.

Identity	Molecular Mass (Da)	
	Observed	*Calculated*
Lysozyme	14305.77	14305.14
Lysozyme^{1CM}	14365.86	14365.14
Lysozyme^{2CM}	14423.65	14423.12

Table 2.2. Identity of the peptides derived by the fingerprinting of LYS^{CM6,127} with trypsin, as determined by mass spectrometry analysis.

Peptide Identity	Amino acid sequence	Molecular Mass (Da)	
		Observed	*Calculated*
6–13	**C^{CM}**ELAAAMK	893.40	893.40
22–33	GYSLGNWV**C**AAK	1267.62	1267.60
62–68	WW**C**NDGR	935.39	935.37
74–96	NL**C**NIP**C**SALLSSDITASVN**C**AK	2336.23	2336.12
115-125	**C**KGTDVQAWIR	1275.76	1275.64
126–129	G**C^{CM}**RL	505.25	505.23

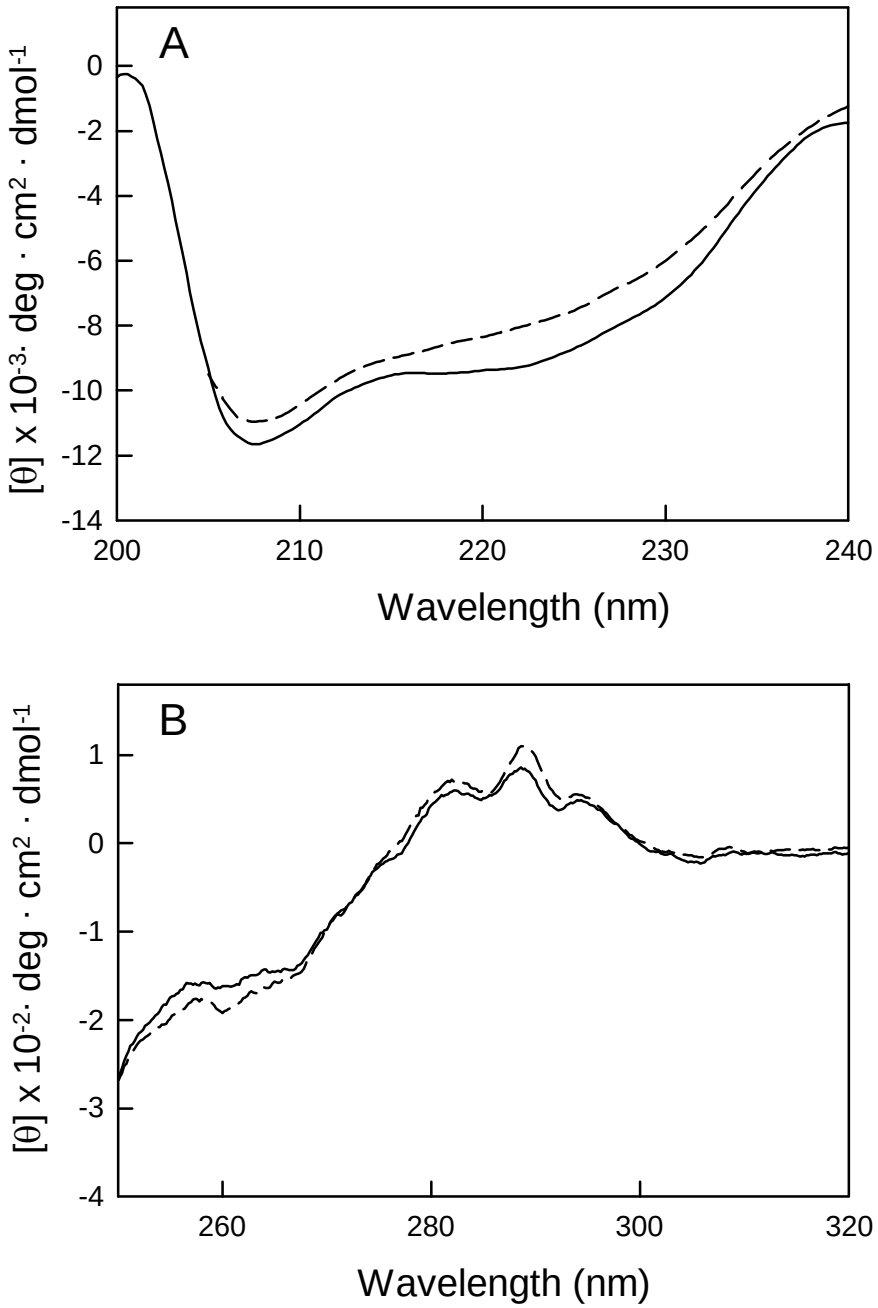


Figure 2.4. Far-UV (A) and near-UV (B) CD spectra of lysozyme (LYS) (continuous line) and its three-disulphide derivative $LYS^{CM6,127}$ (dashed line) in 10 mM phosphate buffer, pH 7.0, containing 50 mM NaCl. The spectra were recorded at 8^{o} C and subtracted for the buffer baseline (see Methods 5.2 for details).

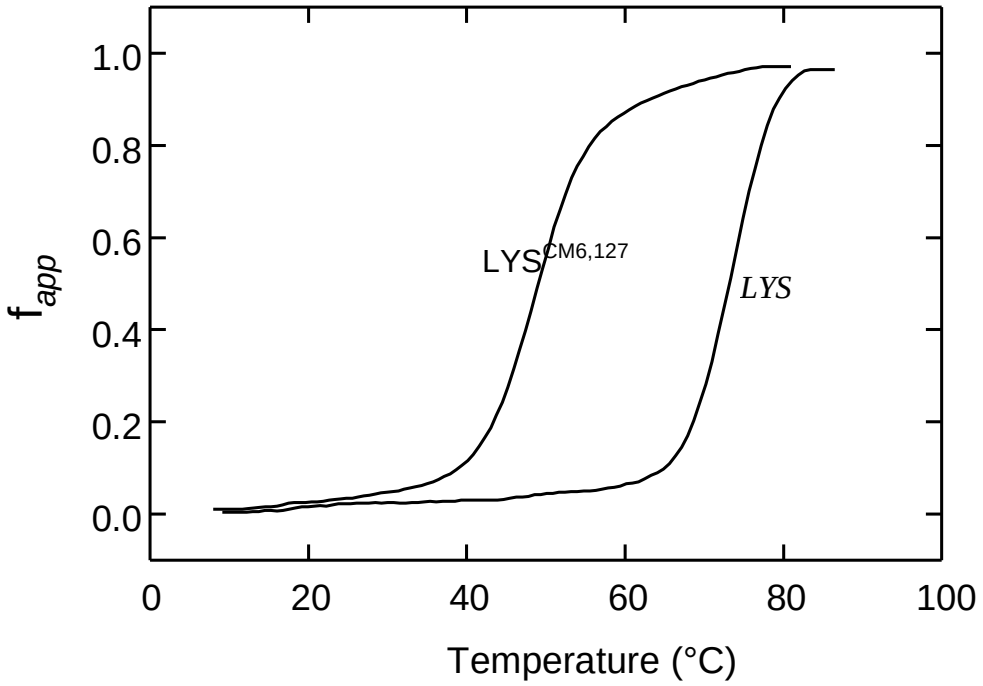


Figure 2.5. Temperature dependence of the mean residue ellipticity at 222 nm of native lysozyme (LYS) and $LYS^{CM6,127}$. The melting profile of the proteins was measured in 10 mM phosphate buffer, pH 7.0, containing 50 mM NaCl (see Methods 5.2 for details). f*app* was calculated using the formulae f*app* = ($[\theta]_U$ - [θ]) / ($[\theta]_U$ - $[\theta]_N$), where [θ] = ellipticity, $[\theta]_U$ = ellipticity of unfolded protein, $[\theta]_N$ = ellipticity of the protein in its native state. The T_m for native lysozyme was calculated to be 74.7 ^{o} C with ~ 95 % reversibility of the transition, whereas the T_m for $LYS^{CM6,127}$ was observed to be 49.5^o C, with ~ 92 % of reversibility of the transition.

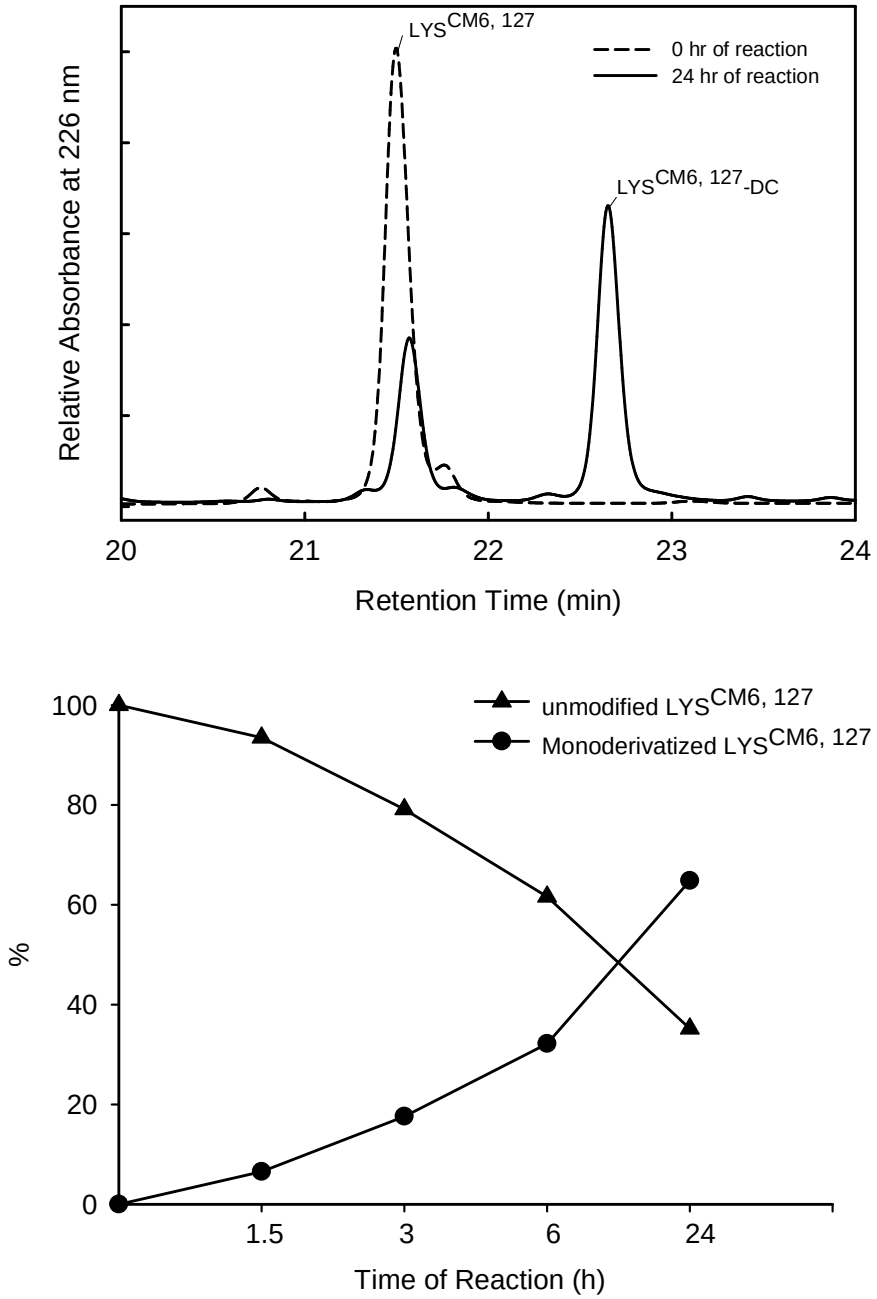


Figure 2.6. TGase-mediated conjugation of $LYS^{CM6,127}$ with dansylcadaverine (DC). (Top) RP-HPLC analysis of the reaction with DC after 0 (dashed line) and 24 h (continuous line). The separation was performed with a linear gradient of acetonitrile containing 0.085 % TFA on a C18 column (see Methods 5.3 for details). (Bottom) Time dependent kinetics for the formation of the $LYS^{CM6,127}$–DC conjugate.

The time dependent kinetics for the formation of the $LYS^{CM6,127}$–DC conjugate is shown in Figure 2.6 (bottom), indicating a gradual decrease of the quantity of $LYS^{CM6,127}$ with time along with the increase of the quantity of the $LYS^{CM6,127}$–DC derivative. A final ~70 % yield of formation of mono-conjugated derivative was observed after 24 h under the described conditions.The protein fingerprinting technique associated with mass spectrometry was employed for characterizing the TGase conjugation site. A tryptic peptide corresponding to the sequence 117–125 of $LYS^{CM6,127}$ has a mass value difference corresponding to one dansyl cadaverine molecule (Figure 2.7). MS/MS analysis of the tryptic peptide $LYS^{CM6,\ 127}$(117-125) indicated that the conjugation is at the level of Gln121 (Figure 2.7). Therefore, out of three glutamines present in the sequence of $LYS^{CM6,\ 127}$ (Gln41, Gln57 and Gln121), only one (Gln121) was conjugated with dansyl cadaverine by TGase.

4. TGase-mediated conjugation of $LYS^{CM6,\ 127}$ with CBZ-Gln-Gly

TGase-mediated conjugation studies were performed on $lysozyme^{CM6,\ 127}$ with CBZ-Gln-Gly (ZQG), to elucidate the selectivity of the reaction at the level of lysines. The RP-HPLC elution profile (Figure 2.8) indicates the presence of more than one isoforms of the conjugated $LYS^{CM6,\ 127}$-ZQG derivative. The reaction was monitored at 0, 6 and 24 hrs of time. After 6 hr of reaction, a few chromatographic peaks appear at higher retention times if compared to the peak corresponding to the intact protein. These RP-HPLC fractions were analyzed by mass spectrometry (MS) for the identification of the number of CBZ-Gln-Gly molecules attached to $LYS^{CM6,\ 127}$. Further characterization of the site(s) of conjugation was performed by digestion of $LYS^{CM6,\ 127}$-ZQG conjugates with trypsin followed by MS and MS/MS analysis of the modified peptides (Table 2.3). The MS analysis revealed that the conjugation occurred predominantly at Lys13 and Lys116 after 6 hr of reaction, leading to LYS^{ZQG13} and LYS^{ZQG116} derivatives. As the reaction proceeds, the formation of di-, tri- and tetra- conjugated derivatives was observed with conjugation site(s) identified also at Lys33 and Lys97 (Figures 2.9-2.12).

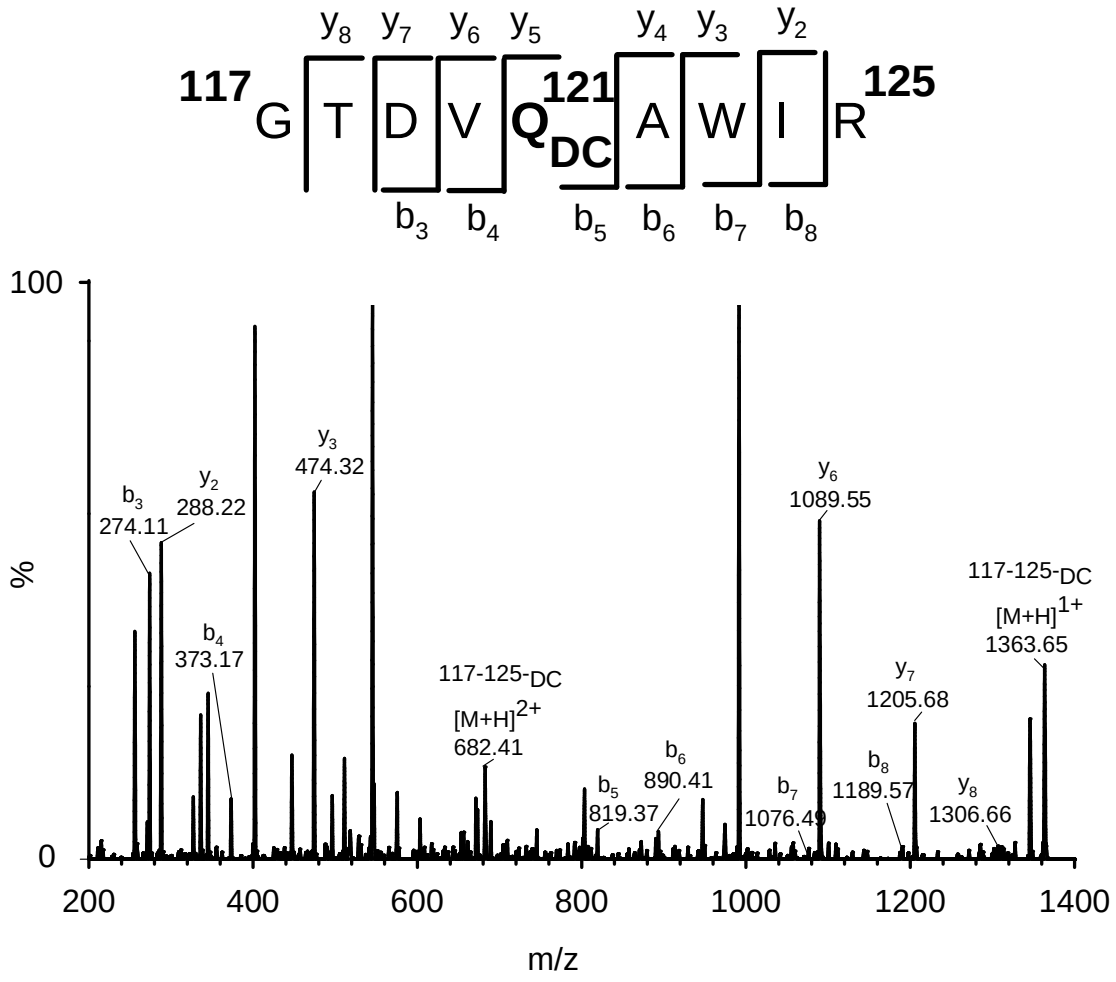


Figure 2.7. Characterization of the dansyl cadaverine (DC) conjugation site on $LYS^{CM6,127}$. (Top) DC conjugated Gln121 residue is shown in bold as Q_{DC} in the amino acid sequence of the peptide 117-125. (Bottom) Electrospray MS/MS spectrum of the ion at 682.41 m/z of peptide 117–125 of $LYS^{CM\ 6,\ 127}$-DC, indicating the assigned ions for fragments of the series b and y.

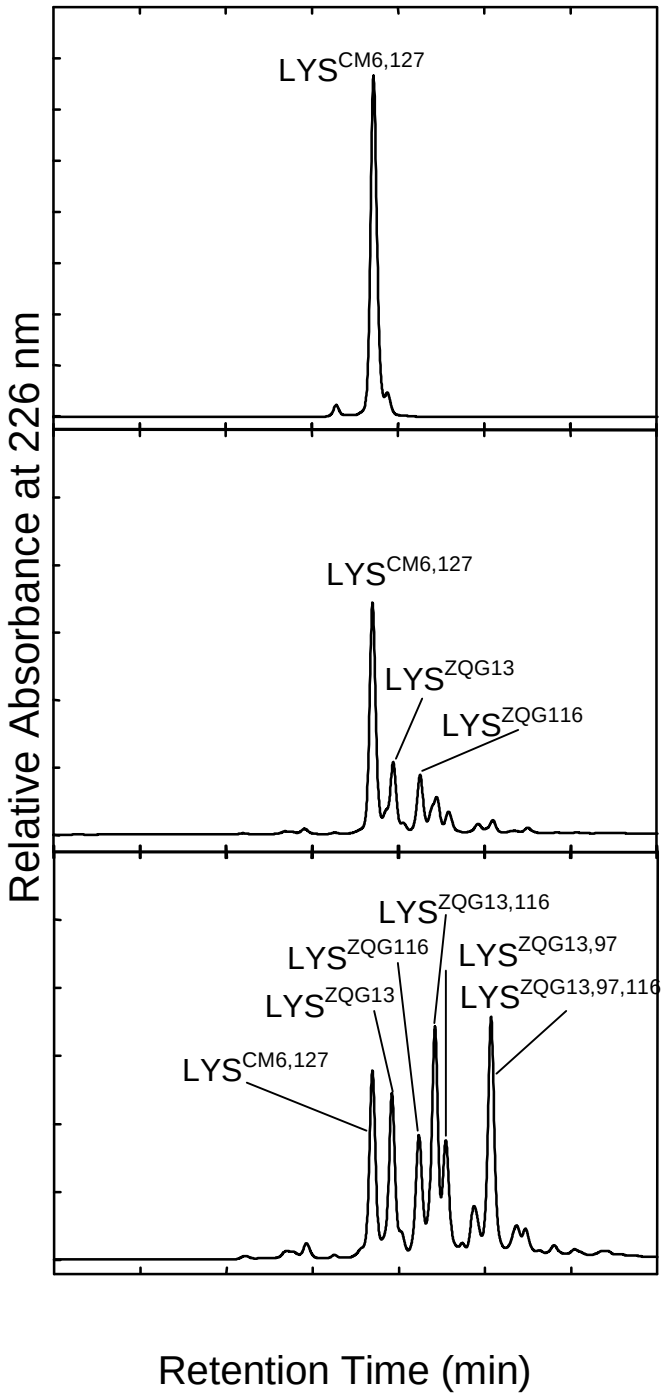


Figure 2.8. RP-HPLC analysis of the Tgase-mediated conjugation of $LYS^{CM6,127}$ with CBZ-Gln-Gly (ZQG). The separation was performed with a linear gradient of acetonitrile containing 0.085 % TFA on a C18 column (see Methods 5.4 for details). The chromatograms corresponding to 0, 6 and 24 h of reaction are shown. The sites of modification of $LYS^{CM6,127}$ at lysine residues with ZQG have been determined by mass spectrometry (see Table 2.3) and are indicated on the corresponding chromatographic peaks.

Table 2.3. Molecular masses of the protein species eluted in the chromatographic peaks relative the TGase mediated conjugation of $LYS^{CM6,127}$ with ZQG, determined by mass spectrometry (Figure 2.8) and of the peptides derived by trypsin digestion of the protein species eluted in each chromatographic fraction, containing the site of ZQG conjugation (see also Figures 2.9-12).

RP-HPLC Retention Time (min)	Protein/Peptide Identity	Conjugation site	Molecular mass (Da)	
			Observed	Calculated
23.5	$LYS^{CM6,127}$	–	14423.44	14423.23
24.0	LYS^{ZQG13} $6–14^{ZQG}$	K_{13}	14743.25 1369.50	14743.23 1369.50
24.4	LYS^{ZQG116} $115–125^{ZQG}$	K_{116}	14743.25 1595.76	14743.23 1595.64
24.8	$LYS^{ZQG13,116}$	K_{13}, K_{116}	15063.31	15063.23
25.0	$LYS^{ZQG13,97}$ $97–112^{ZQG}$	K_{13}, K_{97}	15062.67 2123.10	15063.23 2122.89
25.8	$LYS^{ZQG13,33,97,116}$ $22–45^{ZQG}$	K_{13}, K_{33}, K_{97}, K_{116}	15703.10 2997.52	15703.23 2998.93
26.2	$LYS^{ZQG13,97,116}$	K_{13}, K_{97}, K_{116}	15383.27	15383.23

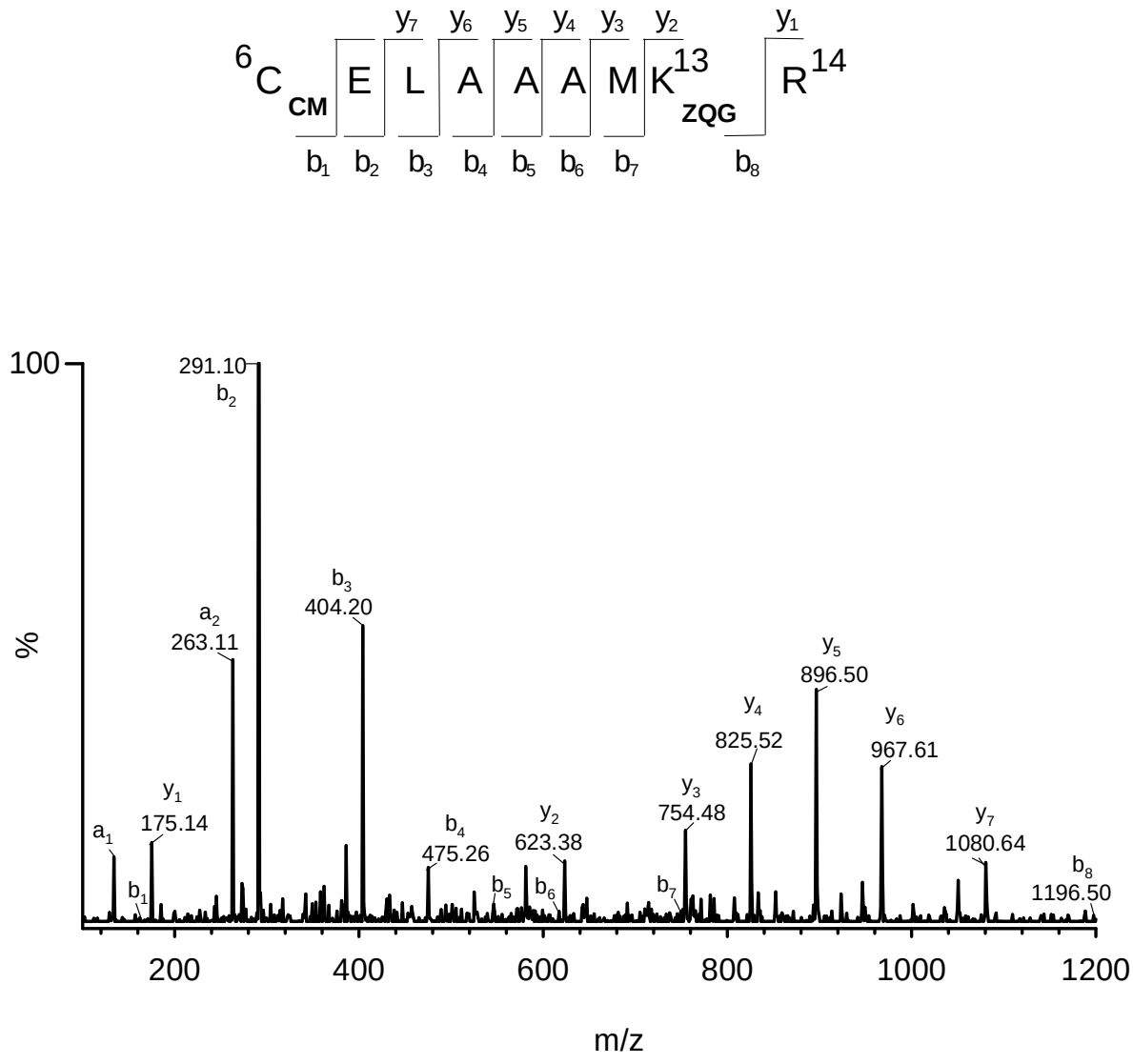


Figure 2.9. Electrospray MS/MS mass spectrum of the ion at 685.84 m/z of peptide 6-14 derived by trypsin digestion of LYS$^{CM6,127-}$ZQG-13 (see Figure 2.8 and Table 2.3). (Top) Fragments of the series *b* and *y* that were identified in the MS/MS spectrum are indicated on the sequence of the peptide. Modified Lys13 is indicated as K_{ZQG}, whereas carboxymethylated Cys6 is indicated as $C_{CM.}$ (Bottom) In the MS/MS spectrum, ions assigned to fragments of the series *a, b* and *y* are indicated.

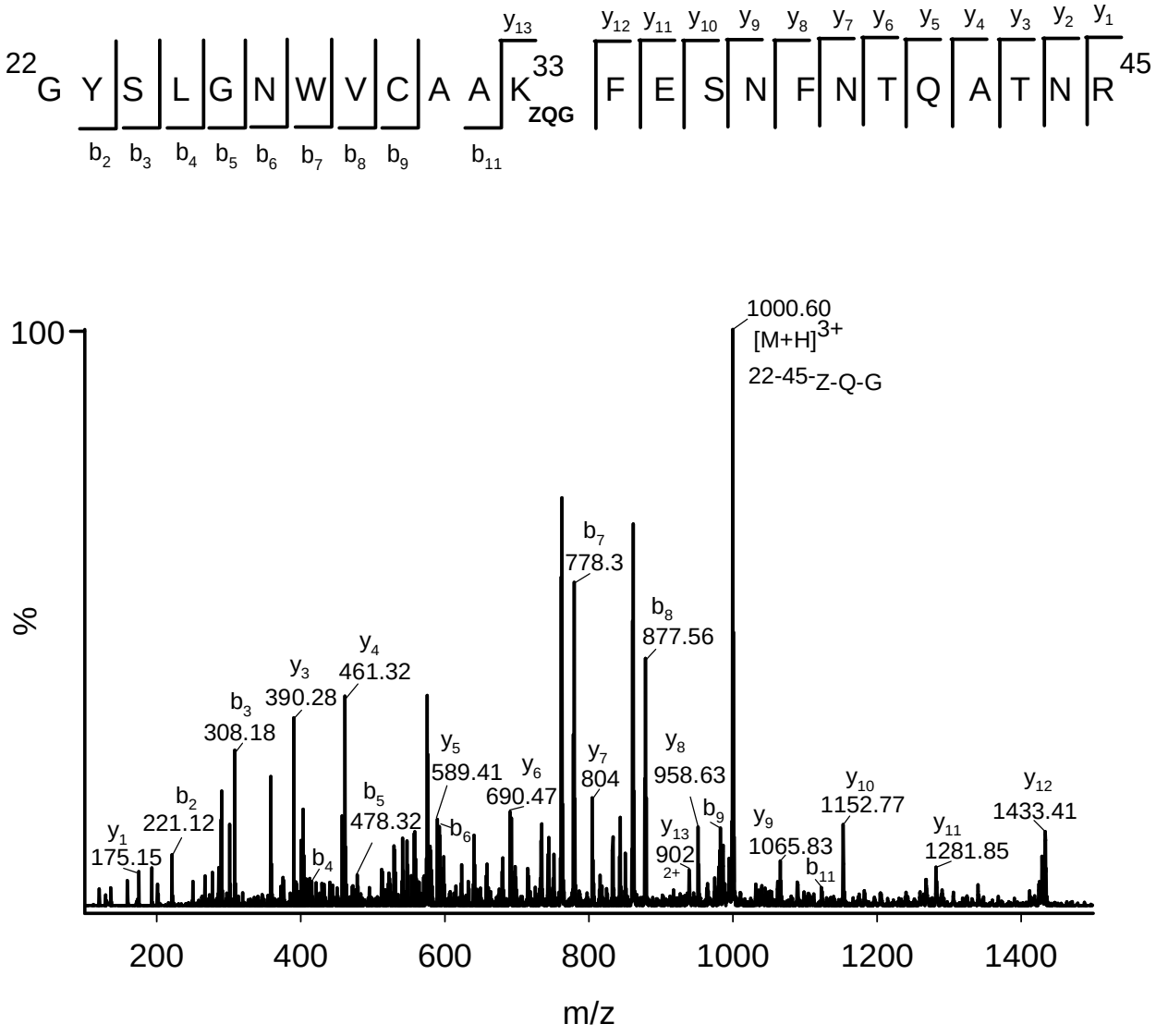


Figure 2.10. Electrospray MS/MS mass spectrum of the ion at 1000.60 m/z of peptide 22-45 derived by trypsin digestion of LYS^{CM6,127}-ZQG-33 (see Figure 2.8 and Table 2.3). (Top) Fragments of the series *b* and *y* that were identified in the MS/MS spectrum are indicated on the sequence of the peptide. Modified Lys33 is indicated as K_{ZQG}. (Bottom) In the MS/MS spectrum, ions assigned to fragments of the series *b* and *y* are indicated.

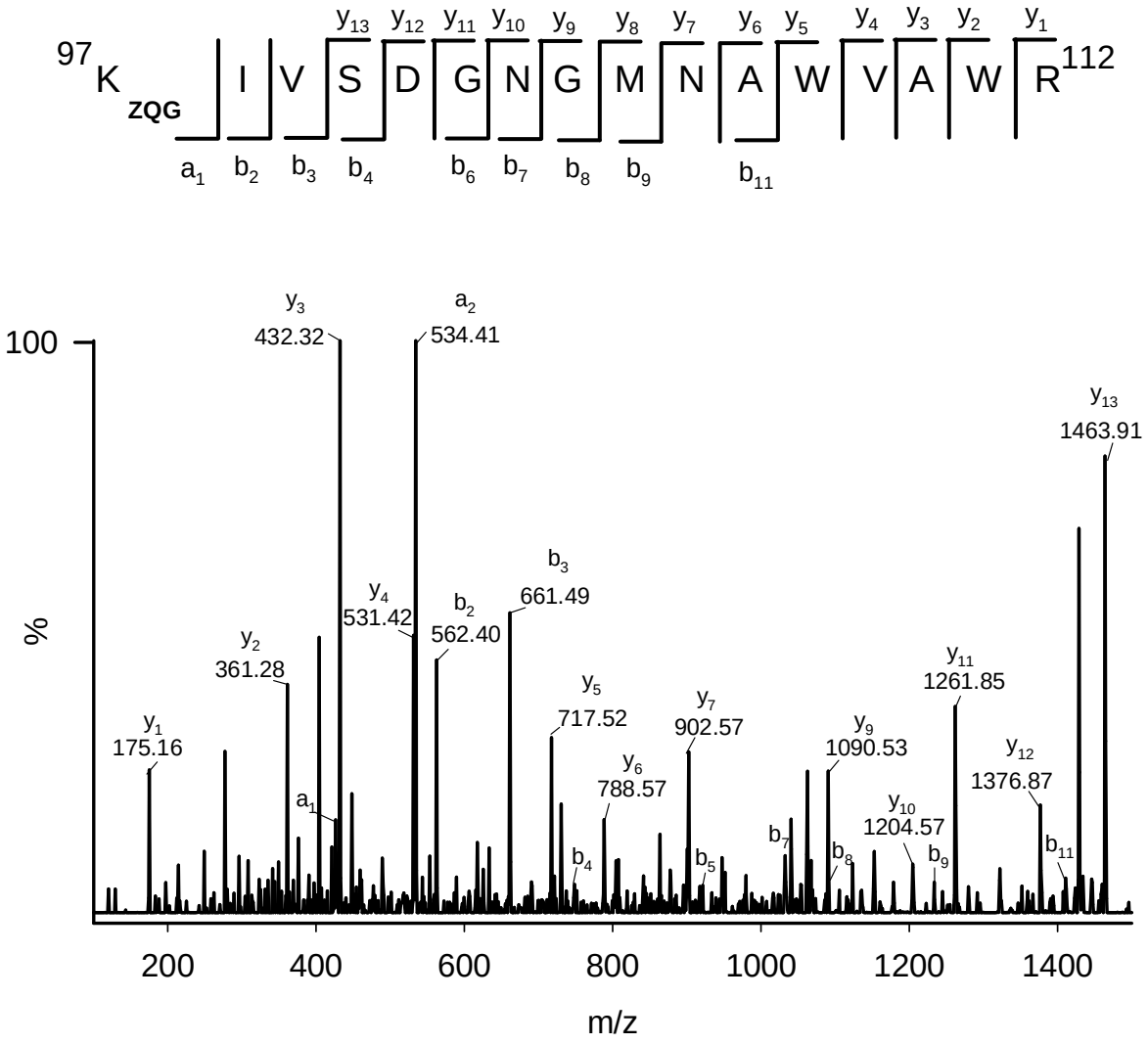


Figure 2.11. Electrospray MS/MS mass spectrum of the ion at 1062.72 m/z of peptide 97-112 derived by trypsin digestion of $LYS^{CM6,127}$-ZQG-97 (see Figure 2.8 and Table 2.3). (Top) Fragments of the series *a*, *b* and *y* that were identified in the MS/MS spectrum are indicated on the sequence of the peptide. Modified Lys97 is indicated as K_{ZQG}. (Bottom) In the MS/MS spectrum, ions assigned to fragments of the series *a*, *b* and *y* are indicated.

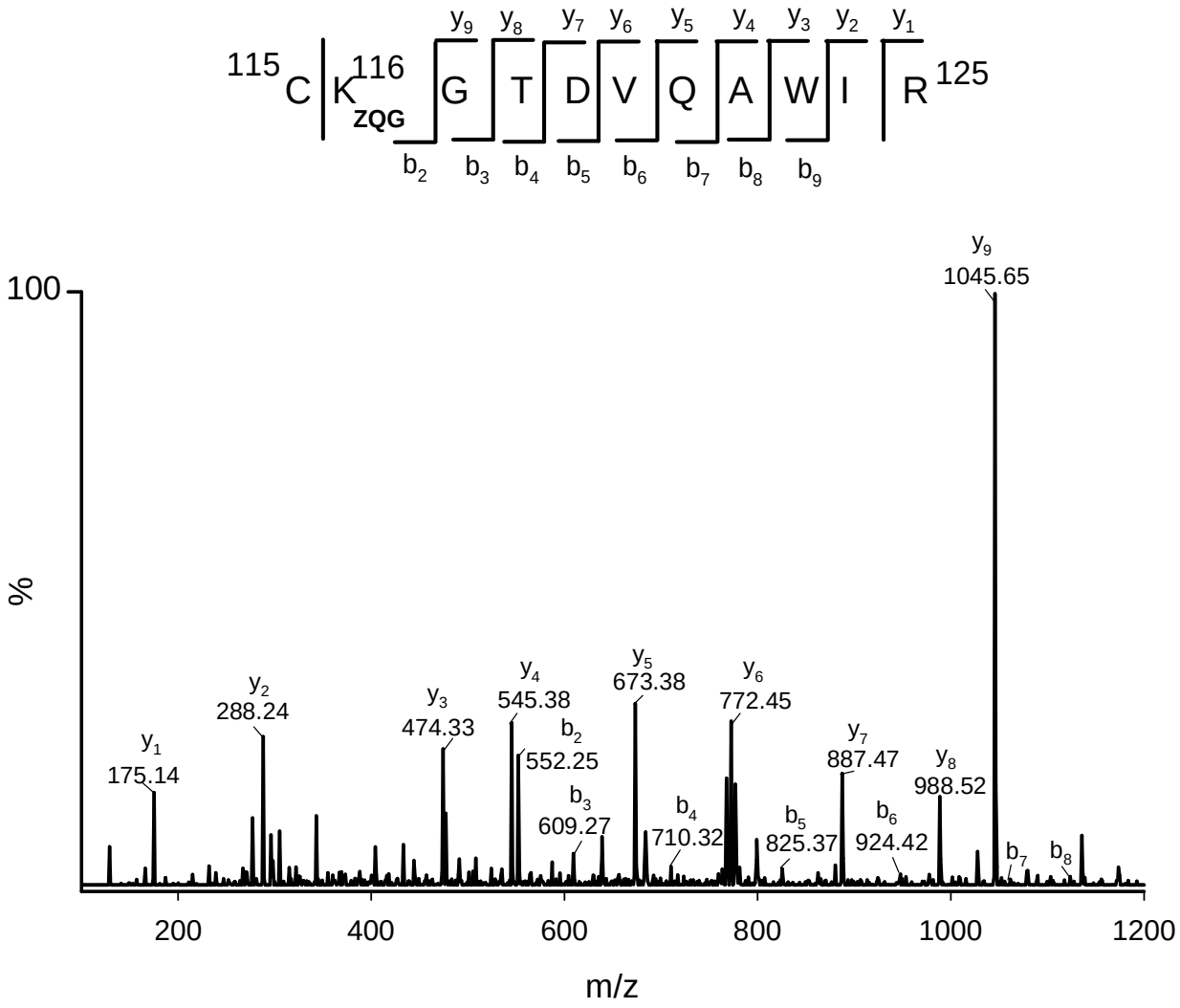


Figure 2.12. Electrospray MS/MS mass spectrum of the ion at 799.02 m/z of peptide 115-125 of LYS^{CM6,127}-ZQG-116 (see Figure 2.8 and Table 2.3). (Top) Fragments of the series *b* and *y* that were identified in the MS/MS spectrum are indicated on the sequence of the peptide. Modified Lys116 is indicated as K_{ZQG}. (Bottom) In the MS/MS spectrum, ions assigned to fragments of the series *b* and *y* are indicated.

5. Limited proteolysis of LYS$^{CM6, 127}$

In order to assess the susceptibility of LYS^{CM6,127} to proteolysis, a limited proteolysis experiment on lysozyme$^{CM6, 127}$ with chymotrysin was performed (Figure 2.13). The reaction was monitored by SDS-PAGE and RP-HPLC, with subsequent identification of the proteolytic peptides by MS.

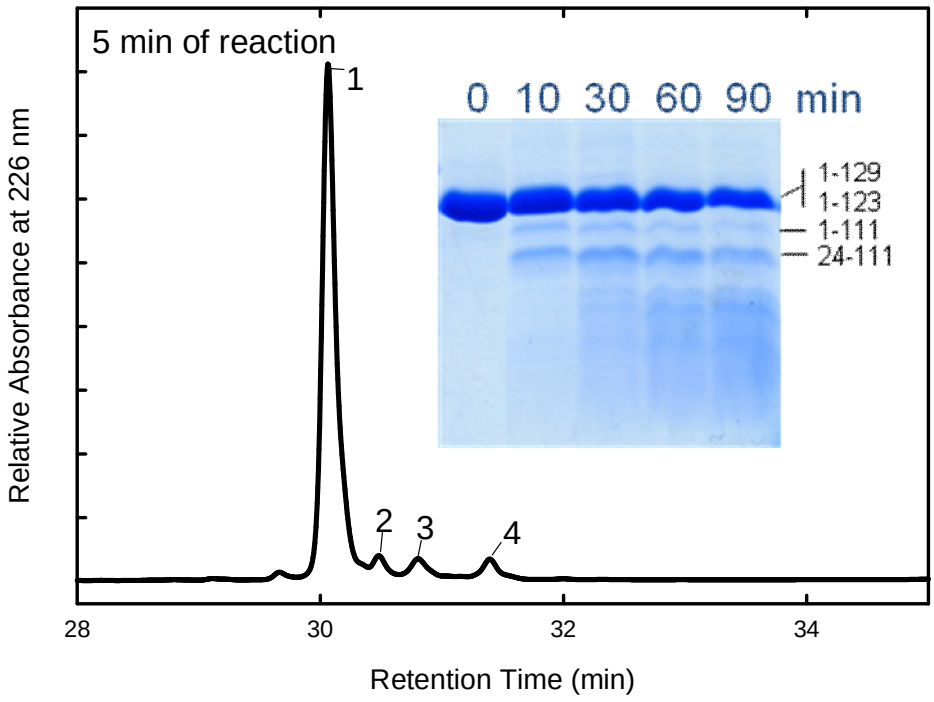


Figure 2.13. Limited proteolysis of LYS^{CM6,127} with chymotrypsin. RP-HPLC chromatogram relative to the 5-min limited proteolysis reaction of LYS^{CM6,127}. The separation was performed with a linear gradient of acetonitrile containing 0.085 % TFA (see Methods 5.5 for details). (Inset) SDS-PAGE analysis of the proteolysis mixture after 0, 10, 30, 60 and 90 min of reaction at 30^{o} C.

An aliquot from the 5-min proteolysis reaction was analysed by RP-HPLC, which shows the presence of a main peak with RT 30 min (Peak 1) and of three peaks with lower intensity at RT 30.5, 30.8 and 31.5 min (Peak 2, Peak 3 and Peak 4) (Figure 2.13). These peaks were analyzed by MS to determine their identities (Table 2.4). The results indicate that the main proteolysis cleavages occur at peptide bonds Tyr23–Ser24, Trp111–Arg112 and Trp123–Ile124, leading to the formation of fragment species 1–123, 1–111/112-123 and 24–111/112–123 (Figure 2.14). The last two fragment species are formed by the protein segments 1–111, 112–123 and 24–111 held together by the three disulfides of $LYS^{CM6,127}$ that remain intact during the course of chymotrypsin digestion. The SDS-PAGE gel shows that after 10 min of reaction the bands corresponding to fragments 1–111 and 24–111 start to appear (Figure 2.13, inset). On the contrary, the band corresponding to fragment 1–123 is not distinguishable from that of intact $LYS^{CM6,127}$. After 30 min of reaction a series of short fragments start to form, indicating a further degradation of the protein. Therefore, the core of the protein (segment 24–111) remains intact for a long time and then gets further degraded.

Table 2.4. Molecular masses and identity of the protein species derived by the proteolysis of $LYS^{CM6,127}$ with chymotrypsin (see Figure 2.13).

RP-HPLC Retention Time (min)	Identity[a]	Molecular mass (Da)	
		Observed	Calculated
30.0	$LYS^{CM6,127}$	14423.62	14423.12
	1–123	13667.00	13666.30
30.5	1–111/112–123	13684.00	13684.30
	24–123	10970.00	10970.19
30.8	24–108/112–123	10631.97	10630.92
31.5	24–111/112–123	10988.36	10988.19

[a]Some fragments are nicked species linked by the disulfide bonds of the protein. In cases where ambiguities existed, the nicked fragments have been reduced with tris(2-carboxyethyl)phosphine (TCEP) in order to cleave the disulfide bridges and the individual peptides were then purified by RP-HPLC and analyzed by ESI MS.

6. Comments

In this Thesis, LYS has been selectively reduced and alkilated to trap a species of LYS with three intact disulphide bridges and two carboxymethyl-(CM) cysteine residues (Cys6 and Cys127) derived by the breakage of Cys6-Cys127 disulfide bond. The disulphide bond located at the position Cys6-Cys127 forms during the end of folding pathway of lysozyme (*8*). Hence, during the unfolding this disulfide is likely to be the first to open at low concentration of reductants (*6*).

As observed from the circular dichroism (CD) studies reported in this Thesis, the loss of Cys6-Cys127 disulfide bridge has no overall effect on the three dimensional structure of LYS. Conversely, thermal denaturation profiles revealed that the thermal stability of $LYS^{CM6,\ 127}$ is dramatically decreased (ΔT_m=25 °C). Moreover, the activity of this derivative was observed to be 40 % lower if compared to that of native LYS (*9*) These observations suggest that the breakage of the Cys6-Cys127 disulfide bond of LYS leads to subtle changes in the physico-chemical interactions, but essential for the function of the protein. Interestingly, native LYS had shown to be resistant towards TGase-mediated conjugation and proteolysis reactions, whereas $LYS^{CM6,127}$ showed an increased reactivity. Out of three Gln present in the sequence, the conjugation of $LYS^{CM6,127}$ with dansyl cadaverine by TGase occurred only at Gln121 (see Figure 2.14). In the case of TGase-mediated conjugation of CBZ-Gln-Gly at the level of lysines, four residues have been conjugated (Lys13, Lys33, Lys98 and Lys116), indicating a lower specificity of TGase towards the lysines. This result could be explained with the fact that long ε-NH_2 side chains of lysines protrude outside and tend to be more flexible than the side chains of other amino acids or glutamines. Of interest, Lys1 is not TGase-reactive in native LYS as well as in $LYS^{CM6,\ 127}$, possibly because of the presence of a conserved small anti-parallel beta sheet involving residues 1-5 and 38-40 (see Figure 2.1). Finally, limited proteolysis experiments of $LYS^{CM6,\ 127}$ conducted with chymotrypsin show that the initial sites of chymotrypsin attack occur at Tyr23, Trp111 and Trp123.

In conclusion, the experiments conducted on LYS in this Thesis show that TGase-mediated conjugation and proteolysis events occur only if the structure of the protein is rendered more flexible by the reduction of a single disulfide bond. A very

interesting correlation between the TGase and the limited proteolysis sites is also evident, as the main sites are located at the N-and C-termini of the protein (Tyr23, Trp111, Gln121 and Trp123). Actually, NMR data showed that slight chemical shift differences between LYS and $LYS^{CM6,127}$ are located only in proximity of the N- and C-termini of the protein (*6*). Therefore, these data suggest that the main determinant of the specificity of TGase, as well as that of proteases (*10*) could be located in some local unfolding or flexibility of the protein chain.

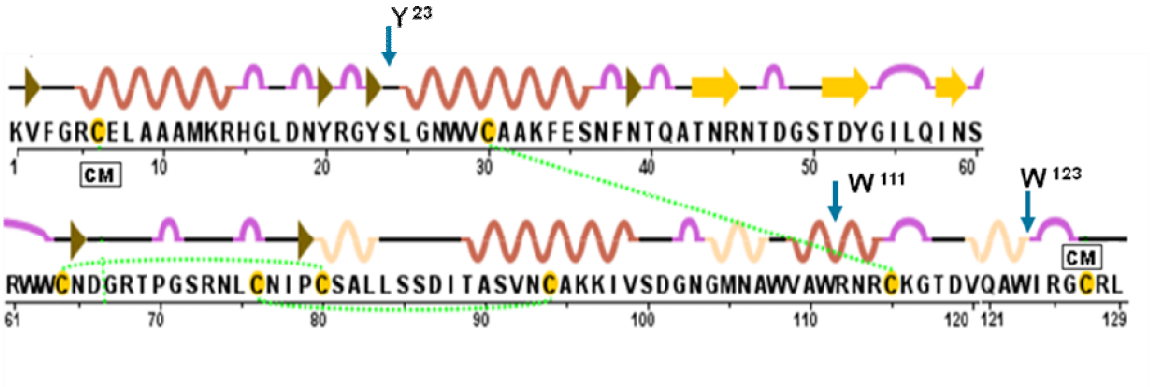


Figure 2.14. Scheme of the amino acid sequence and of the secondary structure of $LYS^{CM6,\ 127}$ (*1-2*). Helical segments and beta-sheet regions along the protein chain are indicated by squiggled red lines and yellow arrows above the amino acid sequence, respectively. Carboxymethylated cysteines at the position Cys6 and Cys127 are indicated with the label CM. The main sites of limited proteolysis by chymotrypsin are indicated by arrows at Y23, W111 and W123.

7. References

1. Blake, C.C., Koenig, D.F., Mair, G.A., North, A.C., Phillips, D.C., and Sarma, V.R. (1965) Structure of hen egg-white lysozyme. A three-dimensional Fourier synthesis at 2 Angstrom resolution, *Nature 206*, 757–761.

2. Canfield, R.E. (1963) The amino acid sequence of egg white lysozyme, *J. Biol. Chem. 238*, 2698–2707.

3. Radford, S.E., Dobson, C.M., and Evans, P.A. (1992) The folding of hen lysozyme involves partially structured intermediates and multiple pathways., *Nature 358*, 302–307.

4. Polverino de Laureto, P., Frare, E., Gottardo, R., Van Dael, H., and Fontana, A. (2002) Partly folded states of members of the lysozyme/lactalbumin superfamily: a comparative study by circular dichroism spectroscopy and limited proteolysis, *Protein Sci. 11*, 2932–2946.

5. 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.

6. Radford, S.E., Woolfson, D.N., Martin, S.R., Lowe, G., and Dobson, C.M. (1991) A three-disulphide derivative of hen lysozyme. Structure, dynamics and stability, *Biochem. J. 273(Pt 1)*, 211–217.

7. DeLano, W.L. (2002) The PyMol Molecular Graphics System, DeLano Scientific, San Carlos, CA.

8. Anderson, W.L., and Wetlaufer, D.B. (1976) The folding pathway of reduced lysozyme, *J. Biol. Chem. 251*, 3147–3153.

9. Denton, M.E., and Scheraga, H.A. (1991) Spectroscopic, immunochemical, and thermodynamic properties of carboxymethyl(Cys6, Cys127)-hen egg white lysozyme, *J. Protein Chem. 10*, 213–232.

10. 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.

3. RIBONUCLEASE A (RNase)

Bovine pancreatic ribonuclease A (RNase) has been widely studied as a model protein to gain insights into the structural characteristics of intermediates in the protein folding process (*1-2*) as well as to analyze enzyme structure-function relationships (*3*). RNase consists of 124 amino acids residues with the overall molecular weight of 13682.30 Daltons. Structurally, RNase has an α/β structure with three α-helices located at the N-terminus of the protein while the C-terminal part consists of β-hairpins arranged in two β sheets (Figure 3.1). The structure of RNase is also stabilized by four disulphide bridges formed between Cys26–Cys84, Cys58–110, Cys40–95 and Cys65–Cys72.

The B-factor profile obtained from the X-ray structure of RNase (Figure 3.2.) shows that the protein has a compact structure with an enhanced flexibility at the level of the N-terminal region. Interestingly, TGase conjugation studies using TGase 2 on the selectivity of the derivatization at the level of lysine residues have reported that only Lys1 of RNase is prone to the TGase mediated conjugation (*4*). Limited proteolysis experiments conducted in our laboratory on RNase demonstrated that hydrolysis with thermolysin in the presence of 50 % trifluoethanol (TFE) produces a single nicked species (RNase Th1) (*5*), which is structurally similar to the native protein even if with a slightly distorted tertiary structure. Therefore, in this Thesis TGase mediated conjugation studies were performed on RNase and on RNase Th1 in order to understand how the structural flexibility of the protein substrate affects the specificity of the TGase reaction.

1. Preparation of RNase Th1

RNase Th1 was prepared using a modified version of the method reported by Polverino de Laureto, *et al.* (*5*). The proteolysis reaction of RNase with thermolysin was conducted in the presence of 50 % TFE and it led to the production of a RNase species selectively hydrolysed at the level of the Asn34–Leu35 peptide bond. The time course of the reaction was monitored at 0, 30, 90 and 180 minutes by FPLC using a cation exchange column (Figure 3.3). It can be observed from the

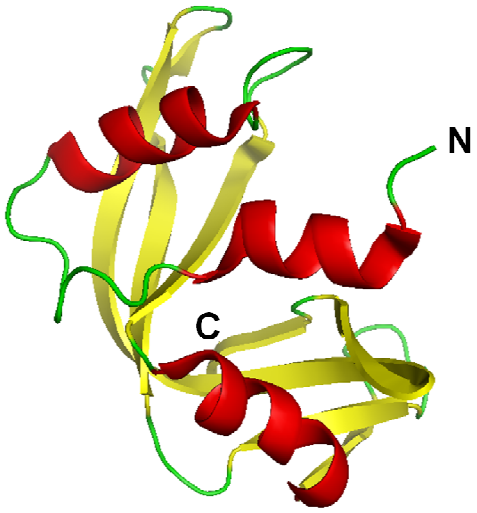


Figure 3.1. 3D structure of bovine ribonuclease A (RNase). The model was constructed from the X-ray structure of RNase (PDB file 7RSA) using the program PlyMOL. In the model, α-helices are colored in red, and β-sheets are coloured in yellow.

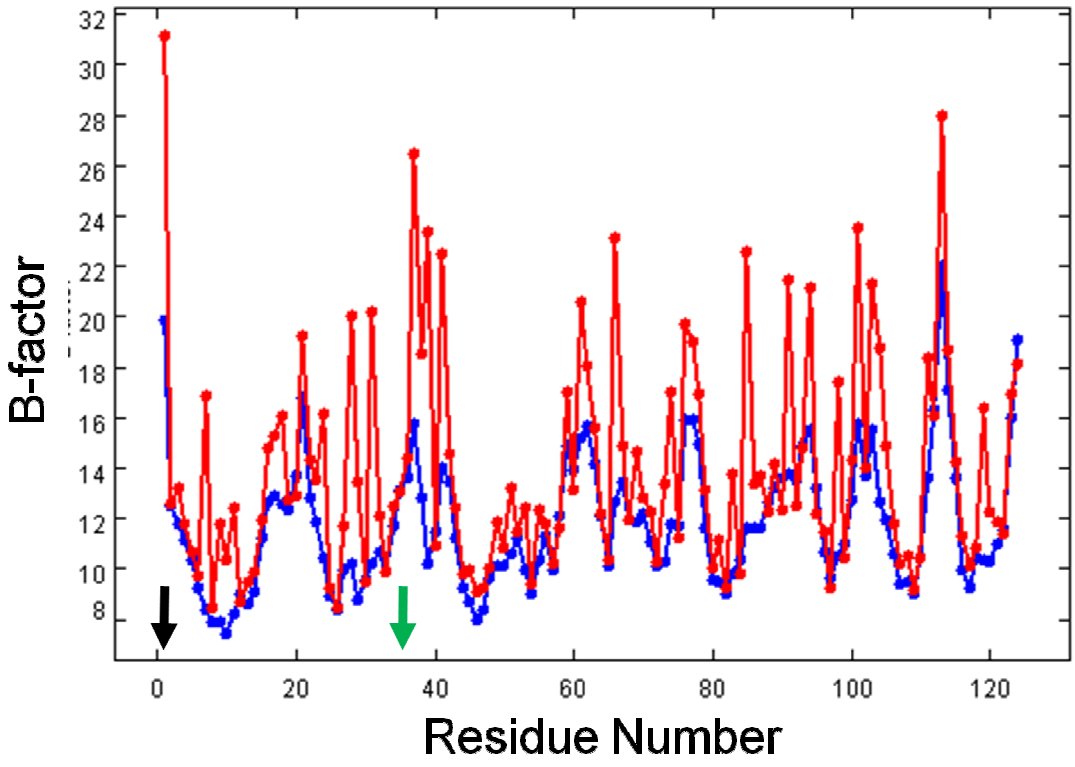


Figure 3.2. B-factor profile of native RNase as determined from the X-ray structure of the protein (PDB file 7RSA). The TGase conjugation site at the level of Lys1 (*4*) is indicated with a black arrow, whereas the thermolysin cleavage site at peptide bond Asn34–Leu35 (*5*) is indicated with a green arrow.

chromatographic profiles that the proteolysis is highly selective leading to the production of a main product, which was identified to be the nicked form of RNase (RNase Th1) on the basis of the MS analysis (Table 3.1). The maximum yield of this nicked form was obtained after 180 min of incubation and this time of the reaction was chosen to perform a quantitative purification of RNase Th1 by cation exchange chromatography (Figure 3.3).

Table 3.1: Molecular masses measured by MS of the protein material eluted in the ion exchange chromatography fractions (Figure 3.3).

Identity	Molecular mass (Da)	
	Observed	*Calculated*
1-124 (RNase)	13681.90	13682.30
1-34/35-124 (RNase Th1)	13700.44	13700.30

2. Structural characteristics of RNase Th1

Studies by Polverino de Laureto. *et al.* (*5*) on the structural properties of RNase Th1 have shown that nicked RNase is structurally similar to the native protein. Indeed, RNase Th1 has far-UV and near-UV CD spectra similar to those of the native protein, indicating that the secondary and tertiary structures of the protein remain unchanged even after the hydrolysis of the Asn34–Leu35 peptide bond (*5*). However, NMR measurements indicated that chemical shift differences of NH resonance between RNase and RNase Th1 (Figure 3.4) are located at the level of the thermolysin cleavage site in RNase Th1, demonstrating the presence of disorder in this region of the nicked protein. Hence, it is expected that such disorder would favor the derivatization of the protein by TGase.

3. TGase-mediated conjugation studies on RNase

TGase mediated conjugation studies on RNase at the level of glutamine residues were performed using dansyl cadaverine as the amine acceptor but no conjugation was observed. Further studies were focused on the TGase mediated conjugation of carbobenzoxy-glutamine-glycine (ZQG) to RNase at the level of lysine residues. The RP-HPLC elution profile showed the presence of two peaks in addition

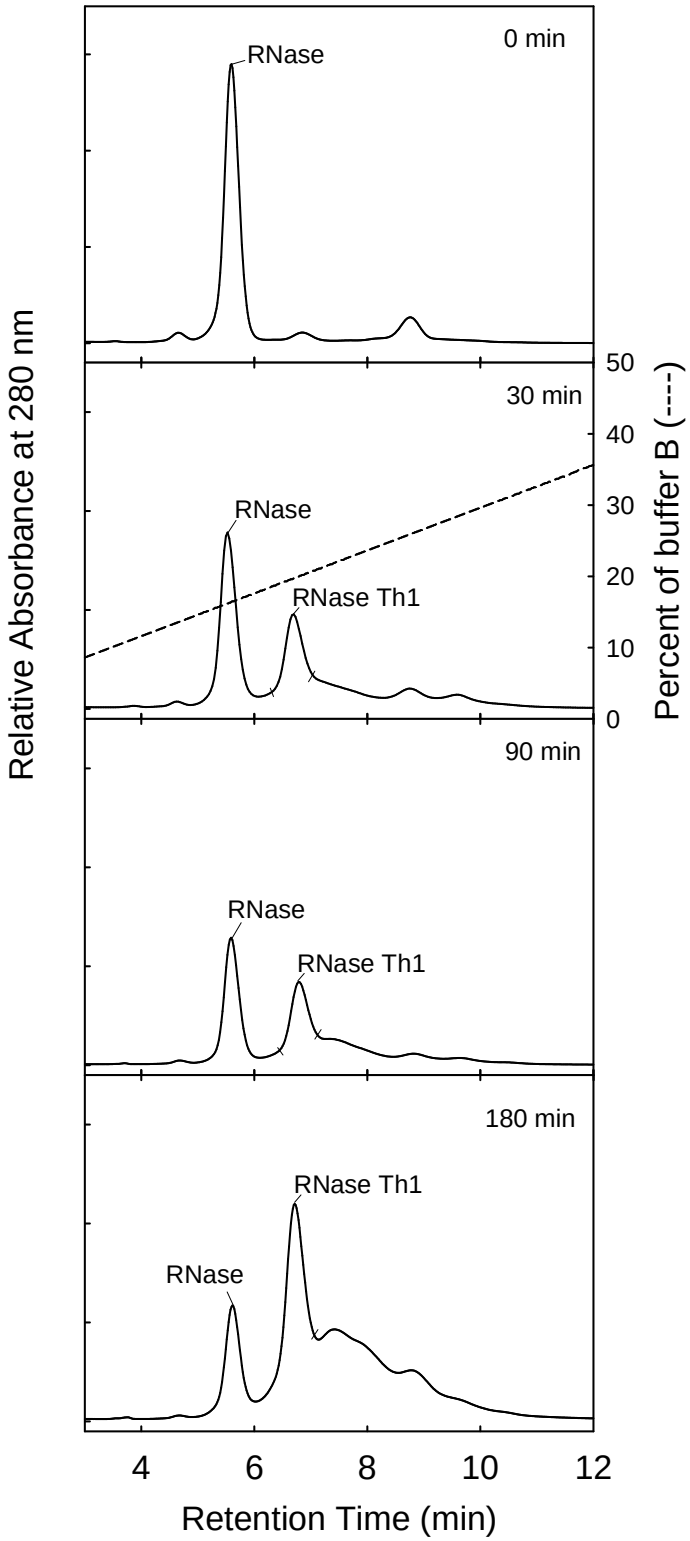


Figure 3.3. Preparation of RNase Th1. RNase Th1 was prepared by thermolysin digestion of native RNase in the presence of 50 % TFE in 50 mM Tris-HCl buffer pH 7.0, containing 5 mM $CaCl_2$. The separation was performed on a Resource S cation exchange column equilibrated with 20 mM phosphate buffer pH 7.0, and eluted with a gradient of the same buffer containing 0.8 M NaCl from 0 to 60 % in 20 min.

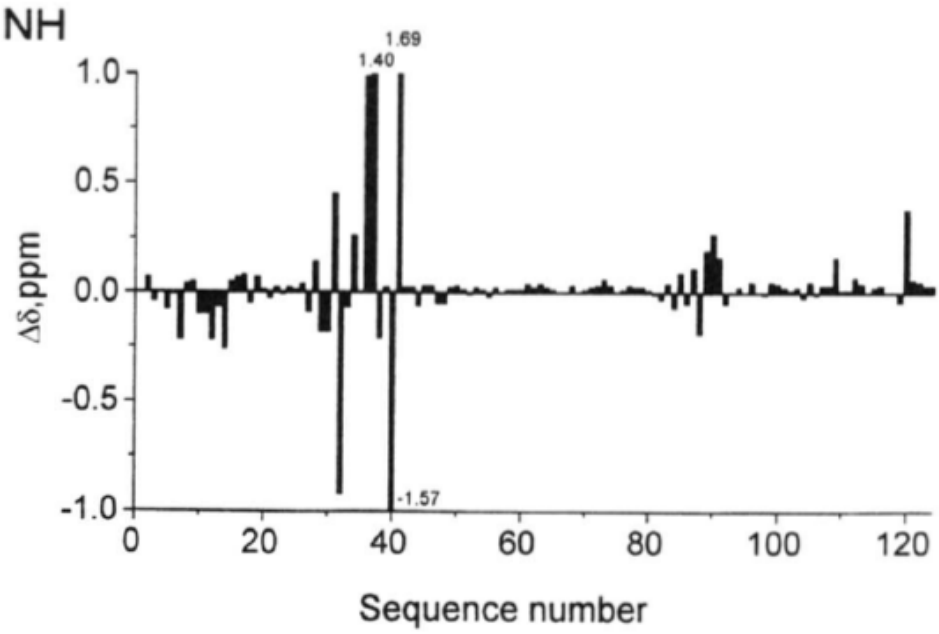


Figure 3.4. Chemical shift plot of NH resonance differences between native RNase and RNase Th1 versus amino acid sequence (adapted from Polverino de Laureto. *et al.* 1997).

to the peak corresponding to native RNase (Figure 3.5). MS based identification of the eluting protein material revealed that the peaks corresponded to the presence of RNase derivatives mono and double conjugated to ZQG (Table 3.2). Characterization of the site(s) of conjugation was performed by digestion of $RNase^{ZQG}$ derivatives with trypsin followed by MS analysis to identify the modified peptides. The sites of conjugation in the modified peptides were identified by MS/MS analysis (Figures 3.7 and 3.8) and they were found to occur at the level of Lys1 and Lys91. From the RP-HPLC profile, the conjugation at the level of Lys1 was seen to be favored in respect to Lys91, since the derivatization at Lys 91 occurred only after that of Lys1.

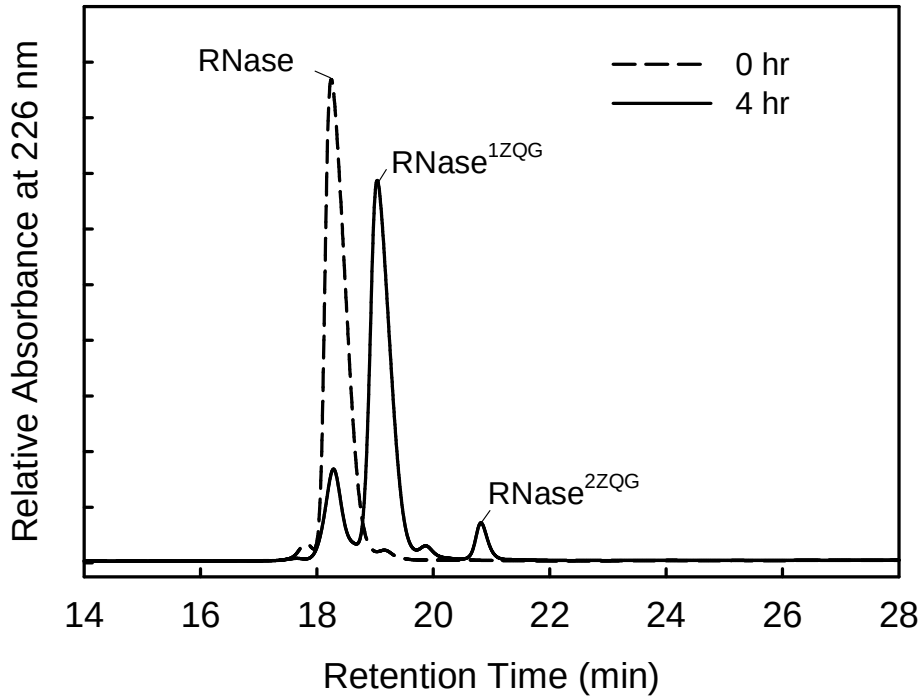


Figure 3.5. TGase mediated conjugation of native RNase with CBZ-Gln-Gly (ZQG). RP-HPLC analyses of the reaction mixture at time 0 minute (dotted line) and after 4 hours (solid line) of incubation. The analysis was performed using a C 18 Phenomenex column eluted with a linear gradient of AcCN containing 0.085 % TFA from 22-40 % in 34 min.

Table 3.2: Molecular masses as observed from the MS analysis of the peptide material contained in the RP-HPLC peaks of the TGase mediated conjugation of RNase with ZQG (Figure 3.5). The sites of ZQG conjugation were determined by overnight trypsin digestion of the conjugated protein in 0.1 M ammonium bicarbonate buffer pH 8.9 with an E:S ratio of 1:100 w/w, followed by MS identification of the modified peptides. Confirmation of the site of conjugation was obtained by MS/MS analysis of the modified peptides (Figure 3.7 and 3.8).

Identity	Conjugation site	Molecular mass (Da) *Observed*	*Calculated*
ZQG	-	-	337.33
1–124	-	13681.90	13682.30
$[1–124]^{1ZQG}$		14002.25	14002.30
$[1–7]^{1ZQG}$	K_1	1037.64	1037.40
$[1–124]^{2ZQG}$		14322.04	14322.30
$[1–7]^{1ZQG}$	$K_{1,}$	1037.62	1037.40
$[86–98]^{1ZQG}$	K_{91}	1767.00	1766.65

4. TGase-mediated conjugation studies on RNase Th1

Conjugation studies on the derivatization of RNase Th1 with TGase at the level of glutamine residues were performed using dansyl cadaverine as the amine acceptor. However, no conjugation was observed even at the level of a glutamine residue close to the site of thermolysin cleavage. Further studies were focused on the TGase-mediated conjugation of RNase with carbobenzoxy-glutamine-glycine (ZQG) at the level of lysines. The RP-HPLC chromatogram showed the presence of many isoforms (Figure 3.6) of the conjugated proteins with different times of elution. The reaction was monitored at 0, 2, 4 and 7 hours. After 2 h of reaction, peaks denoted as 1, 2 and 3 in Figure 3.6 were seen to be of almost the same intensity. MS measurements conducted on the RP-HPLC fractions indicated that peak 1 contains as the major component species RNase Th1 hydrolysed at the peptide bond Thr36–Lys37 (indicated as RNase del-Th1) along with a small amount of RNase Th1. RNase del-Th1 is likely generated by the presence of a minimal protease activity in the reaction mixture. This can be attributed to the fact that the commercially available TGase preparation is only partially pure and contains traces of proteases which caused such side reactions.

MS analysis of peaks 2 and 3 indicated that they contain mono-conjugated RNase del-Th1 derivatives. Identification of the site(s) of conjugation by digestion of the RNase del-Th1^{ZQG} derivatives with trypsin followed by MS analysis revealed that the conjugation occurred at the level of Lys91 and Lys1 for peaks 2 and 3, respectively. The site of conjugation in the modified peptides was confirmed by MS/MS analysis (Figures 3.7 and 3.8). Unlike the TGase mediated conjugation studies on RNase, the conjugation at the level of Lys1 and Lys91 occurred simultaneously and independently of each other. This observation indicates that the disorder caused by nicking of the Asn35–Leu36 peptide bond favors the derivatization of Lys91 by TGase. Moreover, at longer reaction times the formation of double conjugated RNase del-Th1^{ZQG} derivatives were detected (peak 4-6), where it was identified an additional conjugation site at the level of Lys37. Interestingly, this residue lies in the region showing high chemical shift differences in the sequence of

RNase Th1 (Figure 3.4) thus indicating that the increased disorder around the site of nicking favors the derivatization by TGase.

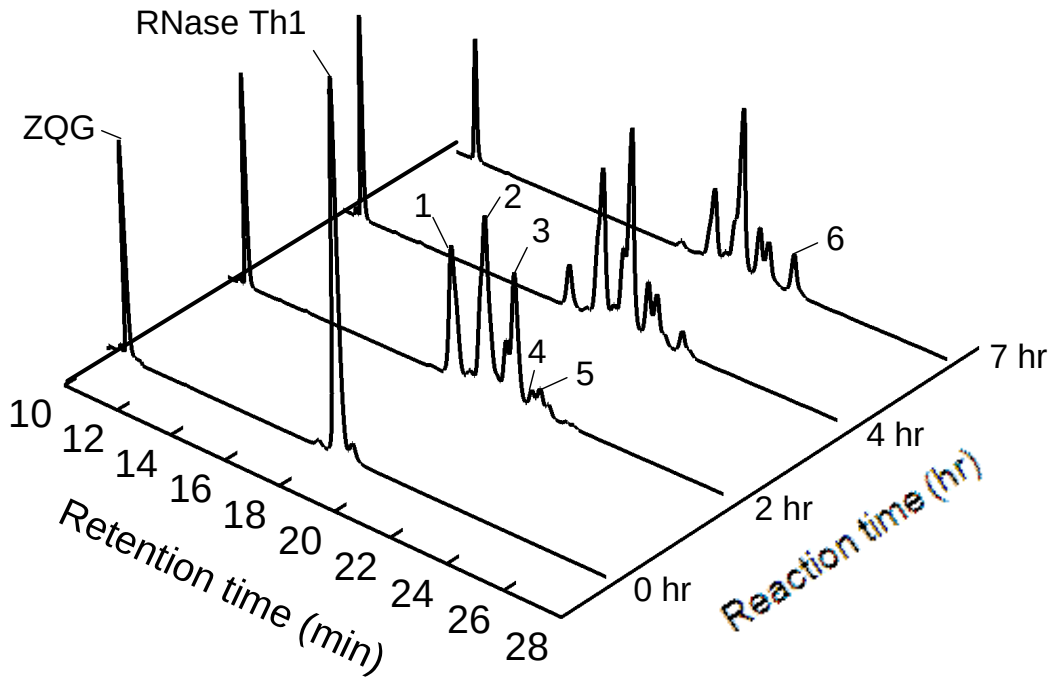


Figure 3.6. TGase mediated conjugation of RNase Th1 with CBZ-Gln-Gly (ZQG). RP-HPLC analysis of the reaction on a C18 Phenomenex column. The separation was performed with a linear gradient of AcCN containing 0.085 % TFA from 22 to 40 % in 34 min.

Table 3.3: Molecular masses measured by MS of the protein material eluted in the RP-HPLC peaks (See Figure 3.6) for the TGase mediated conjugation of RNase A (Th1) with ZQG. Confirmation of the site of conjugation was obtained by MS/MS analysis of the modified peptides.

RP-HPLC peak	Identity	Conjugation site	Molecular mass (Da) *Observed*	*Calculated*
	ZQG	-	-	337.33
1	RNase Th1 [1-34/35-124]	-	13700.00	13700.30
	RNase del-Th1 [1-34/37-124]		13486.67	13486.04
2	RNase del-Th1^{1ZQG}		13806.00	13806.04
	[86–98]^{1ZQG}	K_{91}	1766.18	1766.65
3	RNase del-Th1^{1ZQG}		13806.24	13806.04
	[1–9]^{1ZQG}	K_{1}	1037.54	1037.40
4	RNase del-Th1^{2ZQG}		14126.04	14126.30
	[1–9]^{1ZQG}	K_{1}	1037.58	1037.40
	[37–39]^{1ZQG}	K_{37}	737.38	737.46
5	RNase del-Th1^{2ZQG}		14125.98	14126.30
	[37–39]^{1ZQG}	K_{37}	737.46	737.46
	[86–98]^{1ZQG}	K_{91}	1766.96	1766.65
6	RNase del-Th1^{3ZQG}		14446.32	14446.30
	[1–9]^{1ZQG}	K_{1}	1037.67	1037.40
	[37–39]^{1ZQG}	K_{37}	737.21	737.46
	[86–98]^{1ZQG}	K_{91}	1766.01	1766.65

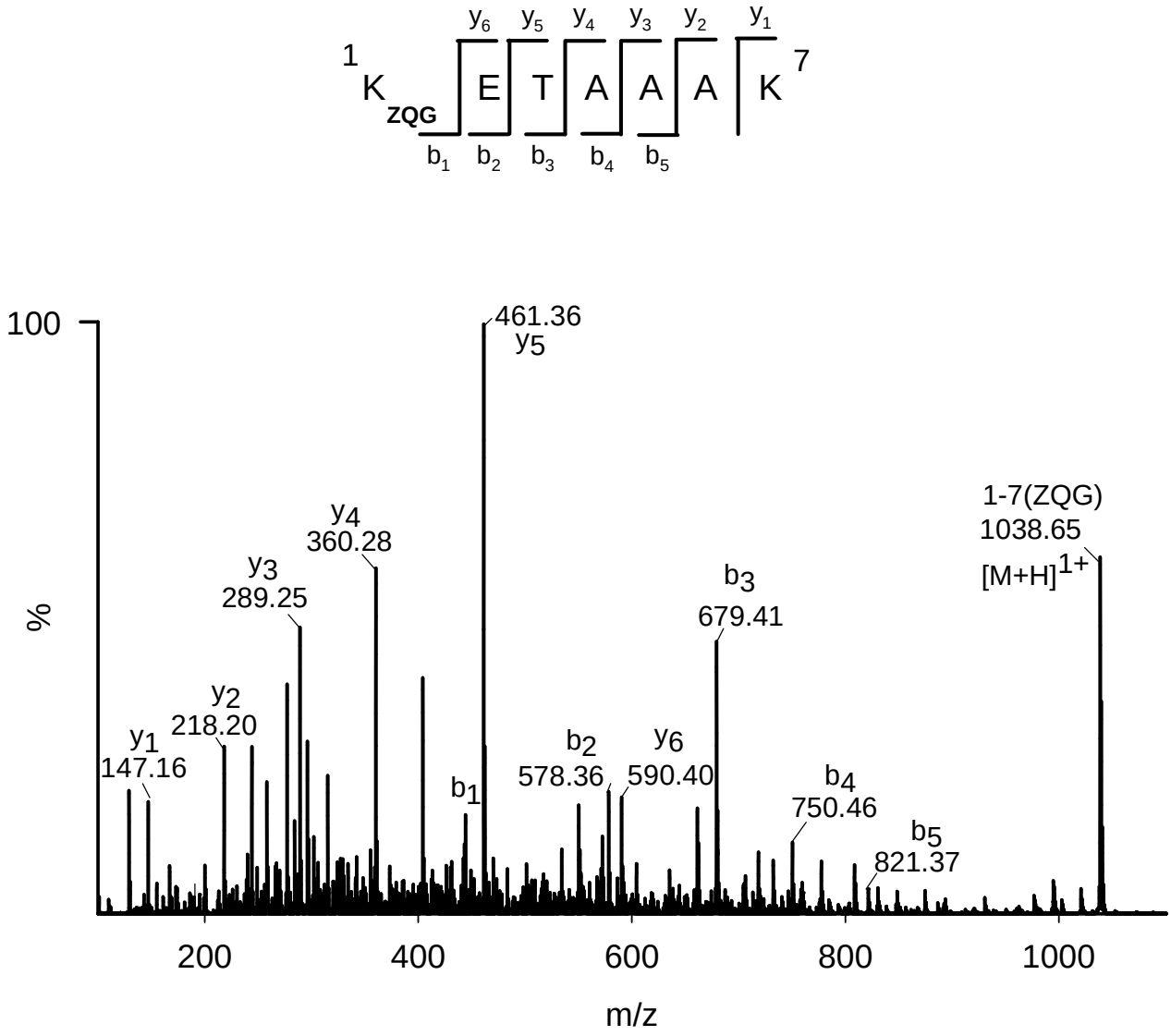
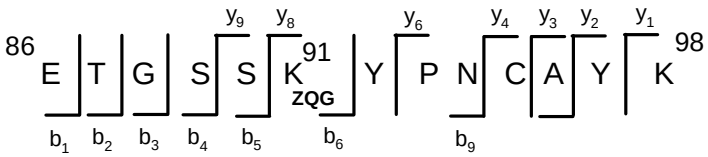


Figure 3.7. Electrospray MS/MS mass spectrum of the ion at 1038.65 m/z of peptide Lys1–Lys7 $[M+H]^{1+}$ of RNase Th1^{1ZQG}. (Top) Fragments of the series *b* and *y* that were identified in the MS/MS spectrum are indicated on the sequence of the peptide. Modified Lys1 is indicated as K_{ZQG}. (Bottom) In the MS/MS spectrum, ions assigned to fragments of the series *b* and *y* are indicated.



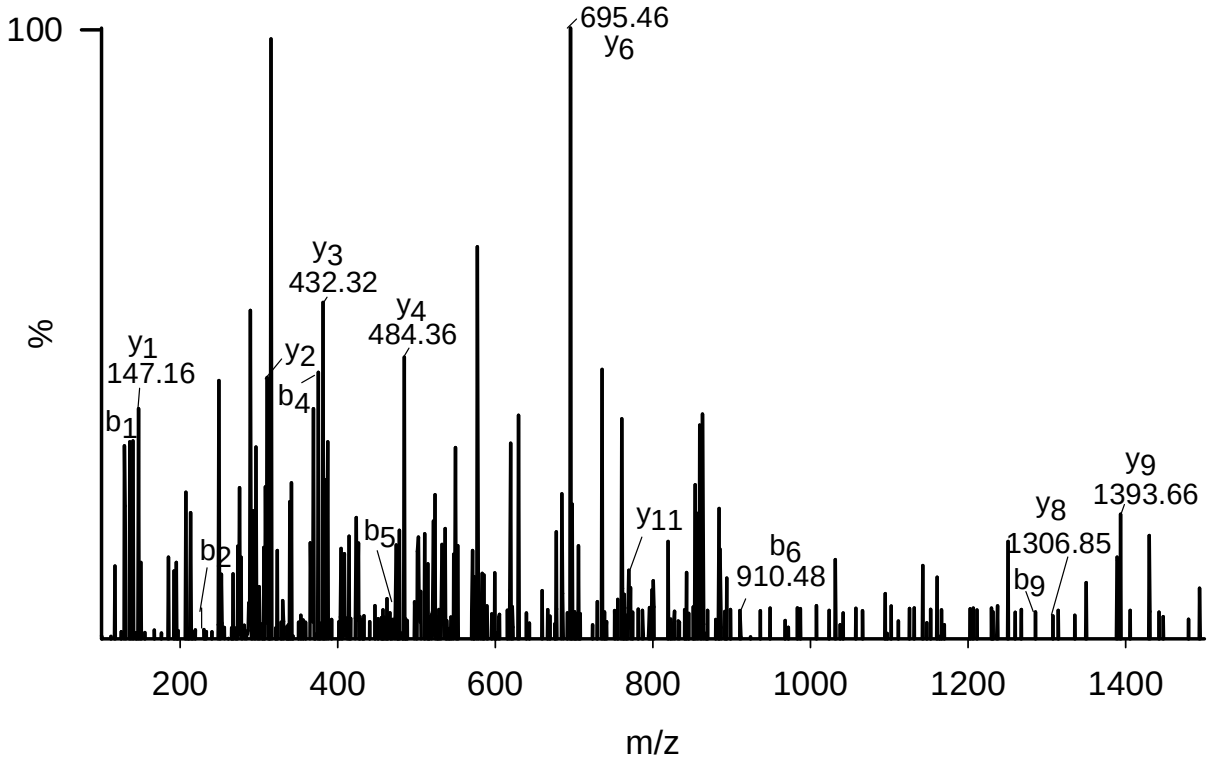


Figure 3.8. Electrospray MS/MS mass spectrum of the ion at 884.45 m/z of peptide Glu86–Lys98 $[M+2H]^{2+}$ of RNase Th1^{1ZQG}. Fragments of the series *b* and *y* that were identified in the MS/MS spectrum are indicated on the sequence of the peptide. Modified Lys91 is indicated as K_{ZQG}.

5. Comments

Proteases are used extensively to probe the structure of proteins. Indeed, it has been demonstrated that limited proteolysis of globular proteins in their native state occurs in regions of flexible loops and never at the regions encompassing elements of secondary structure (*5-8*). Moreover, a clear-cut correlation was found between the sites of limited proteolysis and the regions of high segmental mobility as given by the B-factor values obtained from data of X- ray crystallography (*6*).

RNase is resistant to thermolysin in Tris-HCl buffer, pH 7.0 at 20-42 ° C (*5*). However, in the presence of 50 % TFE the interactions that stabilize the tertiary structure of the protein are severely reduced leading to the initial thermolysin hydrolysis at position Asn34–Leu35, and then to a later cleavage at Thr45–Phe46 peptide bond (*5*). It is evident from previous proteolysis studies on RNase conducted at elevated temperatures that the initial cleavages occur near or at the site of thermolytic cleavage. Indeed, at 60 ° C trypsin cleaves at Lys31–Ser32, Arg33–Asn34 and Lys37–Asp38 peptide bonds (*9*). Chymotrypsin at 60 ° C cleaves rapidly at Tyr25–Lys26, followed by Leu35–Thr36 and Phe46–Val47 (*10*). These data indicate that in the presence of denaturating conditions the chain segment Met30–Phe46 becomes highly flexible favoring the selective proteolysis. Moreover, NMR spectroscopy studies have shown the presence of large dispersion values of the backbone angles located at the flexible loop encompassing the thermolysin cleavage site in RNase (*3*).

As seen from the present study, despite the presence of 10 lysines in the RNase A sequence, only 3 were conjugated by TGase. Interestingly, the TGase conjugation sites at the level of lysine residues (Lys1, Lys37 and Lys91) occur in regions of enhanced flexibility (Figure 3.9). Indeed, Lys1 is present in the region of high flexibility in native RNase as given by the B-factor values (Figure 3.2). Similarly for the Lys91 TGase conjugation site, NMR studies on native RNase in aqueous solution revealed that the loop comprising residues Thr87–Cys95 is flexible (*3*). In the case of RNase Th1, thermolysin cleavage at the Asn34–Leu35 site increases the disorder in the mobile loop Thr87–Cys95, as evident from NMR studies on the chemical shift NH resonance differences between RNase and RNase Th1 (Figure 3.4).

Therefore, the conjugation at Lys91 is faster in RNase Th1 than RNase, which has been clearly demonstrated by the results of our experiments. Moreover, Lys37 was seen to be conjugated only in RNase Th1 and indeed thermolytic cleavage at the position Asn34–Leu35 destabilizes the structure of RNase around the site of hydrolysis as demonstrated by the large NMR chemical shift differences of this region of the nicked protein in respect to native RNase.

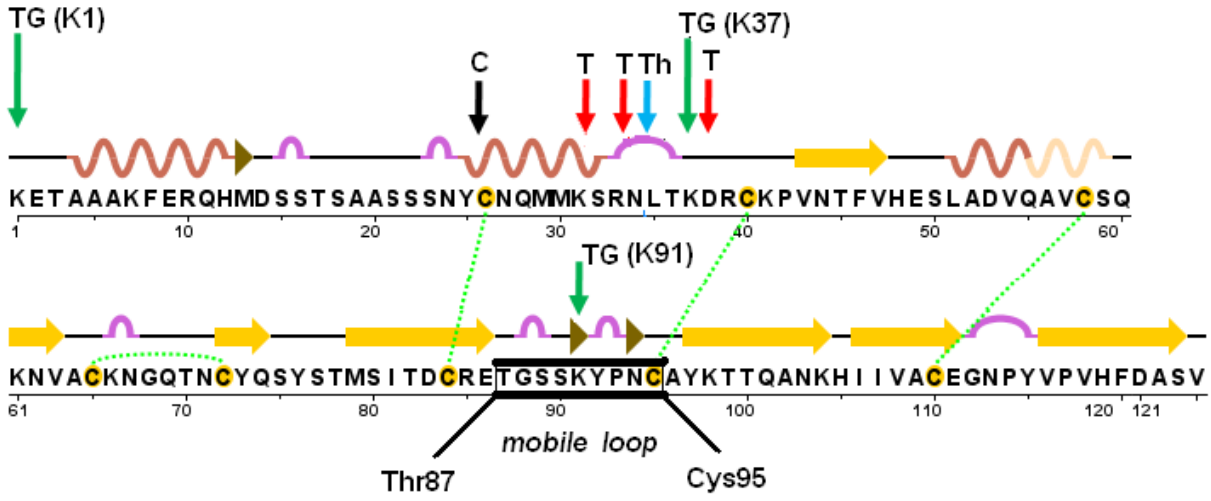


Figure 3.9. Sites of limited proteolysis and TGase conjugation sites in RNase and RNase Th1. Position of proteases cleavage sites are indicated for thermolysin (Th) (blue), trypsin (T) (red) and chymotrypsin (C) (black) by arrows. TGase conjugation sites at the level of lysines are indicated with TG and green arrows along with the position of conjugation. The region corresponding to the mobile loop Thr87–Cys95 as given from the NMR studies on RNase A in solution is indicated between two bold lines (9).

On the other end, TGase mediated conjugation studies at the level of glutamines in RNase as well as RNase Th1 did not lead to any product in the reaction. There is only one glutamine residue (Gln28) that is located in the flexible region Cys26–Phe46. However, this Gln residue is close to Cys26, which is involved in the formation of a disulphide bridge with Cys84. The presence of such disulphide bridge might cause hindrance to the binding of TGase.

Overall, even in the case of RNase the sites of proteolysis occur in the same regions as the sites of TGase derivatization, thus indicating that flexibility determines the selectivity both of TGase and proteases. These results also indicate that flexibility

dictates the selectivity of TGase even at the level of lysine residues. Eventually, it was demonstrated that by inducing disorder in selected regions of a protein as in the case of RNase Th1, it is possible to favor the modification of a protein by TGase..

6. References

1. Udgaonkar, J.B., and Baldwin, R.L. (1988) NMR evidence for an early framework intermediate on the folding pathway of ribonuclease A, *Nature 335*, 694–699.

2. Udgaonkar, J.B., and Baldwin, R.L. (1990) Early folding intermediate of ribonuclease A, *Proc Natl Acad Sci U S A 87*, 8197–8201.

3. Santoro, J., Gonzalez, C., Bruix, M., Neira, J.L., Nieto, J.L., Herranz, J., and Rico, M. (1993) High-resolution three-dimensional structure of ribonuclease A in solution by nuclear magnetic resonance spectroscopy, *J Mol Biol 229*, 722–734.

4. 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.

5. Polverino de Laureto, P., Scaramella, E., De Filippis, V., Bruix, M., Rico, M., and Fontana, A. (1997) Limited proteolysis of ribonuclease A with thermolysin in trifluoroethanol, *Protein Sci 6*, 860–872.

6. 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.

7. Polverino de Laureto, P., De Filippis, V., Scaramella, E., Zambonin, M., and Fontana, A. (1995) Limited proteolysis of lysozyme in trifluoroethanol. Isolation and characterization of a partially active enzyme derivative, *Eur J Biochem 230*, 779–787.

8. Polverino de Laureto, P., Toma, S., Tonon, G., and Fontana, A. (1995) Probing the structure of human growth hormone by limited proteolysis, *Int J Pept Protein Res 45*, 200–208.

9. Ooi, T., Rupley, J.A., and Scheraga, H.A. (1963) Structural Studies of Ribonuclease. Viii. Tryptic Hydrolysis of Ribonuclease a at Elevated Temperatures, *Biochemistry 2*, 432–437.

10. Rupley, J.A., and Scheraga, H.A. (1963) Structural Studies of Ribonuclease. Vii. Chymotryptic Hydrolysis of Ribonuclease a at Elevated Temperatures, *Biochemistry 2*, 421–431.

4. CRITICAL EVALUATION OF CASE STUDIES

During my PhD work, I had the opportunity to interact and collaborate with other members of the Laboratory of Protein Chemistry of CRIBI involved in studies dealing with the use of TGase for the modification of proteins of pharmaceutical interest. These studies aimed to unravel molecular features of enzymatic reactions mediated by microbial TGase and, in particular, to use an amino-derivative of poly(ethylene)glycol (PEG-NH_2) as an acyl-acceptor in a TGase-mediated reaction, thus leading to PEGylation of a protein (*1-3*). These studies were conducted on human growth hormone (hGH) and granulocyte colony-stimulating factor (G-CSF) and other proteins. Here, a short presentation and discussion of the results obtained will be given, together with an evaluation of related results obtained in other laboratories.

1. Human growth hormone (hGH)

Human growth hormone (hGH) is an important growth factor, which is required in various tissues during the maintenance and development of physiological processes. Recombinant hGH is commercially available as a pharmaceutical protein and is used to treat deficiencies related to the growth hormone inadequacy, such as short stature and renal failure in children (*4*). The 191-residue chain of hGH adopts a four-helix bundle architecture and the helical segments comprise chain segments 9–34, 72–92, 106–128 and 155–184 (*5, 6*). The long chain loop approximately from residue 129 to 154 that connects helix C to helix D displays high *B*-factor values (see 3D structure in PDB code 1BP3) or no electron density (see 3D structure 1HUW), implying that this loop is highly flexible or even disordered (7-*10*). Also the region approximately from residue 30 to residue 60 displays high *B*-factor values (Figure 1).

TGase-mediated conjugation of PEG to hGH. The protein hormone has a short functional half-life *in vivo* and must be administered daily by subcutaneous injection for maximum effectiveness. Improved versions of hGH were obtained by PEGylation of the hormone at the level of α- and ε-amino groups using the N-hydroxysuccinimidyl ester derivative of PEG (*11*). Eight to nine lysine residues and the N-terminal amino acid were modified to varying extents and the various hGH

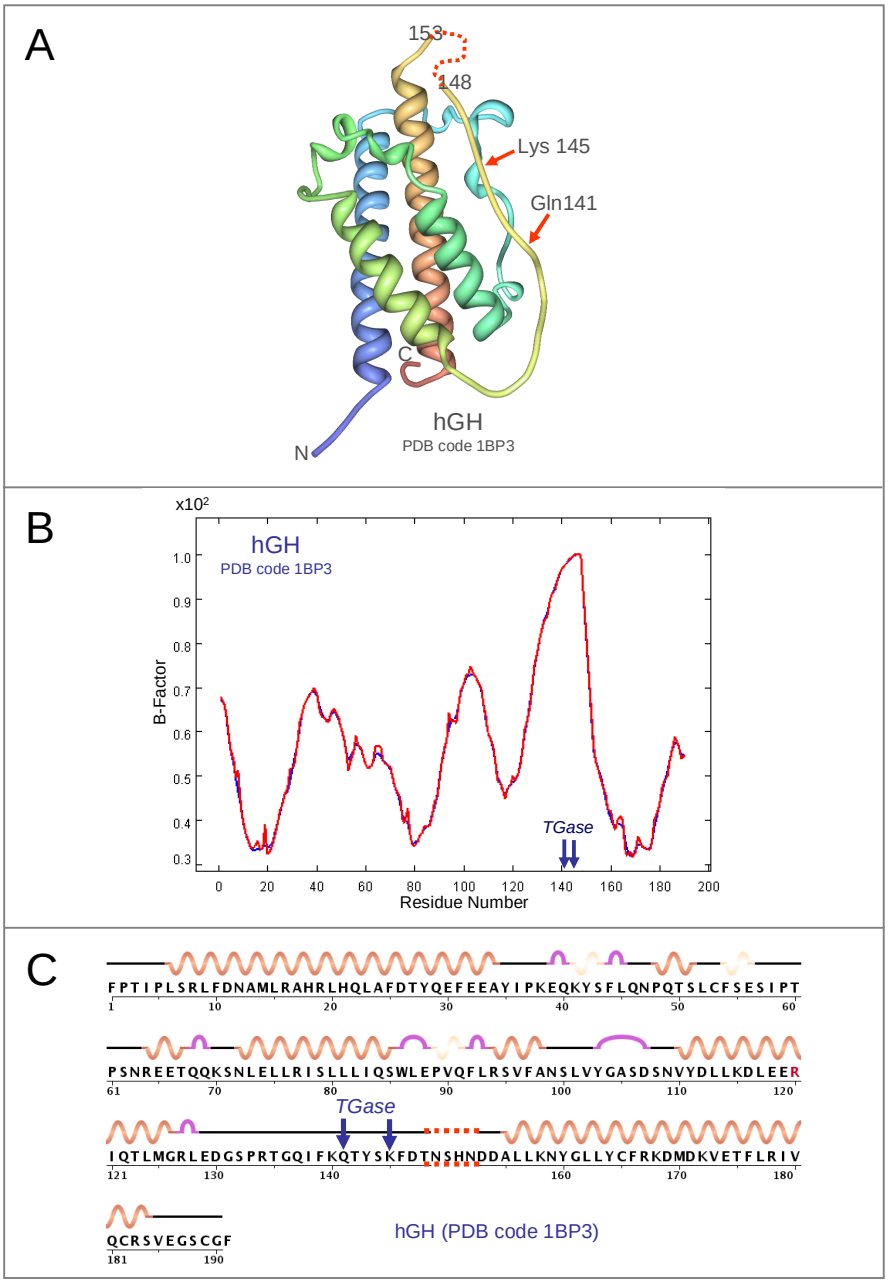


Figure 1. Site-specific modifications of hGH mediated by TGase. (*A*) 3D structure of the hormone (PDB code 1BP3) (*15*). The chain region 148-153 does not show electron density and thus is disordered. (*B*) Plot of *B*-factor values along the 191-residue chain of human hGH (PDB code 1BP3). (*C*) Amino acid sequence of human hGH. Arrows indicate the (mostly) selective reactions mediated by TGase at the level of Gln141 and Lys145 (see text).

derivatives differed from one another for their *in vitro* bioactivity. To achieve a selective protein modification, a Cys residue was introduced by site-directed mutagenesis into the hormone at position 3 in place of a threonine residue (Thr3) and then the mutant hGH (Thr3Cys) was reacted with a PEG derivative containing an alkylating moiety for the thiol group of Cys3 (*12*).

More recently, a TGase-mediated reaction was used for the covalent binding of PEG-NH_2 to the protein hormone (*1–3*). It was shown that the covalent conjugation of PEG-NH_2 occurs mostly at Gln141 and to a minor extent to Gln40. In another study, 1,3-*bis*-aminoxy-propane (H_2N-O-CH_2-CH_2-CH_2-O-NH_2) was conjugated by TGase at the level of Gln residues and then the newly introduced aminoxy group was further reacted with an aldehyde derivative of PEG (PEG-CHO) to yield the condensation product hGH-...–O-N=CH-PEG conjugate (*13*). Using this chemoenzymatic method it was possible to produce and isolate to homogeneity after chromatography PEGylated derivatives of hGH labelled at Gln141 or Gln40. In this study, it was clearly shown that the TGase reaction leads to mono-conjugated products at Gln141 or Gln40, as well to a bis-derivative labelled at both Gln141 and Gln40. At any rate, the derivative at Gln141 was shown to be the major product and it was isolated to homogeneity by chromatography of the reaction mixture in 48% yield (*13*). Moreover, N-carbobenzoxy-Gln-Gly (ZQG) derivatives have been conjugated selectively at the level of Lys145 in hGH (*13*). Therefore, TGase-mediated reactions occur mostly at the level of Gln141 and Lys 145, despite the fact that the hormone contains 13 Gln and 7 Lys residues. Of interest is the fact that the Gln and Lys conjugation site(s) are present in a chain region displaying high flexibility, as given by an analysis of the plot of *B*-factor values *versus* residue number of the protein (see Figure 1). Moreover, in the X-ray structure of hGH (3HHR), the chain region 129-154 was shown to be even fully unfolded (*14*). Therefore, these results provide a strong evidence that the site-specific modifications of a protein mediated by TGase occur only at flexible/unfolded regions of the chain.

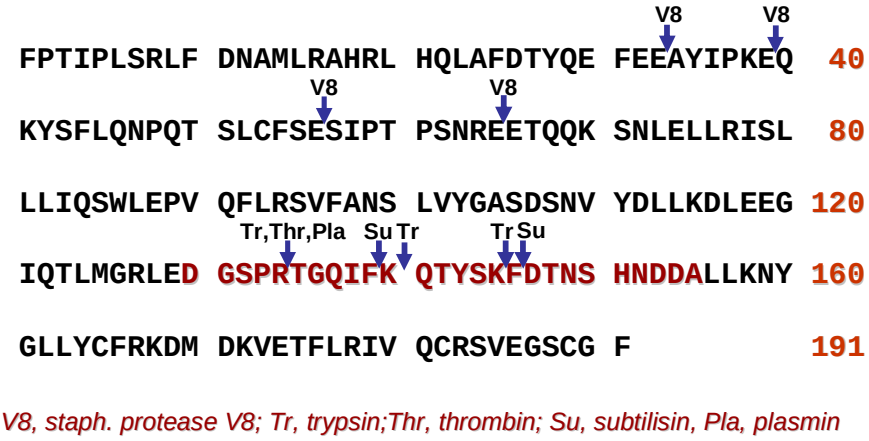


Figure 2. Limited proteolysis of human hGH. Arrows indicate the sites of proteolysis by the indicated enzymes (*4, 16, 17*).

Correlation of TGase-mediated reactions with limited proteolysis. Previous limited proteolysis experiments on hGH with plasmin, trypsin, thrombin and subtilisin have demonstrated that the flexible/disordered region 130-150 is highly prone towards proteolytic cleavages by several proteases (Figure 2) (*4*). Staphylococcal V8-protease cleaves at Glu33, followed by a slower cleavage reaction at Glu56 and Glu66 (*16*). Moreover, pepsin at acidic pH cleaves the Phe44-Leu45peptide bond, generating two complementing fragments 1-44 and 45-191 (*17*). These cleavage sites occur at chain region 30-60 of hGH sequence, which displays an enhanced flexibility, as given by the high *B*-factor values of this region (see Figure 1).

Since the TGase-mediated reactions occur mostly at Gln141 and Lys145 and to a minor extent to Gln40, it can be concluded that both TGase and a protease require a flexible/unfolded polypeptide substrate for their selective reactions, as it can be deduced by examining the plot of *B*-factor values along the protein chain (Figure 1). The analogy between TGase and a protease is in line with the consideration that TGase acts as a reverse protease (synthesis of an amide moiety) and that TGase‘s active site is given by a Cys-His-Asp triad in analogy to a Cys-protease (*18*). Therefore, it can be anticipated that the molecular features of limited proteolysis phenomena (*19-22*) would apply also to the often observed site-specific protein modifications by TGase.

2. Granulocyte-colony stimulating factor (G-CSF)

Granulocyte colony stimulating factor (G-CSF) is a protein cytokine, which stimulates the proliferation, survival and differentiation of neutrophil granulocyte progenitor cells and mature neutrophils (*23*). Recombinant G-CSF is nowadays available for the treatment of several kinds of neutropenia and also in bone marrow transplantation. In particular, G-CSF is administered in the late-stage HIV infections or for myeloablation followed by bone marrow transplant. Since the recombinant G-CSF is rapidly eliminated from the body by a renal clearance mechanisms and receptor mediated endocytosis, the therapeutic use of G-CSF requires multiple injections for efficacy. In order to avoid these shortcomings, in several reports it has been demonstrated that PEGylation of G-CSF is an effective procedure for obtaining a safer and better protein drug (*24-28)*.

The G-CSF protein is a single 174-residue polypeptide chain arranged in a four-helix bundle architecture and stabilised by two disulphide bridges (Cys36-Cys42 and Cys64-Cys72). Helical segments comprise segments 10-39, 72-91, 100-120 and 138-171 (Figure 3) (*29*). Of note, there are chain regions for which no electron density is obtained, so that regions 1–9, 61–71, 126–137 and 171–174 are considered to be unfolded/disordered (*7-10*). Glycosylation at Thr133 is present in the wild-type G-CSF, whereas in the sample of recombinant G-CSF glycosylation is missing.

TGase-mediated PEGylation of G-CSF. Microbial TGase has been used to conjugate a PEG-NH_2 to recombinant G-CSF to generate a PEGylated form of G-CSF (*2, 3, 30, 31*). The reaction proved to be selective and producing a mono-PEGylated G-CSF derivative at the level of Gln134. This selectivity is undoubtedly striking, if one considers that G-CSF contains 17 Gln residues. Since this Gln residue is embedded in a chain segment for which there is no electron density (see Figure 3), it can be concluded that local unfolding of the site of TGase attack is required for the observed selectivity. The non-susceptibility of other Gln residues to a TGase attack can be explained as follows. Residues Gln131 and Gln173 are located in a flexible/unfolded region, but these residues are followed by a proline (Pro) residue (see Figure 3).

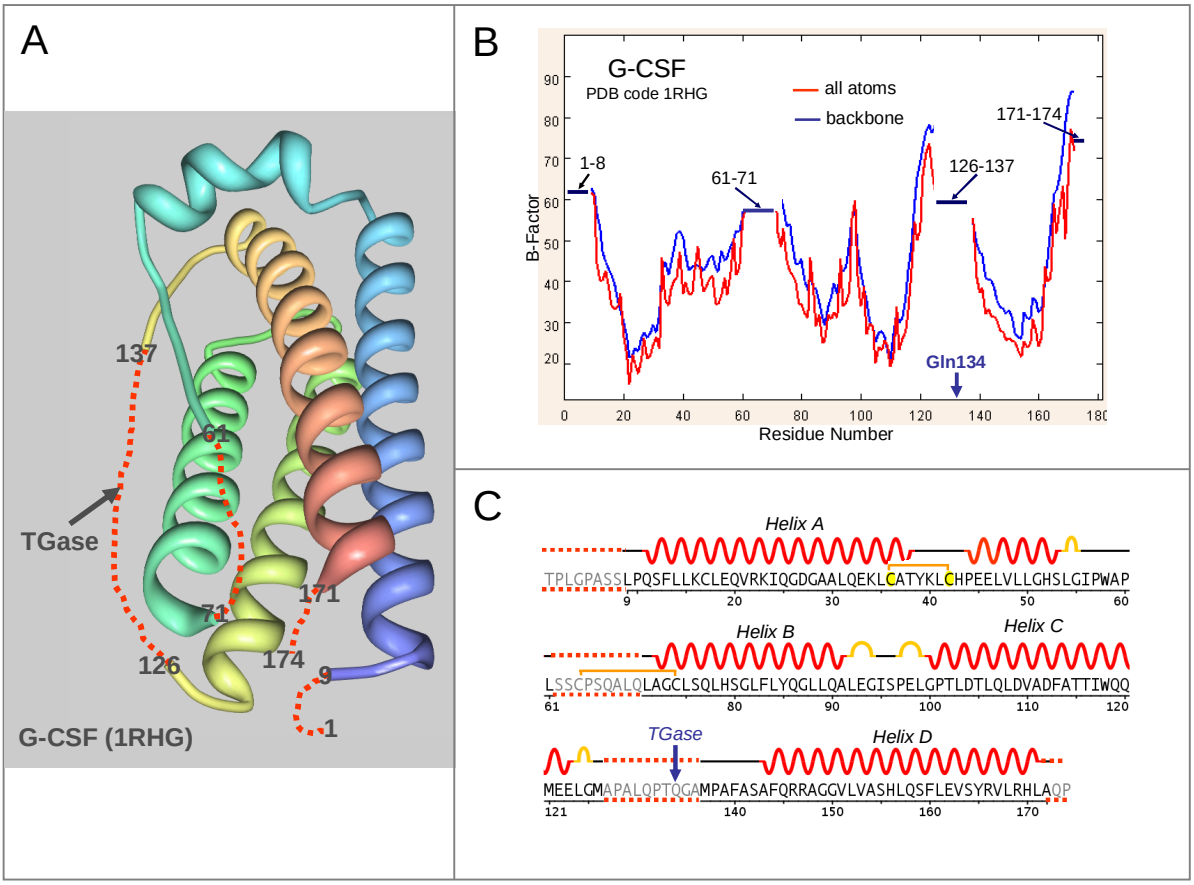


Figure 3. PEGylation of G-CSF. (*A*) Structure of the 174-residue chain of G-CSF determined crystallographixcally (PDB code 1RHG). (*B*) Plot of the *B*-factor values along the polypeptide chain of G-CSF. Missing electron density indicates disordered chain regions (1-8, 61-71, 126-137 and 171-174). (*C*) Amino acid sequence and secondary structure of G-CSF (PDB code 1RHG). A waved red segment above the sequence indicates location of the hour helical segments. The arrow indicates the location of the TGase-mediated PEGylation site at Gln134.

It could be well that TGase cannot attack at these Gln residues if one accepts the view that TGase is related to a protease and by considering that indeed trypsin that does not cleave Lys-Pro or Arg-Pro peptide bonds (the reader is referred to the web site http://www.expasy.ch/tools/peptidecutter). Furthermore, the lack of TGase attack at residues Gln67 and Gln70, which lie in the flexible region 61-71, can be explained by proposing that the disulphide bridge Cys64-Cys72 can hinder substrate binding at the TGase's active site. In the N-terminal disordered region 1-9 there are no Gln residues and hence no TGase reaction can occur.

We may observe here that the TGase reactive Gln134 is located near the glycosylation site Thr133 in the flexible/disordered region 126-137 of G-CSF (*32*). The possibility that site-specific enzymatic reactions on proteins (*e.g.*, phophorylation, glycosylation) could share the same requirements of substrate flexibility as TGase and proteases was already advanced (*3*).

Mobility of G-CSF determined by NMR. Most of our studies were conducted with the aim to correlate sites of TGase reactions with structural data of protein substrates obtained by X-ray methods, since the crystallographic data can be analysed not only in terms of structure, but also in terms of segmental mobility, or flexibility or even protein disorder (*7-10*). These analyses allowed us to conclude that for both TGase- and protease-mediated reactions the critical parameter dictating the site-specific reactions was the flexibility or even full unfolding of the site of enzymatic reaction (*19-22*). It is well known that NMR methods can provide reliable data regarding protein dynamics, since these methods can provide protein structural data in solution, while the X-ray data on protein mobility can be damped by crystal lattice effects or from intermolecular protein contacts in the crystal. Therefore, we considered of interest to examine structural data of G-CSF obtained by NMR methodologies.

Figure 4 (part *A*) shows the structure of G-CSF determined by NMR (*33*), together with an evaluation of its dynamics using a recently published method (*34*). First of all, it is seen that even the NMR-determined structure provides evidence that the reactive Gln134 is located in a flexible segment of the protein chain (Figure 4, parts *B-D*). Indeed, Gln134 is located in the long flexible loop connecting helix C and D (part *C*) and embedded in one of the disordered chain segments (denoted as M-mobile, part *B*), these last automatically determined by the algorithm (*34*). Ther are other Gln residue that could be reactive towards TGase, but we consider that Gln68 and Gln71 are embedded in a disulfide crosslinked loop that can hinder an enzymatic attack. Moreover, since Gln171 and Gln173 are both followed by a Pro residue, we can speculate that, in analogy to trypsin and chymotrypsin, likely also TGase cannot attack a Gln residue followed by a Pro. Therefore, chain mobility of the G-CSF

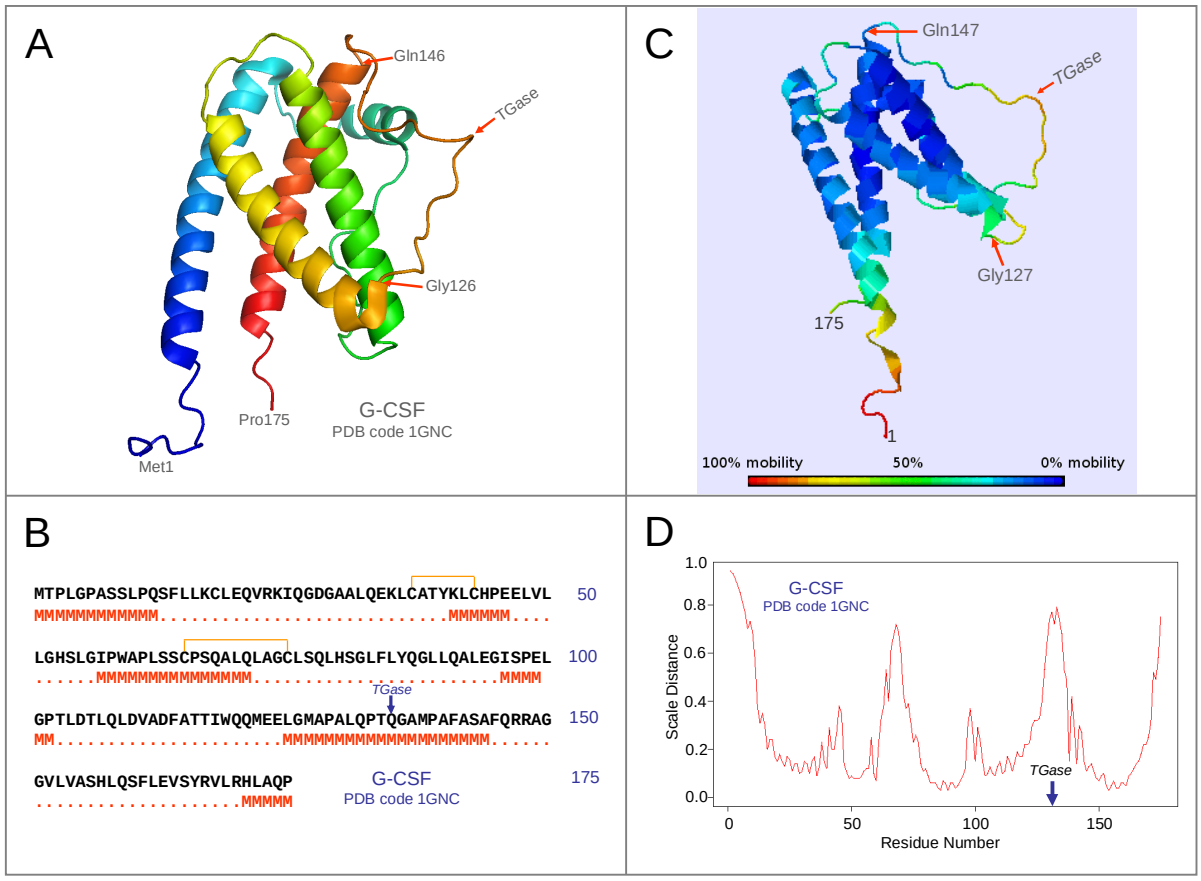


Figure 4. Correlation of the site of TGase-mediated modification (PEGylation) with protein dynamics. (*A*) 3D structure of G-CSF determined by NMR (*33*). Please note that the amino acid sequence of G-CSF given in this figure contains the N-terminal Met. (*B*) Evaluation of the mobile (M) regions along the polypeptide chain of G-CSF determined by the MOBY algorithm (*34*).The connectivities of the two disulfide bridges are indicated by a yellow line (*C*) Mobility of the 175-residue chain of G-CSF colored with rainbow colours, from blue (low mobility) to red (high mobility). (*D*) Scaled graph mobility of the various NMR structures G-CSF (for details on the computing methods see http://protein.bio.unipd.it/mobi/).

molecule determined by NMR appears to be the main characteristic that allows specific PEGylation at Gln134.

Site-specific enzymatic reactions are not explained in terms of site's exposure. In a number of studies originating from the Laboratory of Protein Chemistry at CRIBI it was emphasised that exposure of sites of limited proteolysis of proteins is not at all sufficient to explain the specificity of proteolysis (*19-22*). Actually, the concept of exposure is used quite often in current literature to explain limited proteolysis of globular proteins, but exposure is not at all sufficient to explain the phenomenon, since it is easily understood that in a globular protein there are many exposed sites (the all protein surface), but usually only one site (or very few) has such a peculiar characteristic that dictates the selective enzymatic attack (*22*).

In a recent study, it was attempted to explain the site-specific TGase-mediated PEGylation at the level of Gln 134 of G-CSF in terms of an enhanced exposure of this Gln residue (*31*). As above commented, this explanation unlikely would be valid, considering that many Gln residues in a protein substrate can be surface exposed. Indeed, as shown in Figure 5, the Gln residues of G-CSF show erratic values of exposure and by no means this parameter could explain why TGase selectively reacts with Gln134. Therefore, we can firmly conclude that flexibility or unfolding is the key parameter dictating the site-specific enzymatic modification of a protein substrate by TGase.

3. Conclusion

The above discussed examples document the possibility to achieve a site-specific TGase-mediated modification of valuable proteins of pharmaceutical interest. The procedure utilised is mild and can allow production of homogeneous protein bioconjugates, thus offering a significant advantage over the chemical methods of PEGylation in current use. An interesting and useful aspect of the TGase reaction is that the site of PEGylation of a protein drug can be anticipated, provided that the structure and dynamics of the protein is known. Recently, the TGase procedure was

applied to few protein drugs in the Laboratory of Protein Chemistry at CRIBI and homogeneous PEGylated proteins were thus obtained.

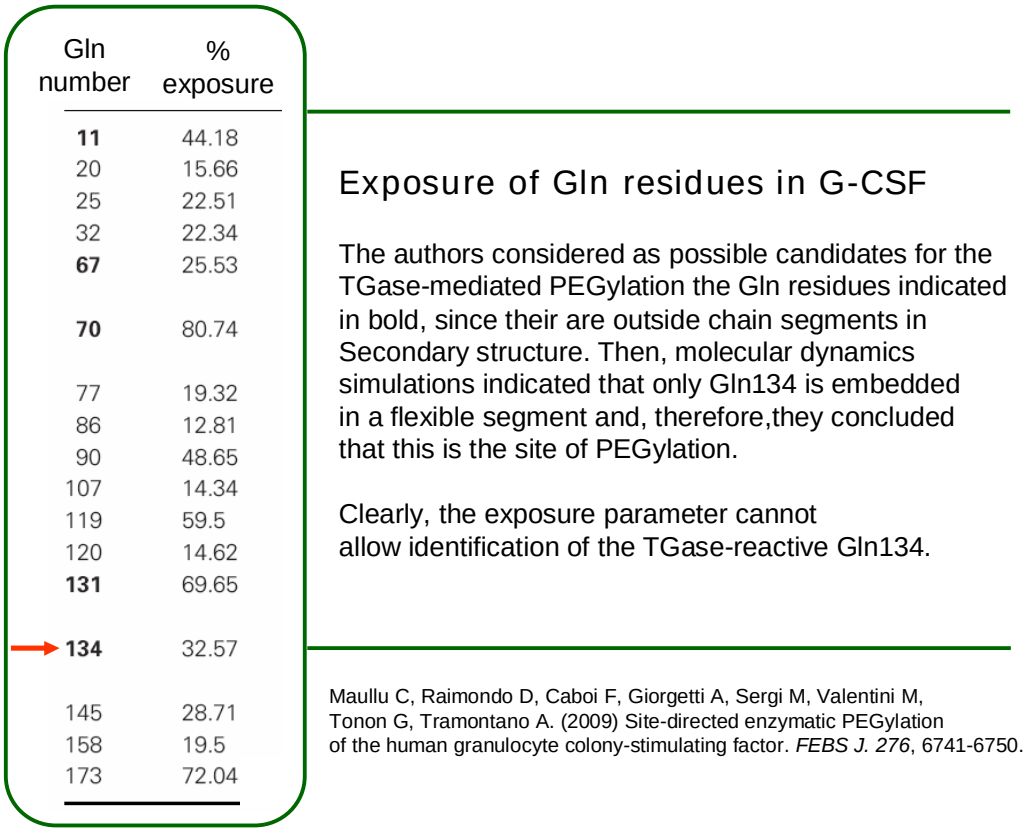


Figure 5. Exposure of Gln residue in a protein substrate is not at all sufficient to explain the TGase-mediated site-specific modification of a protein at the level of Gln residue(s).

4. References

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

2. 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, *Bioconjug. Chem. 20*, 384–389.

3. Fontana, A., Spolaore, B., Mero, A., and Veronese, F. (2009) The site-specific TGase-mediated PEGylation of proteins occurs at flexible sites, in *PEGylated*

Protein Drugs: Basic Science and Clinical Applications (Veronese, F., ed.), pp. 89–112, Birkhäuser Verlag, Basel.

4. Li, C. H. (1982) Human growth hormone: 1974-1981, *Mol. Cell Biochem. 46*, 31–41.

5. de Vos, A. M., Ultsch, M. H., and. Kossiakoff, A. A. (1992) Human growth hormone and extracellular domain of its receptor: Crystal structure of the complex, *Science 255*, 306–312.

6. Ultsch, M. H, Somers, W., Kossiakoff, A. A., and de Vos, A. M. (1994) The crystal structure of affinity-matured human growth hormone at 2 Å resolution, *J. Mol. Biol. 236*, 286–299.

7. Ringe, D., and Petsko, G. A. (1985) Mapping protein dynamics by X-ray diffraction, *Prog. Biophys. Mol. Biol. 45*, 197–235.

8. Ringe, D. and Petsko, G. A. (1986) Study of protein dynamics by X-ray diffraction. *Methods Enzymol. 131*, 389–433.

9. Kundu, S., Melton, J. S., Sorensen, D.C., and Phillips Jr., G. N. (2002) Dynamics of proteins in crystals: Comparison of experiment with simple models, *Biophys. J. 83*, 723–732.

10. Smith, D. K., Radivojac, P., Obradovic, Z., Dunker, A. K., and Zhu, G. (2003) Improved amino acid flexibility parameters, *Protein Sci. 12*, 1060–1072.

11. Clark, R., Olson, K., Fuh, G., Marian, M., Mortensen, D., Teshima, G., Chang, S., Chu, H., Mukku, V., Canova-Davis, E., Somers, T., Cronin, M., Winkler, M., and Wells, J. A. (1996) Long-acting growth hormones produced by conjugation with poly(ethylene glycol), *J. Biol. Chem. 271*, 21969–21977.

12. Cox, G. N., Rosendahl, M. S., Chlipala, E. A., Smith, D. J., Carlson, S. J., and Doherty, D. H. (2007) A long-acting mono-PEGylated human growth hormone analog is a potent stimulator of weight gain and bone growth in hypophysectomized rats, *Endocrinology 148*, 1590–1597.

13. Buchardt, J., Selvig, H., Nielsen, P. F., and Johansen, N. L. Transglutaminase-mediated methods for site-selective modification of human growth hormone, *Biopolymers 94*, 229–235.

14. de Vos, A. M., Ultsch, M., and Kossiakoff, A. A. (1992) Human growth hormone and extracellular domain of its receptor: crystal structure of the complex, *Science 255*, 306-312.

15. Somers, W., Ultsch, M., De Vos, A. M., and Kossiakoff, A. A. (1994) The X-ray structure of a growth hormone-prolactin receptor complex, *Nature 372*, 478-481.

16. Polverino de Laureto, P., Toma, S., Tonon, G., and Fontana, A. (1995) Probing the structure of human growth hormone by limited proteolysis, *Int. J. Pept. Protein Res. 45*, 200–208.

17. Spolaore, B., Polverino de Laureto, P., Zambonin, M., and Fontana, A. (2004) Limited proteolysis of human growth hormone at low pH: isolation, characterization, and complementation of the two biologically relevant fragments 1-44 and 45-191, *Biochemistry 43*, 6576–6586.
18. Kashiwagi, T., Yokoyama, K., Ishikawa, K., Ono, K., Ejima, D., Matsui, H., and Suzuki, E. (2002) Crystal structure of microbial transglutaminase from *Streptoverticillium mobaraensis, J. Biol. Chem. 277*, 44252–44260.

19. 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.

20. Fontana, A., Polverino de Laureto, P., De Filippis, V., Scaramella, E., and Zambonin, M. (1999) Limited proteolysis in the study of protein conformation. In *Proteolytic Enzymes: Tools and Targets* (Sterchi, E. E., and Stöcker, W., Eds.), pp. 257–284, Springer Verlag, Heidelberg.

21. 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.

22. 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.

23. Akbarzadeh, S., and Layton, J. E. (2001) Granulocyte colony-stimulating factor receptor: Structure and function, Vitam. Horm. 63, 159–194.

24. Morstyn, G., and Dexter, T. (1994) Neopogen (r-metHuG-CSF) in clinical practice, Marcel Dekker, New York.

25. Piedmonte, D. M., and Treuheit, M. J. (2008) Formulation of Neulasta (pegfilgrastim), *Adv. Drug Deliv. Rev. 60*, 50–58.

26. Molineux, G. (2004) The design and development of pegfilgrastim (PEG-rmetHuG-CSF, Neulasta), *Curr. Pharm. Des. 10*, 1235–1244.

27. Lubenau, H., Bias, P., Maly, A. K., Siegler, K. E., and Mehltretter, K. (2009) Pharmacokinetic and pharmacodynamic profile of new biosimilar filgrastim XM02 equivalent to marketed filgrastim Neupogen: single-blind, randomized, crossover trial, *BioDrugs 23*, 43–51.

28. Herman, A. C., Boone, T. C., and Lu, H. S. (1996) Characterization, formulation and stability of Neupogen (Filgrastim), a recombinant human granulocyte-colony stimulating factor, *Pharm. Biotechnol. 9*, 303–328.

29. Hill, C. P., Osslund, T. D., and Eisenberg, D. (1993) The structure of granulocyte-colony-stimulating factor and its relationship to other growth factors, *Proc. Natl. Acad. Sci. USA 90*, 5167–5171.

30. Tonon, G., and Orsini, G. (2008) G-CSF site-specific mono-conjugates, Patent WO/2008/7017603, Int. Application No. PCT/EP2007/057824.

31. Maullu, C., Raimondo, D., Caboi, F., Giorgetti, A., Sergi, M., Valentini, M., Tonon, G., and Tramontano, A. (2009) Site-directed enzymatic PEGylation of the human granulocyte colony-stimulating factor, *FEBS J. 276*, 6741–6750.

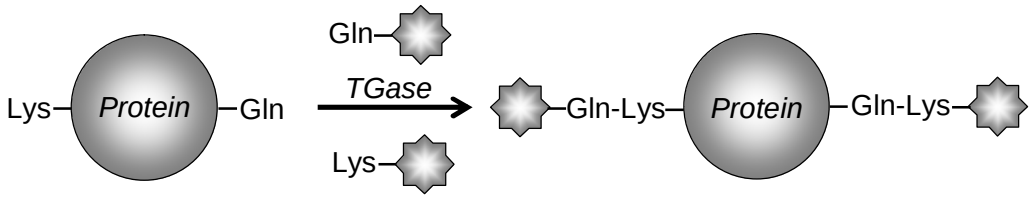
32. DeFrees, S., Wang, Z. G., Xing, R., Scott, A. E., Wang, J., Zopf, D., Gouty, D. L., Sjoberg, E. R., Panneerselvam, K., Brinkman-Van der Linden, E. C., Bayer, R. J., Tarp, M. A., and Clausen, H. (2006) GlycoPEGylation of recombinant therapeutic proteins produced in *Escherichia coli*, *Glycobiology 16*, 833–843.

33. Zink, T., Ross, A., Luers, K., Cieslar, C., Rudolph, R., and Holak, T. A. (1994) Structure and dynamics of the human granulocyte colony-stimulating factor determined by NMR spectroscopy: Loop mobility in a four-helix-bundle protein, *Biochemistry 33*, 8453–8463.

34. Martin, A. J., Walsh, I., and Tosatto, S. C. (2010) MOBI: a web server to define and visualize structural mobility in NMR protein ensembles, *Bioinformatics 26*, 2916–2917.

IV. GENERAL DISCUSSION AND CONCLUSIONS

The main aim of this study was to deduce molecular features of the TGase-mediated reactions using well known protein substrates, as apomyoglobin (apoMb), lysozyme (LYS) and ribonuclease (RNase). Using both amino- and amido-ligands mimicking the ε-amino group of lysine (Lys) and glutamine (Gln) residues, respectively, it was possible to analyse the TGase-mediated reactions in terms of molecular and dynamic properties of the investigated protein substrates. The Scheme below (Figure. 1) illustrates the permutations of substrates of the TGase reactions herewith investigated, so that it was possible to identify both TGase-reactive Gln and Lys residues in a protein and to relate the observed reactive residues in terms of molecular properties of the protein substrate. In general, we have here convincingly shown with a number of TGase reactions, performed under different experimental conditions and various protein substrates, that local flexibility or even unfolding dictates the site-specific modification of a Gln residue in a protein, while a less stringent specificity was observed for Lys residues.



TGase-mediated modification of proteins

Two approaches are possible: either a ligand derivatised with Gln-residue can be coupled to a Lys-residue of the protein or a Lys-mimicking ligand can be attached to a Gln-residue. The ligand can be a fluorescent probe, biotin, a drug entity or a chelating agent.

Figure 1. Scheme of the TGase-mediated reactions.

Mechanism of TGase catalysis.The catalytic mechanism of TGase is related to that of Cys-proteases and the same catalytic triad given by the active site residues Cys, Asp and His is involved in both TGase- and protease-mediated catalysis (*1, 2*). The TGase's reaction involve first formation of a thiol ester at the level of the active site Cys residue of the enzyme by a protein/peptide bound Gln residue with concomitant release of ammonia. Subsequently, the reactive thiol ester reacts with a nucleophile that acts as an acyl acceptor. Usually, these nucleophile moieties *in vivo* are the ε-amino groups of lysine residues, leading to the formation of isopeptide bonds linking the side chains of Gln and Lys residues and consequent protein crosslinking. However, besides Lys residues acting as amine donors, a great variety of other nucleophiles can react with the thiol ester intermediate, namely many primary amines, such as cadaverin, dansyl- or biotin-cadaverin, methyl amine, glycine ethyl ester, hydroxylamine, hydrazine and other amine-containing reagents (*3, 4*), as well as polymers bearing an amino group as PEG-NH_2 (*5, 6*). When there is no amine donor available for acting as a nucleophile, water can react with the thiol ester, leading to hydrolysis of the reactive Gln residue to a glutamic acid (Glu) residue.

Considering that Cys64 in microbial TGase is located in a deep cleft of the protein fold (*1, 2*) (Figure. 2), it can be anticipated that the most critical step in the TGase-mediated reaction(s) is the formation of a thiol ester at the level of a protein-bound Gln residue and that there is no specificity of the nucleophile attacking the reactive thiol ester intermediate. Indeed, it has been found that a large variety of small nucleophiles can react with the intermediate thiol ester intermediate, including water, hydroxylamine, alkyl amines and others. On the other hand, when the nucleophile is the ε-amino group of a Lys residue embedded in a protein substrate, it can be anticipated that steric requirements, surface accessibility, nucleophilicity, as well as mobility of the ε-amino group of Lys residues can play a role in the aminolysis of the reactive thiol ester. This view of the catalytic mechanism of TGase is consistent with the fact that, in general, there is little specificity for protein-bound Lys residues as amino-donors and, at the same time, it appears to explain the often observed stringent specificity for protein bound Gln residues as acyl-donors.

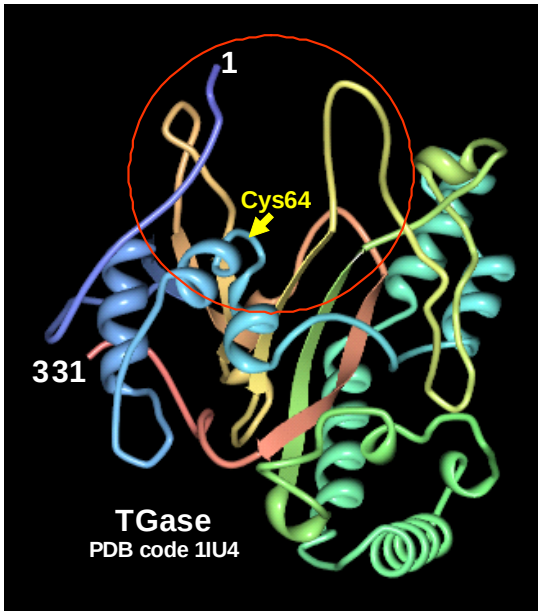


Figure 2. Three-dimensional structure of microbial TGase from *Streptomyces mobaraense*. The protein model was prepared from the X-ray structure of the enzyme using the software MBT (Molecular Biology Toolkit; http://mbt.sdsc.edu) available in PDB (code 1IU4). The 331-residue chain of TGase is depicted in rainbow colours from the N-terminus (blue) to the C-terminus (red). The active site area of the enzyme is circled with a red line and the location of the deeply buried Cys64 is indicated by an arrow.

Specificity for Gln. In several instances it has been demonstrated that TGase reacts with Gln-residues located at the chain ends of a protein substrate. Clearly, the N- and C-terminal regions of a globular protein usually are rather flexible or locally unfolded and thus able to interact easily at the TGase's active site. We may mention here that the TGase-reactive Gln residues of the 466-residue protein vimentin are Gln453 and Gln560 (*7*). The Gln-residues acting as acyl-donors sites in three β-crystallin chains have been found Gln9 in crystallin βB2, Gln21 in crystallin βB3 and Gln23 and

Gln24 in crystallin βA3 (*8*). All these Gln residues are located in the N-terminal extensions of the crystallins chains and extend out from the compact two-domain structure of β-crystallin. The 317-residue chain of the nuclear factor KB (NF-KB) forms polymers in the presence of TGase (*9*). A very detailed analysis of the molecular identity of the NF-KB polymers conducted by fingerprinting experiments and MALDI mass spectrometry revealed that the TGase-reactive residues are Gln266, Gln267 and Gln317, all residues located at the C-terminal region of the protein. On the other hand, these analyses revealed that there is no clear preference for reactive Lys residues in the polymer formation by NF-KB, even if a minor preference for the crosslinking reactions was for Lys21, Lys22 and Lys177. Of note, the C-terminal region of NF-KB encompassing the TGase-reactive Gln residues contains a PEST sequence, this last being a site of preferential proteolytic attack and thus usually associated with rapid protein turnover (*10*).

Here, we can add strength to the proposal that chain flexibility dictates TGase's reactions by commenting also the results of other experimental studies that demonstrate that TGase can react with globular proteins at the level of Gln residues embedded in flexible/unfolded regions of polypeptide chains. First of all, in a number of studies, relatively short Gln-peptides were added to globular proteins as an N- and C-terminal tails using recombinant methods (*11-13*). It was found that TGase mediates enzymatic reactions selectively at the protein end(s) of these chimeric proteins. The added short peptide tags, lying outside the core of the globular protein, certainly are flexible or disordered, considering that the N- and C-terminal ends of globular proteins are usually rather flexible or even disordered, as amply demonstrated by the profile along the polypeptide chain of *B*-factor values in many protein structures reported in PDB (*14*). In general, the protein chain ends display high *B*-factor values or are even fully disordered, since no electron density is available for the N- and C-termini.

There are numerous examples of requirements of protein flexibility for TGase-mediated reactions. For example, TGase does not react with a native rigid protein, while it can react on a protein substrate being partly or fully denatured/unfolded. The apo-form of bovine α-lactalbumin in its partly folded or molten globule state at neutral pH was shown to react with TGase, while the native calcium-loaded protein was not a substrate for TGase (*15-17*). A variety of natively or intrinsically disordered proteins (*18-21*) are substrates for TGase *in vivo*, including α-synuclein (*22*), tau protein (*23*) and huntingtin (*24*). Of interest, these proteins are involved in severe diseases as those of Alzheimer, Parkinson and Huntington and consequently a role of TGase in these diseases has been proposed (*25*).

TGase is a reverse protease. We have shown here also that the sites of TGase-mediated reactions occur at the level of regions prone to suffer also limited proteolysis phenomena, as elegantly demonstrated by the apoMb case (*26, 27*). Of interest, we interpret the experimental data herewith reported in terms of chain flexibility or local unfolding as dictating the TGase's selectivity, in analogy to previous explanations for the site-specific limited proteolysis of globular proteins (*28-31*). Therefore, a mechanism of local unfolding of the site of enzymatic attack appears to be the critical parameter dictating the site-specific enzymatic reactions for both

TGases and proteases. Considering the molecular similarities between TGase-mediated reactions and proteolysis, it seems appropriate to use also for the TGase reaction the nomenclature of Schechter and Berger (*32*) previously utilized for describing the primary and secondary binding sites in protease-substrate interactions, with the residues at the N-terminal side of the residue of enzymatic attack being P1, P2, P3.... and those at the C-terminal side P', P2', P3' .... As shown in Figure 3, a stretch of at least 10-12 amino acid residues encompassing the site of enzymatic reaction should be in a flexible conformation to allow binding at the enzyme's catalytic site in an extended backbone geometry. Indeed, a recent survey of the X-ray structures of a number of proteases with bound peptidic inhibitors and substrates revealed that these last in their bonded state adopt an extended conformation, without intra-chain hydrogen bonding, but extensive hydrogen bonding between their backbones and the backbones or side chains of the protease (*33*).

Clearly, local unfolding in the TGase's reaction is the most critical parameter when globular proteins are used as substrates. However, when short flexible peptides are used in a TGase's reaction, chain flexibility is no more controlling the reactivity of individual Gln residues and the rate of reaction. In this case, the amino acid sequence near a Gln-residue is expected to play an important role in dictating the proper and favourable interaction between the enzyme and the peptide substrate, as illustrated by the scheme shown in Figure 3.

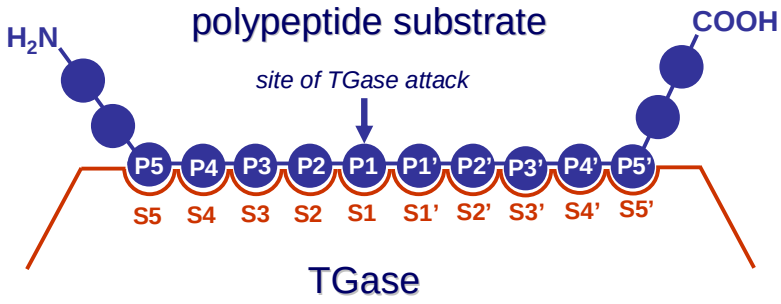


Figure 3. Schematic representation of the binding of a polypeptide substrate at the active site of TGase. A 10-12 residue segment of a polypeptide chain interacts with its side-chain residues (P) at a series of subsites (S) of TGase. The interaction of the substrate at the TGase's active requires a specific stereochemical adaptation of the substrate and thus likely a significant degree of chain mobility. It is suggested that the polypeptide substrate binds at the TGase's active site in an extended conformation. The P1 side-chain residue interacting with the S1 binding site of TGase is the carboxamido side-chain group of a Gln residue. 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 (*32*).

Therefore, the kinetics of the enzymatic reaction at the level of Gln residue(s) will be influenced by the amino acid residues neighbouring the P1 site af attack (see Figure 3). Indeed, in a recent study, the preferred peptide substrate sequences of microbial

TGase have been identified by using a phage-display library and some peptide sequences were shown to react faster with TGase than others (*34*). Considering the features of the microenvironment of the active site of microbial TGase, it can be anticipated that charge and hydrophobicity of a polypeptide substrate will play a role in the enzyme's kinetics. In particular, since the active site of microbial TGase is negatively charged by the presence of several carboxylate groups near the reactive Cys64 (*1, 2*), likely negatively or positively charged residues near a Gln-residue will hinder or facilitate the enzymatic reaction on a polypeptide substrate, respectively. Indeed, the flexible peptidyl linkers containing a positively charged arginine (Arg) residue near the Gln-residue exhibited the highest reactivity in a TGase-mediated protein crosslinking reaction (*35*). Similarly, in our laboratory we have found that a negatively charged glutamic acid (Glu) residue near a Gln-residue strongly hinders the TGase reaction in a Gln-peptide substrate (unpublished).

Reactivity of Lys. The results herewith obtained and also reported in other studies (*3,4*) have indicated that that there is only a moderate preference for some Lys residues in protein substrates in TGase-mediated reactions. This is in keeping with the view of the general mechanism of the TGase's enzymatic reaction (see above) and also by considering that the ε-amino group of a protein-bound Lys residue is anyway rather exposed and flexible due to the relatively long arm (spacer) of the $-(CH_2)_4-$ alkyl chain. Nevertheless, we have herewith described the specific case of a Lys residue identically located at the N-terminal end of RNase and LYS being reactive in one case (RNase) and fully resistant to the TGase attack in the other (LYS) (see also ref. *36*). Of interest, this opposite behaviour in the TGase reaction can be elegantly explained by examining in detail the structural and dynamic features of both RNase and LYS deduced from their crystallographically determined structures. *Fig. 4* shows the schematic 3D structures of bovine pancreatic RNase and egg-white LYS, together with their profiles of *B*-factor values along their 124- and 129-residue chain, respectively. While the N-terminal segment of LYS displays very low *B*-factor values and thus is very rigid, the N-terminal segment in RNase is instead very mobile, as given by the high *B*-factor values (*37, 38*) of this segment. Therefore, we can conclude that also for a protein-bound Lys reactivity in a TGase-mediated reaction there is a requirement of substrate's flexibility.

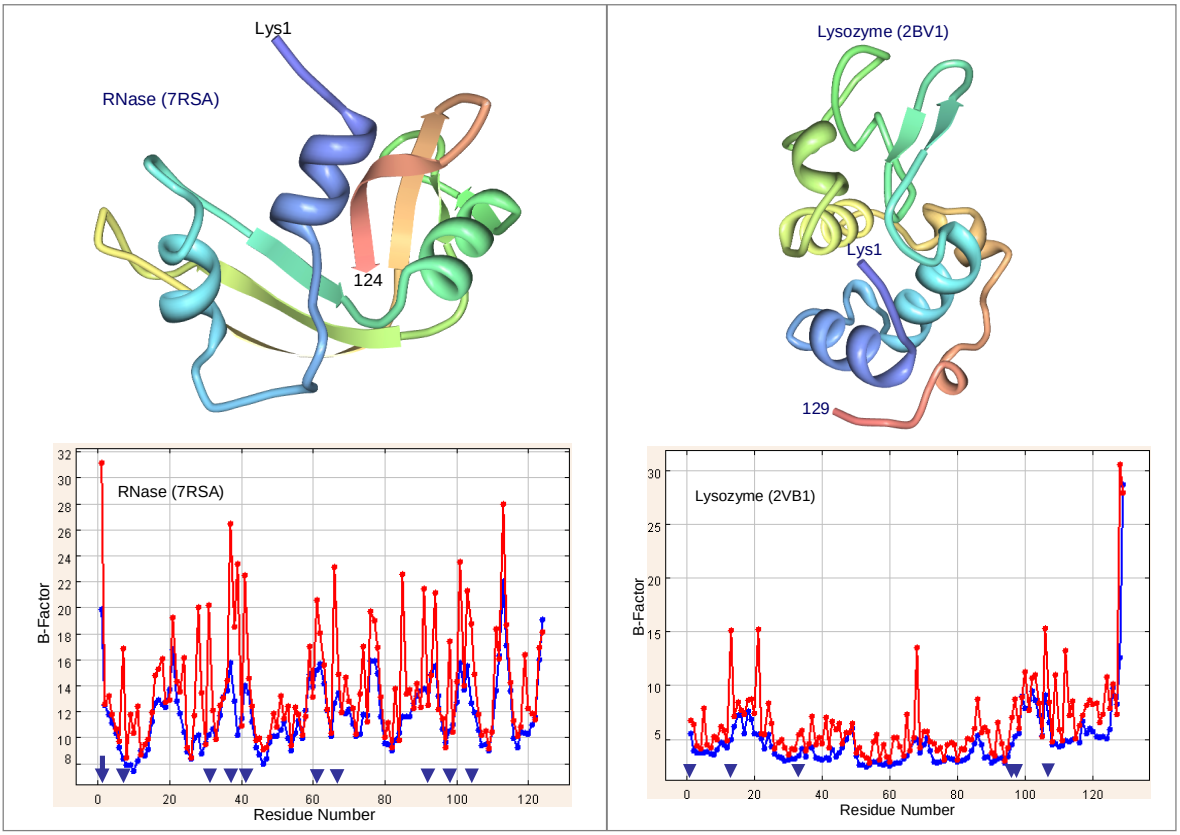


Figure 4. Three-dimensional structures and segmental mobility of bovine pancreatic ribonuclease A (RNase) and egg-white lsozyme (LYS). (***Top***) Schematic models of RNase (PDB code 7RSA) and egg-white lysozyme (PDB code 2BV1). The protein models were prepared from the X-ray structures of the proteins using the software MBT (Molecular Biology Toolkit; http://mbt.sdsc.edu) available in PDB. The polypeptide chains of the two proteins are depicted in rainbow colours from the N-terminus (blue) to the C-terminus (red). (***Bottom***) Plot of the *B*-factor values along the polypeptide chains of ribonuclease (124 residues) and lysozyme (129 residues). Arrows at the bottom of the diagram indicate the location of Lys residues along the polypeptide chains of the two proteins.

TGase-mediated PEGylation. Among the techniques so far explored for the development of safer and more useful protein drugs, undeniably the protein surface modification by covalent attachment of poly(ethylene glycol) (PEG) became an extremely valuable technique for producing protein drugs more water-soluble, non-aggregating, non-immunogenic and more stable to proteolytic digestion (*39-41*). The broad applicability and comparably low cost of PEG will maintain this polymer in a leading position and PEGylation will find more and more useful applications for the successful development of pharmaceutical drugs. The results here obtained and those of previous studies (*5*, *6*) have indicated that the PEGylation reaction can be achieved by means of TGase and using an amino-derivative of PEG (PEG-NH_2) as a substrate for the enzymatic reaction. A specific advantage of the TGase method lies in the possibility of achieving a site-specific PEGylation, while the chemical methods so far

in use lead to over-labelling and heterogeneity of products (*39*). Considering the increasing relevance of protein pharmaceuticals and the high regulatory demands for their approval, it can be anticipated that the innovative TGase method for the site-specific PEGylation of proteins will be further investigated for the years to come.

Conclusions. Here, we have presented data that indicate that the main features dictating the site-specific modification of a protein-bound Gln residue by TGase in a globular protein is the flexibility or local unfolding of the chain region encompassing the reactive Gln residue. In particular, TGase appears to act on a polypeptide substrate in analogy to a protease and, in fact, often the same region of the polypeptide chain of a globular protein that suffers limited proteolysis is also the site of a specific TGase attack. Therefore, it is possible to predict the site(s) of TGase-mediated modification of a protein on the basis of its 3D-structure and dynamics and, consequently, the likely effects on its physicochemical and functional properties. An important development of these studies resides in the TGase-mediated approach for obtaining a site-specific PEGylation of valuable protein drugs using PEG-NH_2. The enzymatic PEGylation method appears to be mild and, more importantly, leads to homogeneous protein conjugates. Summing up, we can anticipate that the TGase-mediated reactions will be further explored as a versatile method of enzymatic protein modification for the purpose of studies of protein structure and function, as well as a method to improve the properties of protein drugs (*42*).

References

1. Ando, H., Adachi, M., Umeda, K., Matsuura, A., Nonaka, M., Uchio, R., Tanaka, H., and Motoki, M. (1989) Purification and characterization of a novel transglutaminase derived from microorganisms, *Agric. Biol. Chem. 53*, 2613–2617.

2. Kanaji, T., Ozaki, H., Takao, T., Kawajiri, H., Ide, H., Motoki, M., and Shimonishi, Y. (1993) Primary structure of microbial transglutaminase from *Streptoverticillium sp.* strain s-8112, *J. Biol. Chem. 268*, 11565–11572.

3. Folk, J.E. (1980) Transglutaminases, *Annu. Rev. Biochem. 49*, 517–531.

4. Lorand, L., and Conrad, S. M. (1984) Transglutaminases, *Mol. Cell. Biochem. 58*, 9–35.

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

6. 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, *Bioconjug. Chem. 20*, 384–389.

7. Clément, S., Velasco, P. T., Murthy, S. N., Wilson, J. H., Lukas, T. J., Goldman, R. D., and Lorand, L. (1998) The intermediate filament protein, vimentin, in the lens is a target for cross-linking by transglutaminases, *J. Biol. Chem. 273*, 604–7609.

8. Berbers, G. A., Bentlage, H. C., Brans, A. M., Bloemendal, H., and de Jong, W.W. (1983) Beta-crystallin: endogenous substrate of lens transglutaminases: Characterization of the acyl-donor site in the beta Bp chain, *Eur. J. Biochem. 135*, 315–320.

9. Park, S. S., Kim, J. M., Kim, D. S., Kim, I. H., and Kim, S.Y. (2006) Transglutaminase-2 mediates polymer formation of NF-KB through C-terminal glutamine cluster. *J. Biol. Chem. 281*, 34965-34972.

10. Rechsteiner, M, and Rogers, S. W. (1996) PEST sequences and regulation by proteolysis, *Trends Biochem Sci. 21*, 267–271.

11. Taki, M., Shiota, M., and Taira, K. (2004) Transglutaminase-mediated N- and C-terminal fluorescein labelling of a protein can support the activity of the modified protein, *Protein Eng. Des. Select. 17*, 119–126.

12. Tanaka, T., Kamiya, N., and Nagamune, T. (2004) Peptidyl linkers for protein heterodimerization catalyzed by microbial transglutaminase, *Bioconjug. Chem. 15*, 491–497.

13. Meusel, M. (2004) Synthesis of hapten-protein conjugates using microbial transglutaminase, *Methods Mol. Biol. 283*, 109–123.

14. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000) The protein data bank, *Nucleic Acids Res. 28*, 235–242.

15. Gu, Y.S., Matsumura, Y., Yamaguchi, S., and Mori, T. (2001) Action of protein-glutaminase on α-lactalbumin in the native and molten globule states, *J. Agric. Food Chem. 49*, 5999–6005.

16. Nieuwenhuisen, W.F., Dekker, H.L., De Koning, L.J., Groneveld, 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.

17. Lee, D.S., Matsumoto, S., Matsumura, Y., and Mori, T. (2002) Identification of the ε-(γ-glutamyl)lysine crosslinking sites in α-lactalbumin polymerized by mammalian and microbial transglutaminases, *J. Agric. Food Chem. 50*, 7412–7419.

18. Dyson, H. J., and Wright, P .E. (2002) Coupling of folding and binding for unstructured proteins, *Curr. Opin. Struct. Biol. 12*, 54–60.

19. 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.

20. Uversky, V. N. (2002) Natively unfolded proteins: A point where biology waits for physics, *Protein Sci. 11*, 739–756.
21. Tompa, P. (2002) Intrinsically unstructured proteins, *Trends Biochem. Sci. 27*, 527–533.

22. Junn, E., Ronchetti, R. D., Quezabo, M. M., Kim, S. Y., and Mouradian, M. M. (2003) Tissue transglutaminase-induced aggregation of α-synuclein: Implications for Lewy body formation in Parkinson's disease and dementia with Lewy bodies, *Proc. Natl. Acad. Sci. USA 100*, 2047–2052.

23. Prasana Murthy, S. N., Wilson, J. H., Lukas, T. J., Kuret, J., and Lorand, L. (1998) Crosslinking sites of the human tau protein probed by reactions with human transglutaminase, *J. Neurochem. 71*, 2607–2614.

24. Karpuj, M. V., Garren, H., Slunt, H., Price, D. L., Gusella, J., Becker, M. W., and Steinman, L. (1999) Transglutaminase aggregates huntingtin into non-amyloidogenic polymers and its enzymatic activity increases in Huntington's disease brain nuclei, *Proc. Natl. Acad. Sci. USA 96*, 7388–7393.

25. Karpuj, M., and Steinman, L. (2004) The multifaceted role of transglutaminase in neurodegeneration, *Amino Acids 26*, 373–379.

26. 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.

27. Picotti, P., Marabotti, A., Negro, A., Musi, V., Spolaore, B., Zambonin, M., and Fontana, A. (2004) Modulation of the structural integrity of helix F in apomyoglobin by single amino acid replacements, *Protein Sci. 13*, 1572–1585.

28. 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.

29. Fontana, A., Polverino de Laureto, P., De Filippis, V., Scaramella, E., and Zambonin, M. (1999) Limited proteolysis in the study of protein conformation. In *Proteolytic Enzymes: Tools and Targets* (Sterchi, E.E., and Stöcker, W., Eds.), Springer Verlag, Heidelberg, pp. 257–284.

30. 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.

31. 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.

32. Schechter, I., and Berger, A. (1967) On the size of the active site in proteases. I. Papain, *Biochem. Biophys. Res. Commun. 27*, 157–162.

33. Tyndall, J. D. A., Nall, T., and Fairlie, D. P. (2005) Proteases universally recognize beta strands in their active site, *Chem. Rev. 105*, 973–999.

34. 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.

35. Tanaka, T., Kamiya, N., and Nagamune, T. (2004) Peptidyl linkers for protein heterodimerization catalyzed by microbial transglutaminase, *Bioconjug. Chem. 15*, 491–497.

36. 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.

37. Ringe, D. and Petsko, G.A. (1986) Study of protein dynamics by X-ray diffraction, *Methods Enzymol. 131*, 389–433.

38. Kundu, S., Melton, J.S., Sorensen, D.C., and Phillips Jr., G.N. (2002) Dynamics of proteins in crystals: Comparison of experiment with simple models, *Biophys. J. 83*, 723–732.

39. Veronese, F. M., and Pasut, G. (2005) PEGylation, successful approach to drug delivery, *Drug Discovery Today 10*, 1451–1458.

40. Harris J. M., and Veronese, F. M. (Eds.) (2002) Peptide and protein PEGylation, *Adv. Drug Deliv. Rev. 54*, 453–610.

41. Harris, J. M., and Chess, R. (2003) Effect of PEGylation on pharmaceuticals, *Nature Rev. Drug Discov. 2*, 214–221.

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