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
XXIV CICLO

Transglutaminase-mediated modification of proteins: Molecular mechanisms and metal-labelling of proteins

Direttore della scuola: Ch.mo Prof. Giuseppe ZANOTTI
Coordinatore d'indirizzio: Ch.mo Prof. Giorgio VALLE
Supervisore: Ch.mo Prof. Angelo FONTANA

Dottorando : Nunzio DAMIANO

In general, scientific progress calls for no more than the absorption and elaboration of new ideas and this is a call most scientists are happy to heed.

Werner Heisenberg

Dedicated to

My Mother…

Acknowledgements

After three years that I spent to work at the Proteomics facility of the CRIBI, I met many people, very different but united by their diligence and expertise. To thank someone is always difficult. It is difficult because the risk to be inopportune is always possible, but I will try in this with the hope to be descreet.

First of all, I wish to thank Prof. Angelo Fontana for giving me the opportunity to perform my doctorate in his group. Working under his supervision allowed me to gain deep insight into protein chemistry. I would like to thank him for his immense patience and for providing me the opportunities to attend the congresses and summer schools.

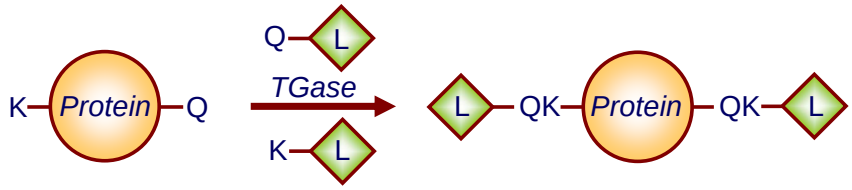
I would like to thank Barbara Spolaore that always helped me when I needed to be supported. I would like to thank her for the encouragement and the scientific guidance. I am really deeply grateful to the researchers Patrizia Polverino de Laureto, Giorgia Defranceschi and Erica Frare for their support in the laboratory and for teaching me from their huge experience. Finally, I wish to extend my thanks to the whole group members and all the visiting PhD and thesis students, Abhijeet Ajit Satwekar, Micaela Pivato, Chiara Fecchio, Giacomo Forzato, Odra Pinato, that gave me many nice and happy moments during these years.

I wish to thank all my friends from Napoli, in particular Fabio Burzo and Rosario Finamore that always gave me their support during this long time faraway from my city. Finally, I wish to thank my mother, for her love and support in every situation.

Table of Contents

Abstract

Transglutaminase (TGase) is a 331-residue monomeric microbial enzyme produced by recombinant methods, easily available from a commercial source and well characterized in structural terms. TGase 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. This enzymatic approach appears to be mild and specific for glutamine (Gln) and lysine (Lys) residues of a protein. Indeed, initial experiments have already established that TGase can be used for an effective modification of proteins. In this PhD project we have used microbial TGase for the site-specific modification of proteins at the level of Gln or Lys residue(s) by using different substrates for the enzymatic reaction. In particular, we aimed also to prepare protein derivatives containing covalently-bound peptide moieties capable of chelating metals and radionuclides.



TGase-mediated modification of a protein

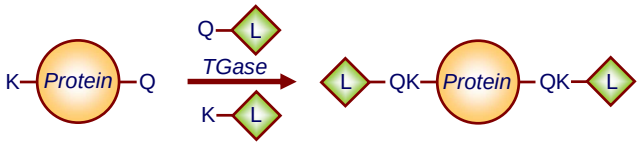
Two approaches are possible, *i.e.*, (1) a ligand (L) derivatised with a glutamine (Q) residue can be coupled by TGase to a protein-bound lysine (K) residue or (2) a lysine-mimicking ligand (*e.g.*, a primary alkylamine) can be attached to a glutamine residue. The ligand can be a fluorescent probe, a biotin derivative, a drug entity or even a polymer (*e.g.*, PEG-NH_2).

The TGase-mediated labelling of proteins has been examined in detail using protein substrates of known 3D structure and dynamics in order to unravel the conformational and dynamic features that dictate preferential or specific protein modification by TGase. Avidin and apo-α-lactalbumin were respectively modified with carbobenzoxy-glutaminyl-glycine (ZQG) and dansyl-cadaverine. In both cases, modified proteins were purified by RP-HPLC and analyzed by mass spectrometry (MS) in order to identify the sites of the modification. The results of these analyses indicated that in the case of avidin there is a double modification of the protein, while in the case of apo-α-lactalbumin there is only one site of modification. Interestingly, both proteins are derivatised by TGase at the level of Gln or Lys residues located at flexible regions of their structure, as deduced from the crystallographically determined *B*-factor profile along the polypeptide chain of the protein. Moreover, it was shown that there is a correlation between sites of TGase attack and sites of limited proteolysis, indicating that both TGase and a protease require some local unfolding of the protein substrate for their selective enzymatic reactions.

We have also designed and synthesised two peptides that can function as suitable metal-chelating agents. These chelating peptides contain a cysteine (Cys) residue at the C-terminus that can coordinate metals as technetium (99mTc) or rhenium (186/188Re) with the thiol group and C-terminal carboxylic group acting as ligands for the metal. Moreover, these peptides contain a Lys or Gln residue that can be exploited for their TGase-mediated conjugation to proteins. The complexes of these peptides with 99mTc and 185/187Re were purified and characterised in terms of metal binding properties. Experiments are now in progress in order to use TGase for the site-specific conjugation of these metal-chelating peptides to proteins that can be used as imaging agents.

RIASSUNTO

La transglutaminasi (TGase) microbica è un enzima di 331 residui amminoacidici prodotta con metodi ricombinanti, facilmente reperibili da una fonte commerciale e ben caratterizzata in termini strutturali. La TGase catalizza la reazione tra il gruppo γ-ammidico di un residuo di glutammina (Gln) (-$CONH_2$, accettore) ed un gruppo amminico (-NH_2, donatore) di una alchil-ammina, ad es., il gruppo ε-amminico di una lisina (Lys). Le reazioni mediate da TGase avvengono in condizioni blande ed inoltre con notevole specificità, soprattutto per i residui di Gln. La TGase può essere utilizzata per la modifica enzimatica di proteine a livello di Gln, ma anche di Lys. Lo Schema illustra la possibilità di utilizzare nella reazione di una proteina con TGase un ligando in grado di mimare sia la catena laterale di Gln che Lys, potendo in tal modo, con opportuni ligandi, introdurre covalentemente in una proteina un gruppo fluorescente, un farmaco o anche un complesso metallo-peptide.



TGase-mediated modification of a protein

In questa Tesi di dottorato la TGase è stata usata per la modifica di proteine modello, tra cui α-lattalbumina, avidina ed apomioglobina (apoMb). Lo scopo della ricerca è stato quello di comprendere i motivi strutturali che determinano la specificità di azione della TGase. Inoltre, la modifica di proteine con TGase è stata condotta utilizzando substrati peptidici in grado di legare metalli e radionuclidi.

La modifica di proteine con TGase è stata esaminata in dettaglio utilizzando proteine di struttura 3D e dinamiche note, al fine di individuare le caratteristiche conformazionali e dinamiche che determinano la modifica preferenziale o specifica. L'avidina e l'apo-α-lattalbumina sono state modificate utilizzando carbobenzoxy-glutaminyl-glicina (ZQG) (donatore) e dansyl-cadaverina (accettore). Le proteine modificate sono state purificate mediante RP-HPLC ed analizzate mediante spettrometria di massa (MS) al fine di individuare i siti di modifica. I risultati di queste analisi hanno indicato che le modifiche di queste proteine avvengono a livello di residui di Gln e Lys localizzati in regioni flessibili della loro struttura, come si deduce dai valori di fattore-*B* (parametro di flessibilità) della loro catena polipeptidica. Inoltre, è stato dimostrato che esiste una correlazione tra i siti di attacco di TGase ed i siti di proteolisi limitata, indicando che sia la TGase che una proteasi richiedono una certa flessibilità del substrato polipeptidico affinchè la reazione possa avvenire.

Sono stati anche progettati e sintetizzati due peptidi in grado di fare complessi con metalli. Questi peptidi, contenendo una cisteina (Cys) quale residuo al C-terminale, sono in grado di coordinare metalli con il gruppo tiolo ed il gruppo carbossilico, tra cui tecnezio (99mTc) o renio (186/188Re). Inoltre, questi peptidi contenevano un residuo di Gln o Lys, ai fini di un riconoscimento da parte della TGase per la loro coniugazione covalente ad una proteina. I complessi di questi peptidi con 99mTc e 185/187Re sono stati purificati e caratterizzati in termini di proprietà leganti per i metalli ed utilizzati per la modifica di proteine. Ulteriori esperimenti sono in corso al fine di utilizzare la TGase per la modifica sito-specifica di proteine con questi metallo-peptidi, al fine di sviluppare una strategia di produzione di proteine marcate da usare per analisi di *imagin*.

List of abbreviations

aa	amino acid
ACN	acetonitrile
AEBSF	4-(2-aminoethyl)benzenesulfonyl fluoride hydrochloride
apoMb	apomyoglobin
apoMb	apomyoglobin
BFCA	bifunctional chelating agent
CM	carboxymethylation
DC	dansyl-cadaverine
DCM	dichloromethane
DIPEA	N,N-diisopropylethylamine
DMF	dimethylformamide
E:S	enzyme to substrate ratio
ESI	electrospray ionization
Et_2O	diethyl ether
Et_3N	tryethyl amine
Fmoc	9-fluorenylmethyloxycarbonyl
Gdn-HCl	guanidine hydrochloride
GM-CSF	granulocyte macrophage colony stimulating factor
HATU	O-(7-azabenzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate
LA	α-lactalbumin
LC-MS/MS	liquid chromatography tandem mass spectrometry
m/z	mass to charge ratio
MS	mass spectrometry
MS/MS	tandem mass spectrometry
MW	molecular weight
NH_4HCO_3	ammonium bicarbonate
NMP	N-methyl-2-pyrrolidone
PAGE	polyacrylamide gel electrophoresis
PEG	poly(ethylene) glycol
PET	positron emission tomography
PNP	amino diphosphine
Q-Tof	quadrupole-time of flight
RCY	radiochemical yield
ReK	($^{185\text{-}187}$Re(N)Cl2(PNP))(Biotin-GKGC)
ReQ	($^{185\text{-}187}$Re(N)Cl2(PNP))(Biotin-GQGC)
RP-HPLC	reverse phase high pressure liquid chromatography
RT	retention time
SDH	succinyldihydrazide
SDS	sodium dodecyl sulphate
SPECT	single photon emission computed tomography
TEMED	tetra(methylethylene)diamine
TFA	trifluoroacetic acid
TFE	trifluoroethanol
TGase	transglutaminase
Tris	tris(hydroxymethyl)aminomethane
UV	ultraviolet
wt	wild type
ZQG	Carbobenzoxy-Glutamyl-Glycine (CBZ-Gln-Gly)

Abbreviations of amino acids

	Three letter code	*Single letter code*
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

I. INTRODUCTION

1. Transglutaminase (TGase)

A transglutaminase (TGase; EC 2.3.2.13) is an enzyme that catalyzes the isopeptide bond formation between glutamine (Gln) and lysine (Lys) residues in proteins, thus leading to protein crosslinking (*1–7*). The term TGase was first introduced for the transamidating activity detected in the liver of guinea-pig. In numerous subsequent studies it has been demonstrated that the TGase-mediated crosslinking occurs through an acyl transfer reaction between the γ-carboxamide group of a protein/peptide bound Gln residue and the ε–amino group of a protein/peptide bound Lys residue, the product of the reaction being an ε–(γ-glutamyl) lysine isopeptide bond. The TGase-mediated reaction involves the concomitant release of ammonia (A-$CONH_2$ + $_2$HN-B → A-CONH-B + NH_3). The formation of isopeptide bonds results in both intra-and inter-molecular cross-linking of proteins, the inter-molecular one leading to protein polymerization (*7,8*). The side-chain amide group of a Gln residue acts as an acyl-donor via the intermediate formation of a reactive thioester, which in turn reacts with an alkylamine as the ε-amino side chain of a Lys residue. Alternatively, other nucleophiles can react with the thioester, such as primary alkyl-amines or even water. In this last case, the protein bound Gln residue is deamidated to a glutamic acid (Glu) residue (*9–14*). The Scheme in Fig. 1 illustrates the various reactions catalysed by TGase.

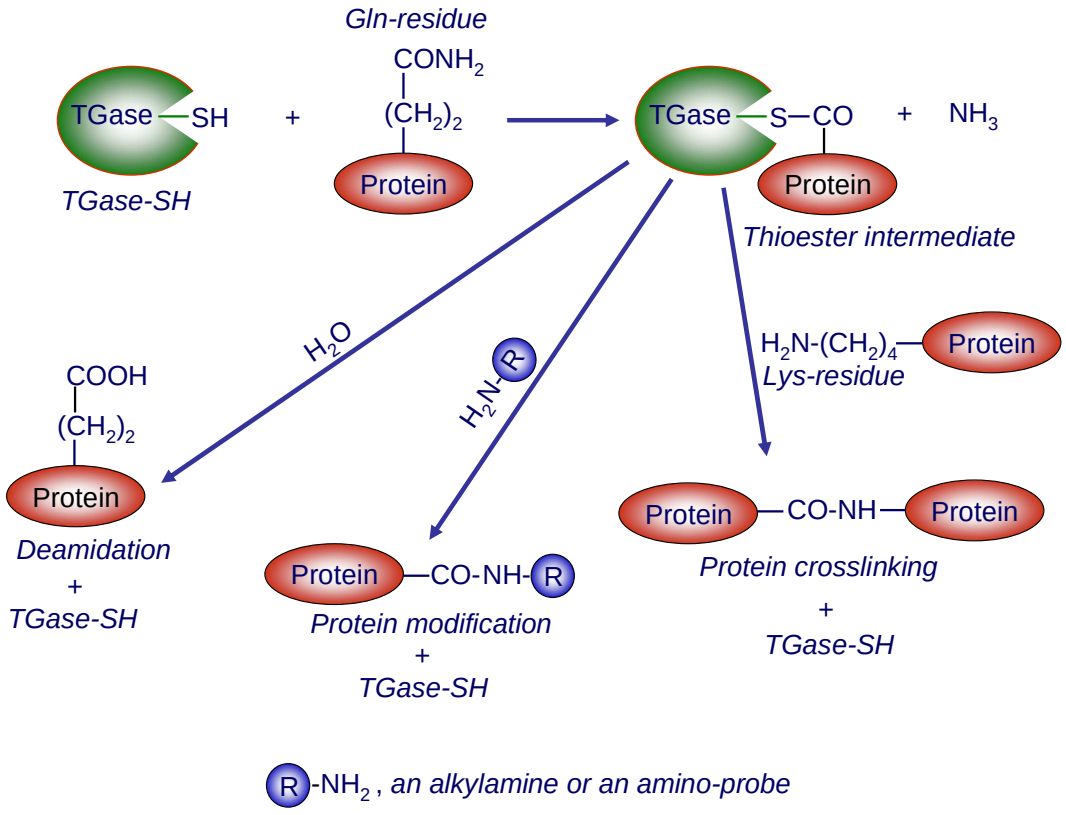


Figure 1. Scheme of the reactions catalysed by TGase. The critical step in the catalysis involves formation of a reactive thioester moiety at the level of a Cys residue, followed by reaction with a nucleophile. Protein crosslinking results from the reaction with the thioester of an ε-amino group of a protein-bound lysine (Lys) residue. Alternatively, a primary alkyl-amine or an amino-probe can react, leading to the incorporation into a protein of a suitable probe. If an amino nucleophile is not available, the thioester can be hydrolysed and deamidation occurs, so that a former glutamine (Gln) is transformed into a glutamic acid (Glu) residue.

TGases are widely distributed in various organisms, including vertebrates, invertebrates, plants and microorganisms. 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 and are distributed in different tissues such as liver, lung, intestine, epidermis, placenta and blood (*7–14*). 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 (*8–10*).

Up to eight members of the mammalian TGase family were identified from the genomic data, of which six isozymes have been purified and characterized as calcium- and thiol-dependent enzymes. However, these TGases differ in their molecular weight and biochemical properties. The human TGases family consists of eight enzymes designated factor XIIIa, TG1, TG2, TG3, TG4, TG5, TG6 and TG7, that catalyze a variety of Ca^{2+}-dependent post-translational protein-modifying reactions. All members of the mammalian TGase family are structurally homologous and arise from different genes, due to the rearrangement and duplication mechanisms in cell. The amino acid sequences encompassing the Cys314 of the active site of TGases is highly conserved, while in other regions of the chain there is little conservation of sequence (Fig. 2).

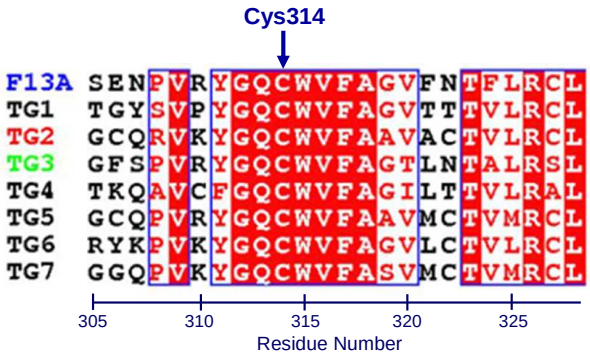


Figure 2. Amino acid sequences near the active site Cys314 of human TGases. The conserved residues are indicated as white letters in a red background, while the conservative replacements in red letters in a white background. The active site region appears to be highly conserved, while in other regions of the polypeptide chain there are large changes in amino acid sequence. The data are taken and adapted from ref. *15*.

Mammalian TGases are comprised of four structurally distinct domains (an N-terminal β-sandwich, an α,β catalytic core and two C-terminal β-barrel domains) that adopt a compact conformation in the absence of calcium (*16–20*). Two family members, FXIIIa and TGase-1, have an additional N-terminal propeptide sequence that is cleaved to produce the active enzyme. The TGase enzymatic activity resides in an active site catalytic triad consisting of a cysteine, histidine and aspartate residue (Cys-His-Asp), in analogy to a Cys-protease. There is also a structural protein, named protein 4.2, which shows structural homology with TGases, but that lacks catalytic activity. Protein 4.2 is evolutionarily related to the other TGases, as is evident from its high degree of sequence conservation (37–51% overall sequence similarity), but none of the amino acid residues critical for catalysis is conserved in this protein (*15*).

Factor XIII

Among the human TGases, the human blood coagulation Factor XIII has been most studied and its molecular and enzymological properties have been analysed in great detail. Factor XIII is unique among TGases in that it is a zymogen (*20–27*). Factor XIII is found both extracellularly in plasma and intracellularly in platelets, liver and prostate tissues. Plasma factor XIII is synthesized in the liver and circulates as a tetramer of 320 kDa, composed of two pairs of nonidentical subunits (A_2B_2). The

concentration of Factor XIII in plasma is approximately 30 mg/ml. The intracellular forms are synthesized in tissues, where they reside as dimers (146 kDa) of two identical A chains (A_2). The A subunits of plasma and intracellular forms of Factor XIII are functionally identical. The A subunit contains 6 free –SH groups, one of which is the active site. The conversion of plasma Factor XIII A_2B_2 to the active Factor XIIIa results from hydrolysis by thrombin of the Arg37-Gly38 peptide bond at the N-terminus of the A subunit, leading to the active Factor XIIIa. Full expression of activity is achieved only after binding Ca^{2+} (*27*).

Factor XIII is a TGase that forms fibrin clots in hemostasis and wound healing by catalyzing the crosslinking between fibrin molecules. It is the last of the zymogens to become activated in the coagulation cascade and it is the only enzyme in this system that is not a serine protease (*20*). Factor XIIIa functions to stabilize the fibrin clot by crosslinking the chains of fibrin. Other proteins known to be substrates Factor XIIIa which may be hemostatically important include fibronectin, α2-antiplasmin, collagen, von Willebrand factor and thrombospondin (*20–26*).

The three-dimensional (3D) structure of the zymogen Factor XIII has been determined by X-ray crystallography (*17*). The zymogen is a homodimeric protein which is folded into four structural domains (Fig. 3). A catalytic triad Cys-His-Asp analogous to that observed in cysteine proteases (*e.g.*, papain) has been identified in the core transamidating domain. The amino-terminal activation peptide of each subunit crosses the dimer interface and partially occludes the opening of the catalytic cavity in the second subunit, preventing substrate binding to the zymogen. The crystallographic structure allowed to propose a mechanism of activation of Factor XIII by thrombin and calcium binding, both being required for an active Factor XIIIa (*25*).

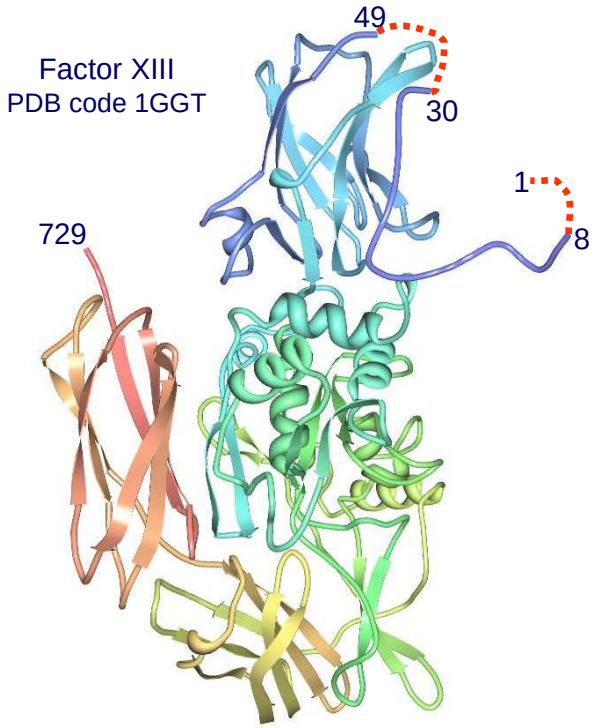


Figure 3. 3D structure of Factor XIII. The protein model was prepared from the X-ray structure of the protein using the software MBT (Molecular Biology Toolkit; http://mbt.sdsc.edu) available in PDB (code 1GGT). The 331-residue chain of TGase is depicted in rainbow colours from the N-terminus (blue) to the C-terminus (red). The 3D structure of Factor XIII clearly reveals the domain architecture of the protein, being formed by for rather well separated domains. In the zymogen, the chain segments 1–8 and 38–43 are not visible in the electron density map and thus are disordered. Of interest, this fact explains the site-specific cleavage of the zymogen by thrombin occurring at Arg37-Gly38 peptide bond and leading to the activation of Factor XIII, formomg Factor XIIIa. Since the cleaved peptide bond is embedded in a disordered region of the chain, it documents the correlation between chain flexibility and site(s) of limited proteolysis.

Transglutaminase-2 (TGase-2)

Human transglutaminase 2 (TGase-2) plays an important role in the extracellular matrix biology of many tissues and is implicated in the gluten-induced pathogenesis of celiac disease (*28–32*). The human TGase-2 is secreted by an unknown non-conventional mechanism, since it lacks a signal sequence. Although potential glycosylation sites are present, no evidence of glycosylation has been reported. Crystallographic analysis has revealed that TGase-2 has four domains, i.e., an N-terminal β-sandwich domain (1-139), a transamidation core domain (140–454) and two C-terminal β-barrels (479–585 and 586–687, respectively). All four domains have particular roles. For example, the N-terminal domain interacts with fibronectin. The transamidation core domain is involved in GTP binding (Ser171 and Lys173), has an active triad site (Cys277, His335 and Asp358) and a calcium binding region. The C-terminal domain regulates transamidation activity, GTPase activity, as well as blocking the active site from contact with possible substrates. Upon calcium binding, the interaction between the active site and C-terminal domains is released, allowing the onset of transamidating activity of the enzyme. TGase-2 has been the most studied member of the human TGase family. Indeed; much of our knowledge about TGase reactions comes from the studies of TGase-2 (*19*, *28–33*), besides Factor XIII (*17*).

A remarkably large conformational change accompanies activation of TGase-2 and likely other TGase family members (*19*). Upon binding of substrate, there is a large conformational change from a compact form to an extended ellipsoid structure that exposes the TGase active site (Fig.es 4 and 5). An inhibitor was found to stabilise TGase-2 in an extended conformation that is dramatically different from other TGases, including the Factor XIII structure. Upon binding the inhibitor, which acts similarly to a substrate, a large domain movement leads to the exposure of the active site. These structural studies revealed that catalysis takes place in a tunnel, where stabilisation of the tetrahedral reaction intermediate occurs (*19*). The structure of the activated conformer of TGase-2 allowed a clearer understanding of aspects of the catalytic mechanism of TGase-2.

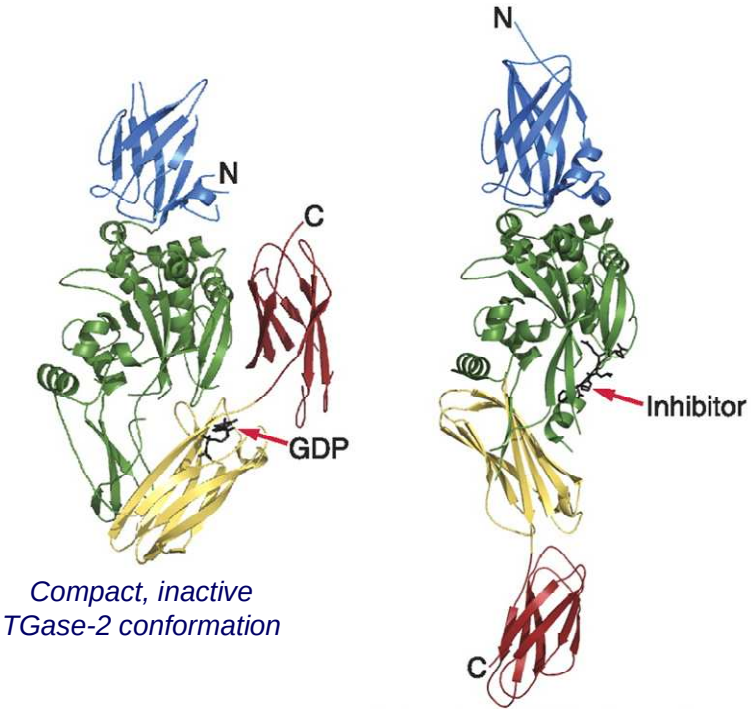


Figure 4. A large conformational change occurs in TGase-2 upon binding the inhibitor. This change leads to the exposure of the active site and thus of an active protein species. It is seen that the C-terminal β-domain covers the active site in the closed and inactive form of TGase-2. Arrows point to the site of binding of GDP and inhibitor. This figure is adapted from ref. *19*.

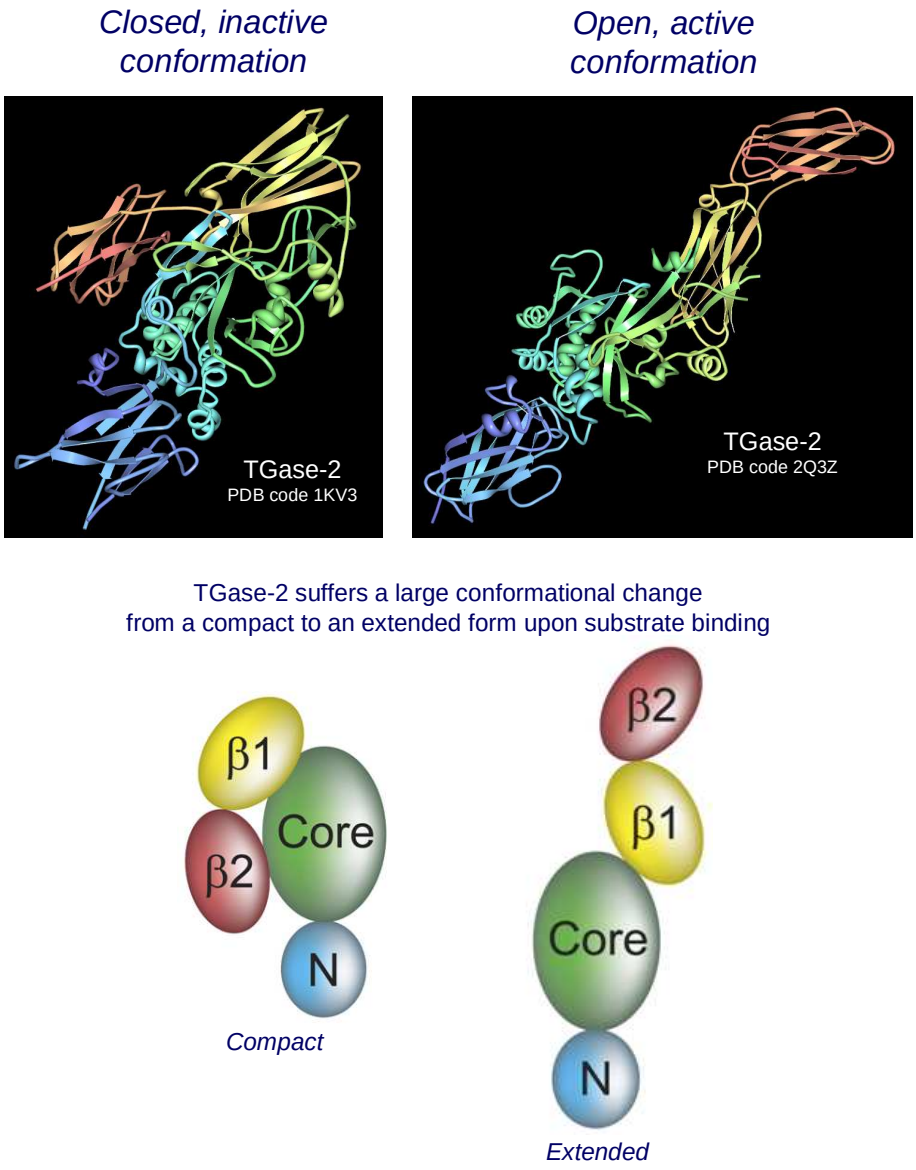


Figure 5. Schematic view of the large conformationl change of TGase-2. The protein adopts an open (active) and closed (inactive) conformation. The two conformations of the protein are observed in the GDP-bound (closed) or inhibitor-bound (open) state. The two states of TGase-2 are shown at the top of the figure, *i.e.*, the closed (PDB code 1KV3) and the open (PDB code 2Q3Z). Below a schematic cartoon of the compact and extended forms of TGase-2 are shown. This figure is an adaptation of data presented in ref. *19*.

Microbial TGase

TGase activity has also been reported in bacteria, in particular *Streptoverticillium* sp., *Streptomyces* sp. and *Bacillus subtilis* (*34–40*). The first microbial TGase was discovered in *Streptomyces mobaraensis* and the amino acid sequence of its zymogen is shown in Fig. 6. Subsequently, other microbial strains that produce TGase were identified. In contrast to many other TGases, the activity of microbial TGase is Ca^{2+}-independent. The optimum pH for TGase activity is between 5 and 8, but the enzyme displays some activity also at pH 4 or 9. The optimum temperature for enzymatic activity is 55°C (for 10 min at pH 6.0) and the enzyme maintains full activity for 10 min at 40°C, but looses activity within a few minutes at 70°C. At variance from other mammalian TGases, including the well-characterized guinea pig liver enzyme, microbial TGase is totally independent of calcium ions. Heavy metals such as Cu^{2+},

Zn^{2+} and $Pb2^{+}$ are strong inhibitors of the enzyme, since they bind to the thiol group of the active-site Cys64. Several molecular and functional characteristics of microbial TGase, including calcium-independence, protein stability, higher reaction rate and small molecular size, are advantageous for its applications for research and biotechnology (*41–44*).

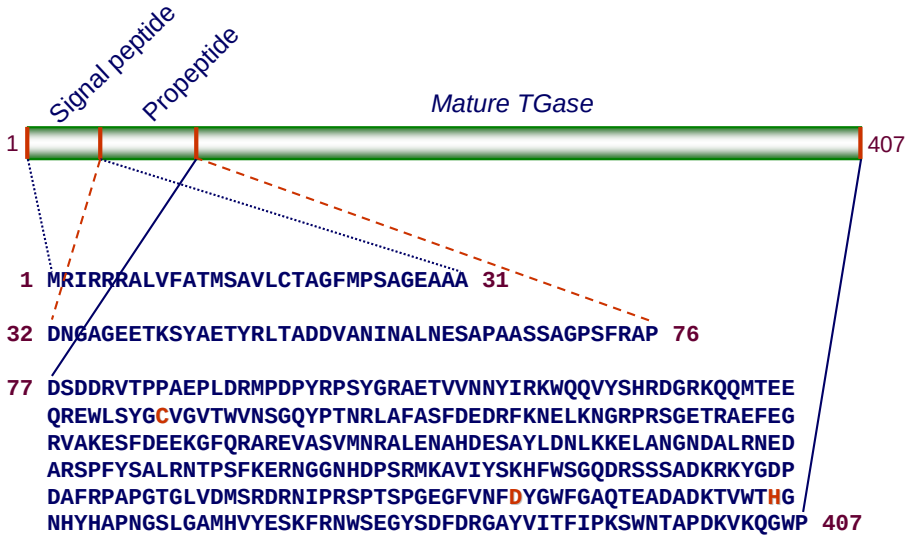


Figure 6. Amino acid sequence of the zymogen of microbial TGase. The signal peptide comprises the segment 1–31 and the propeptide the sequence 32–76. The mature enzyme is given by a polypeptide chain of 331 amino acid residues. The amino acids forming the catalytic triad are labelled in red. The triad is analogous to that of Cys-protease as papain. The amino acid sequence of the zymogen is that reported in PDB (code 1IU0).

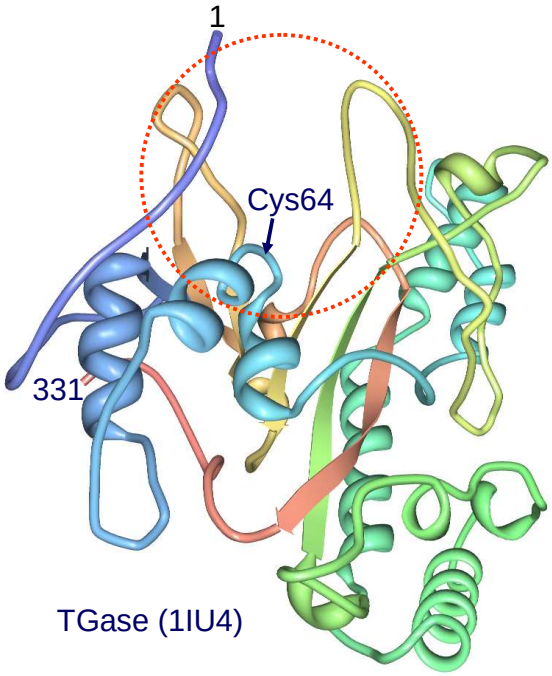


Figure 7. Three-dimensional structure of microbial TGase from *Streptomyces mobaraense*. The protein model was prepared from the X-ray structure of the enzyme using the software MBT (Molecular Biology Toolkit; http://mbt.sdsc.edu) available in PDB (code 1IU4). The 331-residue chain of TGase is depicted 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 Cys64 is indicated by an arrow. It is seen that the active site Cys64 is rather buried in a deep cleft, explaining the fact that microbial TGase often does not react with a globular protein substrate or reacts with a strict selectivity at one (or very few) protein-bound Gln residue(s).

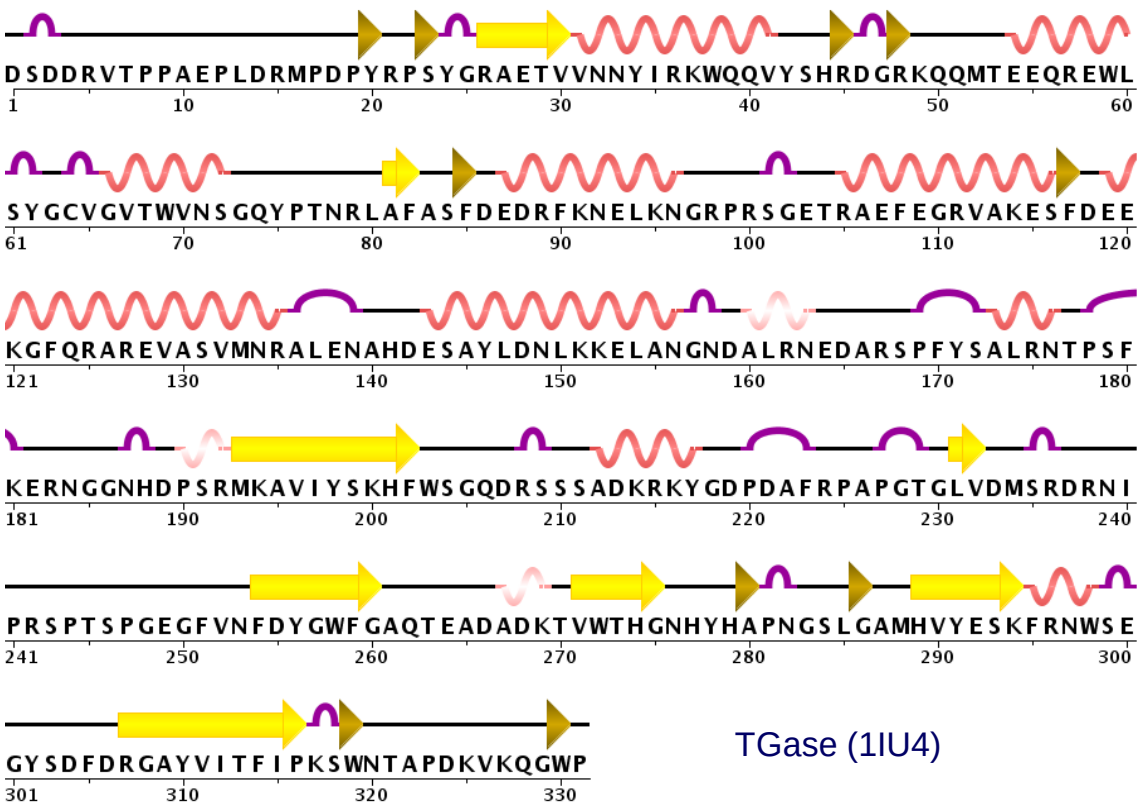


Figure 8. Amino acid sequence of the 331-residue chain of microbial TGase. The location of the helical segments are indicated by a waved red line and β-sheets by yellow arrows above the sequence. The sequence and the structural data are derived from PDB (code 1IU4).

The 3D structure of microbial TGase isolated from the culture medium of *Streptoverticillium sp.* S-8112, which has been identified as a variant of *S. mobaraense* (Fig. 7), is very different from that of other TGases (*37*). The crystal structure of this microbial TGase consists of a compact domain with overall dimensions 65 x 59 x 41 Å. The cysteine (Cys) residue in position 64 of the 331-residue chain of the enzyme, essential for the catalytic activity, is located at the bottom of a deep cleft, its depth being 16 Å. The crystal structure of microbial TGase revealed that the overall fold of this enzyme is different from that of Factor XIII-like TGases. Nevertheless, a similar cysteine protease-like catalytic mechanism for the microbial TGase has been proposed. Indeed, the Cys-His-Asp triad of cysteine-proteases is conserved, being given in the microbial TGase by Cys64, Asp255 and His274 residues. Of interest, the catalytic triad of microbial TGase almost superimposes the same triad of Factor XIII, implying a similar mechanism of catalysis. There are a number of acidic residues (Asp1, Asp3, Asp4, Glu249, Asp255 and Glu300) in the TGase's active site cleft and aromatic residues (Trp59, Tyr62, Trp69, Tyr75, Tyr278, Tyr291 and Tyr302) on the surface around the cleft. These characteristic distributions of the acidic and aromatic residues appear to have an effect on the substrate specificity of microbial TGase. Additional models of microbial TGase (Fig.es 9 and 10), derived from the 3D structure deposited in PDB (code 1IU4) (*37*), could help the reader to visualise some properties of this TGase, such as the location of secondary structure elements or hydrophobicity along the 331-residue chain of the protein.

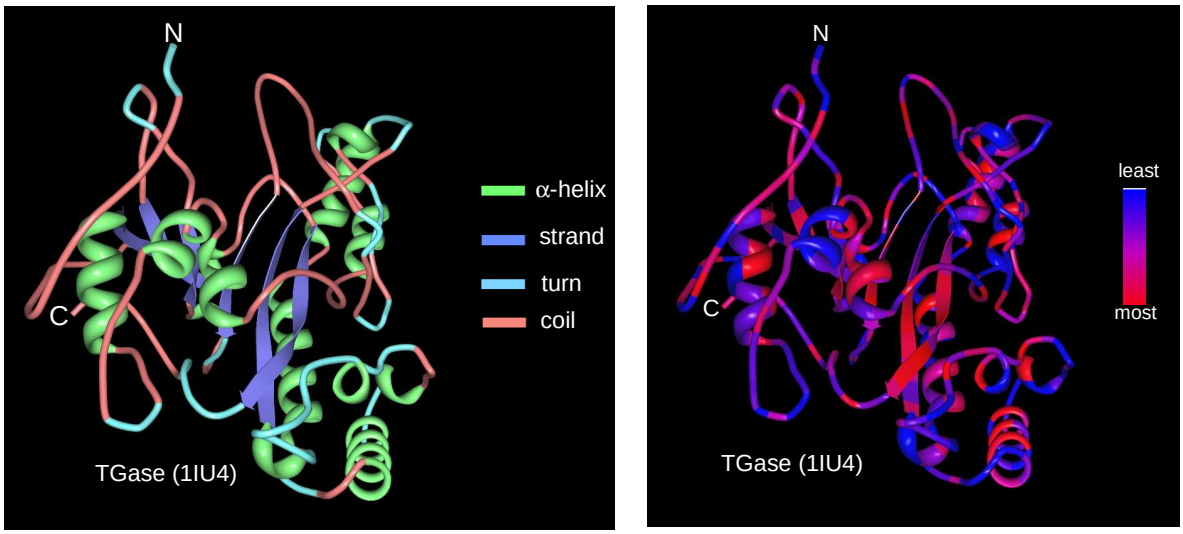


Figure 9. 3D models of microbial TGase. The models highlight the secondary structure segments (left) and hydrophobicity (right) along the 331-residue chain of the protein (PDB code 1IU4).

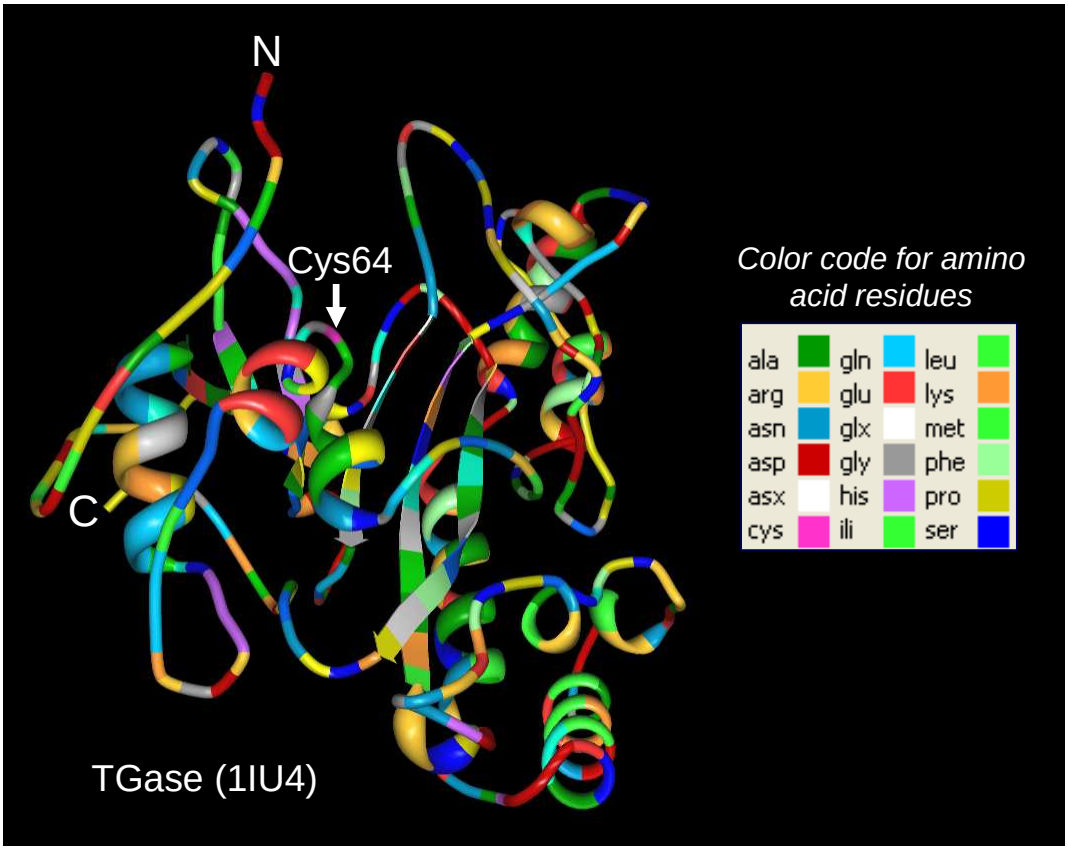


Figure 10. 3D model of the 331-residue chain of microbial TGase with color-coded amino acid residues. The arrow indicated the location of the buried Cys64 at the enzyme's active site. The model is derived from the crystallographic structure of the protein (PDB code 1IU4).

The 3D structure of the zymogen of microbial TGase from *Streptomyces mobaraense* has been determined recently (Fig. 11) (*38*). The overall structure of the zymogen (for its sequence see Fig. 6) is similar to that of the mature form, consisting of a single disk-like domain with a deep active cleft at the edge of the molecule. A major portion of the propeptide (45-residues) folds into an extended N-terminal segment linked with a one-turn short helix and a longer α-helix. Two key residues in the short helix of the prosequence, Tyr12 and Tyr16, are located on top of the catalytic triad Cys-Asp-His of the mature enzyme and thus block the access of the substrate acyl donors and acceptors. Removal by proteolysis of the propeptide leads to the activation of the enzyme (sse also legend to Fig. 12). Of interest, the selective cleavage of the propeptide by proteolytic enzymes occurs at a disordered site of the polypeptide chain of the zymogen.

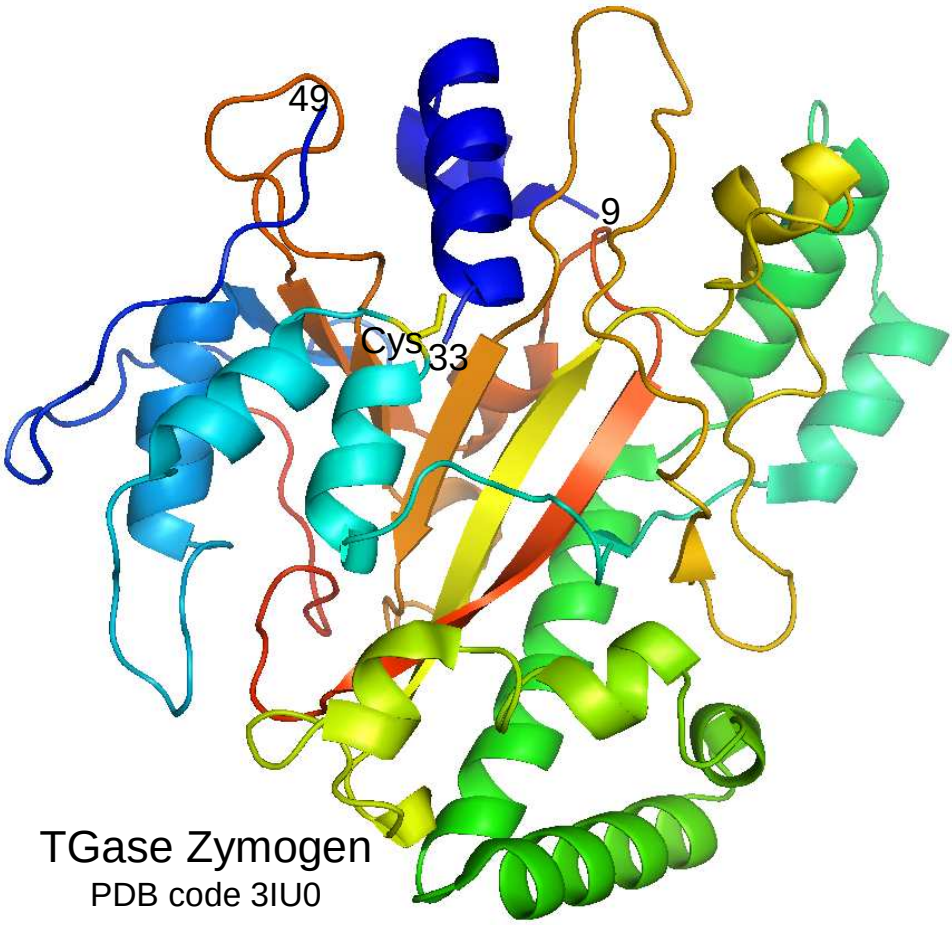


Figure 11. Three-dimensional structure of the zymogen of microbial TGase from *Streptomyces mobaraense*. The protein model was prepared from the X-ray structure of the enzyme using the software MBT (Molecular Biology Toolkit; http://mbt.sdsc.edu) available in PDB (code 1IU0). The chain of TGase is depicted in rainbow colours from the N-terminus (blue) to the C-terminus (red). The active site area of the enzyme is blocked by the propeptide (colored in blue), interacting with the active site Cys64 (yellow stick). The active site occlusion explains why the zymogen is inactive. The sequences 1–9 and 33–43 of the propeptide are not visible in the electron density map and therefore are disordered.

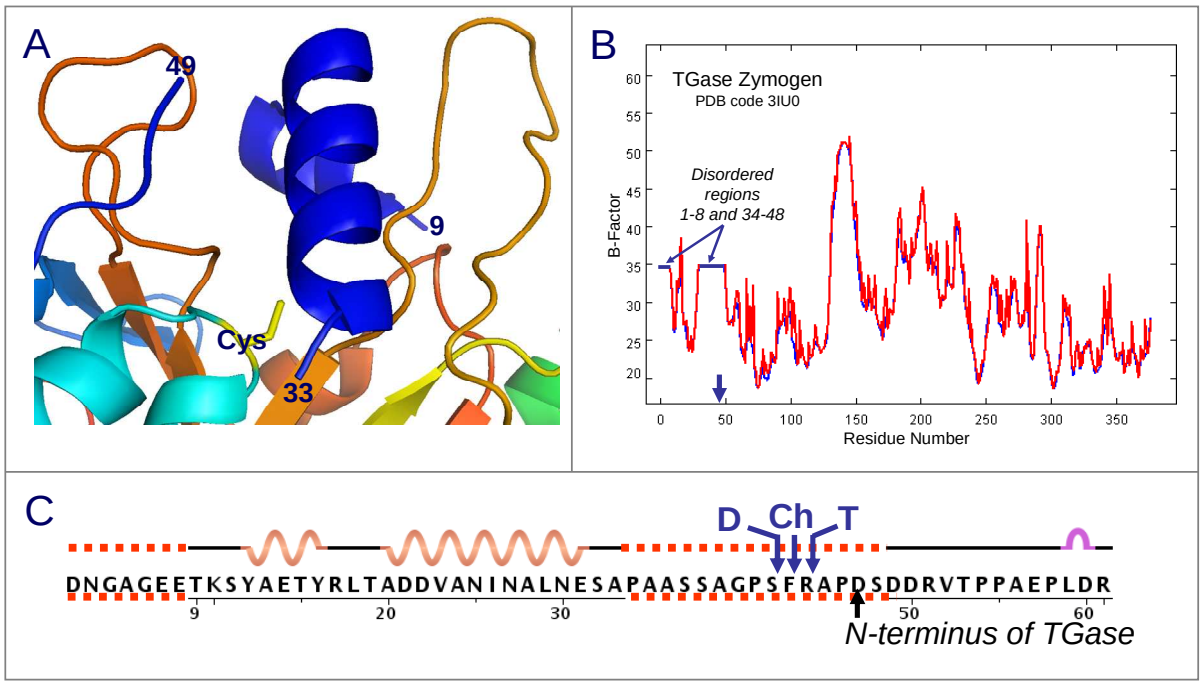


Figure 12. Analysis of the structural features of the propeptide in the zymogen of microbial TGase. (**A**) 3D structure of the N-terminal region of the zymogen (PDB code 1IU0). Only the 9–33 region of the 46-residue propetide is structured. (**B**) *B*-factor profile along the polypeptide chain of the zymogen. There are two regions of disorder at the N-terminus of the protein. The arrow indicate the sites of proteolytic processing of the zymogen, resulting in its activation. (**C**) Sites of specific proteolytic cleavages observed during the proteolytic processing and activation of the zymogen by chymotrypsin (Ch), trypsin (T) and dispase (D) (*35*). Proteolysis occurs selectively at a disordered region of the protein chain, in keeping with the notion that limited proteolysis sites occur at flexible/disordered sites in proteins.

2. Biological role of TGase

The eight TGases encoded by the human genome can be distinguished by their physical properties, tissue distribution, localization and mechanisms of activation, as well as by differences in their substrates and, therefore, in the enzymatic reactions that they catalyze (*6–8*, *14*, *15*, *28*, *29*). The fibrin-stabilizing factor XIII is probably the most prominent human TGase, since it functions in crucial physiological processes such as haemostasis, wound healing and the maintenance of pregnancy. Dysfunction of factor XIII can cause life-threatening bleeding disorders. Human TGases play an important role in biological processes which are dependent on the rapid covalent crosslinking of proteins, *e.g.* blood coagulation, skin-barrier formation and extracellular-matrix assembly. TGase-mediated reactions result in supra-molecular protein structures with high rigidity and stability, since these enzymes catalyze covalent protein crosslinking by forming isopeptide bonds between a Lys donor residue of one protein and an acceptor Gln residue of another protein. All human TGases, except plant and bacterial TGases, require Ca^{2+} for activation.

3. Protein substrates

TGases are widely distributed in various organisms, including vertebrate, invertebrates, mollusks, plants and microorganisms. Over the years numerous studies have been conducted on TGases with the view to identify their protein substrates. A database for TGase substrate proteins, called TRANSDAB

(http://genomics.dote.hu/wiki/index.php/Main_Page), has produced useful structural information on all TGases and provides a list of previously identified protein substrates (see Table 1) (see ref. *45*, for a presentation of this database). However, in spite of these studies on the structural properties and substrate specificity listed in TRANSDAB, the mechanism of substrate recognition by TGases remains poorly defined. Indeed, the amino acid sequences near the Gln residue being site of TGase reaction in several proteins (*46*) offer little information regarding possible consensus sequences recognised by TGase (see the list of Gln-sequences of protein substrates given in Fig. 13).

Table 1. **Protein substrates of TGases**[a]

Protein Substrates of Transglutaminase-2		
• 40S ribosomial protein SA • Acidic proline rich protein • Aconitase • ACTH • Actin • Aldolase A • Alpha lactalbumin • Alpha synuclein • Alpha-2 macroglobulin receptor • Alpha-2 plasmin inhibitor • Alpha-2-HS-glycoprotein • Alpha-ketoglutarate dehydrogenase • Amyloid beta A4 peptide • Androgen receptor • Angiocidin • Ankyrin • Annexin I • Arginase I • AT-rich interactive domain-containing protein 1A • Ataxin-1 • ATP synthase • Band 3 anion transport protein • Band 4.1 protein • Bcr - Breakpoint cluster region • Beta casein • Beta endorphin • Beta tubulin • Beta-2-microglobulin • Betaine-homocysteine S-methyltransferase • BiP protein • Bone sialoprotein • C-CAM • C1 inhibitor • Calbindin • Calpain	• Envelope glycoprotein gp41 • Ephrin A • Eucaryotic initiation factor 4F (eIF-4F) • Exendin 4 • Ezrin-Radixin-Moesin binding phosphoprotein 50 • F-box only protein • Fatty acid synthase • Fibrinogen alpha chain • Fibrinogen gamma chain • Fibronectin • Filamin 1 • Fructose 1,6-bisphosphatase • Galectin 3 • Gliadin • Glucagon • Glutathione S-transferase • Glyceraldehyde-3-phosphate dehydrogenase • Hepatitis C virus core protein • Histamine • Histatin • Histone H1 • Histone octamer • HIV-1 aspartyl protease • Hsp 27 • Hsp60 • Hsp70 • Hsp70/90 organizing protein • Hsp90 • Human Clara-cell 10 kDa protein • Huntingtin • Hyphal wall protein–1 • Ig kappa chain C region	• Nuclease sensitive element binding protein-1 • Nucleophosmin • Orexin B • Osteonectin • Osteopontin • Parkin • Periphilin • Periplakin • Phosphoglycerate dehydrogenase • Phospholipase A2 • Phosphorylase kinase • Plasminogen • Plasminogen activator inhibitor-2 • Plasmodium falciparum liver stage antigen-1 • Procarboxypeptidase B/U • Protein synthesis initiation factor 5A • RAP-Alpha-2 macroglobulin related protein • Retinoblastoma protein • Rho associated, coiled coil, containing protein kinase 2 • RhoA • S100A10 • S100A11 • S100A7 • Seminal vesicle secretory protein IV • Serotonin • SNAP-25 • SP1 transcription factor • Spectrin • Statherin • Substance P • Suprabasin • Synapsin 1

- Caspase-3
- CD38
- Clathrin heavy chain
- Collagen
- Crystallin
- Cytochrome C
- Deoxyribonuclease γ
- Dihydropyrimidinase-like 2 protein
- Dual leucine zipper-bearing kinase (DLK)
- EGF Receptor
- Elafin
- Elongation factor 1α
- Elongation factor 1γ
- Enolase
- Envelope glycoprotein gp120
- Importin β1 subunit
- Template:Infobox
- Insulin
- Insulin-like growth factor-binding protein-1
- Insulin-like growth factor-binding protein-3 (transglutaminase)
- Inter-alpha-inhibitor
- Lamin A, C
- Latent transforming growth factor beta binding protein
- Lipoprotein A
- Melittin
- Microfibril-associated glycoprotein (MAGP)
- Midkine
- Myelin basic protein
- Myosin
- Neurofilament proteins
- Neuropeptide Y
- NF-kappa-B inhibitor α
- Synapsin I
- T-complex protein 1ε subunit
- Tau protein
- Thymosin beta 4
- Thyroglobulin
- Troponin
- Tumor rejection antigen-1
- Ubiquitin
- Uteroglobin
- UV excision repair protein RAD23 homolog B
- Valosin
- Vasoactive intestinal peptide
- VEGFR-2
- Vigilin
- Vimentin
- Vitronectin

Protein Substrates of Transglutaminase FXIIIa

- Actin
- Alpha-2 macroglobulin
- Alpha-2 plasmin inhibitor
- AT1 receptor
- Coagulation factor V
- Collagen
- Fibrinogen alpha chain
- Fibrinogen gamma chain
- Fibronectin
- Fibronectin binding protein A
- Histidine rich glycoprotein
- HIV-1 aspartyl protease
- Inter-alpha-inhibitor
- Laminin
- Lipoprotein A
- Myosin
- Osteopontin
- Phospholipase A2
- Plasminogen
- Procarboxypeptidase B/U
- Protein synthesis initiation factor 5A
- Semenogelin I, II
- Thrombospondin
- Uteroglobin
- Vinculin
- Vitronectin
- Von Willebrand factor

Protein Substrates of Microbial Transglutaminase

- Alpha lactalbumin
- Bacteriorhodopsin
- Dispase autolysis inducing protein (DAIP)
- Gelatin
- Myosin heavy chain subfragment 1
- Phaseolin
- Streptomyces Subtilisin and TAMEP Inhibitor

---

[a]Adapted from TRANSDAB, the data base of protein substrates of transglutaminases (http://genomics.dote.hu/wiki/index.php/Main_Page). See the presentation of Transdab in:
Csősz, E., Meskó, B., and Fésüs, L. (2008) TRANSDAB wiki: the interactive transglutaminase substrate database on web 2.0 surface. *Amino Acids 36*, 615–617; DOI: 10.1007/s00726-008-0121-y.

Transglutaminase substrate protein	Protein sequence
RhoA	LAKM**K**Q**E**P**V**KP
RhoA	EVDG**K**Q**V**EL**A**L
Vitronectin	KGNP**E**Q**T**P**V**LK
Synapsin	RPSL**S**Q**D**E**V**KA
Osteonectin	VAAE**D**Q**R**P**I**NY
Midkine	NAQC**Q**E**T**I**R**V
Lipocortin	IENE**E**Q**E**Y**V**QT
IGFBP1	ALPG**E**Q**Q**P**L**HA
Fibrinogen α	GSTG**N**Q**N**P**G**SP
Elafin	A**Q**E**P**V**K**G
Actin	GRPR**H**Q**G**V**M**VG
α-Synuclein	VTAVA**Q**K**T**V**E**G
α-Synuclein	FVKK**D**Q**L**GKNE
Amyloid β A4 protein	YEVH**H**Q**K**L**V**FF
β-Casein	VLSL**S**Q**S**K**V**LP
β-Endorphin	TSEK**S**Q**T**P**L**VT
α-2 antiplasmin	M**Q**AL**V**LL
Phospholipase A2	ALW**Q**FRSMI
Bacteriorhodopsin	QA**Q**I**T**G**R**P
Glucagon	H**S**Q**G**T**F**TS
Fibronectin	EA**Q**Q**I**V**M**
β B3 crystallin	A**E**Q**H**S**T**PE

Figure 13. Amino acid sequences encompassing the reactive glutamine (Gln) residue in known protein substrates of TGase. The Gln residues attacked by TGase are shown in italic and are boxed in red. The sequences presented here do not allow to derive a consensus sequence being recognised by TGase. The data are taken from ref. *46*.

In the past, emphasis was given to the analysis of amino acid sequences near the reactive Gln residues (*47–53*), but more recently it became clear that the structure and dynamics of the protein substrate is actually controlling and dictating the site of TGase attack. One major observation that can be derived from TRANSDAB data (*45*) is that TGase are very site-specific enzymes, often attacking a protein substrate at one or very few Gln residues, despite the fact that the protein substrate contains numerous Gln residues. If only one Gln is reactive among te total of 20-50 Gln residues in a globular protein substrate, it can be anticipated that the sequence near Gln should play a very minor role and that the tertiary structure of the protein instead is playing an overhelming role in the TGase reaction(s).

It has been suggested that conformational preferences may be the most important factor and that the Gln residues should be located in a flexible region of the polypeptide chain, such as flexible loops (*54–56*). Several proteins (*e.g.*, apomyoglobin, α-lactalbumin, human growth hormone and interleukin-2) are specifically modified by TGase and the modified Gln residues are located in regions of unfolded or disordered regions of the polypeptide chain of protein substrates. Interestingly, proline (Pro) seems to play a role in the specificity of TGases. This can be rationalized by the fact that Pro has a structure-breaking effect on proteins and favors a local unfolding of the protein chain.

Although considerable work has been done to examine the sequence specificity of TGase for Gln residues in proteins, no conclusive amino acid sequence embedding the reactive Gln residue has been derived, nor a clear-cut proposal for explaining the often observed site-specificity of TGase-mediated

reactions (see ref.es *54* and *56*, for a discussion). The TGase-mediated reaction involves the critical step of formation of a thioacyl intermediate at the level of a deeply buried Cys-residue of the active site, followed by an aminolysis of the reactive thioester (see Fig. 1). This first step is the most demanding for the protein substrate, since it requires binding and adaptation of the substrate at the specific (and buried) stereochemistry of the TGase's active site. This is difficult to achieve with globular protein substrates and, therefore, only a locally flexible or disordered polypeptide substrate can bind to the TGase's active site and form the intermediate thioester.

4. Mechanism of catalysis

The mechanism of the reaction catalyzed by TGase-2 has been studied in the past by numerous authors and more recently crystallographic analyses allowed to delineate the enzymatic process in great detail. The reactions catalyzed by TGases occur by a two-step mechanism (see also Fig. 1). The transamidating activity of human TGases is activated by the binding of Ca^{2+}, which exposes an active-site Cys residue, which in turn reacts with the γ-carboxamide group of an incoming Gln residue of a protein/peptide substrate to yield a thioacyl-enzyme intermediate (thioester) and ammonia. The thioacyl-enzyme intermediate then reacts with a nucleophilic primary amine substrate, resulting in the covalent attachment of the amine-containing donor to the substrate Gln acceptor and regeneration of the Cys residue at the enzyme's active site. If the primary amine is donated by the ε-amino group of a Lys residue in a protein, an Nε-(γ-glutamyl)-lysine isopeptide bond is formed. On the other hand, if a polyamine or another primary amine (*e.g.*, cadaverine) acts as the amine donor, a γ-glutamylamine residue is formed. If there is no primary amine present, water may act as the attacking nucleophile, resulting in the deamidation of Gln residues to glutamic acid (Glu) residues (*1–10, 15*).

During the TGase reaction, if the attacking group is a primary amine (either a small biological amine or the ε-group of a peptide bound lysine), the reaction is called transamidation; if it is a water molecule it is called deamidation (see Fig. 1). While TGase-2 reacts with significant selectivity towards Gln residues, the enzyme is much less demanding for an acyl-acceptor, since a wide range of primary amines can act as acceptors (*57*). Small lysine-mimicking primary amines (*e.g.*, 5-pentylamine or cadaverine) can efficiently and covalently bind to the acyl-enzyme intermediate and, in this case, a pseudo-isopeptide bond with the Gln-containing protein is formed. Following aminolysis of the reactive thioester intermediate, the SH group at the active site of TGase is formed again and the enzyme is ready for another catalytic cycle.

Both human and microbial TGases have a Cys-His-Asp catalytic triad, in analogy to a Cys-protease as papain, even if there is no sequence homology near the amino acid residues flanking the catalytic triad within TGases. Nevertheless, the catalytic mechanism of microbial TGase is very similar to the one of other TGases. It should be emphasised that the formation of the covalent acyl-enzyme intermediate (thioester) is the rate-limiting step in TGase's reactions. Since the active site of microbial TGase is deeply buried in groove (*37*) and TGase-2 requires a large conformational change for binding a substrate (*19*), it can be anticipated that binding of Gln-protein or peptide should be difficult. This view would explain why TGases usually are very selective for specific Gln residues in a proytein substrate. On the other hand, Gln-peptide react much more easilty with TGase, as expected from the fact that relatively small peptiodes are srìtructureless and thus able to bind and adapt to the TGase's active site.

5 TGase-mediated modifications proteins

Protein crosslinking is the most common modification of proteins catalysed by TGases. In this case, the acyl-acceptors are the ε-amino groups of lysine residues within proteins, thus leading to covalent aggregates of proteins. In addition to protein crosslinking by transamidation, TGases are also capable

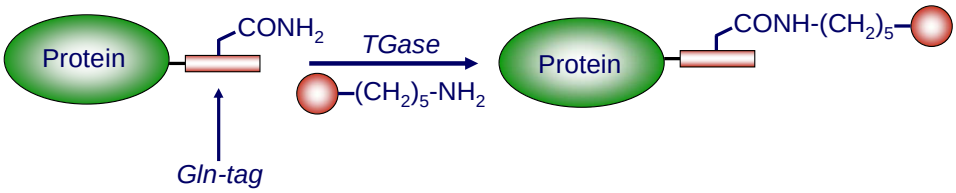
of catalyzing irreversible deamidation of glutamine to glutamic acid (Gln → Glu) with water serving as the nucleophilic acyl acceptor substrate (see Fig. 1). It is worth noting that the hydrolysing activity of TGase can cause a Gln → Glu formation and that likely such a reaction, in the case of the protein gliadin present in wheat, can cause the formation in a protein of new antigenic epitopes, which are responsible of immunological reactions during the celiac disease (*4–6, 28, 32*).

Microbial TGase from *Streptomyces mobaraensis* is widely used in the food industry, *e.g.*, for the crosslinking of proteins in meat and fish and to support gelling of yogurt and cheese (*41–43*). The incorporation of covalent cross-links with microbial TGase into food proteins, such as boiled fish paste and sausage, improves their physical and textural properties. Microbial TGase-mediated modification of proteins leads to improved physical properties of food proteins, including legume globulins, wheat glutens, egg yolk and albumin proteins, actins, myosins, fibrins, milk caseins, α-lactalbumin, β-lactoglobulin and other proteins. Enzymatic modification of proteins results in changes in protein structure and may therefore result in changes of the functional properties, such as solubility, gelation, emulsion formation and stabilisation. These functional properties are of importance to the behaviour of proteins in a food matrix. Recent studies have suggested that TGase-mediated cross-linking also has great potential for tissue engineering, textiles and leather processing, biotechnological tools and other non-food applications. Because of the usefulness and generality of the enzymatic reaction mediated by TGase, additional and novel applications of this enzyme in protein research and technology are expected to be further explored (*43*).

Fluorescent labelling

TGase has been often used for the enzymatic labelling of proteins with fluorescent probes, as alternative to chemical methods (*58–66*). Most chemical probes developed over the years for labelling proteins involve the attachment of synthetic electrophilic moieties to the nucleophilic side chains of amino acid residues, *e.g.*, -SH or -NH_2 groups. However, these chemistries rarely are site-specific, since there are multiple nucleophilic groups in a protein that can react with the labelling reagent and, therefore, usually heterogeneous mixtures of protein derivatives are produced (*61*).

An enzymatic method for protein labelling based on the use of TGase, as outlined in Fig. 14, can be advantageous. This strategy involves the use of a chimeric (fused) protein given by a tag or handle Gln-peptide fused to the target protein. TGase is used to ligate to the amide group of Gln an amino-probe via an isopeptide bond and concomitant release of ammonia. This technique has been used often to attach covalently to proteins a fluorescent label such as dansyl, fluoresceine or Alexa-568. The advantage of the TGase-mediated protein modification resides in the variability of useful probes that can be attached to a protein and, in particular, in the site-specificity of the reaction (*60, 63*).



A cadaverine-functionalised probe (red circle) can be linked to a Gln-tag fused to a protein by TGase

Figure 14. **TGase-mediated fluorescent labelling of a protein chimera**. Th e scheme shows that the strategy involves labelling a chimeric protein containing a short peptide tag containing a Gln-residue. The tag can be linkjed to the N- or C-terminal end of the target protein. The handle or tag is expected to lie outside the core of the globular protein and thus optimally exposed and flexible for acting as a TGase substrate.

TGase appears to offer several advantages for the site-specific modification of target proteins, since on one side the TGase reaction can be specific for the peptide only (*63*). It can be easily anticipated that the peptide tag protrudes from the globular protein scaffold, beings exposed to solvent and flexible and not embedded in the protein core. Thus, the success of the TGase method of modification outlined in Fig. 14 resides in the flexibility of the tag, thus favoring an easy and selective enzymatic reaction.

As an experimental proof of the proposal that the handle or tag attached to the N-or C-terminal end of a globular protein should be exposed and flexible and thus amenable to react with TGase, we can recall the results of an experimental study conducted on a chimera of lysozyme (*66*). Recombinant methods have been used to produce a chimera of egg-white lysozyme (129-residues) linked at his C-terminus to a 13-residue peptide containing the reactive Gln residue(s) of the γ-chain of fibrinogen. The authors were able to crystallise the protein chimera and solve its 3D structure by X-ray crystallography (PDB code 1LSG). It was found that the C-terminal peptide lies outside the core of lysozyme, while the protein chain is folded into the native and catalytically active conformation. Moreover, the C-terminal peptide was highly flexible, as given by the high values of *B*-factor of its chain. Of interest, the peptide tag attached to lysozyme was reactive towards Factor XIIIa and was also easily attacked by proteolytic enzymes.

PEGylation

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. 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 (*67–70*). 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. Recently, the covalent binding of PEG by TGase was reported as a promising method for protein surface modification to make better protein drugs (*54–56, 71, 72*). It was shown that TGase can be successfully used 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 TGase-mediated PEGylation of proteins of pharmaceutical interest shows the advantages in being conducted under physiological conditions and that the protein conjugates can be very homogeneous.

Metal-labelling

The TGase reaction allows a noteworthy versatility of protein substrates and ligands, since it is possible to use, in theory, a large variety of alkyl-amino ligands to be covalently bound to Gln residues of a protein by the enzymatic reaction (see Fig. 1). An interesting possibility is to use a ligand moiety capable of metal binding, so that it will be possible to make a protein derivative capable to bind a suitable metal ion. To this aim, a bifunctional reagent (BFCA) should be used, containing both an alkyl-amino group and a chelating moiety. This strategy based on the use of TGase has been recently used with success for the enzymatic radio-labelling of antibodies (*72* ,*73*). Indeed, it was possible to prepare protein conjugates capable of binding radionuclides such as technetium or rhenium (99mTc or 186/188Re). These protein conjugates were prepared for their use in SPECT (Single-Photon Emission Computed Tomography)/PET (Positron Emission Tomography) diagnosis and radiotherapy.

Previous chemical methods used for preparing radio-labelled proteins mostly involved the use of acylating agents, which suffer from the disadvantage to react with a variety of nucleophilic groups of the protein, leading to heterogeneous mixtures of protein conjugates. Thus, the overall utility of these chemical methods can be limited, since a well-defined radio-labelled protein will be more easily

characterized in terms of chemical identity and diagnostic or therapeutic efficacy. Consequently, it is evident that TGase can be considered a very useful and versatile reagent for radio-labelling proteins, since the enzymatic reaction can be performed under physiological conditions and, moreover, the reaction can occur selectively at one (or very few) protein-bound Gln residues.

6: Aims of the Thesis

The main aim of this PhD work is to examine the reactions mediated by microbial TGase 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 at the level of glutamine (Gln), as well as lysine (Lys) residue(s). To this aim, experiments were conducted on well-studied proteins such as avidin, α-lactalbumin (LA) and apomyoglobin (apoMb). The TGase-mediated reactions made use of primary amines as amino-donors, such as dansyl-cadaverine. Moreover, N-carbobenzoxy-Gln-Gly-OH (ZQG) was used as acyl-donor for the TGase reactions. In this last case, we could analyse the reactive Lys residues in a protein substrate. Indeed, the TGase reaction allows for permutation between acyl donor and acceptors, so that it is possible to modify a protein at the level of a Gln or Lys, as schematically shown in Fig. 15.

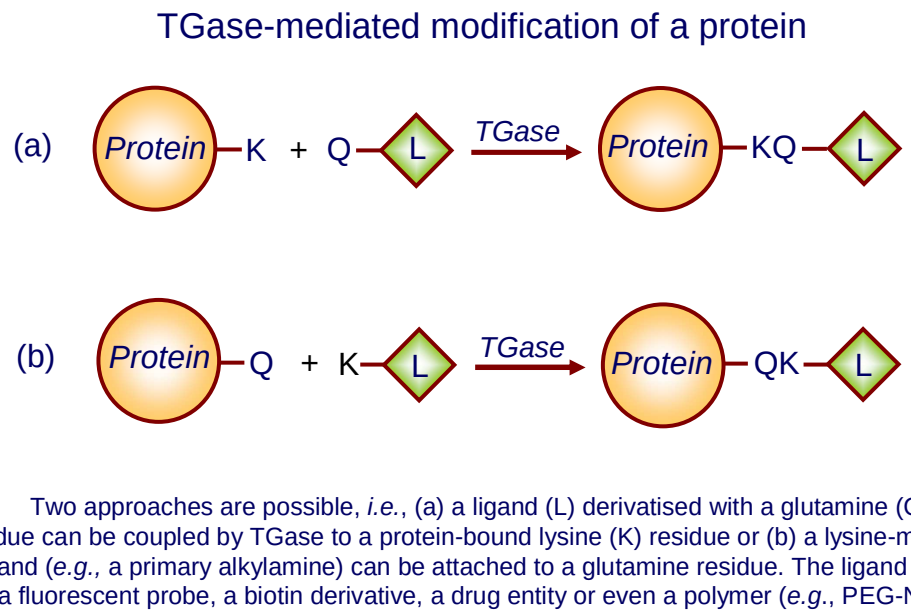


Figure 15. TGase-mediated modification of a protein at the level of a Gln or Lys residue.

Microbial TGase will be used for all experiments of protein modification. Preliminary data for the the site-specific binding of metal-chelating moieties to proteins has been also obtained. Since TGase catalyses the formation of an amide bond between the side-chain of a protein-bound Gln residue and an alkylamine moiety, in a protein a novel -CONH-R moiety can be introduced, where R can be an amino-containing metal-chelating group. Peptidic compounds bearing the alkylamine moiety required for the TGase reaction, as well as capable of chelating metals and radio-nuclides, will be synthesized and used for the TGase-mediated metal- or radio-labelling of a protein. We plan to evaluate and test the use apoMb as a model protein and to perform TGase reactions under various reaction conditions for efficiency and site-specificity of protein modification. The prospects of a successful outcome of the proposed research are based on the fact that initial experiments conducted in our laboratory have already shown that TGase reacts site-specifically with apoMb.

The TGase reactions are being explored in our laboratory using several protein models, as well as proteins of therapeutic interest, such as human growth hormone (hGH), interferon (INF) and granulocyte-colony stimalting factor (G-CSF). The research of this PhD Thesis was focused on the use of α-lactalbumin, avidin and apomyoglobin (apoMb), proteins that have been used before by us for studying aspects of their structure, folding and stability. Therefore, the idea was to take specific advantage and benefit of our own knowledge and experience on these well-characterised proteins.

References

1. Gorman, J.J., and Folk, J.E. (1980) Structural features of glutamine substrates for human plasma factor XIIIa (activated blood coagulation factor XIII, *J. Biol. Chem. 255*, 419–427.

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

3. 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.

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

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

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

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

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

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

10. Facchiano, A., and Facchiano, F. (2009) Transglutaminases and their substrates in biology and human diseases: 50 years of growing, *Amino Acids* 36, 599–614.

11. Folk, J.E., and Finlayson, J.S. (1977) The ε-(γ-glutamyl)lysine crosslink and the catalytic role of transglutaminases, *Adv. Protein Chem. 31*, 1–133.

12. Lorand, L. (2007) Crosslinks in blood: transglutaminase and beyond. *FASEB J. 21*, 1627-32.

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

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

15. Komáromi, I., Bagoly, Z., and Muszbek, L.J. (2011) Factor XIII: novel structural and functional aspects, *Thromb. Haemost. 9*, 9–20.

16. Ichinose, A., Bottenus, R.E., and Davie, E.W. (1990) Structure of transglutaminases, *J. Biol. Chem. 265*, 13411–13414.

17. Yee, V.C., Pedersen, L.C., Le Trong, I., Bishop, P.D., Stenkamp, R.E., and Teller, D.C. (1994) Three-dimensional structure of a transglutaminase: human blood coagulation factor XIII, *Proc. Natl. Acad. Sci. USA 91*, 7296–7300.

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

19. Pinkas, D.M., Strop, P., Brunger, A.T., and Khosla, C. (2007) Transglutaminase 2 undergoes a large conformational change upon activation , *PLoS Biol. 5*, e327.

20. Muszbek, L., Yee, V.C., and Hevessy, Z. (1999) Blood coagulation factor XIII: structure and function, *Thromb. Res. 94*, 271–305.

21. Komáromi, I., Bagoly, Z., and Muszbek, L. (2011) Factor XIII: novel structural and functional aspects, *J. Thromb. Haemost. 9*, 9–20.

22. Muszbek, L., Bereczky, Z., Bagoly, Z., Komáromi, I., and Katona, É. (2011) Factor XIII: a coagulation factor with multiple plasmatic and cellular functions, *Physiol. Rev. 91*, 931–972.

23. Pedersen, L.C, Yee, V.C., Bishop, P.D., Le Trong, I., Teller, D.C., and Stenkamp, R.E. (1994) Transglutaminase factor XIII uses proteinase-like catalytic triad to crosslink macromolecules, *Protein Sci. 3*, 1131–1135.

24. Muszbek, L., Bereczky, Z., Bagoly, Z., Komáromi, I., and Katona, É. (2011) Factor XIII: a coagulation factor with multiple plasmatic and cellular functions, *Physiol Rev. 91*, 931–972.

25. Muszbek, L., Yee, V.C., and Hevessy, Z. (1999) Blood coagulation factor XIII: structure and function, *Thromb. Res. 94*, 271–305.

26. Adany, R., and Bardos, H. (2003) Factor XIII subunit A as an intracellular transglutaminase. *Cell. Mol. Life Sci. 60*, 1049–1060.

27. Kristiansen, G.K., Andersen, M.D. (2011) Reversible activation of cellular factor XIII by calcium, *J. Biol. Chem. 286*, 9833–9839.

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

29. Gundemir, S., Colak, G., Tucholski, J., Johnson, G.V. (2011) Transglutaminase 2: A molecular Swiss army knife, *Biochim Biophys Acta 1823*, 406–419.

30. Király, R., Demény, M., and Fésüs, L (2011) Protein transamidation by transglutaminase 2 in cells: a disputed Ca^{2+}-dependent action of a multifunctional protein, *FEBS J. 278*, 4717–4739.

31. Kuo, T.F., Tatsukawa, H., and Kojima, S. (2011) New insights into the functions and localization of nuclear transglutaminase 2, *FEBS J. 278*, 4756–4767.

32. Belkin, A.M. (2011) Extracellular TG2: emerging functions and regulation, *FEBS J. 278*, 4704–4716.

33. Liu, S., Cerione, R.A., and Clardy, J. (2002) Structural basis for the guanine nucleotide-binding activity of tissue transglutaminase and its regulation of transamidation activity, *Proc. Natl. Acad. Sci. USA 99*, 2743–2747.

34. 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.

35. Zotzel, J., Keller, P., and Fuchsbauer, H.L. (2003) Transglutaminase from *Streptomyces mobaraensis* is activated by an endogenous metalloprotease, *Eur. J. Biochem. 270*, 3214–3222.

36. 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.

37. 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.

38. Yang, M.T., Chang, C.H., Wang, J.M., Wu, T.K., Wang, Y.K., Chang, C.Y., and Li, T.T. (2011) Crystal structure and inhibition studies of transglutaminase from *Streptomyces mobaraense*, *J. Biol. Chem. 286*, 7301–7307.

39. Washizu, K., Ando, K., Koiked, S., Hiros, S., Matsuura, A., Akagi, H., Motoki, M., and Takeuchi, K. (1994) Molecular cloning of the gene for microbial transglutaminase from *Streptoverticillium* and its expression in *Streptomyces lividans*, *Biosci. Biotechnol. Biochem. 58*, 82–87.

40. Zhang, D., Zhu, Y., and Chen, J. (2010) Microbial transglutaminase production: understanding the mechanism, *Biotechnol. Genet. Eng. Rev. 26*, 205–222.

41. Zhu, Y., Rinzema, A., Tramper, J., and Bol, J. (1995) Microbial transglutaminases: A review of its production and application in food processing, *Appl. Microbiol. Biotechnol. 44*, 277–282.

42. Yokohama, K., Nio, N., and Kikuchi, Y. (2004) Properties and applications of microbial transglutaminases, *Appl. Microbiol. Biotechnol. 64*, 447–454.

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

44. Menéndez, O., Rawel, H., Schwarzenbolz, U., and Henle, T. (2006) Structural changes of microbial transglutaminase during thermal and high-pressure treatment, *J. Agric. Food Chem. 54*, 1716–1721.

45. Csősz, E., Meskó, B., and Fésüs, L. (2008) TRANSDAB wiki: the interactive transglutaminase substrate database on web 2.0 surface, *Amino Acids 36*, 615–617.

46. Keresztessy, Z., Csosz, E., Hársfalvi, J., Csomós, K., Gray, J., Lightowlers, R.N., Lakey, J.H., Balajthy, Z., and Fésüs, L. (2006) Phage-display selection of efficient glutamine-donor substrate peptides for transglutaminase 2, *Protein Sci. 15*, 2466–2480.

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

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

49. 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.

50. 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.

51. 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.

52. Lajemi, M., Demignot, S., Borge, L., Thenet-Gauci, S., and Adolphe, M. (1997) The use of fluorescein-cadaverine for detecting amine acceptor protein substrates accessible to active transglutaminase in living cells, *Histochem. J. 29*, 593–606.

53. Pasternack, R., Laurent, H. P., Ruth, T., Kaiser, A., Schon, N., and Fuchsbauer, H.L. (1997) A fluorescent substrate of transglutaminase for detection and characterization of glutamine acceptor compounds, *Anal. Biochem. 249*, 54–60.

54. 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.

55. 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.

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

57. 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.

58. Lin, C.W., and Ting, A.Y. (2006) Transglutaminase-catalyzed site-specific conjugation of small-molecule probes to proteins in vitro and on the surface of living cells, *J. Am. Chem. Soc. 128*, 4542–4543.

59. Jäger, M., Nir, E., and Weiss, S. (2006) Site-specific labeling of proteins for single-molecule FRET by combining chemical and enzymatic modification, *Protein Sci. 15*, 640–646.

60. Kamiya, N., and Abe, H. (2011) New fluorescent substrates of microbial transglutaminase and its application to peptide tag-directed covalent protein labelling, *Methods Mol Biol. 751*, 1–94.

61. Kamiya, N., Abe, H., Goto, M., Tsuji, Y., and Jikuya, H. (2009) Fluorescent substrates for covalent protein labelling catalyzed by microbial transglutaminases, *Org. Biomol. Chem. 7*, 3407–3412.

62. 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.

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

64. 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.

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

66. Donahue, J.P., Patel, H., Anderson, W.F., and Hawiger, J. (1994) Three-dimensional structure of the platelet integrin recognition segment of the fibrinogen gamma chain obtained by carrier protein-driven crystallization, *Proc. Natl. Acad. Sci. USA 91*, 12178–12182.

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

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

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

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

71. Sato, H., Yamamoto, Y., Hayashi, E., and Takahara, Y. (2000) Transglutaminase-mediated dual and site-specific incorporation of poly(ethylene glycol) derivatives into a chimeric interleukin-2, *Bioconjug. Chem. 11*, 502–509.

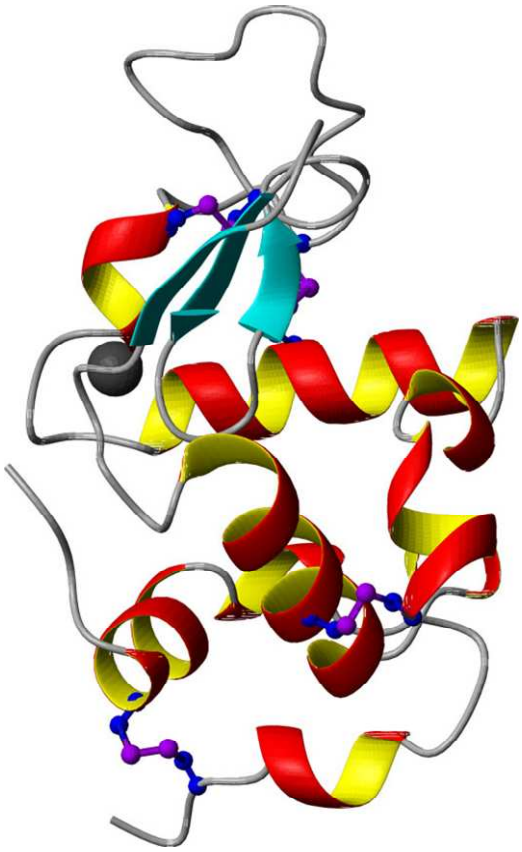
72. Sato, H. (2002) Enzymatic procedure for site-specific PEGylation of proteins, *Adv. Drug Deliv. Rev. 54*, 487–504.

73. Mindt, T.L., Jungi,, V., Wyss S., Friedli, A., Pla, G., Novak-Hofer, I., Grünberg, J., and Schibli, R. (2008) Modification of different IgG1 antibodies via glutamine and lysine using bacterial and human tissue transglutaminases, *Bioconjug Chem. 19*, 271–278.

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

Transglutaminase-mediated modification of proteins: Molecular mechanisms and metal-labelling of proteins

Results



α-Lactalbumin

1. Transglutaminase-mediated modification of α-lactalbumin at the level of lysine and glutamine residues: insights on the determinants of enzyme specifity

INTRODUCTION

Structural features of α-lactalbumin. α-lactalbumin (α-LA) is a globular metalloprotein of low molecular weight (14178.1 Da) expressed exclusively in the mammary gland secretory cells during lactation. It is one of the two components of lactose synthase, which catalyses the final step in lactose biosynthesis, by transferring an UDP-galactose residue on glucose *(1)*. α-LA possesses a single strong Ca^{2+} binding site *(2)* and it is frequently used as a simple model of Ca^{2+} binding protein. The structure of α-LA is also stabilized by four disulfide bridges, between Cys 6-120, 28-111, 61-77, and 73-91. The three-dimensional structure of α-LA consists of two domains: a large α-helical domain and a small β-sheet domain, connected by a calcium binding loop (Figure 1A). The α-helical domain is composed of four major α-helices A, B, C and D (amino acid residues 5-11, 23-34, 86-98 and 105-110) and two short 3_{10} helices (amino acid residues 18-20 and 115-118). The small β-domain is composed by a series of loops, a small three-stranded antiparallel β-sheet (amino acid residues 41-44, 47-50 and 55-56) and a short 3_{10} helix between residues 77-80 *(3-4)*.

α-LA is able to bind several metal cations. It has a single strong calcium binding site, situated in a loop between the two helices B and C. The calcium binding site is formed by oxygen ligands from the carboxylic groups of three Asp residues (at positions 82, 87 and 88) and from the two carbonyl groups of the peptide backbone of amino acid residues Lys79 and Asp84. In addition, one or two water molecules take part in direct coordinating Ca^{2+}. LA has also several zinc binding sites *(5)* and it binds other physiologically significant cations, such as Mg^{2+}, Mn^{2+}, Na^{+} and K^{+}, which can compete with Ca^{2+} for the same binding site *(6)*. The binding of Ca^{2+} to α-LA is very important for the maintenance of the native conformation of the protein. In particular, fluorescence and circular dichroism (CD) data showed that Ca^{2+}-binding causes marked changes in tertiary, but not secondary structure. Calcium-binding increases the thermo-stability of α-LA. From differential scanning calorimetry data, the binding of Ca^{2+} shifts the thermal transition to higher temperatures by more than 40°C *(7)*. The binding of metal cations also increases the stability of α-LA towards the action of

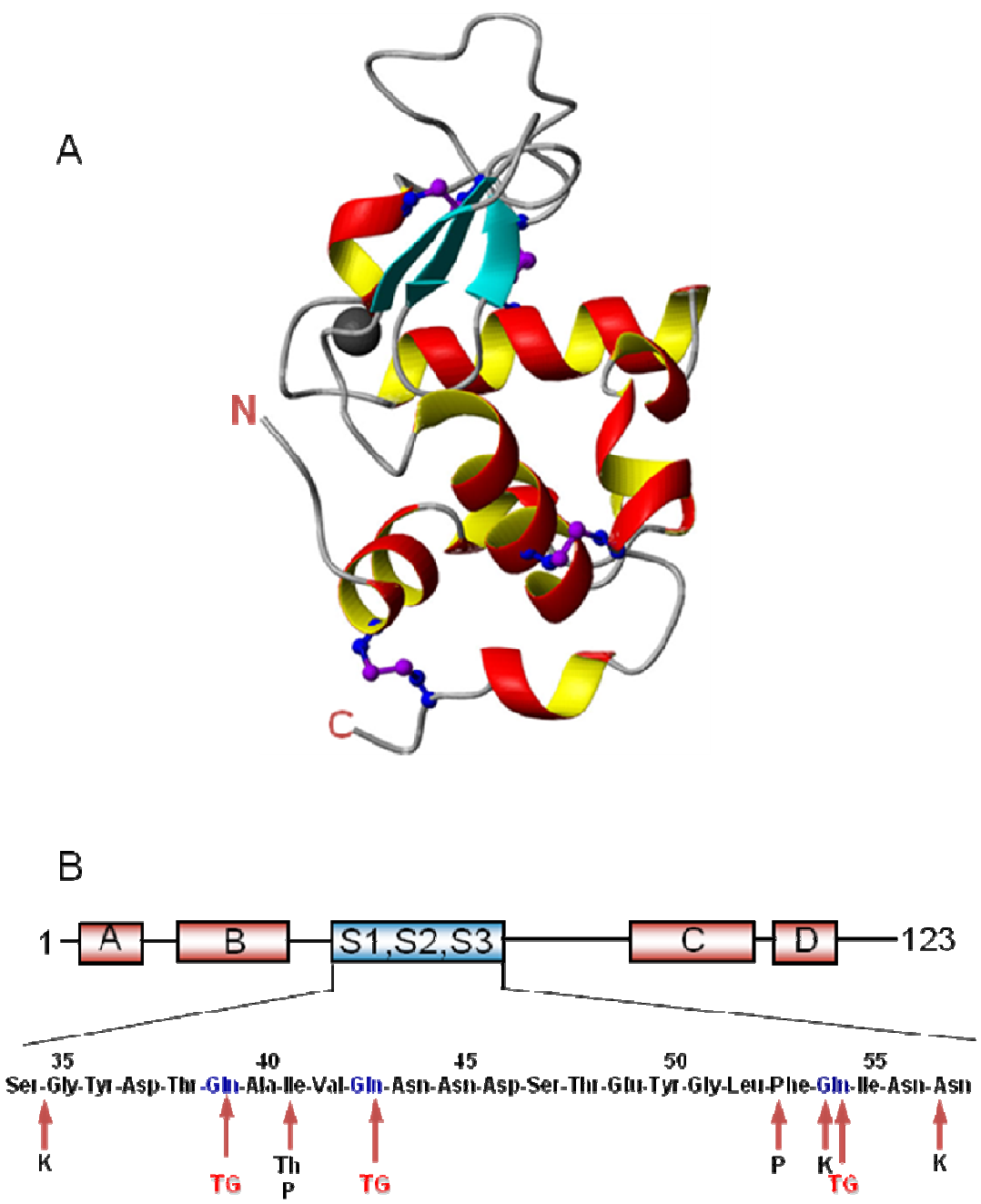


Figure 1. Structure and amino acid sequence of bovine α-lactalbumin. (A) Representation of the 3D structure of α-LA. The diagram was drawn using the PDB file 1HFZ with the program WebLab Viewer Pro 4.0 (Molecular Simulations Inc.). The chain segments corresponding to α-helices are colored in red and β-strands in blue. The four disulphide bonds are represented by violet sticks while a solid grey sphere represents calcium ion. (B) Scheme of the secondary structure of α–LA. The four α-helices (A to D) along the 123-residue chain of the protein are indicated by red boxes and the β-strands by a blue box. The amino acid sequence of the chain region 34–57 is explicitly shown and the sites of limited proteolysis of α-LA in its molten globule state by pepsin (P), thermolysin (Th) and proteinase K (K) are indicated by arrows. The sites of TGase modification are also indicated (Gln39, 43 and 54).

denaturing agents like urea or guanidine hydrochloride *(6)*. Therefore, any denaturation transition in αLA (temperature/denaturant-induced) depends upon metal ion concentration, especially that of calcium ion. This might be related to some temperature regulation of α-LA stability and function in the mammary gland. The binding of other metal ions induces similar, though smaller, protein structural changes.

Analysis of partly folded states. In recent years, numerous studies have been conducted on the partially folded or molten globule (MG) states that proteins can adopt under specific solvent conditions, such as at low pH, in the presence of moderate concentrations of denaturants or after the removal of protein-bound metal ions or cofactors *(8-10)*. The interest in MGs resides in the fact that they are considered protein folding intermediates. The key characteristics of a MG include a native-like secondary structure, lack of specific tertiary interactions and a more expanded and flexible structure than that of the native protein *(8-10)*. The structural analysis of MGs was expected to be difficult, since these states are not only flexible but can also be an ensemble of conformations, so elucidating their structure by either X-ray crystallography or NMR could be difficult if not impossible *(11)*. Nevertheless, developments in hydrogen/deuterium (H/D) exchange, combined with two-dimensional NMR spectroscopy, have provided quite detailed structural information for MGs *(11-13)*.

The limited proteolysis technique can be used to analyse protein structure and dynamics and, in particular, to identify disordered sites or regions within otherwise folded globular proteins. The approach relies on the fact that the proteolysis of a polypeptide substrate requires its binding and adaptation at the protease's active site and thus enhanced backbone flexibility or local unfolding of the site of proteolytic attack *(14)*. In our laboratory, we have shown that limited proteolysis experiments can be very useful in analysing the structural features of MGs of several model proteins, such as α-LA, lysozyme, apomyoglobin and cytochrome *c*. Overall, the chain regions identified as mobile or unfolded by proteolysis closely correlated with those detected by using other physicochemical and spectroscopic measurements.

The molten globule of α-lactalbumin. The conformational state of α-LA exposed to acid pH 2.0 (A-state) has been investigated in great detail using a variety of experimental approaches and techniques by numerous investigators *(15-18)*. As a result of these efforts, nowadays the A-state of α-LA is regarded as a prototype protein MG *(17)*. Nuclear magnetic resonance (NMR) and H/D exchange measurements revealed that α-LA in acid solution

adopts a partly folded or MG state characterized by a disordered β-domain, whereas the α-domain maintains substantial, albeit dynamic, helical secondary structure *(19-21)* (see Figure 1).

The calcium-depleted form of α-LA, as obtained by dissolving the protein at neutral pH in the presence of EDTA, also adopts a MG state at neutral pH, but a moderate heating is required *(17-18)*. There are discrepancies in the reported experimental results and conflicting proposals regarding the conformational state of apo-α-LA, ranging from a classical MG devoid of a cooperative thermal transition to a partly folded state with some native-like properties and displaying instead cooperativity *(17)*. Often it was assumed that apo-α-LA, as obtained for example by dissolving the protein at room temperature in Tris buffer, pH 8.0, containing a calcium chelating agent, adopts a MG state, but this may not be true without specifying ionic strength and temperature of the protein solution. It became clear that the conformational state adopted by apo-α-LA is strongly influenced by the specific solvent conditions *(17-18)*.

With the view to deduce conformational features of α-LA in its MG or partly folded state, in our laboratory we have conducted a series of limited proteolysis experiments on α-LA exposed to various mild perturbing conditions, such as exposing α-LA in acid solution *(22)* or at neutral pH in the presence of EDTA, trifluoroethanol or oleic acid *(23)*. The effect of the fatty acid on the structural features of α-LA was by us investigated, since it has been reported that an ill-defined oleic acid/α-LA complex displays the unusual, but interesting, property of killing tumour but not healthy cells by an apoptosis-like mechanism *(24-25)*. It was found that the various MGs of α-LA obtained under different solvent conditions all suffered limited proteolysis at a rather short portion of the 123-residue chain of the protein (see Figure 1B). The conclusion reached from our studies *(14, 22, 23, 26,)* was that the chain region approximately from residue 34 to 56, encompassing most of the β-domain, was disordered, while the rest of the protein chain remains folded and sufficiently rigid to resist proteolysis *(27-28)*. Therefore, proteolysis data are in agreement with the results of other physicochemical measurements indicating that the MG of α-LA is characterized by an unfolded β-domain and a native-like α-domain *(15, 19, 20, 21, 29)*. Since the protein in its MG state retains a rather large portion of its polypeptide chain stable and rigid enough to resist proteolysis, it was possible to remove selectively the β-domain from the 123-residue chain of α-LA and to isolate a stable, folded "gapped" protein species given by fragment 1–34 covalently linked to fragment 54/57–123 by disulfide bridges *(27)*.

Aims of the study. Microbial transglutaminase (TGase) is able to catalyse the acyl transfer reaction between the γ-carboxamide group of a protein-bound glutamine (Gln) and the amino group of a protein-bound lysine (Lys). One of the biotechnological applications of this enzyme resides in its ability to produce an extremely site-specific derivatization of proteins at the level of Gln residues with different primary amine containing substrates (Figure 2a). The determinants of the exceptional specificity of TGase-mediated conjugation of protein at the level of Gln residues seems not to respond to sequence requirements but most generally on the conformational features of the protein substrate. In particular, we have previously suggested that the main feature dictating the site-specific modification of protein-bound Gln residues by TGase in a globular protein is the flexibility or local unfolding of the chain region encompassing that residue *(30)*. However, TGase can also catalyse the conjugation of Gln containing substrates to lysines of proteins but the selectivity of this reaction and its potential biotechnological applications have less been studied (Figure 2b).

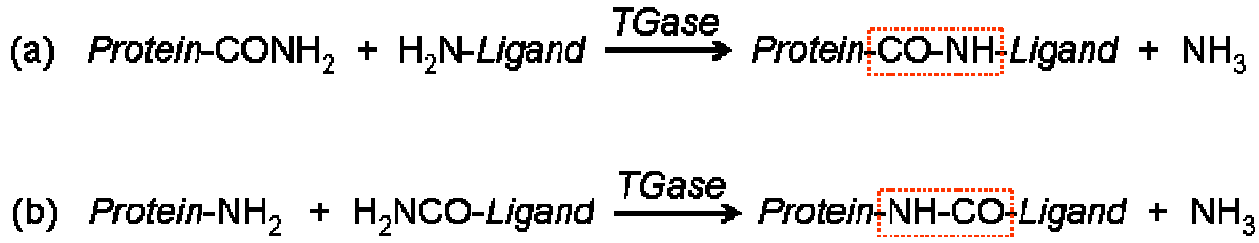


Figure 2. Schematic drawing of the two possibilities (a and b) of TGase-mediated protein derivatization.

In this project, experiments of TGase-mediated labelling both at the level of Lys and Gln residues were conducted on bovine α-lactalbumin, with the aim to compare the specificity of TGase modification at the level of the two residues. α-LA indeed contains a significative number of Gln and Lys residues (6 and 12, respectively) that can be potentially modified by TGase (Fig. 3B). Moreover, it has long been studied in our laboratory using the limited proteolysis approach in order to unravel regions of protein disorder in its molten globule state (see above). The TGase-mediated modification of the apo form of α-LA has already been extensively investigated under different experimental conditions *(31-33)*. It was found that the amino-donor dansylcadaverine was incorporated by TGase at Gln54 when α-LA was reacted in its MG state *(31-34)*. Oligomeric crosslinked α-LA, obtained by incubating the apo-protein with TGase in the presence of the disulfide-reducing agent dithiothreitol, was shown to result

from the covalent binding of Gln54 to several lysine residues of the protein. In other studies, it was found that the apo form of α-LA was modified at Gln residues in position 39, 43 and 54 *(33)* or positions 39, 43, 54 and 65 *(32)*. On the basis of these results reported in the literature, we have already commented that the Gln residues modified by TGase occur a the same chain region encompassing the sites of limited proteolysis by several proteases of α-LA in its MG state *(30)*. In this study, we aim to analyse the reactivity of Gln and Lys residues of the holo and apo-form of α-LA towards TGase under the same solution conditions used in the proteolysis experiments *(23)*. Consequently, the sites of TGase-mediated derivatization can be compared with the sites of hydrolysis by proteases identified in limited proteolysis experiments.

.

MATERIAL AND METHODS

Materials. TGase from *Streptomyces mobaraensis* was purchased from Ajinomoto Co. (Tokyo, Japan), whereas bovine α-LA (α-LA, type I, calcium saturated) and V8-protease from *Staphylococcus aureus* were from Sigma-Aldrich (Milwaukee, WI, USA). Carbobenzoxy-L-glutamyl-glycine (ZQG) and dansylcadaverine (DC) were purchased from Sigma while porcine trypsin was from Promega (Madison, WI, USA). All other chemicals where purchased from Sigma. The activity of TGase solutions was measured as described in Chapter 2 (Material and Methods).

TGase-mediated conjugation of α-LA to DC and ZQG. α-LA was dissolved in buffer 20 mM Tris, 5 mM EDTA pH 7.5 for apo-α-LA and in 20 mM Tris, 8 mM $CaCl_2$ pH 7.5 for the holo from of the protein. The concentration of the protein solutions (~ 0.9 mg/ml) was measured on the basis of the absorbance at 280 nm (extinction coefficient of 2.1 mg/ml, according to Gill and von Hippel, 1989 *(35)*). The apo- and holo-α-LA solutions were mixed with a solution 20 mg/ml of DC (in DMSO) at a molar ratio of 1:30 while TGase was added to the reaction mixture at an enzyme:substrate (E/S) ratio of 1/50 by weight. The reaction was allowed to proceed at 37 °C and aliquots were collected after 0, 1, 5, 10, and 20 min fro apo-α-LA and 0, 30 min and 2 hours for holo-α-LA. The reactions were stopped by addition of an equal volume of an aqueous solution of 1% TFA. Aliquots of the reaction were analyzed by RP-HPLC using an Agilent series 1100 HPLC with an online UV detection from Agilent Technologies (Waldbroon, Germany). RP-HPLC analyses were performed on a C4 Phenomenex column (150 x 4.6 mm) applying a two steps gradient of acetonitrile (AcCN), 0.085 % TFA and water, 0.1 % TFA: from 5 to 30 % of AcCN in 5 min and from 30 to 50% in 20 min. Fractions collected from the RP-HPLC analyses were lyophilized and analyzed by MS.

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

Proteolytic digestion of modified α-LA. Native α-LA and DC- or ZQG-modified α-LA were dissolved in 50 mM NH_4HCO_3 (final protein concentration ~ 25μM) and a stock

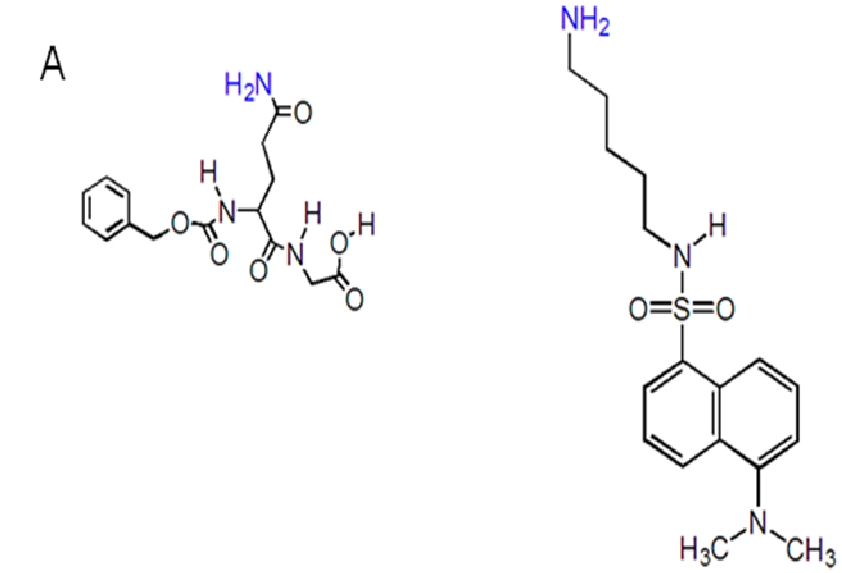
solution containing V8-protease and trypsin was added to obtain final E/S ratios of 1/25 for V8-protease and 1/50 for trypsin (by weight). Proteolyses were let to proceed at 37°C overnight and then acidified with an acqueous solution of 1% formic acid, 7 mM TCEP to obtain a final concentration of 2.5 mM TCEP. Samples were then incubated for 3 hours at 37°C to allow the reduction of the disulphide bonds and then analysed analysed by ESI-MS and MS/MS.

Mass spectrometry analyses. 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 variable collision energy values 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).

RESULTS AND DISCUSSION

2.1. Derivatization at the level of Gln residues using dansyl cadaverine. The presence of Gln residues reactive towards TGase was analysed using dansylcadaverine (DC) as model substrate since it contains a primary amine that is a good amino donor (Figure 3A). The time course of the reaction was followed by RP-HPLC analysis of the reaction mixture at different time of incubation followed by ESI-MS analysis of the protein material eluted in the chromatographic peaks (Figure 4 and Table 1). Apo-α-LA was incubated a 37°C with TGase (E/S 1/50 by weight) in the presence of 30 molar excess of DC. Under these conditions, the derivatization of some Gln residues of the protein occurs very quickly as it can be observed from the RP-HPLC analyses of the reaction mixture after one and ten minutes of incubation (Figure 4, left). Indeed, already after ten minutes the chromatographic peak corresponding to the native protein disappeared and two new peaks are detected. By ESI-MS analysis of the protein material eluted in these peaks, we determined the formation of mono-, bi- and three-DC derivatives of the protein (Table 1). In order to identify the first sites of modification, we analysed by RP-HPLC the reaction mixture after one minute of incubation and we detected the presence of native protein and of a mono-derivative species eluting at a slightly higher retention time. The α-LA^{1DC} derivative collected after one minute of reaction was subjected to trypsin and V8 protease hydrolysis. MS analysis of the digest followed by MS/MS analysis of the modified peptides allowed to identify the first sites of TGase modification in Gln39 and in much lower percent in Gln54 (Table 2; Figure 5). The same analysis was performed on the bi- and tri-DC derivatives isolated after 10 minutes of reaction and Gln43 and Gln65 were identified as secondary sites of modification (Table 2).

Clearly, the TGase-mediated derivatization of the Gln residues of apo-α-LA is quite site-specific and most interestingly, the modified residues are all located in the same chain region encompassing the sites of limited proteolysis by several proteases of α-LA in its MG state (Fig. 1B). We investigated under the same conditions the reactivity to TGase of the calcium bound form of the protein (holo-α-LA). Holo-αLA is resistant to protease hydrolysis and as expected it is also resistant to modification by TGase at the level of Gln residues since after 10 minutes of reaction we did not observed the production of any derivative (data not shown). Therefore, a mechanism of local unfolding of the β-domain in α-LA appears to strongly facilitate both proteolysis and TGase reactions.



B

```
         10          20          30          40          50
EQLTKCEVFR ELKDLKGYGG VSLPEWVCTT FHTSGYDTQA IVQNNDSTEY

         60          70          80          90          100
GLFQINNKIW CKDDQNPHSS NICNISCDKF LDDDLTDDIM CVKKILDKVG

        110         120   123
INYWLAHKAL CSEKLDQWLC EKL
```

Figure 3. Substrates of TGase used in this study. (A) Chemical structures of carbobenzoxy-L-glutamyl-glycine (ZQG) and dansylcadaverine (DC). (B) Amino acid sequence of bovine α-LA (Swiss-Prot entry P00711). Lys and Gln residues potential sites of the TGase-mediated derivatization are shown in bold black. The four α-helices along the 123-residue chain of the protein are indicated by red boxes and the three β-strands that are destabilized in apo-α-LA by dashed blue boxes.

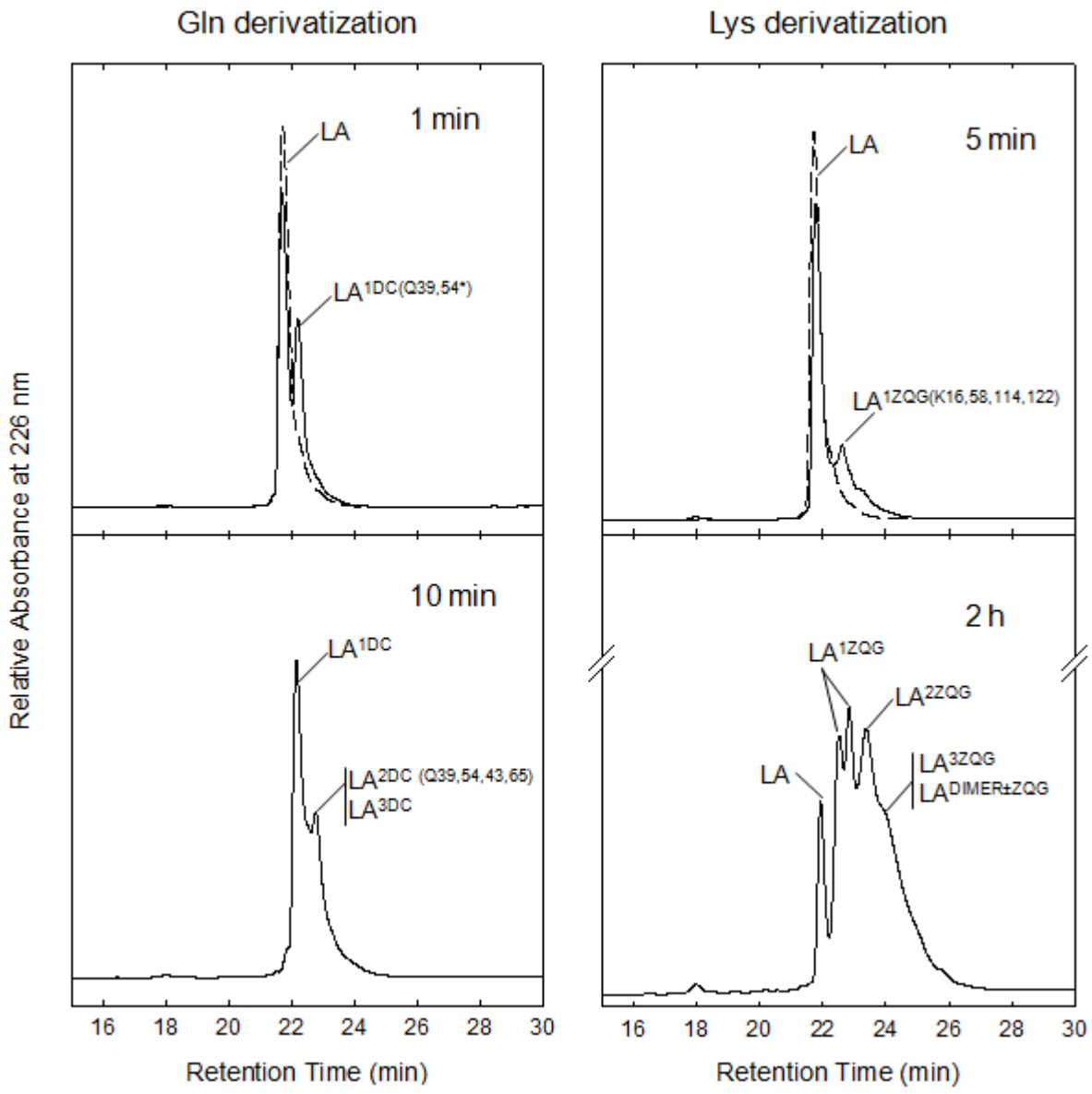


Figure 4. TGase-mediated conjugation at Gln and Lys residues of apo-LA. RP-HPLC analyses of the TGase-conjugated derivatives of apo-α-LA. A dash-dotted line indicates the chromatograms of native apo-α-LA, the straight line the DC- and ZQG-conjugated species at the different times of reaction. * The derivatization of Gln54 after 1 minute of reaction has been detected by MS at trace level.

2.2. Derivatization at the level of Lys residues using carbobenzoxy-L-glutamyl-glycine.

In order to identify Lys residues in the α-LA sequence that can be modified by TGase, the protein was incubated at 37°C with TGase (E/S 1/50, by weight) and carbobenzoxy-L-glutamyl-glycine (ZQG) at a 50 molar excess in respect to α-LA. ZQG is a Gln containing substrate that is efficiently conjugated to Lys residues by the TGase catalysed reaction. RP-HPLC analysis of the reaction mixtures indicated that the derivatization of apo-α-LA with

TGase in the presence of ZQG proceeds much more slowly then with DC. Indeed, after five minutes of reaction a new chromatographic peak with a slightly higher retention time in respect to the native protein is detected but it shows a low intensity (Fig. 4, right). This peak was shown by ESI-MS analysis to contain a monoderivative of α-LA with ZQG (Table 1).

Table 1. Molecular masses of α-LA and α-LA derivatives.

protein species	molecular mass (Da)	
	found[a]	calculated[b]
α-LA	14178.84	14178.06
$\alpha\text{-LA}^{1ZQG}$	14498.85	14498.36
$\alpha\text{-LA}^{2ZQG}$	14819.45	14818.66
$\alpha\text{-LA}^{3ZQG}$	15139.58	15138.97
$\alpha\text{-LA}^{DIMER}$	28340.69	28339.09
$\alpha\text{-LA}^{DIMER,1ZQG}$	28660.93	28659.42
$\alpha\text{-LA}^{DIMER,2ZQG}$	28980.50	28979.75
$\alpha\text{-LA}^{DIMER,3ZQG}$	29301.52	29300.08
$\alpha\text{-LA}^{1DC}$	14496.48	14496.50
$\alpha\text{-LA}^{2DC}$	14815.19	14814.94
$\alpha\text{-LA}^{3DC}$	15133.18	15133.25

[a] Experimental molecular masses determined by ESI-MS. [b] Calculated molecular masses.

Interestingly, trypsin and V8 protease digestion of this conjugation product and MS analysis of the digest allowed to identify four different initial sites of modification and in particular Lys16, Lys58, Lys114 and Lys122 (Table 2). The RP-HPLC chromatogram of the reaction mixture after 2 hours of incubation showed still the presence of the peak corresponding to the native protein and four more poorly separated peaks that by MS were shown to contain mono-, bi and tri-ZQG derivatives of α-LA (Table 1). After this time of incubation, we also detected the presence of dimers of the protein and dimer species conjugated to one to three molecules of ZQG.

Overall, the TGase-mediated derivatization of α-LA at the level of Lys residues proceeds much more slowly and with low specificity in respect to the derivatization at the level of Gln residues. The reason of these results can be envisaged in the absence of lysines in the amino acid sequence encompassing the denaturated β-sheet region. Indeed, among the Lys residues that are modified by TGase, only Lys 58 is located nearby the β-sheet region of the

Table 2. Molecular masses of the fragments obtained upon digestion with trypsin and V8 protease and reduction with TCEP of α-LA and its derivatives.

LA derivative	fragment[a]	site of conjugation	molecular mass (Da)	
			found[b]	calculated[c]
α-LA	26–49		2716.27	2716.17
	32–49		1979.05	1978.85
	50–58		1095.61	1095.57
	59–79		2419.28	2419.04
	63–79		1888.95	1888.77
	115–123		1146.61	1146.57
	14–25		-	1233.62
	50–62		-	1625.84
	109–122		-	1664.79
Apo-α-LA^{1DC}	26–49[d]	Gln39^{DC}	3034.62	3034.31
	32–49[e]	Gln39^{DC}	2297.22	2296.99
	50–58	Gln54^{DC}	1413.90	1413.71
Apo-α-LA^{2DC} (α-LA^{1DC}, α-LA^{3DC})	26–49[d]	Gln39^{DC}	3034.53	3034.31
	50–58	Gln54^{DC}	1413.82	1413.71
	32–49[e]	Gln39^{DC}, Gln43^{DC}	2615.35	2615.13
	59–79	Gln65^{DC}	2737.40	2737.18
	63–79	Gln65^{DC}	2207.07	2206.91
Holo-α-LA^{1ZQG}	115–123[f]	Lys122^{ZQG}	1466.77	1466.68
Apo-α-LA^{1ZQG}	14–25	Lys16^{ZQG}	1553.79	1553.72
	50–62	Lys58^{ZQG}	1945.97	1945.94
	109–122	Lys114^{ZQG}	1984.95	1984.89
	115–123[f]	Lys122^{ZQG}	1466.74	1466.68

[a] Only the molecular masses of the fragments which contain the Gln and Lys residues derivatised with DC or ZQG are reported. [b] Experimental molecular masses determined by ESI-MS. [c] Monoisotopic molecular masses calculated from the amino acid sequence of α-LA. [d] The site of derivatization was determined by MS/MS analysis due to the presence of two Gln residues in this fragment (data not shown). [e] Fragment 32-49 of α-LA is generated by aspecific cleavage of trypsin at the level of the peptide bond Phe31–His32. The identity of this fragment has been confirmed by MS/MS analysis. The site of derivatization was also determined by MS/MS analysis due to the presence of two Gln residues in this fragment (Fig. 5). [f] Since this fragment is isobaric with fragment 8-16 modified with ZQG at Lys13, its identity has been determined by MS/MS (Figure 8).

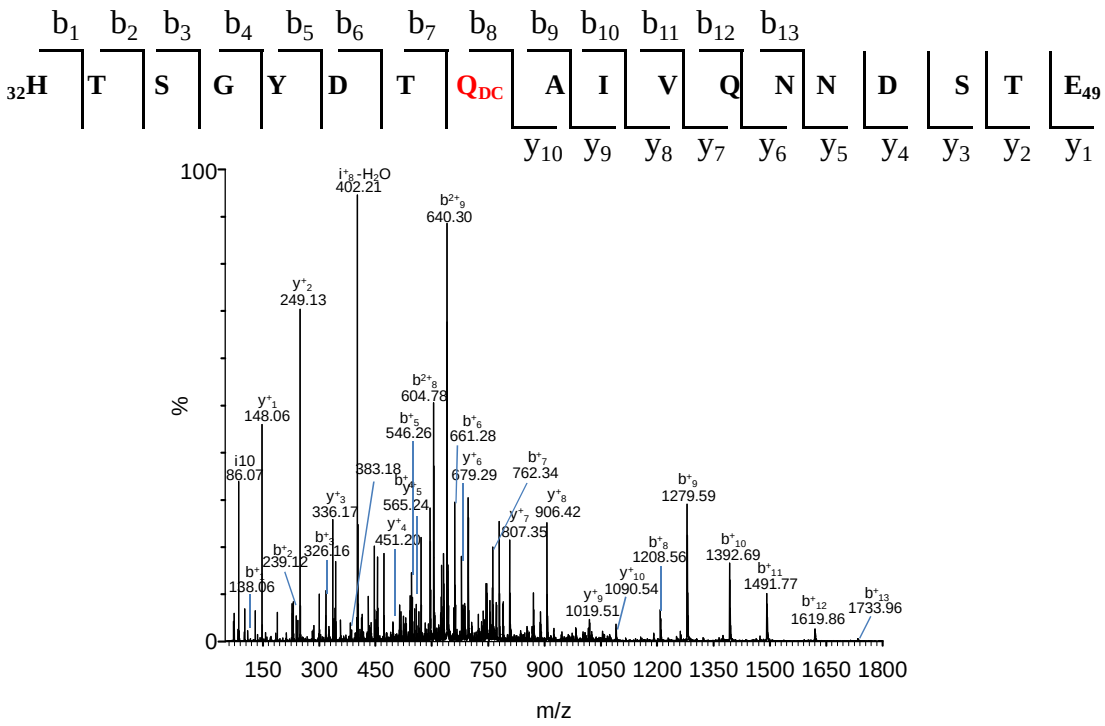


Figure 5. Electrospray MS/MS mass spectrum of peptide 32–49 of α-LA modified with DC (Top). Fragments of the series b and y that were identified in the MS/MS spectrum are indicated on the sequence of the peptide. Gln39 is indicated in red as Q_{DC}. (Bottom) In the MS/MS spectrum, ions assigned to fragments of the series y and b are indicated. The fragment b^{2+}_{8} indentified in the MS/MS spectrum contain Gln39 and show a mass increment corresponding to the conjugation with DC.

protein. The other modified Lys residues are distributed along the sequence of α-LA albeit outside regions of α-helical structure (Figure 3B). We suggest that being apo-α-LA in a relaxed and flexible conformational state, TGase can derivatise Lys residues outside the denaturated β-sheet region but with low efficiency. Interestingly, we performed a limited proteolysis experiments on apo-α-LA with trypsin, a protease that specifically hydrolyses polypeptide chains at the level of Lys and Arg residues (Figure 6, Table 3)). Even with trypsin, we identified seven different initial sites of nicking that are distribute at random along the polypeptide chain of α-LA and we can explain this results with the observation that Lys and Arg residues are absent in the most flexible denatured β-sheet region. We can conclude that even in the case of the derivatization of Lys residues, TGase shows a specifity similar to that observed by proteases.

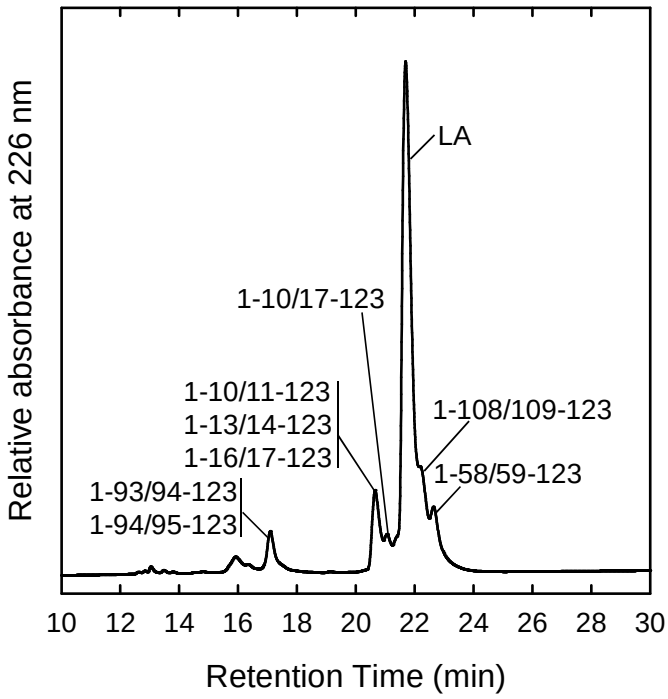


Figure 6. Limited proteolysis of apo α-LA with trypsin. RP-HPLC analysis of the reaction mixture of apo-α-LA with trypsin after 15 min of incubation. In the chromatogram, the identities of the different nicked species are indicated.

Table 3. Molecular masses measured for the nicked forms of α-LA produced by limited proteolysis with trypsin after reduction with TCEP.

nickel species	reduced fragment	molecular mass (Da)	
		found[a]	calculated[b]
1-10/11-123	1-10	1251.13	1251.63
	11-123	12947.56[c]	12951.68
1-13/14-123	1-13	1621.27	1621.85
	14-123	12576.09[c]	12581.23
1-16/17-123	1-16	1977.79	1978.06
	17-123	12220.91[c]	12224.81
1-93/94-123	1-93	10644.96	10644.88
	94-123	3556.78	3556.89
1-94/95-123	1-94	10773.12	10773.05
	95-123	3428.18	3428.80
1-10/17-123	1-10	1249.95	1252.63
	17-123	12220.91[c]	12224.81
1-108/109-123	1-108	12425.01	12425.09
	109-123	1777.66	1777.87
1-58/59-123	1-58	6617.85	6617.35
	59-123	7586.21	7586.79
1-10/11-123	1-10	1251.13	1251.63
	11-123	12947.56[c]	12951.68
1-13/14-123	1-13	1621.27	1621.85
	14-123	12576.09[c]	12581.23

[a] Experimental molecular masses determined by ESI-MS. [b] Calculated molecular masses considering the cysteine residues in a reduced form. [c] The 4 Da difference of the found mass in respect to the calculated mass is due to the presence of 4 cysteine residues in an oxidised form.

We tested also the reactivity of holo-α-LA towards the TGase mediated derivatization of Lys residues. Interestingly, incubation of holo-α-LA with ZQG and TGase leads to the quantitative modification of the protein, as shown by the RP-HPLC analysis of the reaction mixture after 2 hours of incubation (Fig. 7A). ESI-MS analysis allowed to confirm that the product of the reaction is a mono-ZQG derivative of α-LA (Table 1). Mass fingerprinting analysis of this α-LA^{1ZQG} derivative allowed to identify in Lys122 the site of modification (Table 2, Fig. 8). Holo-α-LA is the native stable conformational state of the protein. However, the reactivity of Lys122 can be explained in light of the B-factor profile derived from the X-ray structure of the holo-protein (Fig. 7B; *(3)*). Indeed, the B-factor values of the polypeptide backbone are very high in proximity of the C-terminus (residues Lys122-Leu123) indicating that the polypeptide chain is highly flexible. We can conclude that similarly to the derivatizaion of Gln residues, also the selective modification of holo-α-LA at the level of Lys residues is dictated by the presence of Lys122 in an unfolded/disordered region of the protein.

In conclusion, in this study we demonstrated that similarly to the specificity of protease hydrolysis in limited proteolysis experiments, also the TGase-mediated site-specific derivatisation of Gln and Lys residues of α-LA depends on the presence of these residues at the level highly flexible or unfolded regions of the polypeptide chain of the protein. Indeed, when this requirement is not fulfilled we observe no derivatization, as in the case of the resistance of holo-α-LA to Gln modification. On the other end, in holo-α-LA the presence of only one Lys at the level of the highly flexible C-terminus of the protein leads to the quantitative production of a mono-ZQG α-LA derivative. This result in particular suggests that TGase can be used to selectively modified α-LA at the level of Lys residues. On this basis, in general we suggest that also the derivatization of lysine residues by TGase can have important biotechnological applications since it can be exploited for site-specific protein labelling.

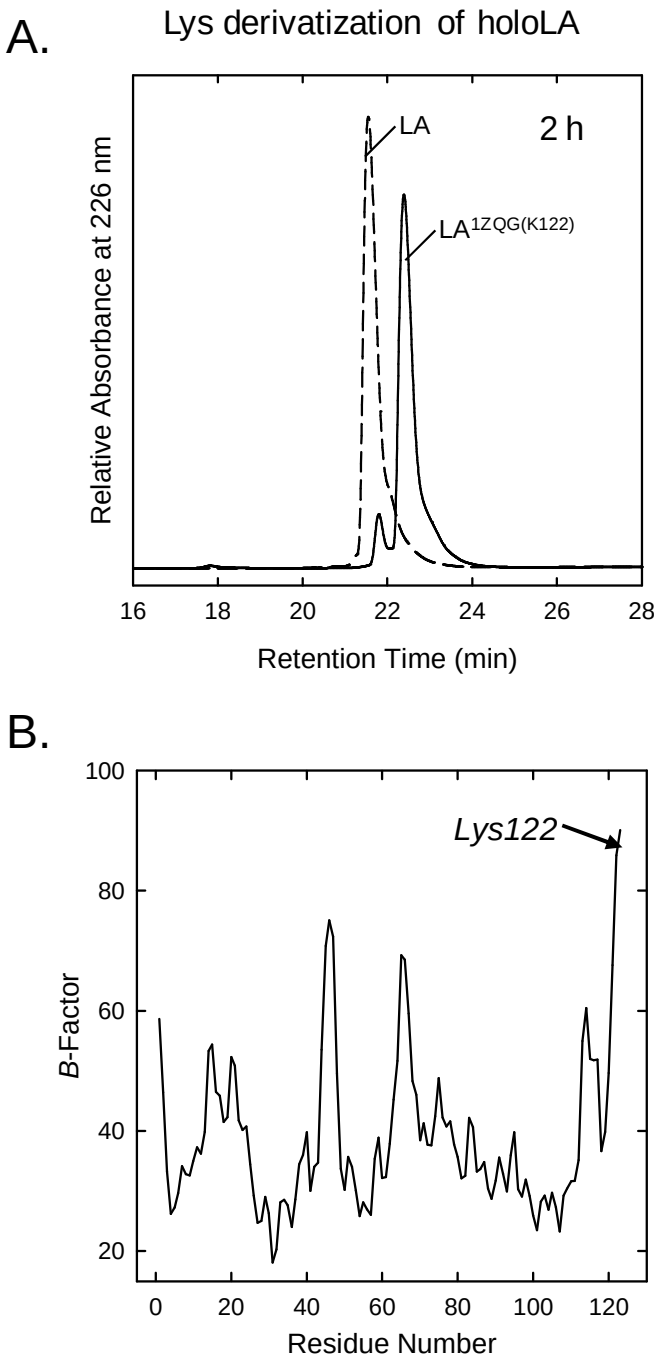


Figure 7. TGase-mediated conjugation at Gln residues of holo-α-LA. (A) RP-HPLC analyses of the TGase-conjugated derivatives of holo-α-LA with ZQG after 0 (dashed line) and 2 (straight line) hours of reaction. (B) Plot of the backbone *B*-factor along the 123-residue polypeptide chain of holo-α-LA (PDB file 1HFZ; Pike et al., 1996). The position of Lys122 is indicated by an arrow.

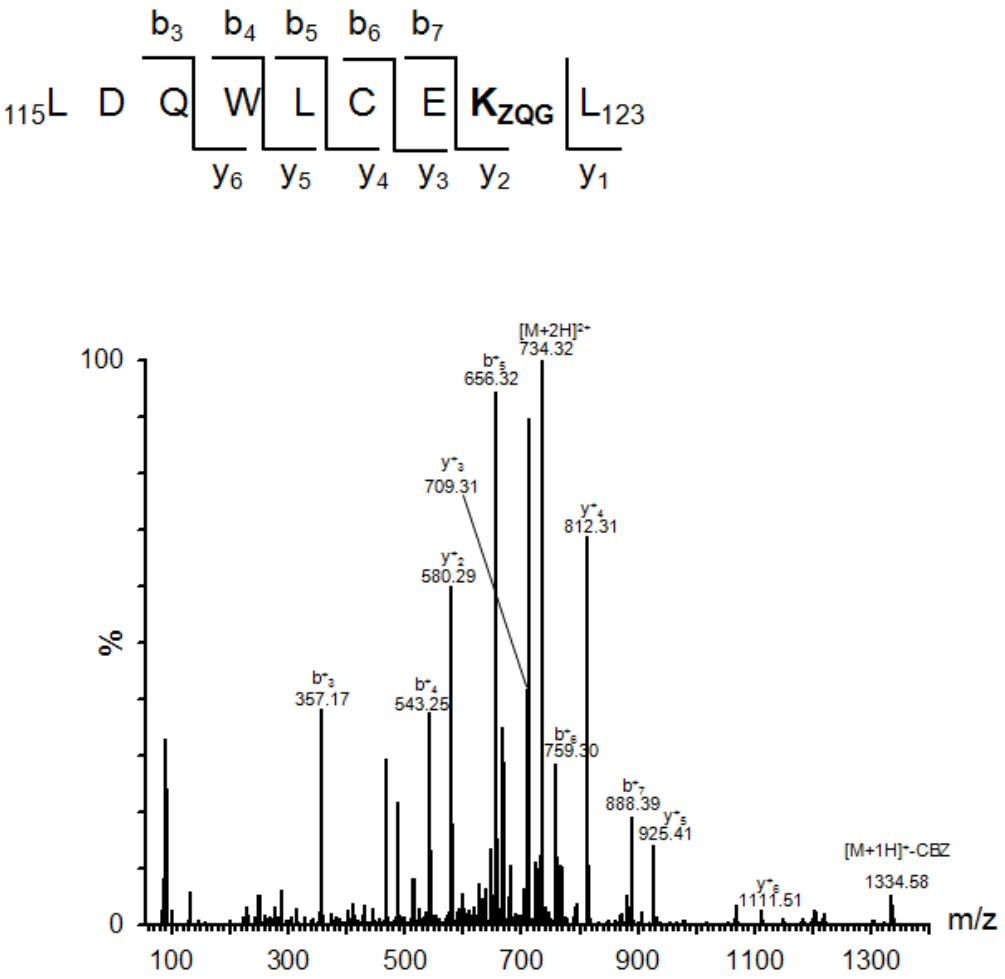


Figure 8. Electrospray MS/MS mass spectrum of the double charged ion at 734.32 m/z of peptide 115-123 of α-LA modified with ZQG. (Top) Fragments of the series b and y that were identified in the MS/MS spectrum are indicated on the sequence of the peptide. Lys122 is indicated in bold as K_{122}^{ZQG}. (Bottom) In the MS/MS spectrum, ions assigned to fragments of the series y and b are indicated. The fragment $y_2$indentified in the MS/MS spectrum contain Lys122 and show a mass increment corresponding to the conjugation with ZQG.

References

1. Hill, R.L., Brew, K. (1975) Lactose synthetase. *Adv. Enzymol.* 43, 411-490.

2. Permyakov, E.A., Yarmolenko, V.V., Kalinichenko, L.P., Morozova, L.A., Burstein, E.A. (1981) Calcium binding to alpha-lactalbumin: structural rearrangement and association constant evaluation by means of intrinsic protein fluorescence changes. *Biochem. Biophys. Res. Commun.* 100, 191-197.

3. Pike, A. C., Brew,W. K., Acharya, K. R. (1996) Crystal structures of guinea-pig, goat and bovine α-lactalbumin highlight the enhanced conformational flexibility of regions that are significant for its action in lactose synthase. Structure 4:691–703.

4. Acharya, K.R., Stuart DI Phillips, D.C., McKenzie, H.A., Teahan, C.G. (1994) Models of the three-dimensional structures of echidna, horse, and pigeon lysozymes: calcium-binding lysozymes and their relationship with alpha-lactalbumins. *J. Protein Chem.* 13, 569-584.

5. Permyakov, E.A., Shnyrov, V.L., Kalinichenko, L.P., Kuchar, A., Reyzer, I.L., Berliner L.J. (1991) Binding of Zn(II) ions to alpha-lactalbumin. *J. Protein Chem.* 10, 577-584.

6. Permyakov, E.A., Morozova, L.A., Burstein, E.A. (1985) Cation binding effects on the pH, thermal and urea denaturation transitions in alpha-lactalbumin. *Biophys. Chem.* 21, 21-31.

7. Veprintsev, D.B., Permyakov, S.E., Permyakov, E.A., Rogov, V.V., Cawthern, K.M., Berliner, L.J. (1997) Cooperative thermal transitions of bovine and human apo-alpha-lactalbumins: evidence for a new intermediate state. *FEBS Lett.* 412, 625-628.

8. Kuwajima, K. (1989) The molten globule state as a clue for understanding the folding and cooperativity of globular-protein structure. Proteins 6:87–103.

9. Ptitsyn, O. B**.** (1995) Molten globule and protein folding. Adv. Protein Chem. 47:183–229.

10. Ptitsyn, O. B. (1995) Structures of folding intermediates. Curr. Opin. Struct. Biol. 5:74–78.

11. Evans, P. A., Radford, S.E. (1994). Probing the structure of folding intermediates. Curr. Opin. Struct. Biol. 4:100–106.

12. Dyson, H. J., and P. E. Wright. (2002) Insights into the structure and dynamics of unfolded proteins from nuclear magnetic resonance. Adv. Protein Chem. 62:311–340.

13. Dyson, H. J., and P. E. Wright. (2004) Unfolded proteins and protein folding studied by NMR. Chem. Rev. 104:3607–3622.

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

15. Arai, M., K. Kuwajima. (2000) Role of the molten globule state in protein folding. Adv. Protein Chem. 53:209–282.

16. Dolgikh, D.A., Gilmanshin, R.I., Brazhnikov, E.V., Bychkova, V.E., Semisotnov, G.V., Venyaminov, S.Y., Ptitsyn, O.B. (1981) Alpha-Lactalbumin: compact state with fluctuating tertiary structure? *FEBS lett.* 136, 311-315.

17. Kuwajima, K. (1996) The molten globule state of α-lactalbumin. *FASEB J.* 10, 102-109.

18. Permyakov, E.A., Berliner, L.J. (2000) alpha-Lactalbumin: structure and function. *FEBS Lett.* 473**,** 269-274.

19. Alexandrescu, A. T., Evans, P. A., Pitkeathly, M., Baum, J., Dobson, C. M. (1993) Structure and dynamics of the acid-denatured molten globule state of α-lactalbumin: A two-dimensional NMR study. Biochemistry 32:1707–1718.

20. Chyan, C. L., C. Wormald, C. M. Dobson, P. A. Evans, and J. Baum. (1993) Structure and stability of the molten globule state of guinea-pig α-lactalbumin: A hydrogen exchange study. Biochemistry 32:5681–5691.

21. Quezada, C. M., B. A. Schulman, J. J. Froggatt, C. M. Dobson, and C. Redfield. (2004) Local and global cooperativity in the human alpha-lactalbumin molten globule. J. Mol. Biol. 338:149–158.

22. Polverino de Laureto, P., V. De Filippis, M. Di Bello, M. Zambonin, and A. Fontana. (1995) Probing the molten globule state of α-lactalbumin by limited proteolysis. Biochemistry 34:12596–12604.

23. Polverino de Laureto, P., Frare, E., R. Gottardo, and A. Fontana. (2002) Molten globule of bovine α-lactalbumin at neutral pH induced by heat, trifluoroethanol and oleic acid: A comparative analysis by circular dichroism spectroscopy and limited proteolysis. Proteins 49:385–397.

24. Håkansson, A., B. Zhivotovsky, S. Orrenius, H. Sabharwal, and C. Svanborg. (1995) Apoptosis induced by a human milk protein. Proc. Natl. Acad. Sci. USA 92:8064–8068.

25. Svensson, M., Sabharwal, H., Hakansson, A., Mossberg, A.K., Lipniunas, P., Leffler, H., Svanborg, C., Linse, S. (1999) Molecular characterization of α–lactalbumin folding variants that induce apoptosis in tumor cells *J. Biol. Chem.* 274**,** 6388–6396.

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

27. Polverino de Laureto, P., D. Vinante, E. Scaramella, E. Frare, and A. Fontana. (2001) Stepwise proteolytic removal of the β-subdomain in α-lactalbumin: The protein remains folded and can form the molten globule in acid solution. Eur. J. Biochem. 268:4324–4333.

28. Polverino de Laureto, P., Scaramella, E., Frigo, M., Wondrich, F.G., De Filippis, V., Zambonin, M., Fontana, A. (1999) Limited proteolysis of bovine α-lactalbumin: isolation and characterization of protein domains. *Protein Sci.* 8, 2290-2303.

29. Schulman, B.A., Peng, Z.Y., Redfield, C., Dobson, C.M., Kim, P.S. (1995) Different subdomains are most protected from hydrogen exchange in the molten globule and native states of alpha-lactalbumin. *J. Mol. Biol.* 253, 651-657.

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

31. Matsumura, Y., Yuporn, C., Kumazawa, Y.. Ohtsuka, T., Mori, T. (1996) Enhanced susceptibility to transglutaminase reaction of α-lactalbumin in molten globule state.*Biochim. Biophys. Acta* 1292: 69–76.

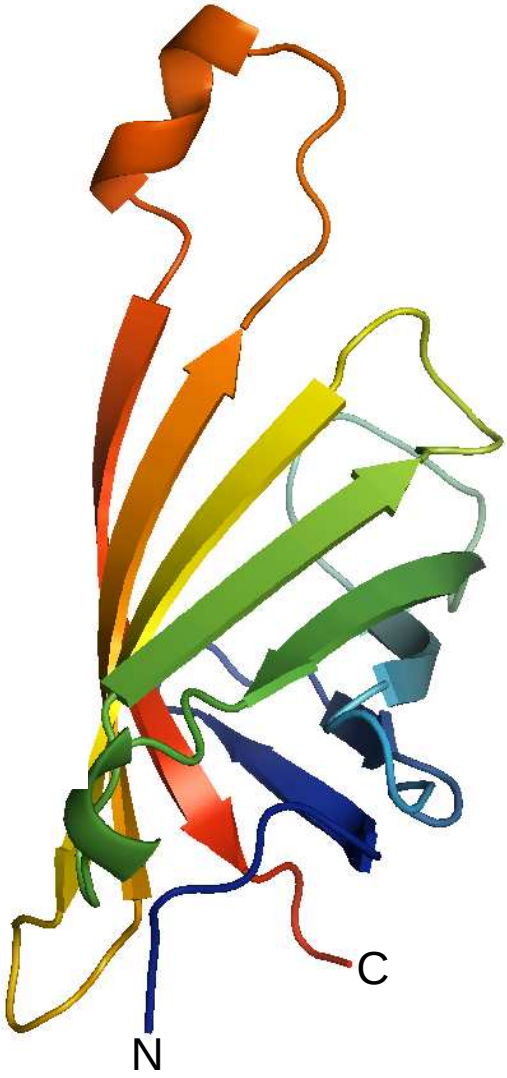
32. Gu, Y.S., Matsumura, Y., Yamaguchi, S., Mori, T. (2001) Action of protein-glutaminase on α-lactalbumin in the native and molten globule states. *J. Agric. Food Chem*. 49: 5999–6005.

33. Nieuwenhuisen, W.F., Dekker, H.L., De Koning, L.J.. Groneveld, T., De Koster, C.G., 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.

34. Lee, D.S., Matsumoto, S., Matsumura, Y., 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.

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

Results



Avidin

2. Site-specific avidin-protein conjugation using microbial transglutaminase

INTRODUCTION

Avidin: structural features and biotin binding properties. Avidin is a homotetrameric glycoprotein found in the chicken egg-white and whose name originates from the high affinity of this protein for biotin **(1)**. The avidin tetramer (molecular weight ~ 64 kDa) is composed of four identical polypeptide chains of 128 aminoacids with a disulphide bond between residues Cys4 e Cys83 and glycosylated at the level of Asn17 (Fig1A). Each subunits can bind biotin with the highest affinity known in nature for a protein-ligand interaction (*Kd* ~10^{-15}M) and under physiological conditions, the binding can essentially be considered irreversible (2-3). This interaction is indeed unaffected by extremes of pH, temperature, organic solvents and other denaturing agents requiring 6-8 M guanidinium chloride pH 1.5, 120°C for 15 minutes in order to be dissociated *(4).* This tight interaction has been utilized during the past three decades for various applications in life sciences to purify, probe, and target various materials both *in vitro* and *in vivo (5).*

X-ray structures of both the apo form *(6-8)* and the holo form of avidin *(6, 9)* are known at high resolution. The 3D crystal structure of the avidin monomer is characterised by the presence of eight antiparallel successive β-strands and the overall shape is that of a β-barrel (Fig. 1B). The quaternary structure of avidin and its bacterial analogue streptavidin consists of a dimer of dimers (Fig. 2) *(10).* Each avidin dimer is stabilised by an intimate set of intramolecular interactions that occur at the interface between two monomers, referred to as the 1–4 monomer–monomer interaction. The strong interaction at the 1–4 interface is a product of the cooperativity between extensive polar and hydrophobic interactions, in particular between residues Asn54 and Asn69, each from opposite monomers. In contrast, the dimer–dimer interface is less intricate and involves two types of monomer–monomer interfaces, named the 1–2 and 1–3 interfaces. The interface between monomers 1 and 3 in avidin is maintained by only three interacting residues from each monomer (Val115, Met96 and Ile117), while the 1–2 interface involves mainly Trp110 contributed by each

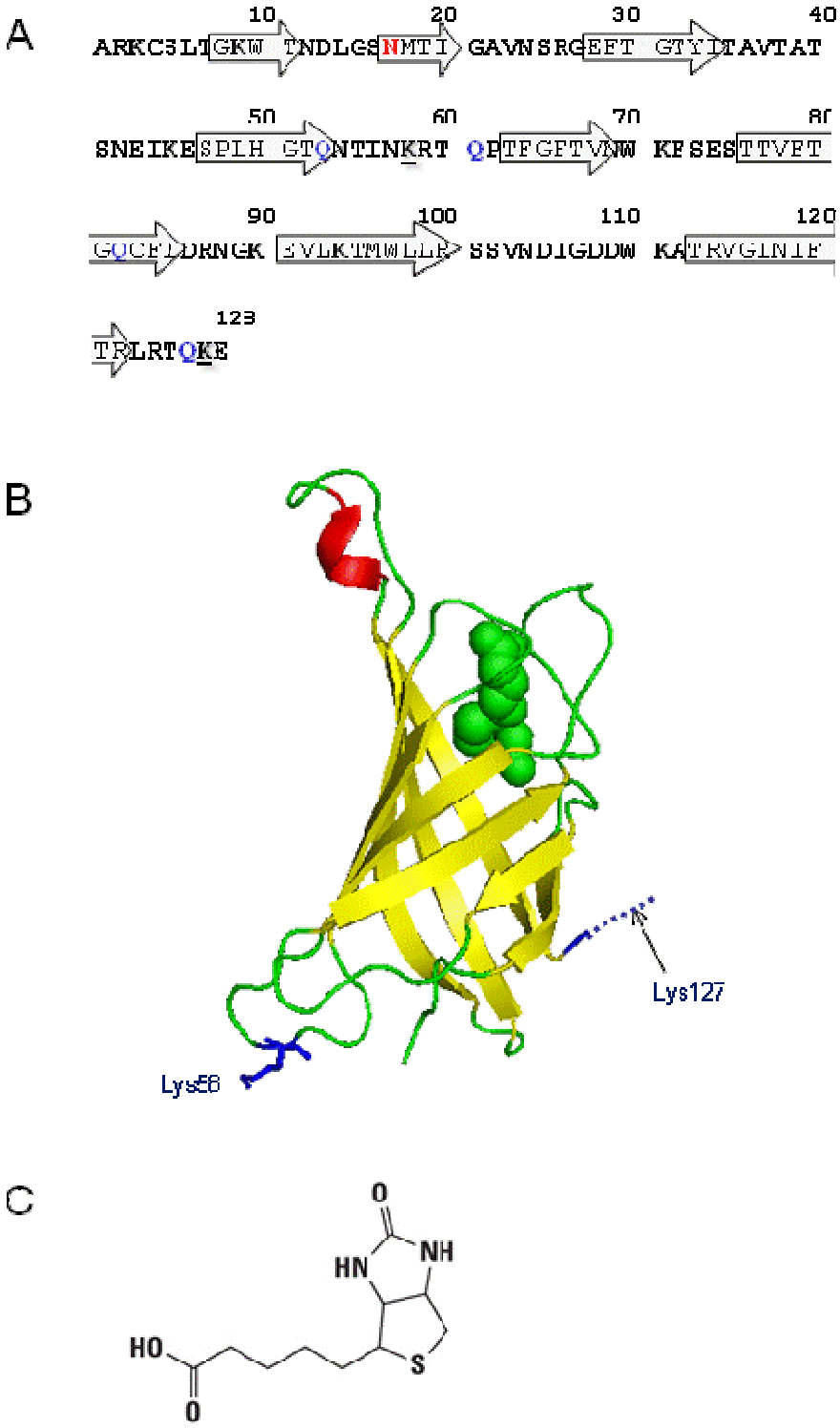


Figure 1. Sequence and 3D structure of chicken avidin. (A) Aminoacid sequence of chicken avidin (Swiss-Prot entry P02701). Residues potential sites of the TGase-mediated derivatization are shown in bold black for Lys residues and in bold blue for Gln residues. Lys58 and Lys127 that are the sites of TGase modification are underlined. Asn17 which is the site of glycosilation is indicated in bold red. The eight β-strands along the 128-residue chain of the protein are indicated by grey arrows. (B) 3D structure of Avidin (PDB file 1AVD, Pugliese et al., 1993). The biotin molecule is shown in green, while Lys58 and Lys127 are coloured in blue. (C) Chemical structure of biotin.

monomer and located on loop between strands 7 and 8. This last interaction becomes stronger upon interaction of the avidin tetramer with biotin, stabilising the quaternary structure of the protein. Introducing mutation W110K and N54A in the avidin sequence produced a monomeric avidin form that binds biotin in a reversible manner (K_d ~10^{-8}M) and that can be usefull in some applications *(11).*

The biotin binding site is positioned at one end of the avidin barrel (Fig.1B) and is characterised by several aromatic and polar residues that are involved in the tight binding. This binding pocket maintains a strong shape complementarity for the ligand also in the apo form of the protein, a characteristic that is common to many high affinity interactions *(6, 10).* Indeed, in the absence of biotin five water molecules are situated in the biotin-binding pocket that are then replaced by the ligand. After biotin binding, some modulations of the tertiary and quaternary structure of avidin also take place. In particular, the loop between strands 3 and 4 (residues 36–44) that is flexible in the apo-form of the protein becomes ordered and locks the ligand in the binding site *(6)* since three aminoacid residues of this loop provide additional interactions with biotin. Interestingly, limited proteolysis experiments confirmed this conformational change. Indeed, hydrolysis of apo-avidin with the fungal proteinase K occurs at this loop between β-strands 3 and 4, specifically cleaving peptide bonds Thr40–Ser41 and Asn42–Glu43 *(12),* while biotin binding makes the protein fully resistant to proteolysis. Overall, binding of biotin does not lead to major alterations of protein structure, but it significantly protects avidin against the action of denaturing agents. In particular, the heat stability of the avidin-biotin complex is significantly enhanced (Tm 132°C) in respect to that of the apo form (Tm 85°C) *(13).*

Biotechnological applications of the avidin-biotin interaction. Avidin and its bacterial analogue streptavidin are widely used protein in different fields of biotechnology where their extremely high affinity toward biotin is exploited *(14).* The avidin-biotin technology has indeed several applications due to the ease with which biotin can be coupled to several molecules without compromising the avidin-biotin interaction or the function of the target molecule (Fig. 3A). Among the diverse applications of this protein-ligand interaction there is affinity chromatography where avidin is immobilsed on a matrix and it allows to selectively purify biotinilated peptides and proteins. Several applications are also in protein detection since biotinylated primary or secondary antibodies can be detected

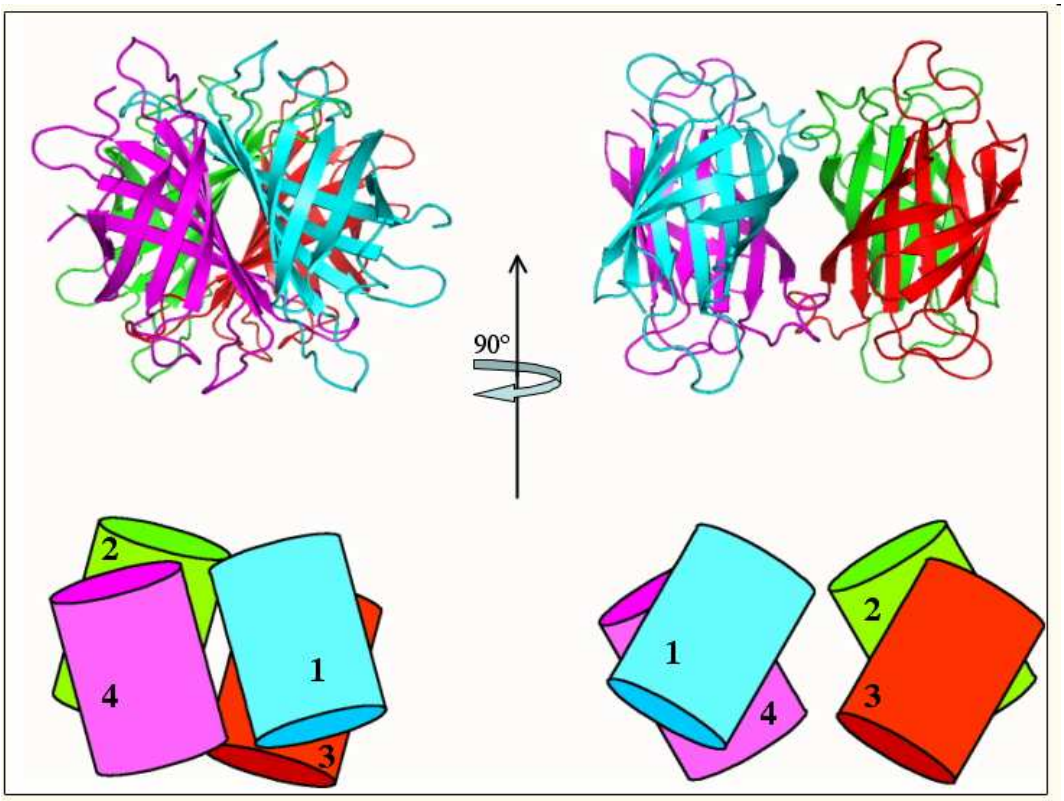


Figure 2. Scheme of the homotetrameric structure of avidin. The ribbon diagram (top,) and the cartoon (bottom) show the relative arrangement of the four monomers in the avidin quaternary structure. The tetramer at right is rotated clockwise by 90° along the vertical axis. The intimate interaction of the 1-4 monomer-monomer interface is clearly visible in the figure at left. The contact surface at the dimer-dimer interface is shown on the right (this image is from Eisenberg-Domovich et al., 2005)

with avidin conjugated to either fluorescent or enzymatic detection reagents (Fig. 3B). Common assays using this format include immunohistochemistry, Western blotting and ELISA. The main advantage of this approach is the possibility to amplify the signal of the original protein enabling the detection of protein expressed at low levels. Nowadays, avidin has also clinical applications *(15).* For example it is used for pre-targeted radioimmunotherapy, which is based on the injection in patients of avidin-antibody conjugates, following by injection of biotinilated radionuclide complex *(16).*

Aims of the study. In this PhD project we aimed to study the reactivity of avidin towards TGase. Clearly, the possibility to selectively modify avidin at the level of Gln and Lys residues can have important implications for the several biotechnological applications of this protein. Nowadays, the production of avidin-protein conjugates is achieved by chemical methods based on the use of bifunctional reagents or by molecular biology techniques *(17-20).* While chemical methods are not specific and they often lead to heterogeneous products with

lowered activity, the DNA based techniques are time consuming and they need a great effort in their optimization. The site-specific modification of avidin mediated by microbial TGase can thus be an alternative and usefull method for the conjugation of avidin to proteins and small molecules.

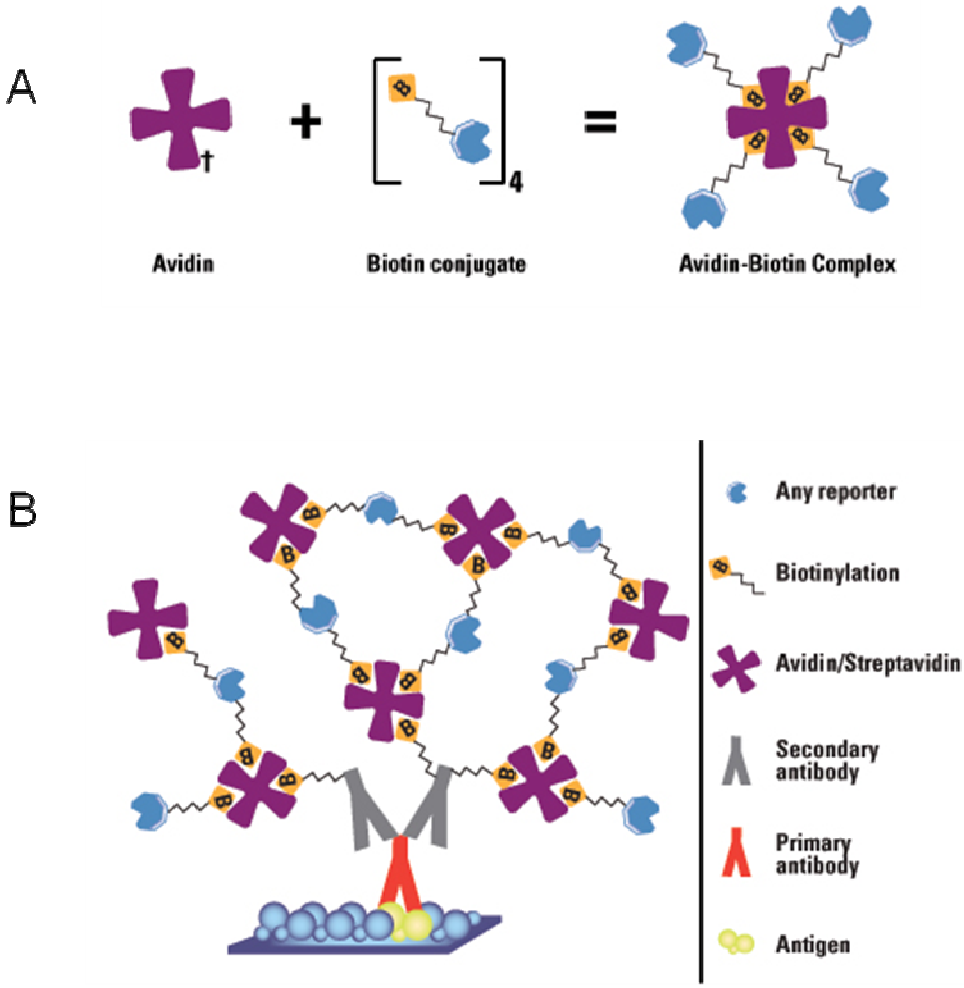


Figure 3. Avidin-biotin technology. (A) Schematic of the avidin-biotin interaction. Avidin can bind up to four biotin molecules, which are normally conjugated to an enzyme, antibody or target protein to form an avidin-biotin complex. † denotes that avidin is also often conjugated to an antibody, target protein or immobilized support. (B) Schematic of signal amplification by avidin-biotin complex fromation. This image is taken from Pierce website (www.piercenet.com).

MATERIALS AND METHODS

Materials. Transglutaminase (TGase) (1000 units/mg protein) isolated from *Streptoverticillium mobaraense* was purchased from Ajinomoto Co. (Tokyo, Japan) as TGase MP. Carbobenzoxy-L-glutamyl-glycine (ZQG) and dansylcadaverine (DC) were purchased from Sigma-Aldrich (Milwaukee, WI, USA). Porcine trypsin was from Promega (Madison, WI, USA). All other chemicals were of analytical reagent grade and were obtained from Sigma or Fluka. Avidin was purchased from Belovo (Bastogne, Belgium).

Activity of TGase solutions. The activity of the solutions of TGase was determined by hydroxamate formation in the enzyme catalysed reaction between the substrate 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 *(21)* (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.

Mass spectrometry analyses. 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 variable collision energy values 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).

TGase-mediated conjugation of avidin to DC and ZQG. Avidin was dissolved in buffer 0.1M NaH2PO4 pH 7.0 and the concentration of the protein solution was measured on the basis of the absorbance at 280 nm (extinction coefficient of 2.01 mg/ml, according to Gill and von Hippel *(21)*). The avidin solution was mixed with a solution 20 mg/ml of DC (in methanol) at a molar ratio of 1:50 and to obtain a final protein concentration of 0.8 mg/ml. TGase was added to the reaction mixture at an enzyme:substrate (E/S) ratio of 1/25 by weight and the reaction was allowed to proceed at 37 °C for 24 hours. Aliquots of 150 µg were collected after 0, 6 and 24 hours and the reaction was stopped by addition of an equal volume of an aqueous solution of 1% TFA. Since DC interferes with the RP-HPLC analyses, it was partially removed from the reaction mixture using VIVASPIN 500 with a membrane cut-off of 3000 MW (VIVASCIENCE, Stonehouse, UK). Aliquots of the reaction were analyzed by RP-HPLC using an Agilent series 1100 HPLC with an online UV detection from Agilent Technologies (Waldbroon, Germany). RP-HPLC analyses were performed on a C4 Phenomenex column (150 x 4.6 mm) applying a three steps gradient of acetonitrile (AcCN), 0.085 % TFA and water, 0.1 % TFA: from 5 to 22 % of AcCN in 5 min, from 22 to 50% in 17 min and from 22 to 95% in 3 min. Fractions collected from the RP-HPLC analyses were lyophilized and analyzed by MS.

For the derivatization with ZQG, a solution of avidin in 0.1M NaH2PO4 pH 7.0 was mixed with a solution 34 mg/ml of ZQG (in DMSO) at a molar ratio of 1:50 obtaining a final concentration of avidin of 0.7 mg/ml. The reaction mixture was incubated at 37 °C with TGase at an E/S ratio of 1/5, by weight. Aliquots of 100 µg were removed after 0 and 5 hours of incubation and the reaction was stopped by addition of an equal volume of 100 µM iodacetamide aqueous solution. RP-HPLC analyses of the reaction aliquots were performed as described above for the reaction with DC.

Trypsin digestion of avidin modified with ZQG. Avidin modified with ZQG was dissolved in 50 mM Tris-HCl pH 8.5, 8 M urea, 5 mM TCEP at a concentration of 2.5 mg/ml and incubated for 15 min at RT. Iodacetamide was added at a final concentration of 25 mM and the solution was further incubated in the dark for 20 min at RT. The avidin sample was then diluted with 50 mM NH4HCO3 pH 8.5 to obtain a final urea concentration of 2 M. Trypsin was added at an E/S 1/100 and the digestion was allowed to proceed at 37° C overnight. The proteolysis reaction was stopped by adding an equal volume of an aqueous solution of 1 % TFA and analyzed by RP-HPLC as described above for the reaction with DC. The peptide material collected from the chromatographic separation was analysed by ESI-MS and MS/MS.

TGase mediated conjugation of avidin to GM-CSF. GM-CSF was dissolved in buffer 0.1M NaH2PO4 pH 7.0 and protein concentration was measured on the basis of the absorbance at 280 nm (extinction coefficient of 0.983 mg/ml (Gill and von Hippel *(21)*)). A solution of avidin in 0.1M NaH2PO4 pH 7.0 was mixed with a solution of GM-CSF in the same buffer at a molar ratio of 1:2 obtaining final concentrations of 1 mg/ml for avidin and 2 mg/ml for GM-CSF. After addition of TGase at an E/S ratio of 1/20, by weight, the reaction mixture was incubated at 37 °C. Aliquots of 30 µg (total protein content) were removed from the reaction mixture after 0, 1 and 5 hours and diluted with an equal volume of an aqueous solution of 1% TFA. 5 µg of each aliquots were analyzed by SDS-PAGE using a 15% gel prepared according to Laemmli *(22)*.

Coomassie-stained protein bands were excised and in-gel digested, as previously described *(23)*. Protein digests were resuspended in 0.1% formic acid and analysed by LC-MS/MS. LC-MS/MS analyses were performed on a Micromass CapLC unit (Waters) interfaced to a Micromass Q-Tof Micro mass spectrometer (Waters) equipped with a nanospray source. Tryptic digests were loaded at a flow rate of 20 µl/min onto an Atlantis dC18 Trap Column. After valve switching, the sample was separated on an Atlantis C_{18} column (150 x 0.075 mm, 3.5 µm particle size) (Waters) at a flow rate of 4.8 µl/min using a gradient from 5% B to 55% B in 37 min and from 55% to 85% B% in 5 min (solvent A: 95% H_2O, 5% acetonitrile, 0.1% formic acid; solvent B: 5% H_2O, 95% acetonitrile, 0.1% formic acid). Instrument control, data acquisition and processing were achieved with MassLynx V4.1 software (Waters). MS/MS data were analysed by the online MASCOT software (Matrix Science, http://www.matrixscience.com) against the entire Swiss-Prot database (release 2011_10). The following parameters were used in the MASCOT search: trypsin specificity; maximum number of missed cleavages: 3; fixed modification: carbamidomethyl (Cys); variable modifications: oxidation (Met); peptide mass tolerance: ± 0.5 Da; fragment mass tolerance: ± 0.5 Da; protein mass: unrestricted; mass values: monoisotopic.

Analysis of the aggregation state of avidin modified with ZQG. The quaternary status of avidin and avidin conjugated to ZQG was determined by fast-flow liquid chromatography (FPLC) performed on a Superdex^{TM} 75 HR10/30 (Pharmacia) a using an AKTA FPLC instrument (Amersham Biosciences, Sweden). Samples (55 µg) of avidin and avidin modified with ZQG with or without biotin were loaded on the gel filtration column and chromatography was performed at a flow rate of 0.5 ml/min in 0.65 M NaCl. The column was calibrated using bovine serum albumin, ovalbumin, carbonic anhydrase and α-lactalbumin as molecular weigth markers. In order to analyze avidin bound to biotin, samples of avidin and

the avidin-ZQG reaction mixture after 5 hours of incubation were mixed with an equal volume of 0.22 mg/ml of biotin solution in 0.1M NaH_2PO_4 pH 7.0, followed by incubation at RT for 20 min under shaking.

Heat stability experiments were performed according to Bayer et al. *(24).* Briefly, samples of 5 μg of avidin and of the reaction mixture containing avidin modified with ZQG with and without biotin were lyophilized and then dissolved in Laemmli sample buffer. The protein solutions were incubated a 100°C for 20 min before being analyzed by SDS-PAGE *(22).*

Biotin-binding assays. The HABA colorimetric assay for avidin was performed according to Green et al. *(25).*

RESULTS

2.1. Characterization by ESI-MS of avidin from egg white. Avidin purify from egg yolk is glycosilated and analysis of the carbohydrate indicated an extensive glycan microheterogeneity *(26).* We analysed by ESI-MS the sample of avidin used in the TGase derivatization reactions (Fig. 4). The mass spectrum confirmed the heterogeneity of the avidin sample since several different molecular masses could be measured. Neither the less, we took as reference the three main isoforms of avidin (isoforms A-C), whose mass values can be assigned to a specific glycan composition in mannose and N-acetylglucosamine (Table 1) *(26).* No m/z signals of deglycosilated avidin were detected.

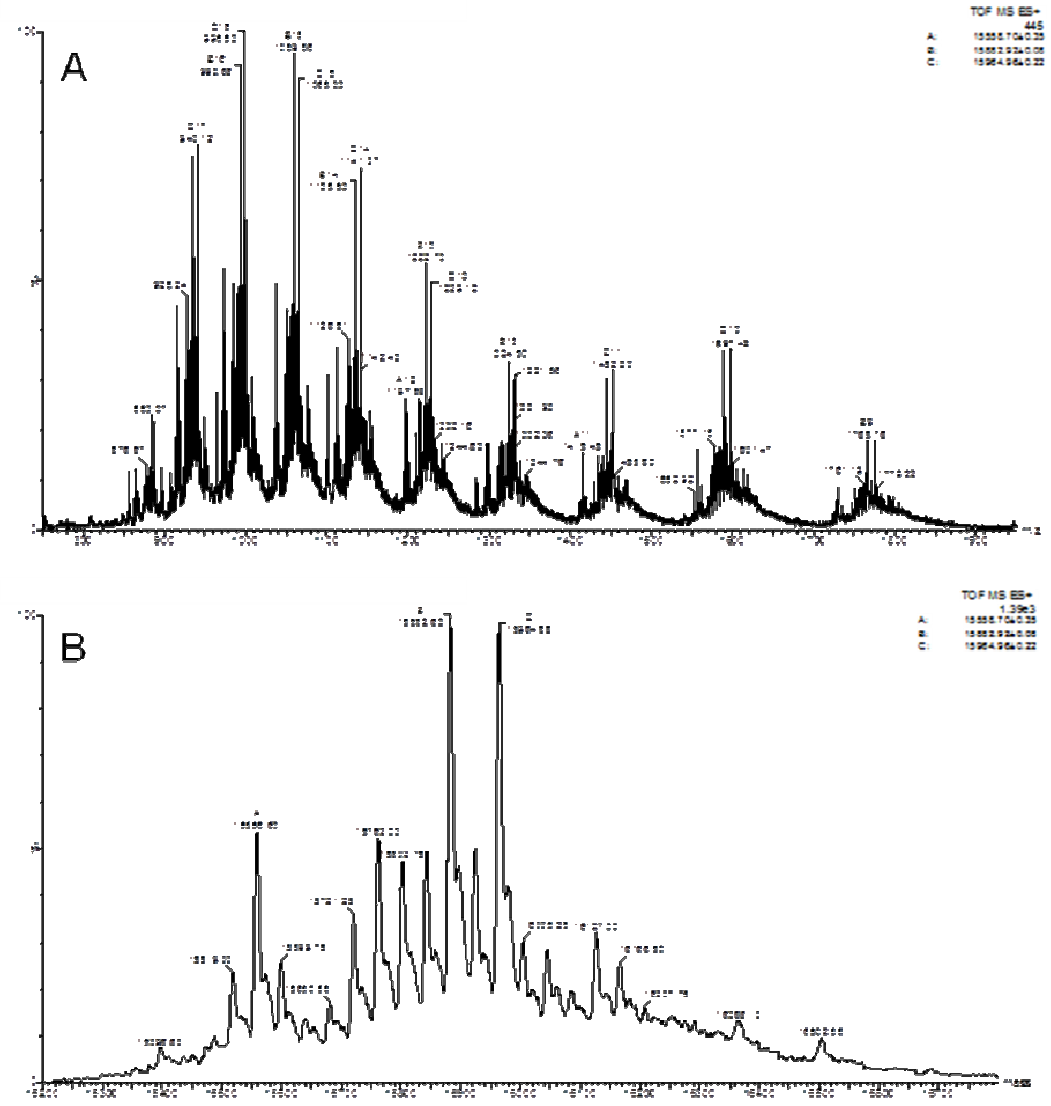


Figure 4. ESI mass spectrum (A) and deconvoluted spectrum (B) of the sample of chicken avidin used in the present study. The main components of the mass spectrum are indicated with the letters A-C (see Table 1).

Table 1. Molecular masses of the main avidin glycosilated isoforms.

Avidin species	Carbohydrate composition	molecular mass (Da)	
		found[a]	calculated[b]
Avidin	—	—	14341.19
A[c]	$Man_5GlcNAc_2$[d]	15558.70	15558.29
B	$Man_7GlcNAc_2$	15882.92	15882.58
C	$Man_5GlcNAc_4$	15964.96	15964.68

[a] Experimental molecular masses determined by ESI-MS. [b] Calculated molecular masses. [c] The different glycosilated isoforms are labeled according to the Components Table of the ESI-MS analysis reported in Figure 4. [d] Carbohydrate composition of the N-glycosilated avidin isoforms. Man is mannose while GlcNAc stays for N-acetylglucosamine.

2.2. Derivatization of avidin at the level of Gln residues using dansyl cadaverine. The presence of Gln residues reactive towards TGase was analysed using dansyl cadaverine (DC) as model substrate since it contains a primary amine that is a good amino donor (Fig. 5A). The time course of the reaction was followed by RP-HPLC analysis of the reaction mixture at different time of incubation followed by ESI-MS analysis of the protein material eluted in the chromatographic peaks. Avidin was incubated with TGase (E/S 1/25 by weight) in the presence of 50 molar excess of DC and aliquots of the reaction mixture were collected after 0, 6 and 24 hours. The RP-HPLC profile of native avidin in the presence of DC (Fig. 5B, dashed line) shows the presence of three main protein peaks that by ESI-MS showed identical mass spectra corresponding to those of native avidin (Fig. 4). We suggest that the three peaks originate from different association state of the protein (e.g avidin is separated in the RP-HPLC analysis as tetramer, dimer and monomer). After 24 hours of incubation, the reaction mixture shows a RP-HPLC chromatographic profile that is identical to the time zero of the reaction (Fig. 5B, straight line). ESI-MS analysis of the eluted protein material confirmed that the conjugation of DC to avidin did not occur.

2.3. Derivatization of avidin at the level of Lys residues using carbobenzoxy-L-glutamyl-glycine. In order to identify Lys residues in the avidin sequence that can be modified by TGase, the protein was incubated with TGase (E/S 1/5, by weight) and carbobenzoxy-L-glutamyl-glycine (ZQG) at a 50 molar excess in respect to avidin. ZQG is a Gln containing substrate that is efficiently conjugated to Lys residues by the TGase catalysed reaction (Fig.

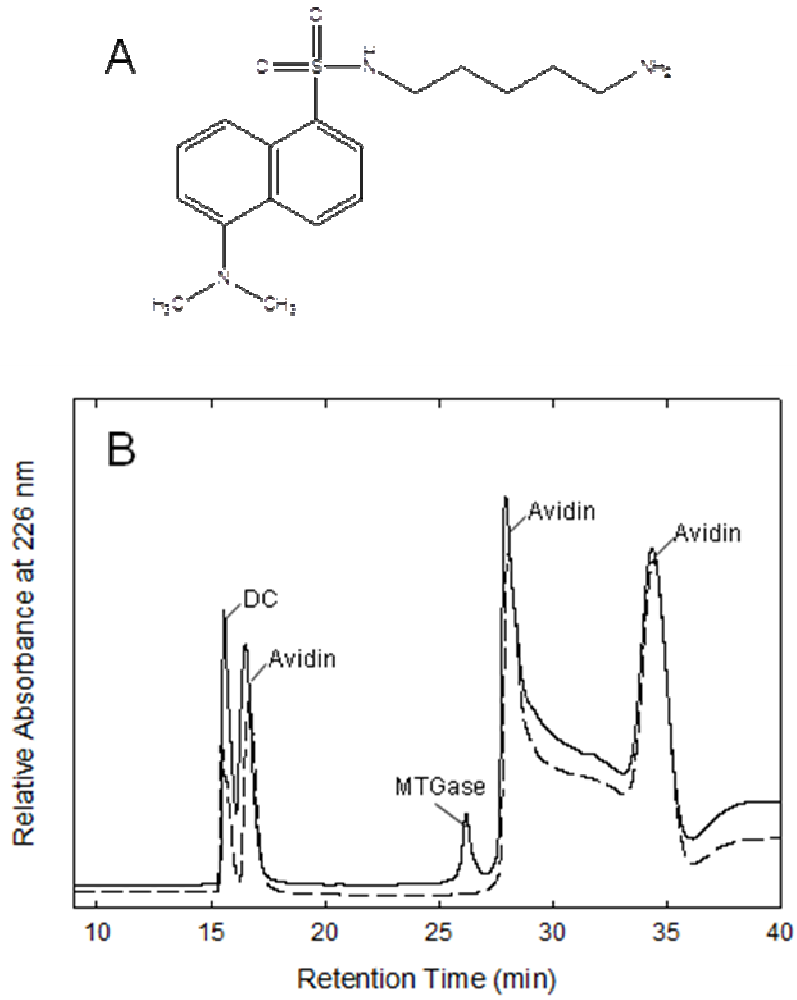


Figure 5. TGase-mediated conjugation at Gln residues of avidin with DC. (A) Chemical structure of DC. (B) RP-HPLC analyses of the reaction mixtures of avidin in the presence of TGase and DC. A dash-dotted line indicates the chromatogram after zero hours of reaction, while the straight line the chromatographic profile after 24 hours of incubation.

6A). The reaction mixture was analysed by RP-HPLC after different time of incubation (Fig. 6B). After 5 hours of reaction, the chromatographic profile of avidin is changed in respect to that of the native protein (Fig. 6B). Analysis of the chromatographic peaks by ESI- MS indicated that avidin present in the reaction mixture was quantitatively conjugated to ZQG and that up to two molecules of this Gln-containing substrate were bound to the protein. Indeed, the mass spectrum of avidin eluted from the RP-HPLC analysis showed for the isoforms A-C of the protein a mass increment corresponding to the addition of 1 and 2 molecules of ZQG (Figure 7, Table 2). On the basis of the MS analysis, the three chromatographic peaks of avidin conjugated to ZQG show identical composition in mono- and bi-derivatives.

In order to identify the Lys residues that are modified by TGase, native avidin and avidin^{1,2ZQG} purified by the RP-HPLC analysis of the reaction mixture were carbamidomethylated and then digested with trypsin. ESI-MS analysis of the two tryptic digests and comparison of their mass spectra allowed to identifiy two peptides that are conjugated to ZQG in the avidin^{1,2ZQG} sample. Indeed, the m/z signals of peptides 46–59 and 125–128 showed a mass increment corresponding to the modification with ZQG and indeed they both contain one Lys residue, Lys58 and Lys127 respectively (Table 3). MS/MS spectra were acquired on the m/z signals of the modified peptides 46–59 and 125–128 and they allowed to confirm the sites of modification (Fig. 8 and 9, respectively).

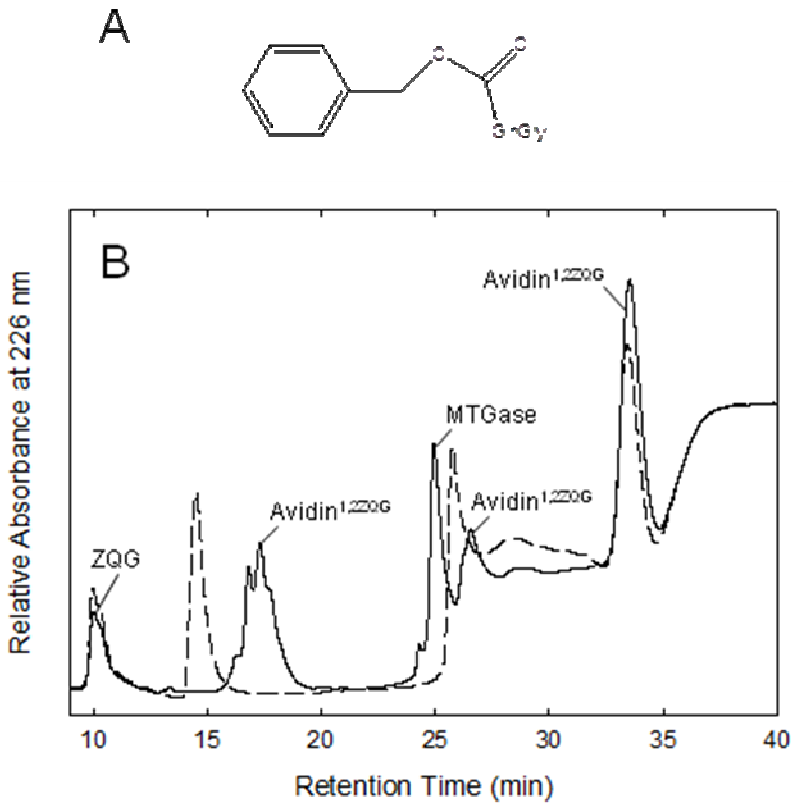


Figure 6. TGase-mediated conjugation at Lys residues of avidin with ZQG. (A) Chemical structure of DC. (B) RP-HPLC analyses of the TGase-conjugated derivatives of avidin. A dash-dotted line indicates the chromatograms of native avidin, the straight line the ZQG-conjugated species after 5 hours of reaction

Table 2. Molecular masses of the main avidin glycosilated isoforms conjugated to ZQG.

Avidin species	Numbers of ZQG	molecular mass (Da)	
		found[a]	calculated[b]
A[c]	1ZQG	15878.85	15878.59
	2ZQG	–[d]	16198.89
B	1ZQG	16201.47	16202.88
	2ZQG	16524.40	16523.18
C	1ZQG	16285.28	16284.98
	2ZQG	16606.18	16605.28

[a] Experimental molecular masses determined by ESI-MS. [b] Calculated molecular masses. [c] The different glycosilated isoforms are labeled according to Table 1 [d] The charge state series of Isoform A^{2ZQG} is expected to have m/z values close to those of Isoform B^{1ZQG} and the two m/z series can not be resolved in the mass spectrum.

2.4. Effect of the modification of Lys58 and Lys127 in avidin^{1,2ZQG} on the biotin-binding properties and on the aggregation state of the protein. The use in biotechnological applications of the TGase-mediated conjugation of avidin at the level of Lys58 and Lys127 residues requires that modification of these amino acid residues does not alter the biotin-binding properties of the protein. We used the HABA assay to determine the number of functional binding sites in avidin^{1,2ZQG} *(25)*. This dye-binding method exploits the spectral changes occurring upon binding of the dye 4-hydroxyazobenezene-2'-carboxylic acid (HABA) to avidin. Indeed, the binding of HABA to avidin is accompanied by a red shift of the maximum of absorbance of the dye from 350nm to 500nm. If biotin is then added to the avidin-HABA complex, the dye is stoichiometrically displaced by biotin and its UV-Visible spectrum returns to be that of unbound HABA. Measurement of the variation in the absorbance of HABA at 500 nm after biotin addition allows to estimate the number of biotin binding sites present in the avidin sample and as a consequence the concentration of the protein. We determined the concentration (mg/ml) of the protein in a sample of native avidin and in one containing avidin^{1,2ZQG} by measuring the absorbance of the protein at 280 nm and by using the HABA colorimetric essay and then we compared the two values obtained for each of the two samples (Table 4). Both for native avidin and for the sample containing avidin^{1,2ZQG} there was a consistency between the values measured with the two different methods. This result is an indication that in avidin^{1,2ZQG} the number of functional binding sites is maintained.

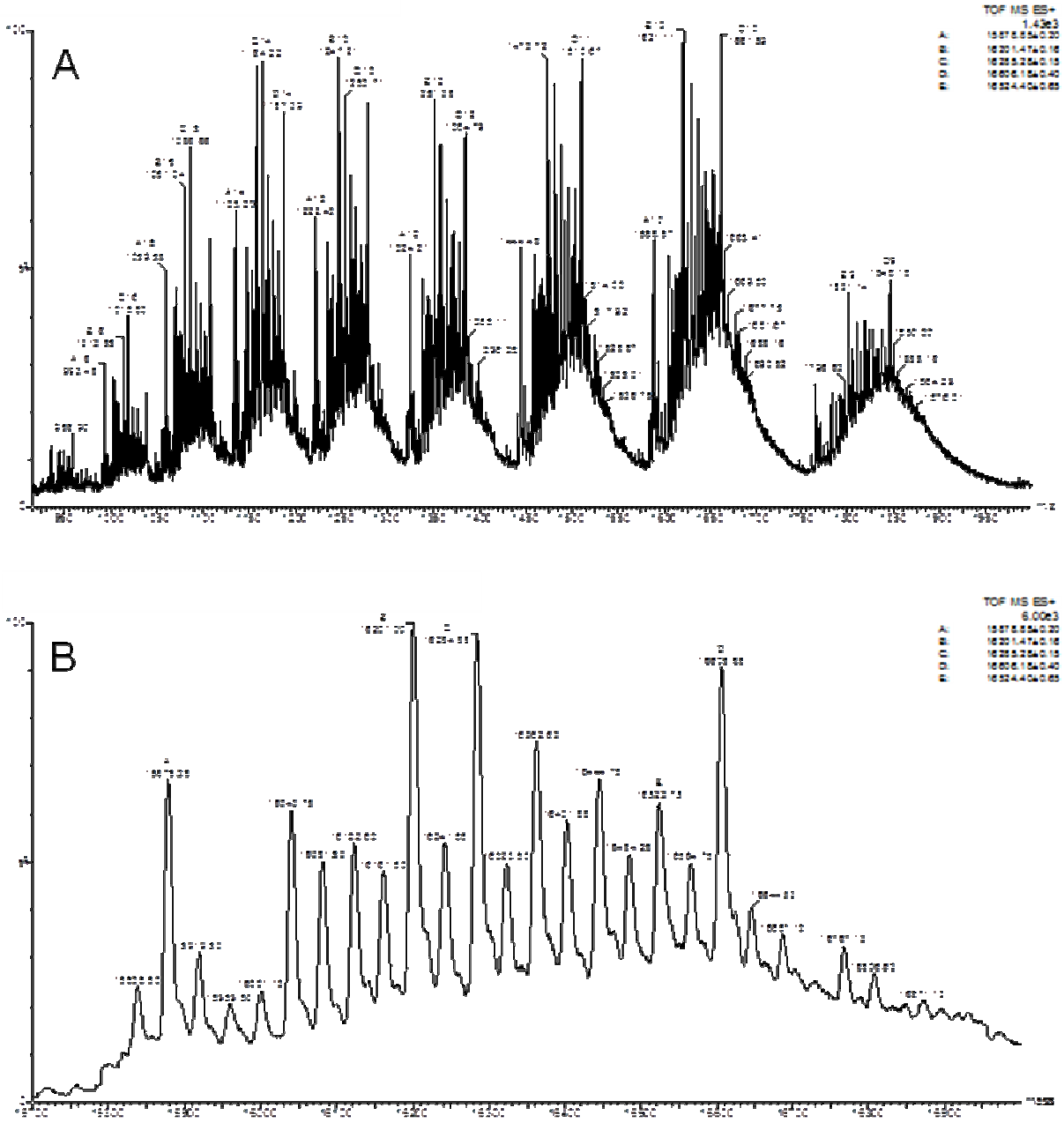


Figure 7. ESI mass spectrum (A) and deconvoluted spectrum (B) of avidin conjugated to ZQG. The ZQG-derivatives of avidin were purified by RP-HPLC (Fig. 6) and the mass spectrum here reported corresponds to the protein material eluted at a retention time of 34 minutes. The main components of the mass spectrum are indicated with the letters A-E (see Table 2), where D and E indicate the found molecular masses of isoforms C and B conjugated to two ZQG, respectively.

Modification of Lys58 and Lys 127 of avidin could also affect the quaternary structure of the protein by destabilizing the homotetramer. The aggregation state of avidin^{1,2ZQG} was analyzed by gel filtration chromatography both in the presence and in the absence of free biotin and compared with a sample of native avidin (Fig. 10A). As it can be observed, avidin

and avidin[1,2ZQG] have similar elution volumes both with and without addition of biotin. Moreover, their migration is consistent with the theoretical molecular mass of the homotetramer of avidin (64 kDa) since the observed masses range from 67.3 kDa (avidin in the absence of biotin) to 60.7 kDa (avidin[1,2ZQG] in the presence of biotin).

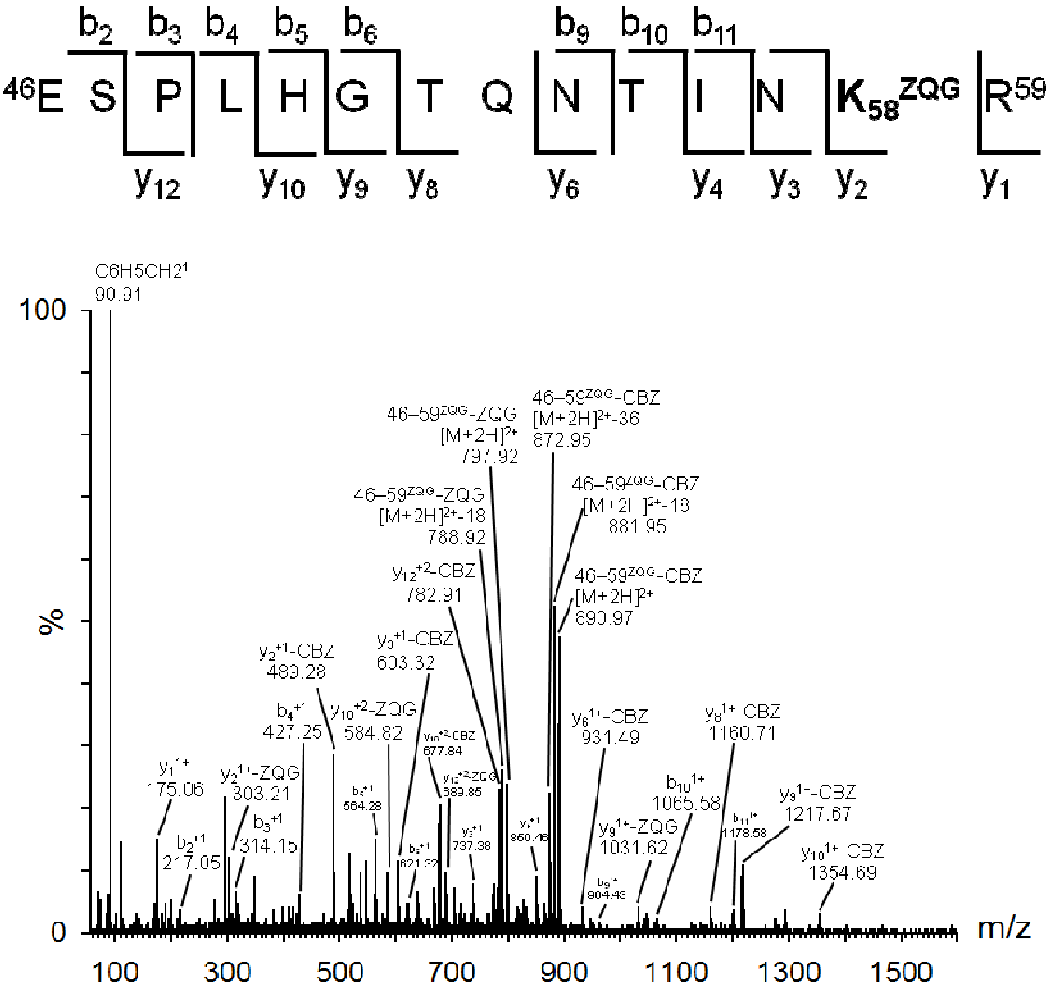


Figure 8. Electrospray MS/MS mass spectrum of the triply charged ion at 638.98 m/z of peptide 46–59 of avidin modified with ZQG. (Top) Fragments of the series b and y that were identified in the MS/MS spectrum are indicated on the sequence of the peptide. Lys58 is indicated in bold as K_{58}^{ZQG}. (Bottom) In the MS/MS spectrum, ions assigned to fragments of the series y and b are indicated. Fragments containing Lys58 show a mass increment corresponding to the conjugation with ZQG without the CBZ moiety since the N–protecting group dissociates in the MS/MS analysis (-CBZ). For some fragments, the products of the dissociation of ZQG were also detected and indicated as -ZQG.

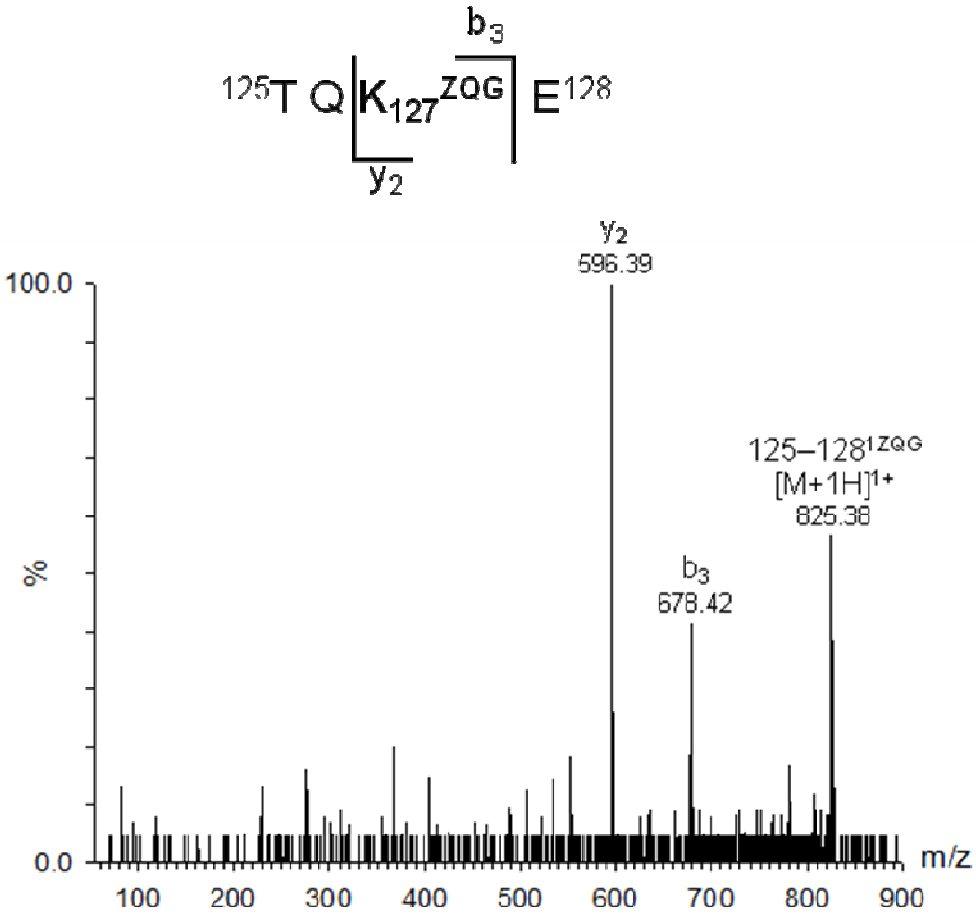


Figure 9. Electrospray MS/MS mass spectrum of the singly charged ion at 825.40 m/z of peptide 125–128 of avidin modified with ZQG. (Top) Fragments of the series b and y that were identified in the MS/MS spectrum are indicated on the sequence of the peptide. Lys127 is indicated in bold as K_{127}^{ZQG}. (Bottom) In the MS/MS spectrum, ions assigned to fragments of the series y and b are indicated. The two fragments indentified in the MS/MS spectrum (y_2 and b_3) contain Lys127 and show a mass increment corresponding to the conjugation with ZQG.

Table 3. Molecular masses measured for the tryptic peptides of avidin conjugated to ZQG.

Tryptic peptide	Peptide sequence	Number of ZQG	molecular mass (Da)	
			found[a]	calculated[b]
46–59	^{46}ESPLHGTQNTIN**K$_{58}$**R^{59}	–	1593.95[c]	1593.82
		1ZQG	1913.91[d]	1913.92
125–128	^{125}TQ**K$_{127}$**E^{128}	–	–[c, e]	504.25
		1ZQG	824.40[d]	824.36

[a] Experimental molecular masses determined by ESI-MS. [b] Calculated molecular masses. [c] Molecular mass measured in the tryptic digest of avidin. [d] Molecular mass measured in the tryptic digest of avidin^{1,2ZQG}. [e] This very short and rather hydrophilic peptide was not detected since likely it was lost in the desalting step of the sample digest (see Materials and Methods).

Table 4. Concentration of avidin measured using the colorimetric HABA assay on samples of native avidin and avidin^{1,2ZQG}.

Sample	Concentration (mg/ml) based on A_{280nm}	Concentration (mg/ml) based on HABA assay
Avidin	0.23	0.22
Avidin^{1,2ZQG}	0.27	0.28

The thermal stability of the tetrameric form of avidin^{1,2ZQG} was also analyzed by SDS-PAGE *(24)* (Fig. 10B). Avidin is a positively charged glycoprotein (p*I* ~ 10.5) and it aggregates extensively when mixed with anionic detergents as sodium dodecyl sulfate (SDS) leading to aggregated tetramers that remain in the stacking gel of SDS-PAGE analyses. However, it has been demonstrated that when avidin is incubated in the presence of SDS at 100°C, it migrates as a monomer (Fig. 10B, lane 2) while avidin bound to biotin forms aggregated tetramers in the presence of SDS that fail to enter the separating gel even after boiling (Fig. 10B, lane 3, *(27)*). Similarly to native avidin, avidin^{1,2ZQG} migrates as a monomer when it is incubated at 100°C in the presence of SDS (Fig. 10B, lane 4). However, avidin^{1,2ZQG} bound to biotin forms a tetramer that do not aggregate after incubation at 100°C in the presence of SDS (Fig. 10B, lane 5). This observation can be explained in light of the fact that derivatization of avidin with TGase modifies two Lys residues of the protein that are not anymore available for SDS binding. Indeed, it has been demonstrated that acetylation of avidin inhibits the aggregation of the biotin-bound tetramer when it is boiled in the presence of SDS *(24)*. We can conclude that avidin^{1,2ZQG} forms tetramers that display stability characteristics similar to those of native avidin.

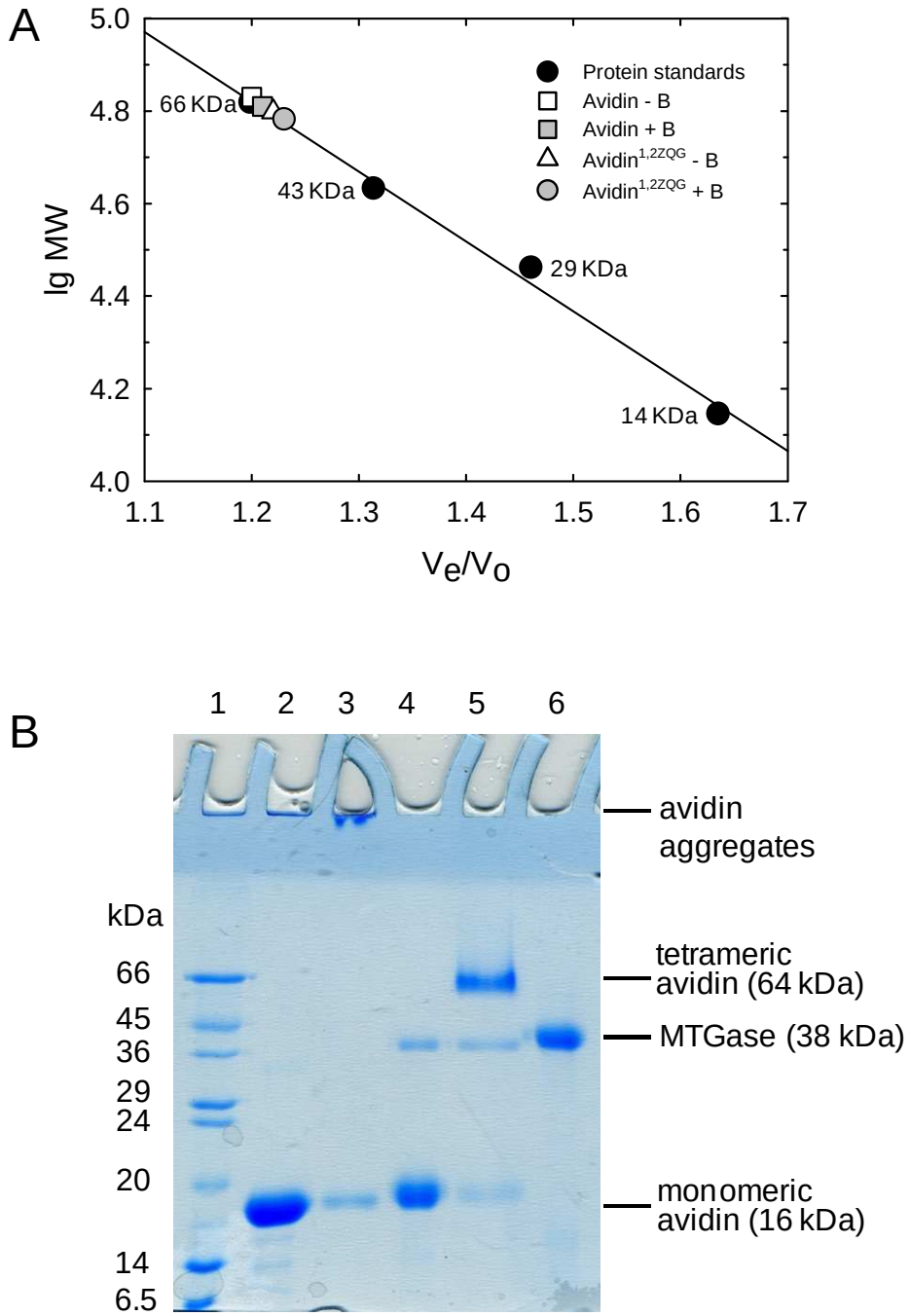


Figure 10. Analysis of the aggregation state (A) of avidin and avidin^{1,2ZQG} and of its thermostability (B) in the presence or in the absence of biotin. (A) Gel filtration profiles of avidin and avidin^{1,2ZQG} in the absence (-B) and in the presence of biotin (+B). A Superdex 75 column was used for the analysis. Serum albumin (66 kDa), ovalbumin (43 kDa), carbonic anhydrase (29 kDa) and lactalbumin (14 kDa) were used as molecular weight standards to calibrate the column. The logarithm of molecular weight (MW) is plotted *versus* V_e/V_o (V_e is the elution volume and V_o the void volume of the column). (B) Native avidin and avidin^{1,2ZQG} (the reaction mixture of avidin with ZQG and TGase after 5 hours of incubation) were preincubated for 20 min at 100°C in presence of SDS without and with biotin (-B and +B, respectively) and analysed by SDS-PAGE. Protein samples were loaded in the following order: MW standards (lane 1); avidin –B (lane 2); avidin +B (lane 3); avidin^{1,2ZQG} –B (lane 4); avidin^{1,2ZQG} +B (lane 5); TGase (lane 6).

2.5. Production of an avidin–GM-CSF conjugate using microbial transglutaminase. Several biotechnological applications require conjugation of avidin to another protein by recombinant techniques or chemical methods. TGase provides an alternative method to produce avidin-protein conjugates since it can catalyse the crosslinking reaction between reactive Lys residues of avidin and Gln residues of another protein. With the aim to test this application, we studied the possibility to conjugate *via* TGase avidin to granulocyte-macrophage colony-stimulating factor (GM-CSF). GM-CSF is cytokine that functions as a hematopoietic growth factor and is used in therapy after bone marrow transplantation and in some forms of leukemia (Fleetwood et al., 2005).

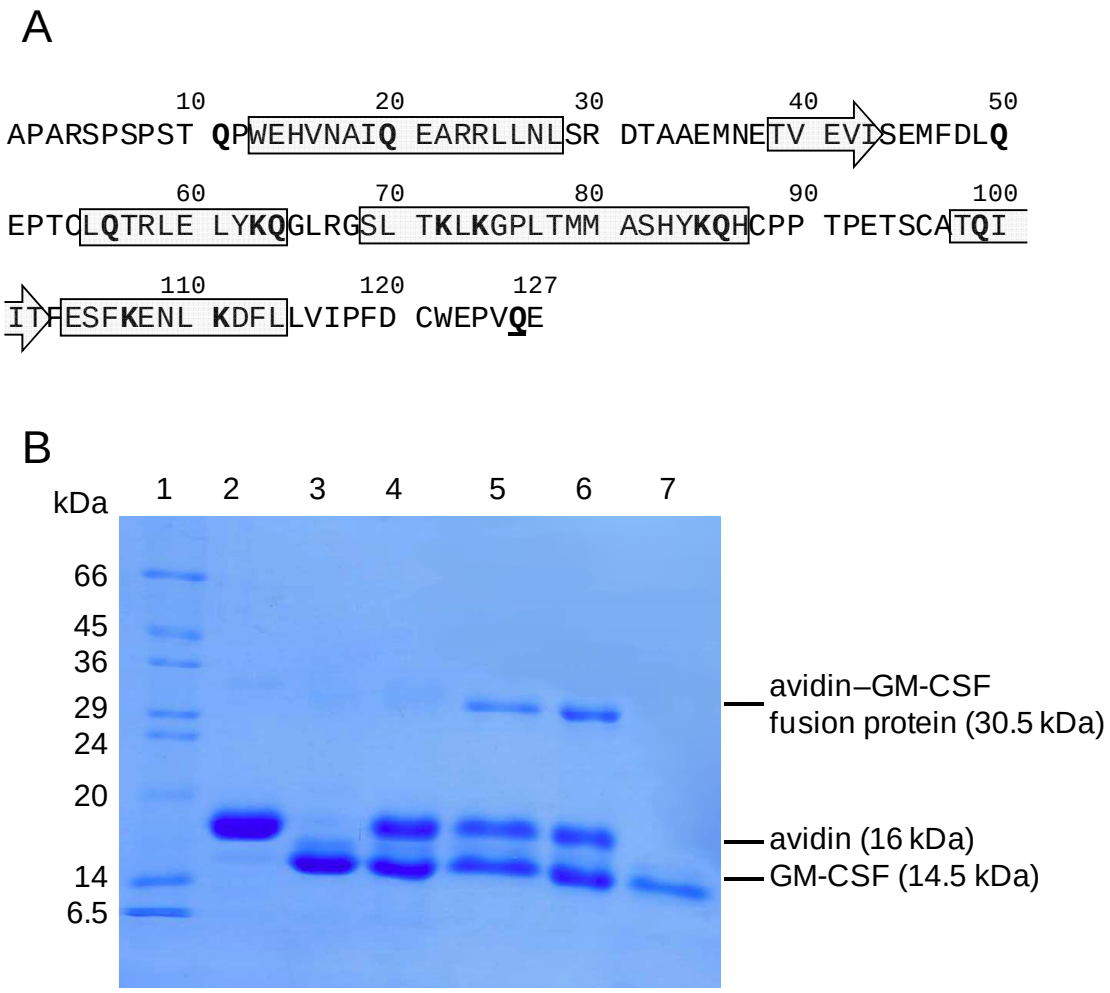


Figure 11. TGase-mediated production of an avidin–GM-CSF fusion protein. (A) Aminoacid sequence of GM-CSF. The four α-helices along the 127-residue chain of the protein are indicated by grey boxes and the two β-strands by grey arrows. The Gln and Lys residues potential sites of TGase conjugation are shown in bold, while Gln126 is also underlined. (B) SDS-PAGE analysis of the conjugation reaction between avidin and GM-CSF catalysed by TGase. Protein samples were loaded in the following order: MW standards (lane 1); avidin (lane 2); GM-CSF (lane 3); Reaction mixture containing avidin, GM-CSF and TGase after 0, 1 and 5 hours of incubation (lane 4, 5 and 6, respectively); GM-CSF after 5 hours of reaction with TGase (lane 7).

Experiments conducted in our laboratory demonstrated that GM-CSF has only one Gln residue that is reactive towards TGase and we identified it as Gln126 (unpublished results, Fig. 11A). Moreover, GM-CSF has no reactive Lys residues and thus in the presence of TGase it can not form dimers that could complicate the reaction pattern.

Analysis by SDS-PAGE of the reaction between avidin and GM-CSF catalyzed by TGase after one and five hours of incubation (Fig. 11B, lanes 5 and 6 respectively) indicated the formation of only one product that migrates at a molecular weight consistent with the formation of an avidin–GM-CSF conjugate. In gel digestion of this protein band followed by LC-MS/MS analysis of the peptide mixture and MASCOT search of the MS data allowed to identify the presence both of avidin and GM-CSF (Table 5), thus confirming the formation of the conjugation product. In order to determine the sites of crosslinking, LC-MS/MS data were manually inspected and m/z signals were found corresponding to the formation of isopeptide bonds between peptides 112–127 of GM-CSF and 46–59 of avidin and 112–127 of GM-CSF and 125–128 of avidin (Table 6). Analysis of the MS/MS fragmentation patterns of these two peptides (Figures 12 and 13) allowed to confirm that crosslinking occurs between residues Gln126 of GM-CSF and either Lys58 or Lys127 of avidin, in agreement with previous data on the reactivity of these proteins to TGase. Interestingly, these results also suggest that Lys58 and Lys127 are both primary sites of TGase derivatization of avidin.

Table 5. List of the proteins identified by LC-MS/MS analysis of the triptic digest obtained from the band at 30kDa in lane 6 of the SDS-PAGE analysis reported in Fig. 11.[a]

Accession No.	Protein name	Sequence Coverage (%)	Sequences of the tryptic peptides
P02701	Avidin (Gallus gallus)	69.5	GEFTGTYITAVTATSNEIK ESPLHGTQNTINKR TQPTFGFTVNWK FSESTTVFTGQCFIDRNGK TMWLLR SSVNDIGDDWK VGINIFTR
P04141	Granulocyte-macrophage colony-stimulating factor (GM-CSF) (Homo sapiens)	51.2	SPSPSTQPWEHVNAIQEAR RLLNLSR DTAAEMNETVEVISEMFDLQEPTCLQTR GPLTMMASHYK

[a] For the electrophoretic band at 30kDa in lane 6 of the SDS-PAGE analysis reported in Fig. 11, the identified proteins with their accession numbers are listed. Protein identification was performed with the MASCOT software searching LC-MS/MS data against the sequences of the SwissProt database (release 2011_10). The fourth column contains a list of the peptides that were sequenced by MS/MS.

Table 6. Molecular masses measured for peptide 112–127 of GM-CSF conjugated to peptides 46–59 or 125–128 of avidin (AVI).

Species	molecular mass (Da)	
	found[a]	calculated[b]
$112–127^{GM\text{-}CSF}/46–59^{AVI}$	3582.99	3582.74
$112–127^{GM\text{-}CSF}/125–128^{AVI}$	2493.33	2493.17

[a] Experimental molecular masses determined by LC-MS/MS analysis of the tryptic digest of the avidin/GM-CSF fusion protein. [b] Calculated molecular masses.

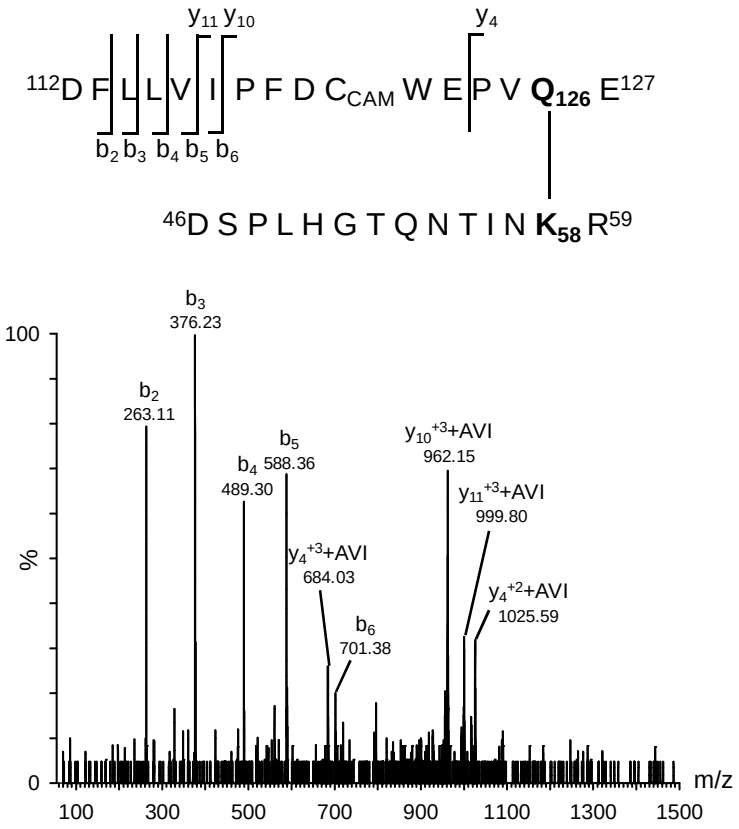


Figure 12. Electrospray MS/MS mass spectrum of the fourthly charged ion at 896.69 m/z of the avidin tryptic peptide 46–59 conjugated to the tryptic peptide 112–127 of GM-CSF. (Top) Sequence of the two peptides. A line connecting Lys58 of avidin and Gln126 of GM-CSF represents the TGase-catalysed formation of an isopeptide bond between two residues. Fragments of the series b and y of peptide 112–127 of GM-CSF that were identified in the MS/MS spectrum are indicated on its sequence. (Bottom) In the MS/MS spectrum, ions assigned to fragments of the series b and y of peptide 112–127 of GM-CSF are indicated. Fragment ions of the y series show a mass increment corresponding to the conjugation to peptide 46–59 of avidin (indicated as +AVI).

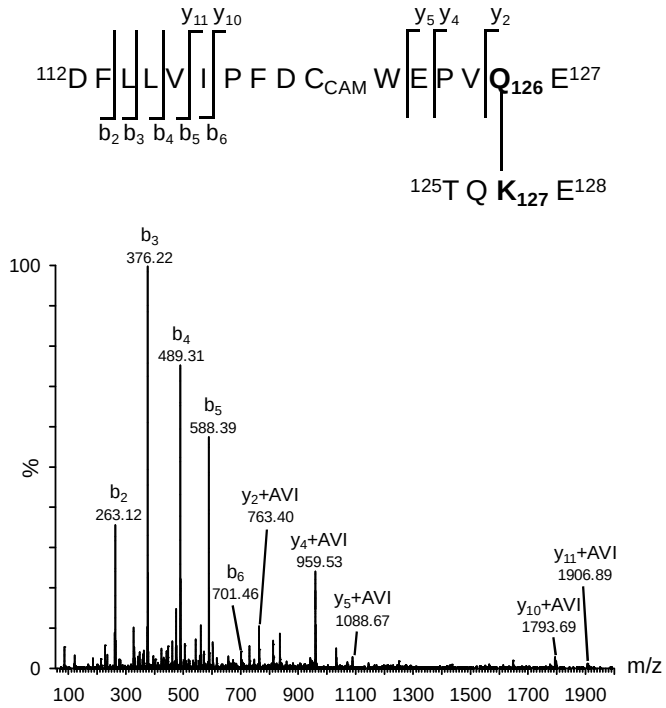


Figure 13. Electrospray MS/MS mass spectrum of the doubly charged ion at 1247.59 m/z of the avidin tryptic peptide 125–128 conjugated to the tryptic peptide 112–127 of GM-CSF. (Top) Sequence of the two peptides. A line connecting Lys127 of avidin and Gln126 of GM-CSF represents the TGase-catalysed formation of an isopeptide bond between two residues. The carbamidomethylation of the cysteine residue of peptide 112–127 of GM-CSF is indicated as Cys_{CAM}. Fragments of the series b and y of peptide 112–127 of GM-CSF that were identified in the MS/MS spectrum are indicated on its sequence. (Bottom) In the MS/MS spectrum, ions assigned to fragments of the series b and y of peptide 112–127 of GM-CSF are indicated. Fragment ions of the y series show a mass increment corresponding to the conjugation to the peptide 125–128 of avidin (indicated as +AVI).

DISCUSSION

Structural determinants of the TGase site specific modification of avidin. In this study, we demonstrated that avidin purified from egg yolk can be specifically modified at the level of only two Lys residues, namely Lys58 and Lys127 while no Gln residue was found to be reactive towards TGase. We can discuss the selectivity of the TGase mediate reaction based on our knowledge of the determinants of TGase selectivity. We have proposed that TGase selectivity at the level of both Gln and Lys residues (see Chapter 2) is determined mainly by the flexibility of the polypeptide chain since a selective modification is observed when Gln and Lys residues are embedded in flexible-unfolded chain segments of the protein. Avidin is a tightly structured protein since even in the apo-form its Tm is of 85°C *(13)*. Limited proteolysis experiments also demonstrated its structural stability. Indeed, the apo-form of the protein is hydrolyzed by proteinase K only upon incubation at an E/S of 1/1 which is

extremely high for limited proteolysis experiments and considering the high activity and broad specificity of this protease *(12).*

Several 3D structures of avidin in the apo-and holo- form have been published with similar results *(6, 8, 9)*. They indeed indicate that the most flexible regions in the apo-form of the protein are located at level of the loop connecting strands 3 and 4 (residues Ala36-Ile44), at the N-terminus (Ala1–Arg2) and C-terminus (Thr125–Glu128) of the protein and in the Asp86–Arg88 segment (8). Avidin sequence contains nine Lys residues and four Gln residues (Fig.1A). Among these residues, only Lys127 and Gln126 are located at the disordered C-terminal region of the protein and indeed Lys127 is modified by TGase. Gln126 is not substrate of the enzyme and we can explain this observation considering the sequence flanking this residue (–$Thr_{125}Gln_{126}Lys_{127}$–). Indeed, its has been reported that Gln residues that have at the C-terminus positively charged Lys residues are not reactive to tissue TGase *(28-29)* and this is the case also of Gln126 in the avidin sequence. However, further experiments on model peptides are needed to confirm if this sequence rule on the reactivity of Gln residues is valid also for microbial TGase. Among the other Lys and Gln residues of avidin, no one is present in one of the disordered regions of the protein as predicted from the X-ray structures. However, Lys58 which is modified by TGase is located in the middle of a loop region (residues 54–62) connecting strands 4 and 5, that is expected to display some level of flexibility.

TGase derivatization does not affect the biotin-binding properties and quaternary structure of avidin. Avidin modified with ZQG at the level of both Lys58 and Lys126 still displays biotin binding properties and maintains its tetrameric quaternary structure. This is expected since residues that participate to the binding of biotin are located on the opposite end of the avidin barrel in respect to the two reactive Lys residues (Fig. 1B) *(6).* Moreover, the interactions that stabilize avidin quaternary structure are not affected by the modification since Lys58 and Lys126 are not among the residues involved in tetramer formation and outside the surfaces of interaction between monomers. This is demonstrated also by the fact that a circular permutated form of avidin where the new N- and C-termini are located at Arg59 and Lys58, respectively maintains the biotin-binding properties of the native protein and forms a quaternary structure resembling that of wt avidin *(30).*

In conclusion, we demonstrated that avidin can be quantitatively modified by TGase at the level of two Lys residues without affecting its biotin-binding property and its quaternary structure. This reaction can be used to produce conjugation products of avidin with other proteins, as demonstrated for GM-CSF or with small molecules. Our study is the first report

on the possibility to produce avidin conjugates by using microbial TGase and we believe it will open to several applications of this reaction in the avidin-biotin technology.

References

1. Eakin, R.E., Snell, E.E. & Williams, R.J. (1941) The concentration and assay of avidin, the injury-producing protein in raw egg-white. *J.Biol.Chem.* 140: 535-543.
2. Green, N.M. (1975) Avidin. *Adv. Prot. Chem.* 295: 85-133.
3. Green, N.M. (1990) Avidin and streptavidin. *Methods Enzymol*. 184: 51-67.
4. Green, N. M. and Joynson M. A. (1970) A preliminary crystallographic investigation of avidin. *Biochem. J.* 118: 71–72.
5. Wilchek, M., Bayer, E.A. (1990) Introduction to avidin-biotin technology. *Methods Enzymol*. 184:5-13.
6. Livnah, O., Bayer, E.A., Wilchek, M. & Sussman, J.L. (1993) Three-dimensional structures of avidin and the avidin-biotin complex. Proc Natl Acad Sci U S A. 90297: 5076-5080.
7. Pugliese, L., Malcovati, M., Coda, A., Bolognesi, M. (1994) Crystal structure of apo-avidin from hen egg-white. *J Mol Biol.* 235(1):42-6.
8. Nardone, E., Rosano, C., Santambrogio, P., Curnis, F., Corti, A., Magni, F., Siccardi, A.G., Paganelli, G., Losso, R., Apreda, B., Bolognesi, M., Sidoli, A., Arosio, P. (1998) Biochemical characterization and crystal structure of a recombinant hen avidin and its acidic mutant expressed in Escherichia coli. *Eur J Biochem*. 256(2):453-60.
9. Pugliese, L., Coda, A., Malcovati, M. & Bolognesi, M. (1993) Three-dimensional structure of the tetragonal crystal form of egg-white avidin in its functional complex with biotin at2.7 A resolution. Journal of Molecular Biology. 231: 698-710.
10. Wilchek, M., Bayer, E.A., Livnah, O. (2006) Essentials of biorecognition: the (strept)avidin-biotin system as a model for protein-protein and protein-ligand interaction. *Immunol Lett.* 103(1):27-32.
11. Laitinen, O.H., Nordlund, H.R., Hytönen, V.P., Uotila, S.T., Marttila, A.T., Savolainen, J., Airenne, K.J., Livnah, O., Bayer, E.A., Wilchek, M., Kulomaa, M.S. (2003) Rational design of an active avidin monomer. *J Biol Chem*. 278(6):4010-4.
12. Ellison, D., Hinton, J., Hubbard, S.J., Beynon, R.J. (1995) Limited proteolysis of native proteins: the interaction between avidin and proteinase K. *Protein Sci.* 4(7):1337-45.
13. Donovan, J.W. & Ross, K.D. (1973) Increase in the stability of avidin produced by binding of biotin. A differential scanning calorimetric study of denaturation by heat. Biochemistry. 12: 512-517.
14. Wilchek, M., and Bayer, E.A. (1990) Introduction to avidin-biotin technology, *Methods Enzymol. 184*, 5–13.
15. Lesch, H.P., Kaikkonen, M.U., Pikkarainen, J.T., Ylä-Herttuala, S. (2010) Avidin-biotin technology in targeted therapy. *Expert Opin Drug Deliv.* 7(5):551-64.

16. Frost, S., Jensen, H., Lindegren, S. (2010) In Vitro Evaluation of Avidin Antibody Pretargeting Using 211At-Labeled and Biotinylated Poly-L-Lysine as Effector Molecule. Cancer. 1101-1110.

17. Hylarides, M.D., Mallett, R.W., Meyer, D.L. (2001) A robust method for the preparation and purification of antibody/streptavidin conjugates. *Bioconjug Chem.* 12:421-427.

18. Jia, F., Shelton, T.D., Lewis, M.R. (2007) Preparation, characterization, and biological evaluation of a streptavidin-chimeric t84.66 conjugate for antibody pretargeting. *Cancer Biother Radiopharm*. 22:654-664.

19. Schultz, J., Lin, Y., Sanderson, J., et al. (2000) A tetravalent singlechain antibody-streptavidin fusion protein for pretargeted lymphoma therapy. *Cancer Res*. 60:6663-6669.

20. Laitinen, O.H., Airenne, K.J., Raty, J.K., Wirth, T., Yla-Herttuala, S.Y. (2005) Avidin fusion protein strategies in targeted drug and gene delivery. *Letters in Drug Design & Discovery*. 2:124-132.

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

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

23. Wilm, M., Shevchenko, A., Houthaeve, T., Breit, S., Schweigerer, L., Fotsis, T., Mann, M. (1996) Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. *Nature* 379: 466–469.

24. Bayer, E.A., Ehrlich-Rogozinski, S., Wilchek, M. (1996) Sodium dodecyl sulfate-polyacrylamide gel electrophoretic method for assessing the quaternary state and comparative thermostability of avidin and streptavidin. *Electrophoresis.* 17(8):1319-24.

25. Green, N.M. (1970). Spectrophotometric determination of avidin and biotin. Methods in Enzymol., Vol XVIII, Academic Press: New York, p. 418.

26. Bruch, R.C., White, H.B. (1982) 3rd. Compositional and structural heterogeneity of avidin glycopeptides. *Biochemistry.*;21(21):5334-41.

27. Laitinen, O.H., Marttila, A.T., Airenne, K.J., Kulik, T., Livnah, O., Bayer, E.A., Wilchek, M. & Kulomaa, M.S. (2001) Biotin induces tetramerization of a recombinant monomeric avidin. A model for protein-protein interactions. J Biol Chem. 276: 8219-8224.

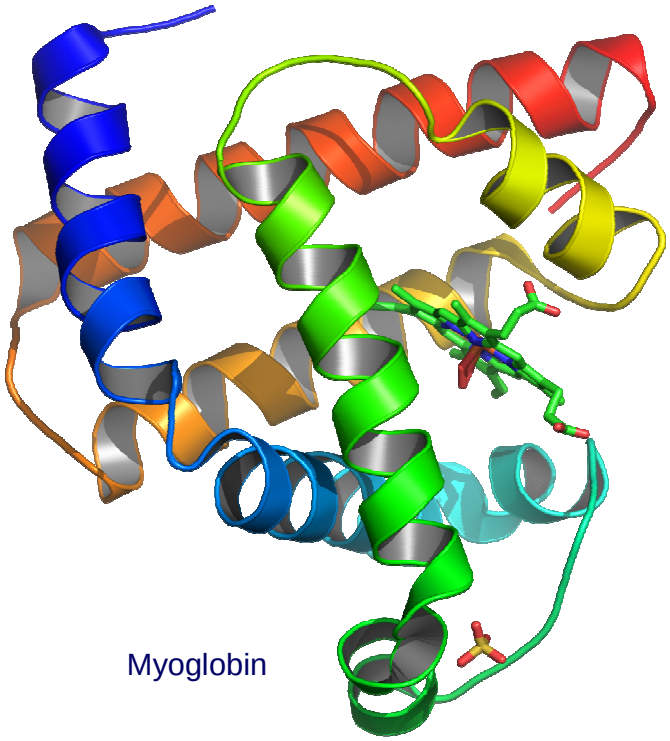
28. Coussons, P.J., Price, N.C., Kelly, S.M., Smith, B., and Sawyer, L. (1992) Factors that govern the specificity of transglutaminase-catalysed modification of proteins and peptides, *Biochem. J*. 282 (Pt 3), 929–930.

29. Pastor, M.T., Diez, A., Pérez-Payá, E., Abad, C. (1999) Addressing substrate glutamine requirements for tissue transglutaminase using substance P analogues. *FEBS Lett*. 28;451(3):231-4.

30. Nordlund, H.R., Laitinen, O.H., Hytönen, V.P., Uotila, S.T., Porkka, E., Kulomaa, M.S. (2004) Construction of a dual chain pseudotetrameric chicken avidin by combining two circularly permuted avidins. *J Biol Chem*. 279(35):36715-9.

Transglutaminase-mediated modification of proteins: Molecular mechanisms and metal-labelling of proteins

Results



Myoglobin

3. Site-specific radiolabelling of proteins using TGase

INTRODUCTION

Nuclear imaging techniques. In order to obtain a structural and functional non invasive evaluation in a living subject several *in vivo* imaging techniques can be used as single photon emission computed tomography (SPECT) and positron emission tomography (PET). SPECT and PET are imaging techniques that use a radiolabelled biochemical compound injected into a patient (Figure 1). These techniques detect photons that are emitted by the radioisotope during its decay from the patient body where the radiolabelled compound was injected, allowing to produce an imagine of the distribution of the radiotracer in the body. The main difference between SPECT and PET is the type of detector that is used to acquire the photons produced by radionuclides decay and the choice of the radionuclide itself. In SPECT γ emitter radionuclides are used and the photons are detected and recorded from different projections. The detector needs to acquire 180° of projection images *(1)*. Instead, in PET positron emitters radionuclides are used. The positron emitters radionuclides have a particular decay characteristic that results in the simultaneous emission of two γ rays in opposite directions. In this case, γ rays are acquired by two opposing detectors *(2-3)*. A SPECT disadvantage, in respect to PET, is that spatial resolution and sensitivity decrease while the dept in the body increases, due to the fact that the γ ray amount produced by γ emitters is half than that produced by β emitters. An advantage of SPECT is that equipment and radionuclides are readily available in every nuclear medicine department and radionuclides can be eluted from a generator, while in the case of PET, radionuclides have to be produced with an on-site cyclotron. Moreover, while in SPECT radionuclides have a long half-life, in PET they have a short half-life *(2)*.

SPECT and PET primarily provide functional and molecular information *(4)*. The radiotracer imaging has the main characteristic that it does not alter or perturb the biological system under investigation. An advantage of the radionuclide imaging is also the possibility to study in real time the living subject in order to increase the knowledgement on physiological mechanisms and to understand the effects of drug administration *(5)*. By using different tracers it is possible to measure different physiological, biochemical and pharmacokinetic parameters as well as blood flow, blood volume, oxygen utilisation, pre- and post-synaptic receptor density and affinity, glucose metabolism, neurotransmitter release, drug delivery and uptake, enzyme activity, gene expression, etc *(2)*.

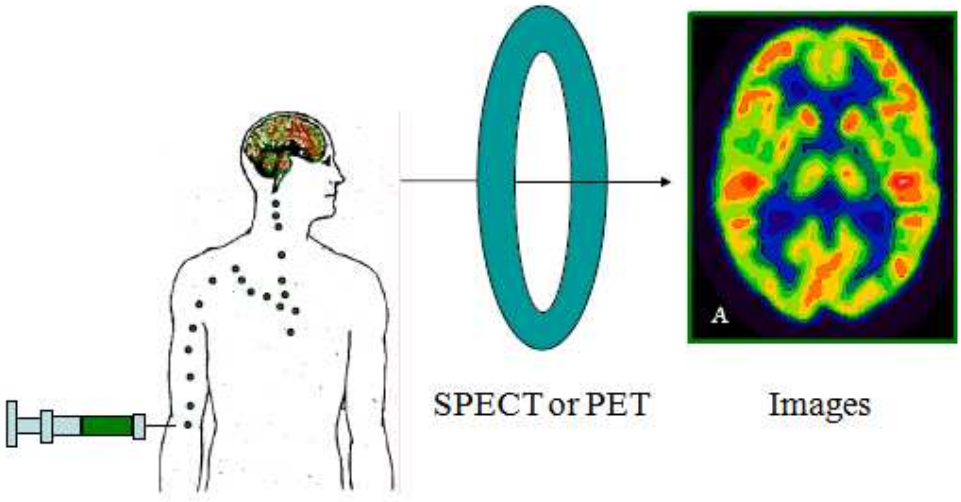


Figure 1. Schematic representation of SPECT and PET imaging techniques.

Proteins radiolabelling. For the development of *in vivo* imaging techniques fundamental is the design of suitable imaging probes and the choice of the radionuclides. The most utilized radionuclides for *in vivo* imaging are: ^{99m}Tc (γ emitter with a half-life of 6 hours) used for SPECT, and ^{64}Cu (β^+ and β^- with a half-life of 12.7 hours) used for PET and radiotherapy. An ideal imaging probe should exhibit several characteristics as high binding affinity toward its cellular target, specific uptake, discrete retention in the target, rapid clereance from nontargeted tissue and high stability *in vivo*. Proteins as monoclonal antibodies are thus ideal to direct the radionuclide towards the target tissue or cells. In order to modify a protein with a radionuclide, particular chelating agents must be designed. Indeed, once the radionuclide is chosen an important aspect is the synthesis of suitable bifunctional chelating agent (BFCA). BFCA are constituted by two different moiety: a metal chelating group and a second functional group that enables incorporation into a targeting protein *(6)*. The efficient bindingof the radionuclide is achieved if the following requirements are fulfilled: charge, matching of cavity size of the chelator with the radius of the radionuclide, chelate denticity or number of donor binding groups, and the character of the donor binding groups *(7)*. Since dissociation of the radiometal is related to toxic effects in the case of therapeutic radiometal complexes and poor image quality for diagnostic radiometal conjugates, the stability of the metal complex is unambiguously the most important parameter that we have to consider during the design and synthesis of suitable BFCA to be used in medicine *(7)*.

The development of methods of conjugation of BFCA to proteins is nowadays of paramount importance for the practical use of proteins in diagnostic and therapy. The chemical procedures developed so far for the radiolabelling of proteins are not site-specific and usually several amino acid side chains are involved in the covalent modification. Thus, the overall utility of these chemical methods can be limited, due to the heterogeneity and even decreased bioactivity of the radiolabelled proteins. A well-defined radiolabelled protein is highly desirable, since this will be more easily characterized in terms of chemical identity and diagnostic or therapeutic efficacy. Few example of site-specific protein radiolabelling are reported in literature as the site-specific labelling of nanobodies at the level of a His-tag with ^{99m}Tc complexes for SPECT imaging *(8)*. The site-specific radiolabelling mediated by TGase was also reported for a mutant of a monoclonal antibody (chCE7agl) (mutation is N297Q) *(9)*. This mutation removes the glycosylation site of the antibody and increases the mobility on loop region between residues Gln295-Thr299, making Gln295 reactive towards the TGase mediated reaction. Modification of this antibody with microbial TGase yielded a highly homogeneous and site-specific radiolabelled product (Figure 2).

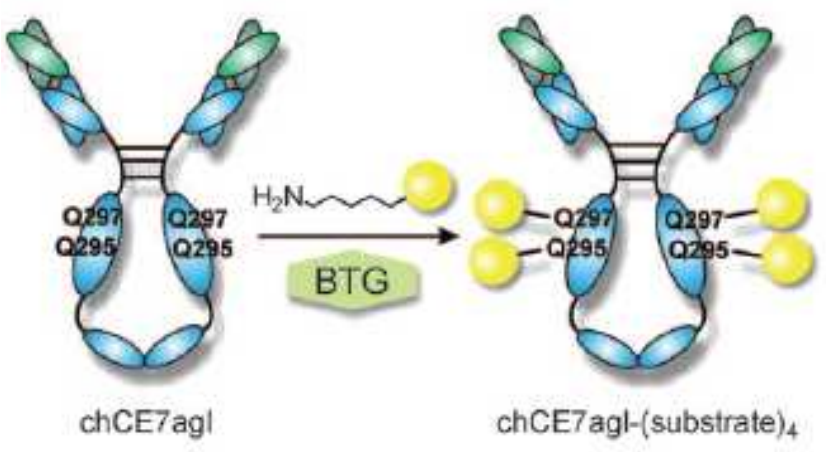


Figure 2. Scheme of site-specific radiolabelling of monoclonal antibody by Tgase mediated reaction *(9)*.

Protein radiolabelling can be obtained by following two different approaches: direct and indirect radiolabelling.

Direct radiolabelling: It is a relatively easy procedure for labeling peptides, its popular appeal stems from simplicity of execution since it does not require peptide conjugation with a bifunctional chelator, a way that may need considerable synthetic expertise.

The technique is based on the coordination of the radionuclide to thiolate groups of cysteine side chains previously generated by reduction with SnCl2 of a disulfide linkage that should be present in the targeting molecule. This approach is easy to perform on antibodies or antibody fragments *(10-15)* but many small peptides do not have disulfide bonds or cannot be reduced without loosing the biological activity.

Indirect radiolabelling: The indirect radiolabelling entails the use of a preformed BFCA to firmly conjugate the peptide to the radiometal. It is important that the labeling conditions do not affect the binding activity of the peptide toward the target receptor. There are two main modalities to conduct the indirect labeling of a peptide: pre-radiolabelling and post-radiolabelling *(16).* The pre-labelling approach implies the initial formation of the BFCA chelate, followed by its conjugation the protein in a separate step. In the post-labelling, the BFCA is initially attached to the protein to form the BFCA-protein conjugate. Once the BFCA-protein conjugate is prepared, radiolabelling is performed by ligand exchange reaction (Figure 3).

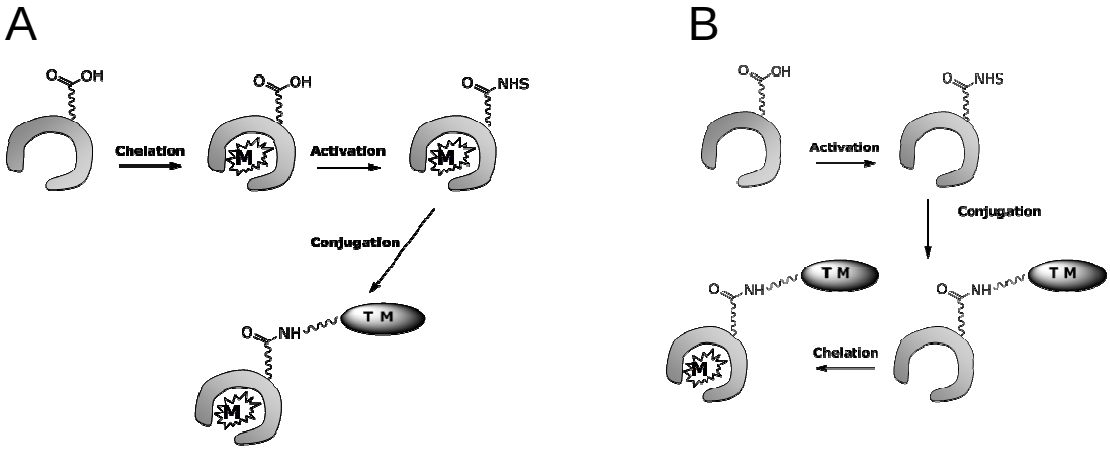


Figure 3. Schematic representation of two different radiolabelling approaches. In pre-labelling the formation of the ^{99m}Tc-BFCA chelate is followed by its conjugation to the target molecule **(A).** in post-labelling the BFCA is conjugated to the target molecule and then the radiolabelling is performed **(B).**

BFCA for ^{99m}Tc and ^{64}Cu radiolabelling via TGase. In this project, we propose to use the TGase mediated reaction to perform the site-specific radiolabelling of a target protein using two new generation of BFCA complexes that bind ^{99m}Tc and ^{64}Cu. These BFCA were design to chelate the specific radionuclide and also to contain a peptide sequence or a primary amine that allow their TGase-mediate conjugation to a protein substrate.

^{99m}Tc complexes. A new class of complexes to be used for *in vivo* imaging is represented by asymmetrical nitrido complexes, based on the chemical properties of the substitution labile [Tc(N)X2(PNP)] complex (PNP=amino-diphosphine), which represent an

interesting opportunity in design receptor-specific ^{99m}Tc agents. The strong electrophylic $[Tc(N)(PNP)]^{2+}$ moiety efficiently reacts with bifunctional ligands (L) carrying π-donors as coordinating atoms to afford asymmetrical nitrido heterocomplexes of the type $[Tc(N)(L)(PNP)]^{0/+}$ *(17)*. It was found that the [Tc(V)≡N]2+ core exhibits very high stability under a wide range of experimental conditions and is able to confer a particular thermodynamic/kinetic stability and a remarkable resistance to red-ox processes to the complexes where it is present. Based on these considerations, extensive studies have been carried out on the synthesis and biological evaluation of different classes of symmetrical and dissymmetrical nitrido technetium complexes. In spite of this, the application of this technology to radiopharmaceutical preparation still remains less investigated, probably due to incorrect beliefs that they are not suitable, for the development of imaging probes using traditional instant kit technologies.

In this project the chelator of Tc was modified with two different peptides with the following sequences: Biotin-GKGC and Biotin-GQGC. These peptides contain a Gln or a Lys residue that can be exploited for the TGase-mediated conjugation to proteins, while the Cys residue on the C-terminus acts as π-donor ligand for $[^{99m}Tc(N)Cl_2(PNP)]^{2+}$ with its free carboxylic group and the thiol group $[O^-,S^-]$ thus forming an asymmetrical nitride complexes. These peptides contain also a biotin moiety at the N-terminus that can be used for a fast and easy purification of the radiolabelled protein (Figure 4A).

^{64}Cu complex. N2-S2-type bis(thiosemicarbazone) ligands form stable neutral and planar complexes with Cu(II) ions. This type of complex has been investigated extensively for the imaging of hypoxia in the form of copper(II)-diacetyl-bis(*N*-4- methylthiosemicarbazone), Cu[ATSM]. Radiolabeling occurs instantly at room temperature and near physiological pH *(18-20)*. This enables an easy labeling after protein conjugation and avoids multi-step radiosynthesis. Another BFCA that we synthesized is a bis(thiosemicarbazone) H2ATSM modified with 6-aminocaproic acid at the level of its azide group, leading to a product named H2ATSM/C *(21)* (Figure 3B). In this BFCA the bis(thiosemicarbazone) plays the role of chelator for ^{64}Cu, while the free amino group of the conjugated 6-aminocaproic acid is used as acyl acceptor for TGase mediated reaction (Figure 4B).

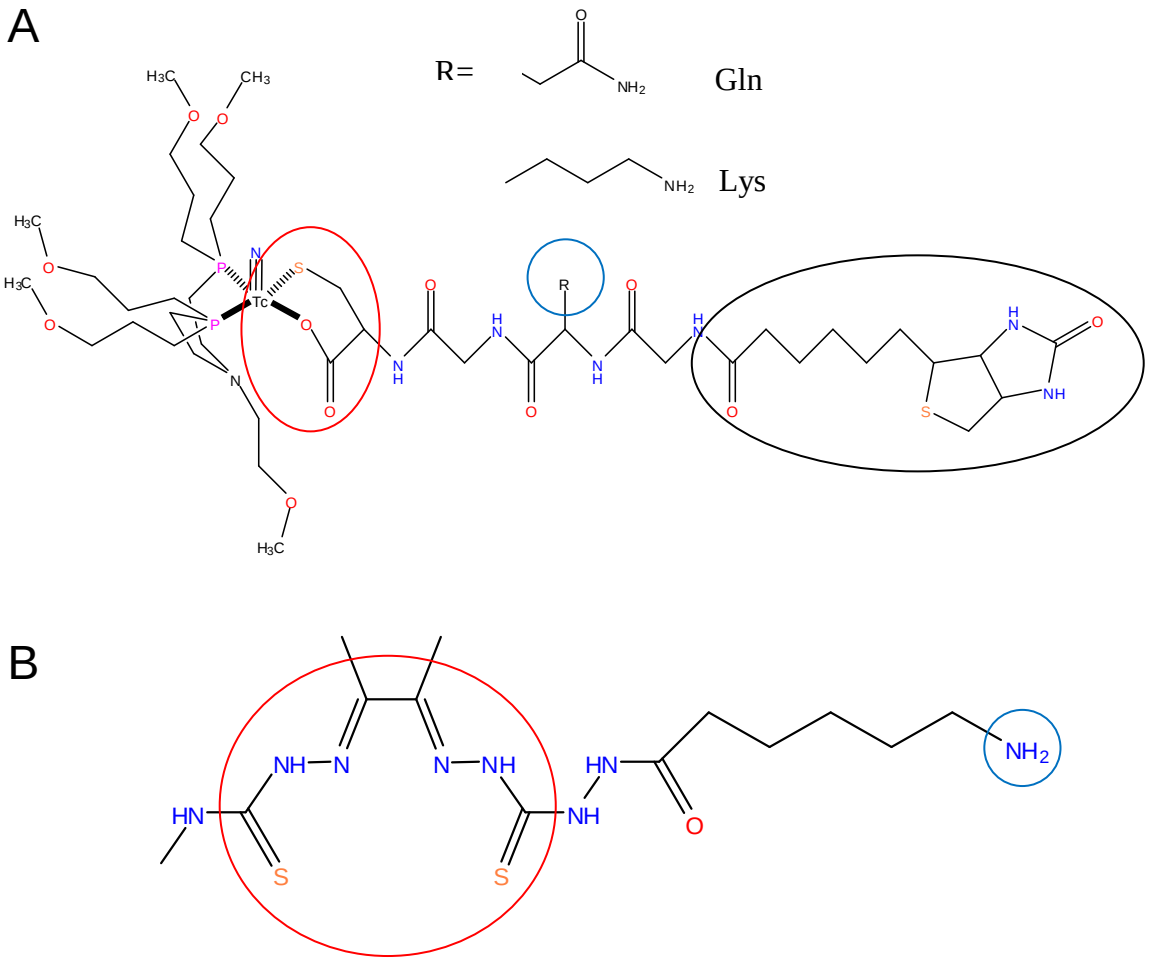


Figure 4. Structure of the BFCA used in the present study. Structure of the $^{185/187}$Re complexes **(A)**. Circles indicate essentials moieties of the designed BFCA. A red circle shows the Cys residue that act as chelator toward ^{99m}Tc. A blue circle shows the Gln or Lys residues (R) that can be exploited in TGase mediated reaction, while a black circle indicates the biotin moiety, that can be used for a fast purification of the modified proteins by using its affinity for avidin. Structure of H2ATSM modified with 6-aminocaproic acid **(B)**. A red circle shows the N2S2-type cheletor system for ^{64}Cu, while a blue circle the free amino group that acts as acyl acceptor in TGase mediated reactions.

Apomyoglobin, a model to study protein radiolabelling mediated by TGase. Apomyoglobin (apoMb), myoglobin without the heme group, was chosen as model protein to test the reactivity toward TGase of the ^{99m}Tc and ^{64}Cu binding BFCA described above. ApoMb is a small protein constituted by 153 amino acid residues without disulphide bonds. It has long been studied as a model protein for protein structure, folding and stability *(22-23)*. While the holo form (containing the heme group) has a secondary structure characterised by eight α-helices (named A through H) arranged into a globular tertiary structure, the apo form display a slightly reduced helical content due to unfolding of the polypeptide chain at the level of helix F (residues 82-97) *(24-25)*, as demonstrated by spectroscopic studies and NMR measurements (Figure 5) *(22)*. The high flexibility of the polypeptide chain encompassing helix F was also demonstrated by limited proteolysis experiments conducted on apoMb since they indicated as this region is selectively hydrolysed by proteases *(26-27).*

The reactivity of apoMb towards TGase has already been reported. Indeed, apoMb is site-specifically conjugated by TGase to a monodisperse amino-derivative of PEG and the site of derivatization was identified in Gln91, which is located in the region of helix F *(28)*. Studies on the modification of apoMb at the level of Lys residues by TGase have also been performed in our laboratory and they demonstrated that only two Lys residues among the 19 residues of the protein sequence were modified. Interestingly, the derivatization of apoMb occurs at the level of the Lys96 and Lys98 that are also located in the region of helix F *(29)*.

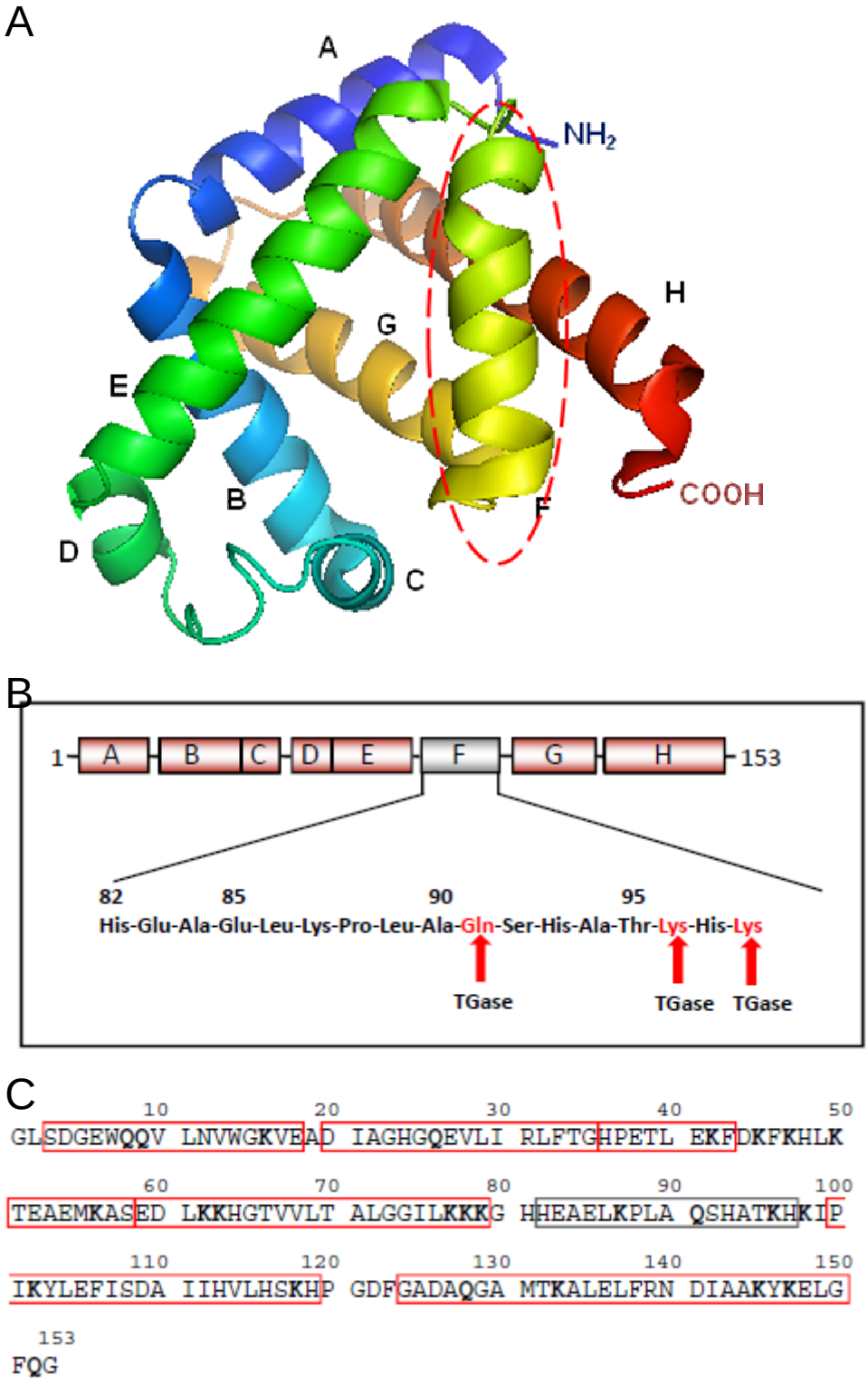


Figure 5. Three dimensional structure of horse heart myoglobin as obtained from the PDB file 1YMB by using PyMOL **(A)**. Helical elements are indicated. Chain segment (residues 82–97) corresponding to the disordered helix F is shown in red dotted circle. Sequence of helix F is shown in **(B)** and the sites of the specific TGase mediated modification are indicated by a red arrow. The sequence of apoMb is shown in **(C)**, where the red rectangles indicate the α-helice and the black rectangle indicates the helix F. The Lys and Gln residues along the polypeptide chain of myoglobin are indicated in bold.

MATERIALS AND METHODS

Materials. Fmoc protected amino acids and resin for solid phase peptide synthesis were purchased from NovaBiochem (Laufelfingen, CH). PNP is a gift of ICIS-CNR. Saline solution was purchased from Novaselect (Tito Scalo, TZ, Italy). SepPak was purchased from Waters (Manchester, UK). Trypsin was purchased from Promega (Madison, WI, USA). V8 protease and AEBSF were purchased from Sigma-Aldrich (Milwaukee, WI, USA) The acetonitrile used for reverse phase HPLC chromatography is from Carlo Erba Reagenti (Italy). Trifluoroacetic acid (TFA) and the other high purity salts used for spectroscopic analyses were obtained from Fluka (Buchs, Switzerland). All other chemicals were of analytical reagent grade and were obtained from Sigma or Fluka.

Mass spectrometry analysis. 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 variable collision energy values 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).

Activity of TGase solutions. 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 (*30)* (extinction coefficient of 1.89 mg/ml). The assay was performed in a total 0.5 ml of an enzyme-substratesolution 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% FeCl3 in0.1 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 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.

Apomyoglbin preparation. 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 of apoMb was determined from its absorbance at 280 nm according to Gill and von Hippel *(30)* (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 0.1 % TFA and then reaction buffer was added to obtain the desired concentration.

Synthesis of Biotin-GQGC and Biotin-GKGC. Solid phase synthesis of biotinilated peptides was performed by using an automated synthesizer Applied Biosystems mod. 431A (Foster City, CA). Amino acids used for the synthesis were protected on the N-terminus with 9-fluorenylmethyloxycarbonyl (Fmoc), while the side chain of Lys, Gln and Cys were protected by trityl. The C-terminal residue (Cys) is bound to a 4-benzyloxybenzyl ester polimer (Wang resin). The successive coupling reactions were based on the activation of the carboxylic group of the next amino acid by using O-(7-azabenzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HATU) with a molar ratio amino acid:HATU of 1:1 and N,N-diisopropylethylamine (DIPEA) with a molar ratio of 1:2 (amino acid:DIPEA). To add the biotin moiety (not soluble) we proceed manually. To the biotin were added 10 eq. of DIPEA and 5 eq. of HATU (both dissolved in NMP). After the activation step all amount of activated biotin became soluble. This solution was then added to the peptides. The solvent used to dissolve amino acid residues was dimethylformamide (DMF), while to wash the reactors in each reaction step dichloromethane (DCM) was used. The cleavage to remove the insoluble resin and the protecting groups on the amino side chain was performed by using a 95% TFA, 3% of triethylsilane and 2% of H2O. This reaction proceed for 3 hours under stir at room temperature. Then, the solution was filtered and precipited in cold Et2O for 20 minutes. The tube containg the precipitate was then centrifuged for 5-10 minutes at 3500 rpm and washed 3 times with cold Et2O. The precipitate was then liofilised. The product was

dissolved in an aqueous solution of 1% TFA and analyzed by RP-HPLC using a C18 Vydac 218TP1010 (10 x 260 mm) column. The separation was performed with a gradient of ACN containing 0.85 % TFA from 15 to 40 % in 35 min and a flow of 2 ml/min. Characterization of the product was performed by MS analysis of the collected peaks by RP-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.

Synthesis of $^{185/187}$Re complexes. A solution of PNP (1.58 mg, 0.0032 mmol, 1.3 eq.) in degassed CH3CH2OH (0.5 ml) was added to a suspension of [Re(N)Cl2(PPh3)2] in degassed DCM (1 ml) under nitrogen atmopsphere. Mixture was left in reflux for a time of 1.5 h (under nitrogen atmpsphere). To the obtained solution it was added a biotinilate peptide solution (2.9 mg in 1 ml of degassed di CH3CH2OH /H2O 1:1) and 6-10 drops of Et3N and left to react for 2 h. The reaction mixture was cooled and the solvent eliminated by nitrogen flow. Then the solid was resuspended in n-esane/Et2O 1:1 (5 ml) and left under shaking for 30 min. The pellet was settled overnight at 4 °C, centrifuged and the liquid phase eliminated. Synthesis of [Re(N)Cl2(PPh3)2] is reported in literature *(31)*. The product was dissolved in DMSO and analyzed by RP-HPLC using a C18 Vydac 218TP1010 (10 x 260 mm) column. The separation was performed with a gradient of ACN containing 0.85 % TFA from 15 to 40 % in 35 min and a flow of 2 ml/min. Characterization of the product was performed by MS analysis of the collected peaks by RP-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.

Synthesis of ^{99m}Tc complexes. In a vial containing 5 mg of SDH, satured with nitrogen, we added 0.1 mg of SnCl2·2H2O suspended in 0.1 ml of aqueous solution of NaCl 0.9% (saline solution) , 0.680 ml of saline solution and 0.250 ml of Na(^{99m}TcO4) (obtained by ^{99}Mo/^{99m}Tc generator (Elumatic III, IBA CIS bio, France)) in saline solution (2-3 mCi). The solution was kept at room temperature for di 15 min. Then, we added PNP (1.0 mg dissolved in 0.5 ml of λ-ciclodestrine 2 mg/ml in saline solution), 0.2 ml of 0.2 M phosphate buffer pH 7.4 and the biotinilated peptide dissolved in 10^{-2} M HCl (0.32 mg in 0.32 ml). The reaction mixture was kept at 100°C for 30 min.

Purification by SepPak and RP-HPLC analysis of ^{99m}Tc complexes. The reaction mixtures were purified by SepPak in order to eliminate not reacted ligand. The reaction mixtures were diluted with 10 ml of deionized H2O and loaded on a SepPak (activated with 5 ml of

CH3CH2OH and 5 ml of H2O), washed with 20 ml of H2O and 5 ml of 25% CH3CH2OH. Elution was performed with 1.5 ml of 50% CH3CH2OH and 1 ml of 90% CH3CH2OH. The two different fractions were mixed. The purified products were then injected in RP-HPLC. The HPLC system used for the RP-HPLC analysis was a Beckman System Gold with a double pump system Model 126, valve injector 210A, UV detector Module 166 and radio detector Model B-FC-3200 Bioscan. Samples were further analyzed by RP-HPLC using a C18 Vydac 218TP (250 x 4.6 mm) column. The separation was performed with a gradient of ACN containing 0.85 % TFA from 15 to 28 % in 18 min, 28% for 3 min, from 28 to 90% in 2 min, from 22 to 95% in 3 min.

Synthesis and analysis of H2ATSM/C. The synthesized H2ATSM *(32)* was modified with 6-aminocaproic acid in correspondence of the azide group in order to add a primary amine to be used in TGase mediated reactions *(15)*. H2ATSM, Fmoc-6-aminocaproic acid, DIPEA and HATU were dissolved in DMF and mixed (in a molar ratio of 1:1:1:1). The precipitate was then stirred in hot CH3CH2OH before being filtered off and washed with cold Et2O to unblock the primary amine (protected by Fmoc). The product was stirred in DMF and piperidine (in a molar ratio of 1:40 product:piperidine) for 45 min. the solution was concentrated and H2O was added until precipitation. The aqueous phase was then concentrated and Et2O was added until precipitation. The product was dissolved in DMSO and analyzed by RP-HPLC using a C18 Vydac 218TP1010 (10 x 260 mm) column. The separation was performed with a gradient of AcCN containing 0.85 % TFA from 15 to 40 % in 35 min and a flow of 2 ml/min. 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. Characterization of the product was performed by MS analysis of the collected RP-HPLC peaks.

TGase mediated conjugation of apoMb with ReQ and ReK. A solution of apoMb in 0.1M NaH2PO4 pH 7 was mixed with a solution 10 mg/ml of ReQ or ReK (in DMSO) in a molar ratio of 1:10 with a final concentration of apoMb of 0.4 mg/ml. 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 150 µg were removed after 0, 1, 2 and 4 hours and the reactions were 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 5 to 40% in 5 min, from 40 to 50 % in 25 min. 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. Fractions collected from the RP-HPLC analysis were collected, lyophilized and analyzed by MS.

Proteolysis of apoMb$^{1/2ReQ}$ and apoMb^{ReK} with trypsin and V8 protease. A solution of modified apoMb (0.7 mg/ ml) 0.1 M NH4HCO3 pH 8.5 was mixed with trypsin and V8 (in a weight ratio of 1/100 E/S). The reaction mixture was incubated at 37° C overnight. The proteolysis reaction was stopped by adding an aqueous solution of AEBSF (4-(2-aminoethyl)benzenesulfonyl fluoride hydrochloride) with a final concentration of 0.1 mM and upon lowering the pH of the reaction mixture to pH 5 by adding HCOOH. Identification of the sites of conjugation on the modified peptides was performed by MS.

TGase mediated conjugation of apoMb with ^{99m}Tc complexes. Pre-radiolabelling in carrier out: in a vial containing 209 µl of [^{99m}Tc(N)(PNP) (peptide)] (3.5 mCi) we added a solution of ReQ or ReK (dissolved in 5 µl PBS), 25 µg of apoMb (dissolved in 33.35 µl of 0.1 M phosphate buffer at pH 7) and 2.5 µg of TGase (dissolved in 2.5 µl of 0.1 M phosphate buffer at pH 7). The reaction mixtures were incubated at 37 °C for 20 hours. *Pre-radiolabelling in carrier free:* in a vial containing 214 µl of [^{99m}Tc(N)(PNP) (peptide)] (3.5 mCi) we added 25 µg of apoMb (dissolved in 0.1 M phosphate buffer at pH 7) and 2.5 µg of TGase (dissolved in 0.1 M phosphate buffer at pH 7). The reaction mixtures were incubated at 37 °C for 20 hours. *Post-radiolabbeling:* 100 µl of [^{99m}Tc(N)(PNP)]$^{2+}$ (15mCi) were added to 25 µl of PBS solution containing 43 µg of apoMb conjugated with Biotin-GQGC or Biotin-GKGC. Reaction mixtures were incubated at 37 °C for 20 hours.

TGase mediated conjugation of apoMb with H2ATSM/C. A solution of apoMb in 0.1M NaH2PO4 pH 7 was mixed with a solution 1.3 mg/ml of H2ATSM/C (in DMSO) in a molar ratio of 1:10 with a final concentration of apoMb of 0.4 mg/ml. The reaction mixture was incubated with TGase at an E/S ratio of 1/20 by weight. The reaction was allowed to proceed at 25 and 37 °C. Sample aliquots was removed after 0, 0.5, 1, 2 and 5 hours 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 5 to 40% in 5 min, from 40 to 50 % in 25 min. 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). Fractions collected from the RP-HPLC analysis were collected, lyophilized and analyzed by MS.

RESULTS AND DISCUSSION

3. 1. Development of ^{99m}Tc complexes to use in TGase mediated reactions. The development of a methodology that allow to carry out the site-specific radiolabelling of proteins mediated by TGase require the synthesis of suitable complexes of the radionuclide (see introduction). In this study we have chose as radionuclide the ^{99m}Tc that is a γ emitter and can be used in SPECT diagnosis. The complexes will be constituted by a complex core of the radionuclide and a BFCA that will play a double role of chelator agent toward the metal ion and substrate for TGase. On the basis of these requirements, we designed BFCA that contains a Cys residue on C-terminus that acts as π-donor chelating agent and a Gln or Lys residues as acyl donor or acceptor respectively, as substrates for TGase. The development of these complexes need also the synthesis of analogous complexes of a not radioactive metal ion in order to characterize on macroscopic scale the behavior of the complexes and to implement the enzymatic reactions catalyzed by TGase. For this reason, we carried out the synthesis and characterization of analogous complexes of $^{185/187}$Re (Figure 6). Since $^{185/187}$Re is chelated by the designed BFCA similarly to ^{99m}Tc.

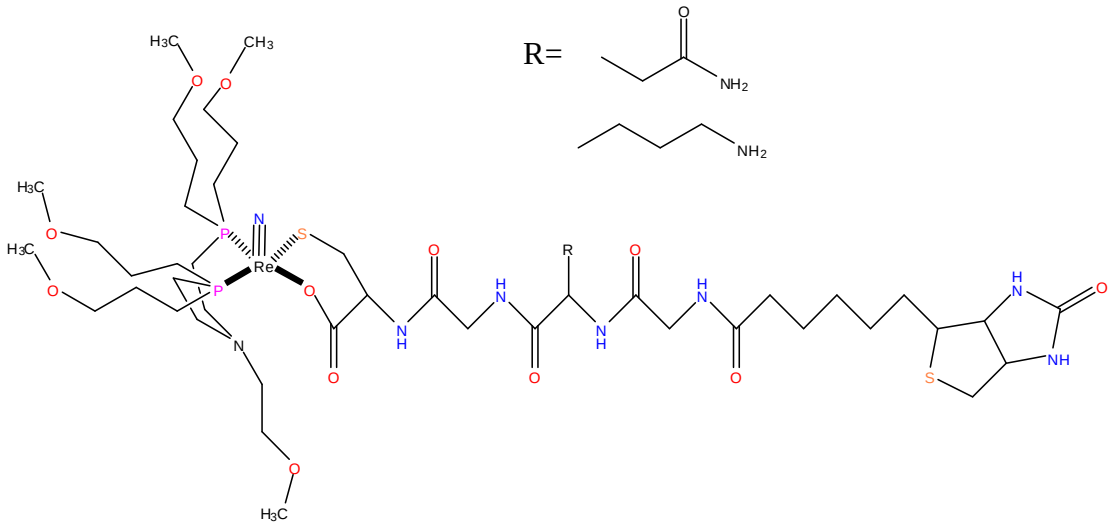


Figure 6. Structure of the $^{185/187}$Re complexes. These complexes are constituted by a Cys residue on C-terminus the act as chelatating agent toward the $[^{185/187}Re(N)(PNP)]^{2+}$ complex, a Lys or Gln residue as acyl acceptor and donor respectively (indicate as R group) and a biotin moiety to use for a fast and easy purification of modified proteins.

3.2. Synthesis and characterisation of Re and Tc chelating peptides. The synthesis of the peptides Biotin-GQGC and Biotin-GKGC, as suitable BFCAs for $^{185/187}$Re and ^{99m}Tc complexes, was performed by classical solid phase synthesis. After the cleavage and the precipitation , these peptides were purified by RP-HPLC and analyzed by ESI-MS and ESI-MS/MS in order to confirm their sequences. In both cases, the chromatograms show the presence of two different peaks that we have analyzed by MS and MS/MS. The MS analysis of the peptide material eluted by RP-HPLC related to the synthesis of Biotin-GQGC indicated the presence of two species, one that corresponds to the desired peptide (calculated mass= 589.19 Da, experimental mass= 589.30 Da) while the second has a higher mass of about 106 Da (695.40 Da). Both species were analyzed by MS/MS confirming their sequences. The same result was obtained in the case of the peptide Biotin-GKGC, where the desired peptide has a mass of 589.25 (theretical mass= 589.23) while a second species gave a mass of 695.38 Da. Interestingly, MS/MS analysis of the adduct (+106 Da) gave an identical sequence to that of the peptide because the modification is very labile when subjected to MS/MS fragmentation.

From the literature it is known that the presence of an increased mass value of 106 Da during peptide synthesis it could be due to a thioanisyl adduct or the combination of oxidation and an anisyl adduct *(33).* Since it was present thioanisol in the cleavage solution, the idea was to remove this scavenger in order to overcome this drawback and obtain a higher yield of peptides by using different scavangers as well as triethylsilane. The cleavage of these two peptides was repeated by using this new condition and a RP-HPLC was performed to analyze the resulting peptides. Despite the thioanisol was replaced by triethylsilane, we still observe the presence of two species, even if the amount of the undesired species is decreased, in the case of the peptide Biotin-GQGC, as shown by the RP-HPLC chromatograms in Figure 7. Moreover, RP-HPLC was repeate at 210 nm in order to decrease the signal of some hypothetical aromatic group bound on the synthesized peptides and in fact we can observe an its decrease but not an absolute absence of its signal (data not shown). This analysis could confirm the presence of an aromatic group on the desired peptides but more detailed analysis are necessary to establish the nature of this type of modification.

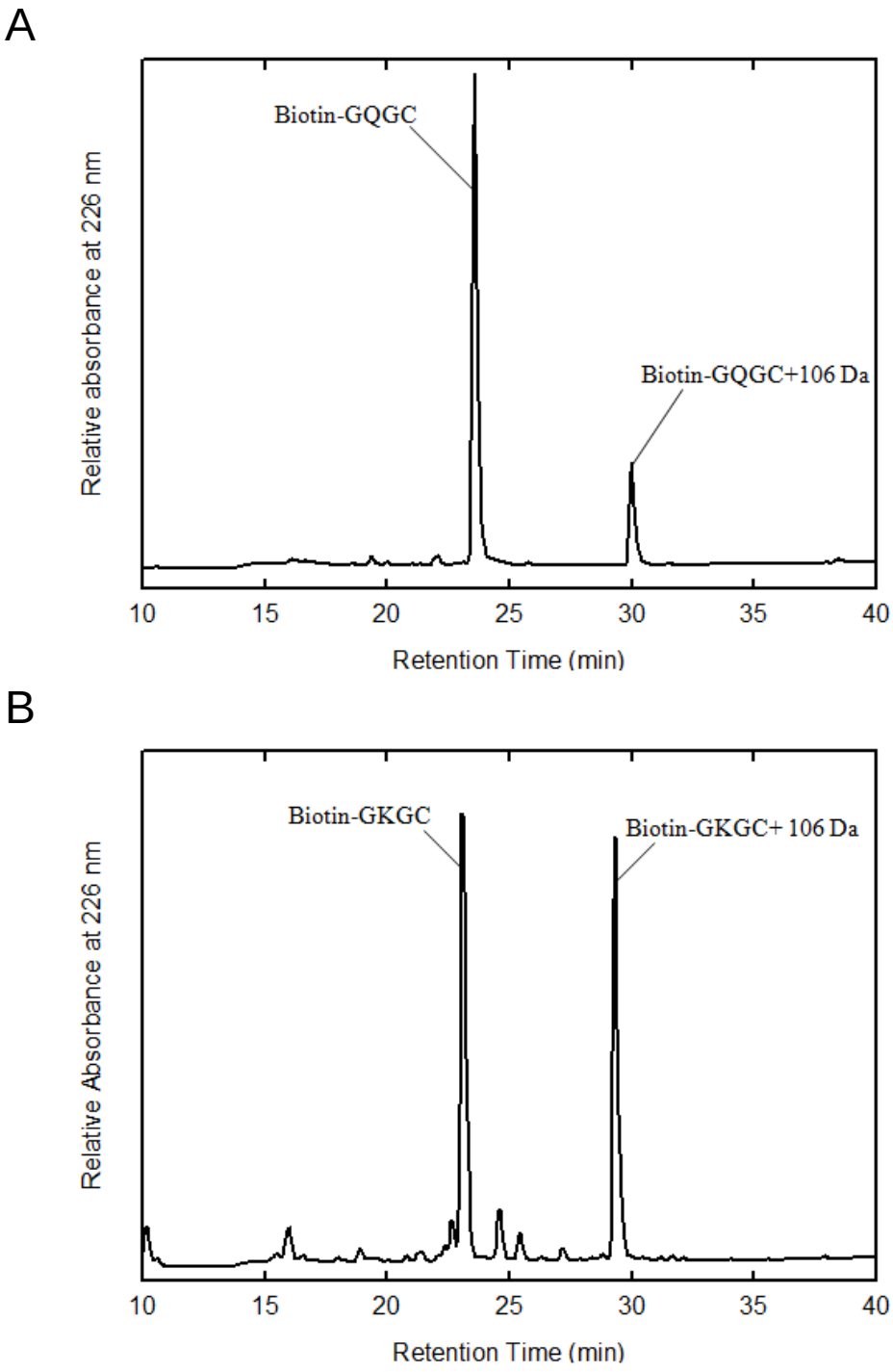


Figure 7. RP-HPLC chromatograms of the peptides Biotin-GQGC **(A)** and Biotin-GKGC **(B)** after the cleavage with triethylsilane as new scavenger. In the case of the peptide Biotin-GQGC is observed a decrease of the signal that corresponds to the modified peptide with a higher mass of 106 Da.

3.3. Synthesis of the $^{185/187}$Re complexes. The synthesis of the complexes ($^{185\text{-}187}$Re(N)Cl2(PNP))(Biotin-GQGC) and ($^{185\text{-}187}$Re(N)Cl2(PNP))(Biotin-GKGC) (Figure 6) that we rename ReQ and ReK respectively, were analyzed by RP-HPLC and ESI-MS. The chromatograms of each complex show more peaks that were analyzed by ESI-MS. The analysis of the complex with the Gln containing peptide shows that the reaction mixture contains mainly the desired product (Figure 7) but also not reacted biotinilated peptide and species that it is possible to consider as a degradation product of the complex (Table 1) (Figure 8A). In fact, these last species present the typical isotopic pattern of the $^{185/187}$Re but on the basis of its mass it seems not contain the biotinilated peptide (Figure 8). The RP-HPLC chromatogram of ReK reaction mixture contains two signals at 20.47 min and 21.30 min (Figure 7B). These two signals give the same m/z signals by ESI-MS analysis. It seems reasonable that they represent two different conformational isomers of ReK. Other signals around 30 minutes are due to not reacted amino diphosphine used to synthesize the complex [$^{185\text{-}187}$Re(N)Cl2(PNP)]$^{2+}$.

Once purified by RP-HPLC, the stability of complexes ReQ and ReK were tested at different incubation times and at different pH values. Stability tests were performed in the conditions used for the TGase mediated conjugation of proteins and the degradation patterns were analyzed by RP-HPLC (Figure 9). The chromatogram shows that after 24 hours incubation at pH 7 and at pH 2 the complexes are not degradated, confirmed by ESI-MS analysis of the material eluted in the peaks confirmed this result.

Table 1. Observed and calculated molecular weight by MS analysis of the RP-HPLC collected peaks of the Re complexes

Molecula weight (Da)		**species**
observed	**calculated**	
1271.60	1271.46	ReQ
589.20	589.30	Biotin-GQGC
1271.52	1271.54	ReK

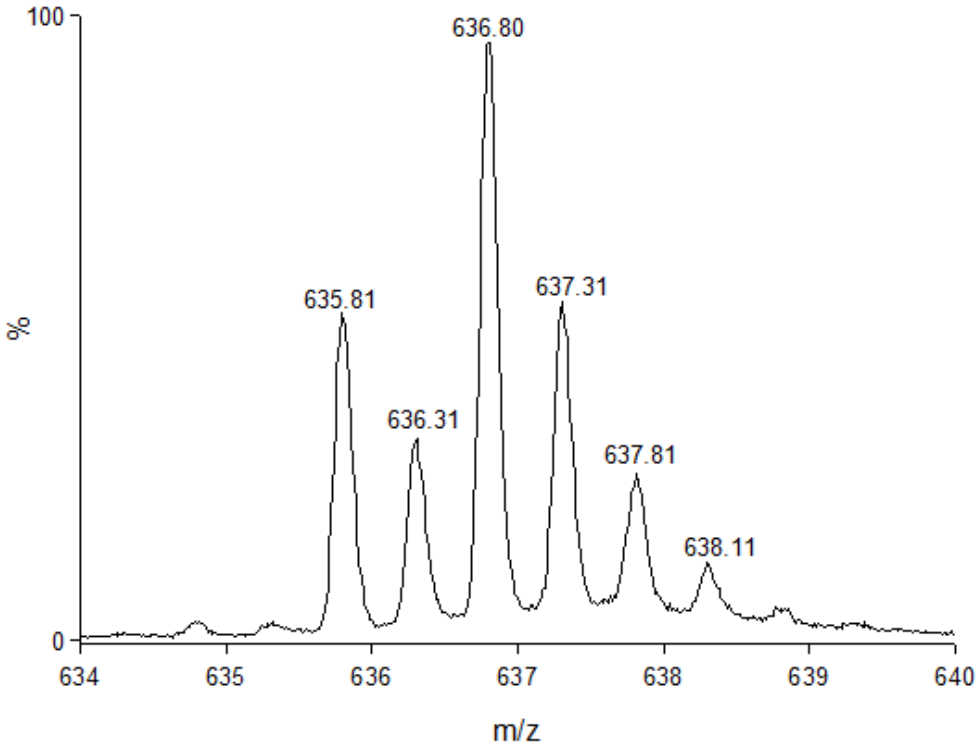


Figure 7. MS analysis of ReQ. It is shown a double charged signal due to ReQ, where it is possible to observe the isotopic pattern of the $^{185/187}$Re.

3.4. Synthesis of the ^{99m}Tc complexes. The synthesis of the complexes (^{99m}Tc(N)(PNP))(Biotin-GQGC) and (^{99m}Tc(N)(PNP))(Biotin-GKGC) were carried out by following a two steps reaction (see Materials and Methods). The chemical identity of the synthesized complexes was established by comparison of their RP-HPLC chromatograms with the chromatograms of the analogous complexes of $^{185\text{-}187}$Re. The ^{99m}Tc complexes showed chromatographic peaks with a retention time shifted toward little higher times if compared with $^{185\text{-}187}$Re complexes, showing a higher hydrophobicity. RP-HPLC chromatograms of ^{99m}Tc complexes show the presence of two peaks that could be representative of two different conformational isomers (Figure 10) and are indicated as isomer a (first eluted peak) and isomer b (second eluted peak).

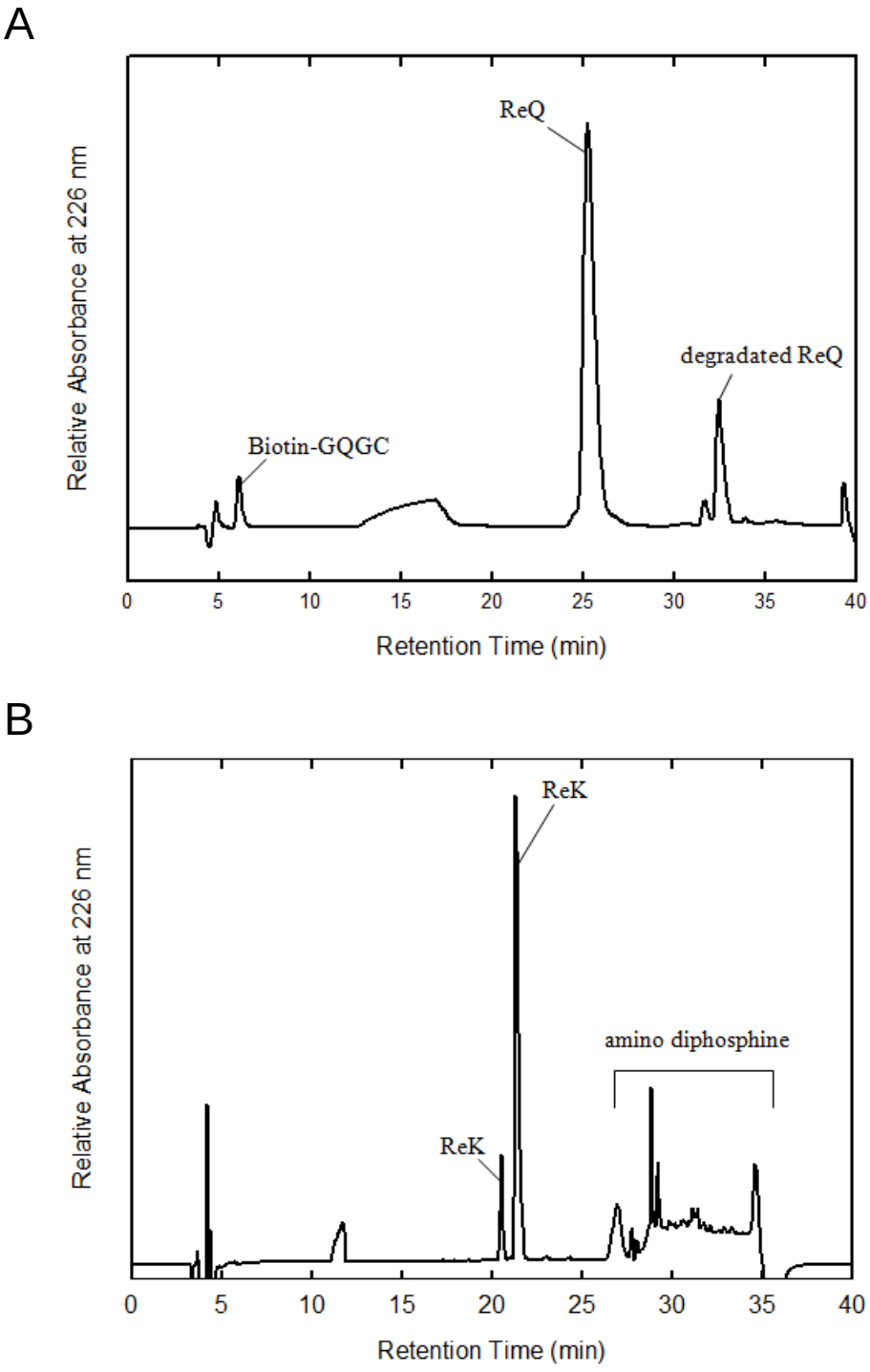


Figure 8. RP-HPLC chromatogram of the complex ReQ . There are three different species: not reacted biotinilated peptide (6.10 min), the ReQ complex (25.22 min) and a degradation product of the complex (32.42 min) **(A)**. RP-HPLC chromatogram of the complex ReK reaction mixture. There are two peaks at retention time 20.47 min and 21.30 min that represent two different conformational isomers (syn and anti) of the complex ReK. Amino diphosphine elutes at higher retention times **(B)**.

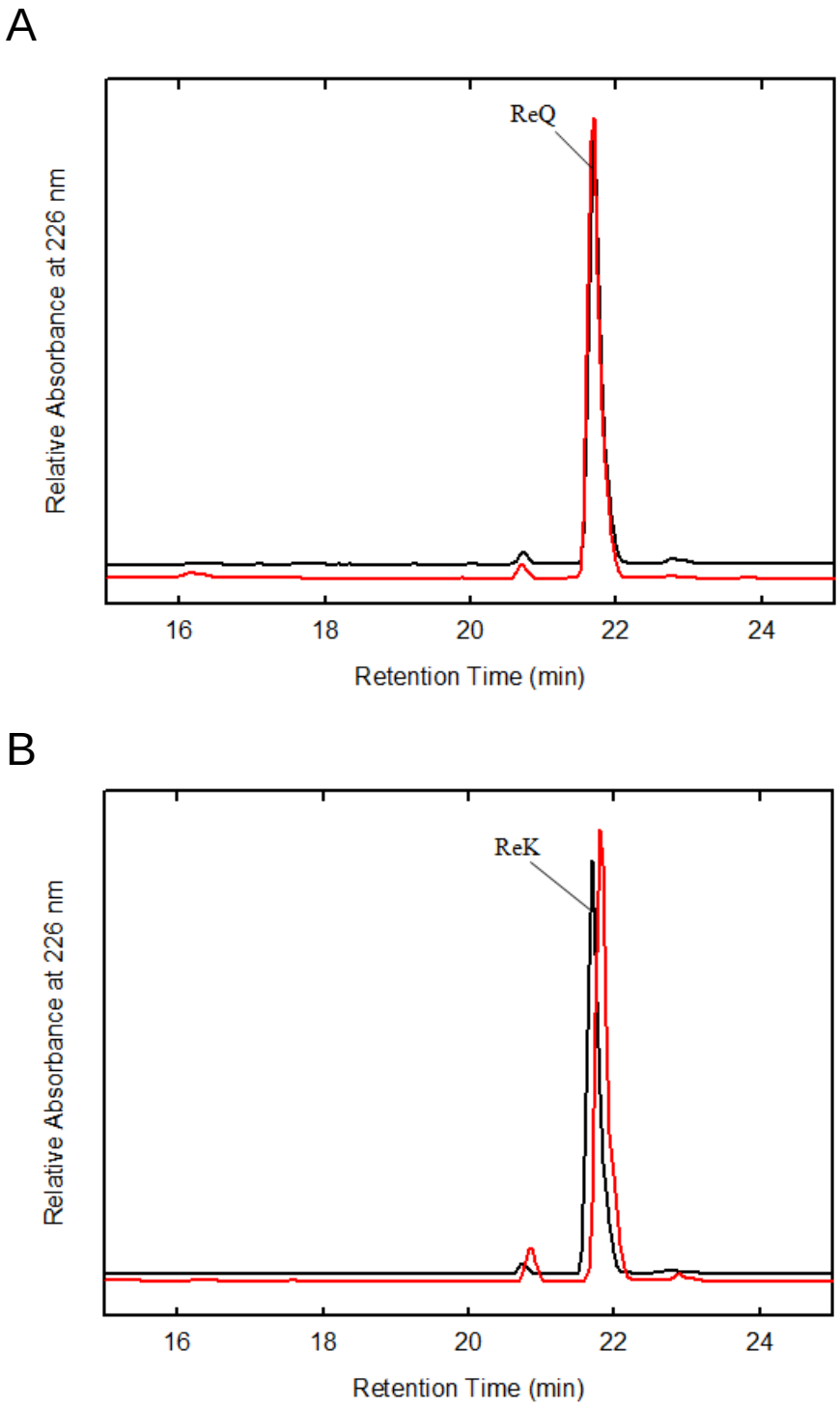
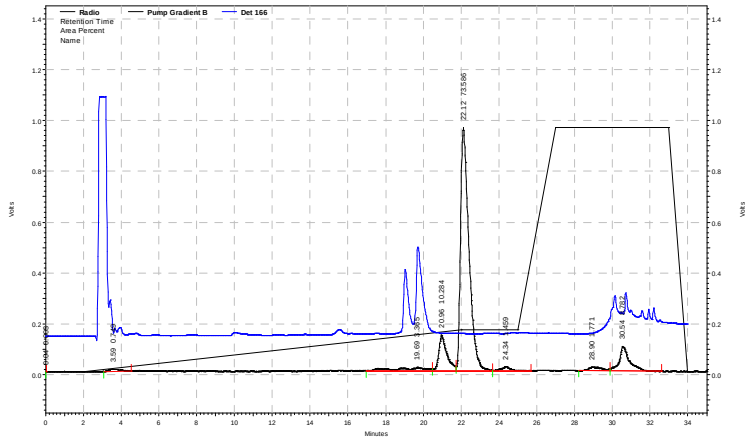


Figure 9. RP-HPLC chromatogram of the complex ReQ after 0 (black line) and 24 hours (red line) in 0.5% TFA and 0.1 M Na2HPO4, pH 7 **(A)**. In **(B)** it is shown the chromatogram of the complex ReK. The chromatogram does not show difference between the complex after 0 hours (black line) and 24 hours (red line) in 0.5% TFA and 0.1 M Na2HPO4, pH 7.

A



B

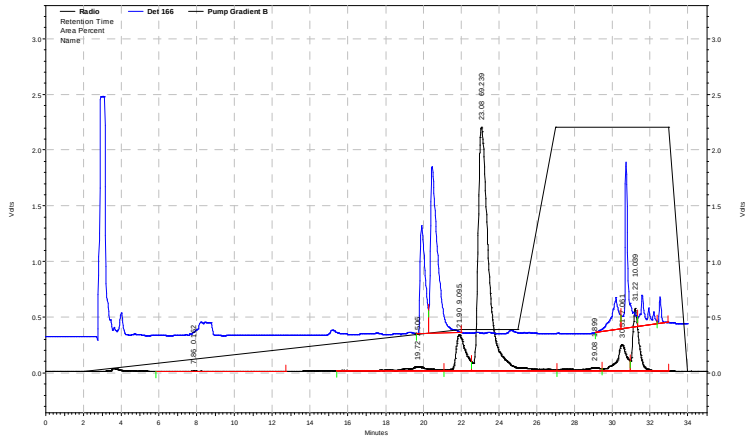


Figure 10. RP-HPLC chromatograms of ^{99m}Tc(N)(PNP)(Biotin-GKGC)] **(A)** and [^{99m}Tc(N)(PNP3)(Biotin-GQGC)] **(B)** obtained with a radio detector. In blue are shown the chromatograms of the anolougs complexes of $^{185-187}$Re (obtained with a UV detector).

The yield of synthesis of these complexes were evaluated at different conditions of temperature, time of reaction and biotinilated peptides concentration. Data are shown in Tables 2, 3 and 4.

Table 2**.** Effect of the biotinilated peptides concentration on the reaction yield. The reaction was performed at 80°C for 30 min.

PNP/ peptide	**RCY%**		
	Isomer “a”	Isomer “b”	Tot.
25.75	11.96	25.25	37.21
12.87	9.41	62.81	72.22
3.81	10.43	63.18	73.61
1.87	15.09	67.22	82.31

Table 3. Effect of temperature on the reaction yield. The reaction was performed for 30 minutes.

PNP/ peptide	**Temp (°C)**	**RCY%**		
		Isomer "a"	Isomer "b"	Tot.
12.87	80°C	9.41	62.81	72.22
12.87	100°C	9.97	56.81	66.78
3.81	80°C	10.43	63.18	73.61
3.81	100°C	11.04	71.68	82.72

Table 4. Effect of the time on the reaction yield. The reaction was performed at 100 °C

PNP/ peptide	**Time (min)**	**RCY%**		
		Isomer "a"	Isomer "b"	Tot.
5.57	30	9.61	69.47	79.08
5.57	60	9.59	70.04	79.63
3.81	30	10.28	73.59	83.87
3.81	60	10.33	73.92	84.25

As expected, the reaction yield shows a dependence from the concentration of the biotinilated peptides and from the temperature, while the time of incubation is not important to determine a high yield of these complexes, underlining the fast kinetic of the reaction. These results show that it is possible to obtain a reasonable yield of the ^{99m}Tc complexes at 100 °C after 30 minutes of incubation. In fact, in these conditions the yield is around 84%. Moreover, it is important to observe that complex formation is shifted toward one of the two different conformational isomers. In fact, after 30 minutes at 100 °C and with a ratio PNP/peptide of 3.81, the yield of the isomer a is 10.28%, while the yield of the isomer b is 73.59% . For the specific aim of this Project, we are not interested to the characterization of these isomers, but to the reactivity of the complexes toward TGase.

3.5. Site-specific derivatization of apoMb with ReQ and ReK. The TGase mediated reaction with apoMb and $^{185/187}$Re complexes was performed at 25 °C with a molar ratio apoMb/ReQ(K) of 1/10, while the weight ratio apoMb/TGase was 50/1. Different time of reaction were analyzed in order to determine the incubation time that yielded the higher amount of modified protein. In this paragraph, we will discuss separately the modification of apoMb at the level of Lys residues with ReQ and at the level of the Gln residues with ReK. Figure 11 shows the RP-HPLC chromatogram of the TGase mediated conjugation of apoMb at level of Lys residues with ReQ. The chromatogram shows the presence of three peaks with a retention time higher than that of the wild type protein. This result suggests the formation of different derivatives of apoMb with a longer retention time due to the presence of a hydrophobic moiety represented by PNP3 on the complex ReQ. The peaks collected from the RP-HPLC analysis were analyzed by MS in order to identify these three different species and the results are shown in Table 5. The mass values clearly demonstrated the incorporation of ReQ in apoMb at two locations. The first and the second peaks correspond both to a monoderivative of apoMb, but the different retention times suggest that the modification occurs at two different Lys residues. MS analysis of the peak at 19.52 min shows that this species corresponds to the biderivative of apoMb where both Lys residues are modified. The peaks collected from the RP-HPLC analysis were digested with trypsin and V8 protease and then analyzed by MS in order to identify the sites of conjugation of apoMb. The results are summarized in Table 6 and the MS spectra are shown in Figure 12. MS analysis indicated that the two monoderivatives of apoMb with ReQ collected at retention time 17.18 min and 18.18 min are modified at the level of Lys96 and Lys98, respectively. While both these residues are modified in the biderivative species.

The main result of these experiments is that the modification of apoMb with ReQ mediated by TGase gives as prevalent product the monoderivative of apoMb, while the biderivative is producted with a very low yield. By integration of the areas of the chromatographic peaks we calculated the % yield of the reaction products. For apoMb^{1ReQ} the yield is of 56.5% while for apoMb^{2ReQ} the yield is of 4.4%.

We also studied the TGase mediated conjugation of apoMb with ReK in order to modify the protein at the level of Gln residues. The RP-HPLC of the reaction mixture after 4 hours of incubation shows the presence of one main reaction product with a retention time higher than the wild type protein. MS analysis of the eluted protein material indicated that in analogy with the chromatogram of the apoMb-ReQ reaction mixture, this peak corresponds to the monoderivative of apoMb (Table 7).

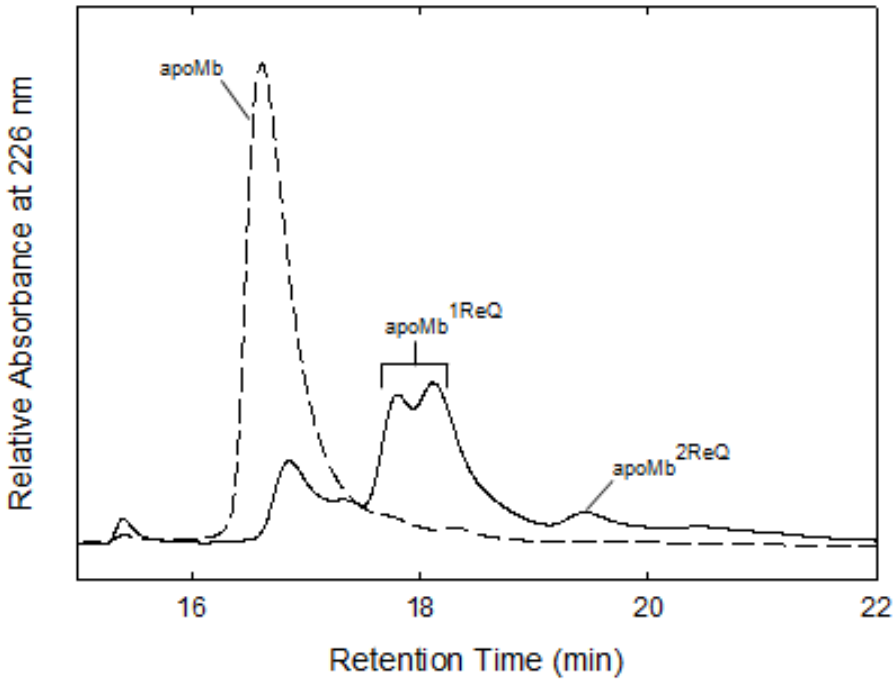


Figure 11. TGase mediated conjugation of apoMb with ReQ. RP-HPLC analysis of the reaction mixture after 0 hour (dotted line) and 4 hours (solid line) of incubation. Analyses were performed on a C18 Phenomenex column using a linear gradient of AcCN containing 0.085 % TFA from 40 to 50 % in 25 min. The 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 5. Molecular masses of apoMb conjugated with ReQ observed from the MS analysis of the RP-HPLC collected peaks.

RT (min)	**Molecula weight (Da)**		**species**
	observed	**calculated**	
16.87	16951.21	16951.50	apoMb
17.78	18205.99	18205.96	apoMb^{1ReQ}
18.18	18206.14	18205.96	apoMb^{1ReQ}
19.52	19461.99	19460.42	apoMb^{2ReQ}

Table 6. Molecular masses as observed from the MS analysis of the apoMb-ReQ Digestion mixture with trypsin and V8 protease.

RT (min)	Molecular weight (Da)		Fragment (modified residue)
	observed	calculated	
17.78	2712.48	2712.30	86-98 (K96)
	3163.92	3164.81	86-102 (K96)
18.18	1989.14	1988.94	97-102 (K98)
	1987.17	1988.94	97-102 (K98)
19.52	4417.60	4419.28	86-102 (K96-K98)
	2712.57	2712.30	86-98 (K96)

Collected peaks were digested with trypsin and V8 protease in order to identify the sites of the modification at the level of Gln residues by MS analysis (Table 8) (Figure 13). Moreover, a second very small peak is present at very high retention time and it corresponds to the apoMb dimer, formed by one or two Lys-Gln crosslinks. Indeed, as indicated by the MS measurements, the dimer species display two mass values that differ of 17 Da, the weight of the ammonia molecule that is released during the TGase mediated reaction. The monoderivative apoMb^{ReK} is derivatized at the level of Gln91, but a little amount of monoderivative where apoMb is modified at the level of Gln152 is also present. The integration of the chromatographic peaks of the reaction mixture gave an yield of 75% in monoderivative, while less than 5 % it corresponds to the formation of the apoMb dimer.

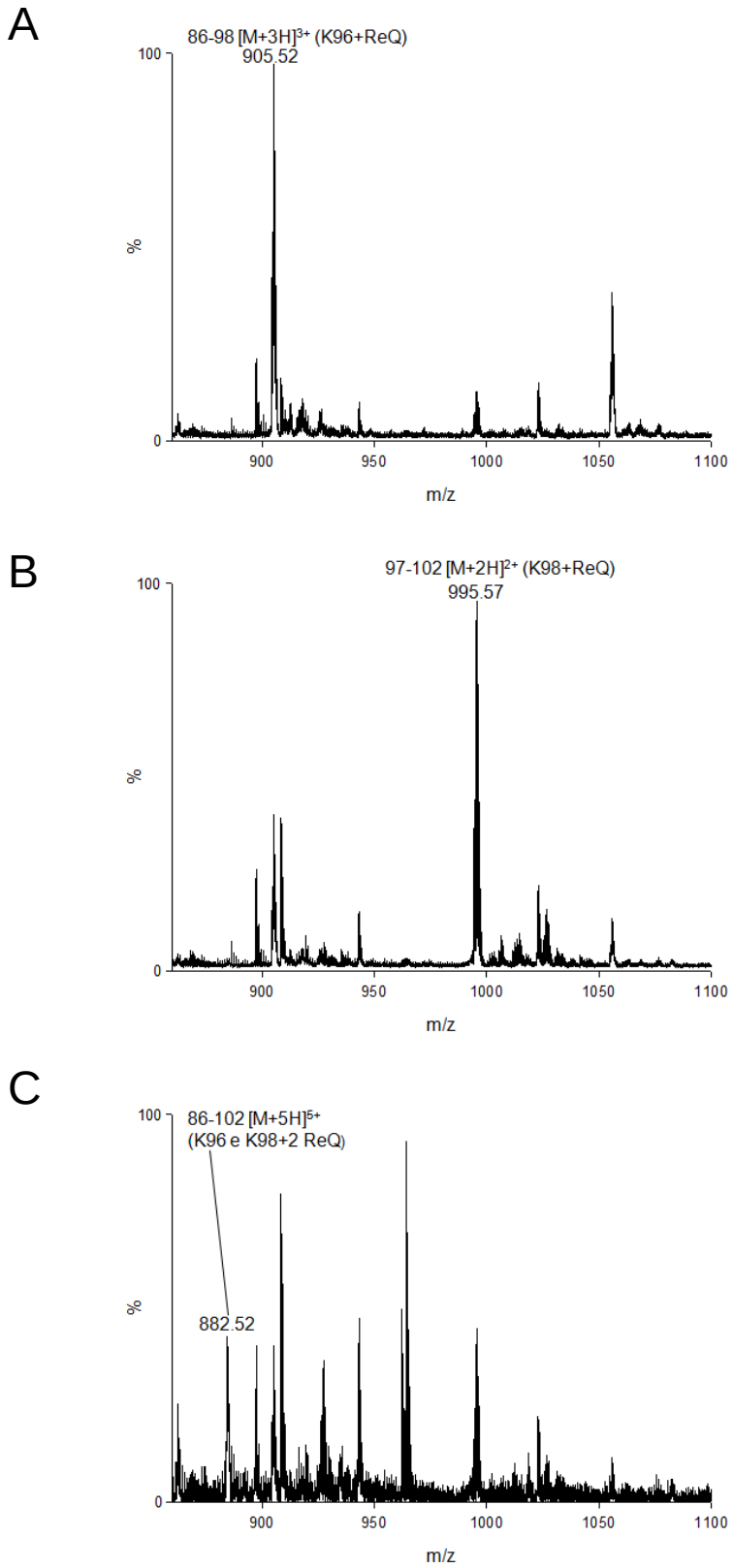


Figure 12. Mass spectra of apoMb-ReQ derivatives digested by trypsin and V8. **A:** digestion of apoMb^{1ReQ} collected at RT 17.78 min; **B:** digestion of apoMb^{1ReQ} collected at RT 18.18 min; **C:** digestion of apoMb^{2ReQ} collected at RT 19.52 mi

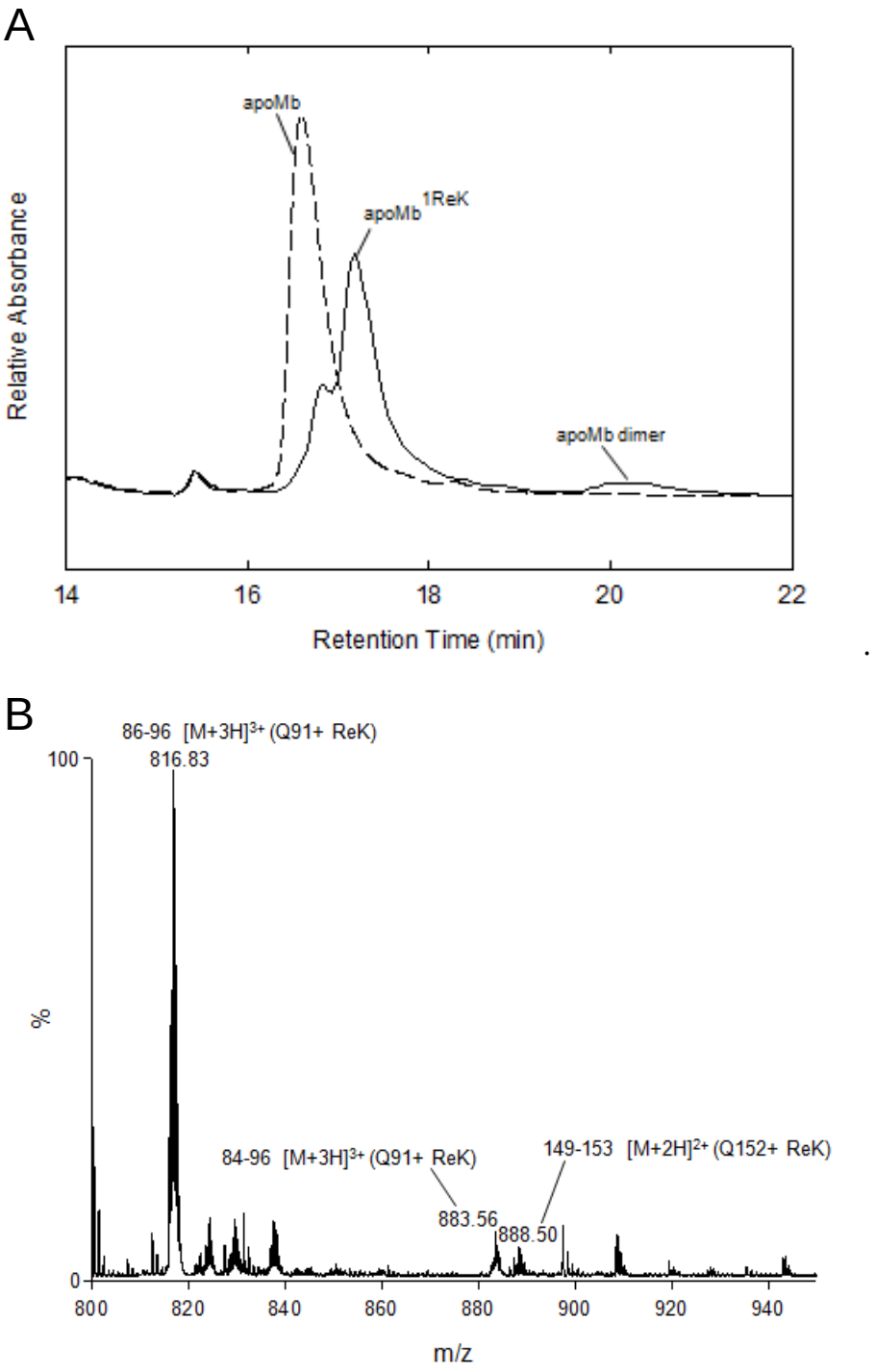


Figure 13. TGase mediated conjugation of apoMb with ReK **(A)**. RP-HPLC analysis of the reaction mixture after 0 hour (dotted line) and 4 hours (solid line) of incubation. Analyses were performed on a C18 Phenomenex column using a linear gradient of AcCN containing 0.085 % TFA from 40 to 50 % in 25 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**(B)**. Mass spectra of the collected chromatographic peak of apoMb^{1ReK} after digestion with trypsin and V8. This spectrum shows the modified peptides m/z value.

Table 7. Molecular masses of apoMb conjugated with ReK observed from the MS analysis of the RP-HPLC collected peaks.

RT (min)	Molecular weight (Da)		species
	observed	calculated	
17.17	18205.61	18205.98	apoMb^{1ReK}
	17523.75	17523.93	apoMb$^{Biotin-GKGC}$
	16950.90	16951.50	apoMb
20.14	33883.18	33886.00	apoMb dimer (-17 Da)
	33867.63	33869.00	apoMb dimer (-34 Da)

Table 8. Molecular masses as observed from the MS analysis of the apoMb-ReK digestion with trypsin and V8 protease.

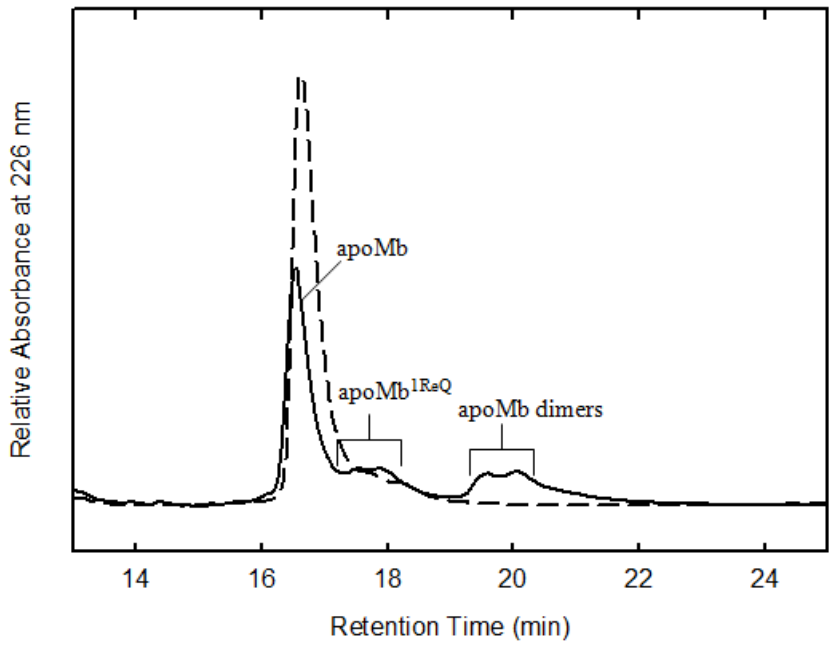
RT (min)	Molecular weight (Da)		Fragment (modified residue)
	observed	calculated	
16.82	2447.52	2447.19	86-96 (Q91+ReK)
	1765.12	1765.16	86-96(Q91+Biotin-GKGC)
17.17	2447.52	2447.19	86-96 (Q91+Biotin-GKGC)
	1765.12	1765.16	84-96 (Q91+ReK)
	1774.96	1774.68	149-153 (Q152+ReK)

3.6. Derivatization of apoMb with the ^{99m}Tc complexes. The results of the experiments of conjugation of apoMb with ReQ and ReK demonstrated that it is possible to obtain the site-specific conjugation of apoMb with $^{185-187}$Re complexes. We thus transfered this methodology to modify the apoMb with the analogous complexes of ^{99m}Tc. Since reaction conditions to modify apoMb with ^{99m}Tc are much more diluted due to the need of use minute amount of ^{99m}Tc, we repeated the site-specific conjugation of apoMb with ReQ and ReK in respect of

these new conditions. Initially, we used the same molar ratio protein-substrate that were used for preliminary experiments and the same temperature, but at a 50 fold diluition of the reactants. There was no derivatization but a loss of proteic material. Then, we increased the molar ratio protein/substrate until 1.2/2 and the temperature of reaction from 25 to 37 °C and carried out the reaction under stirring. In spite of these reaction conditions, results of the reaction were not satisfactory as shown by the RP-HPLC chromatograms (Figure 14).

In order to modify apoMb with ^{99m}Tc (see introduction) two different approaches were followed: pre-labelling and post-labelling. In pre-labelling approach firstly we form the ^{99m}Tc-BFCA chelate then, we proceed with its conjugation to the protein in a separate step. In the post-labelling the BFCA is conjugated to the protein and then the radiolabelling is performed by ligand exchange reaction with the ^{99m}Tc complex BFCA-protein conjugated. The pre-labelling approach was carried out with following two strategies: the carrier added reaction and the carrier free reaction. The carrier added reaction was performed by adding ReQ or ReK complexes to the solution containing the ^{99m}Tc complexes in order to increase the total concentration of the substrate and minimize the formation of the apoMb dimer. The carrier free approach it was performed by adding dirictely the ^{99m}Tc complexes to the apoMb solution. In both these approaches the reactions were carried out at 37 °C and for an incubation time of 20 hours but no modification was observed. The post-labelling approach was performed by adding the ^{99m}Tc complexes to the apoMb conjugated with Biotin-GQGC or Biotin-GKGC chelating agents and incubating for 20 hours at 37 °C. Only analysis it was possible to perform in order to identify the species producted during these reactions it was the RP-HPLC where the eluted species were detected by a radio detector. In both the reactions the presence of three chromatographic peaks is observed. The second and third chromatographic peaks are represented by the ^{99m}Tc complex and the radiolabelled apoMb respectively, while the first is probablily due to the degradation of the ^{99m}Tc complex (Figure 15). It was not possible to perform MS analysis due to the low concentration of the species to analyze and the radioactivity of the ^{99m}Tc complexes. The result of these experiments show a radiochemical yield (RCY) of 27.22% for the radiolabelling of apoMb conjugated with Biotin-GQGC, while in the case of apoMb conjugated with Biotin-GKGC the RCY is 23,97%.

A



B

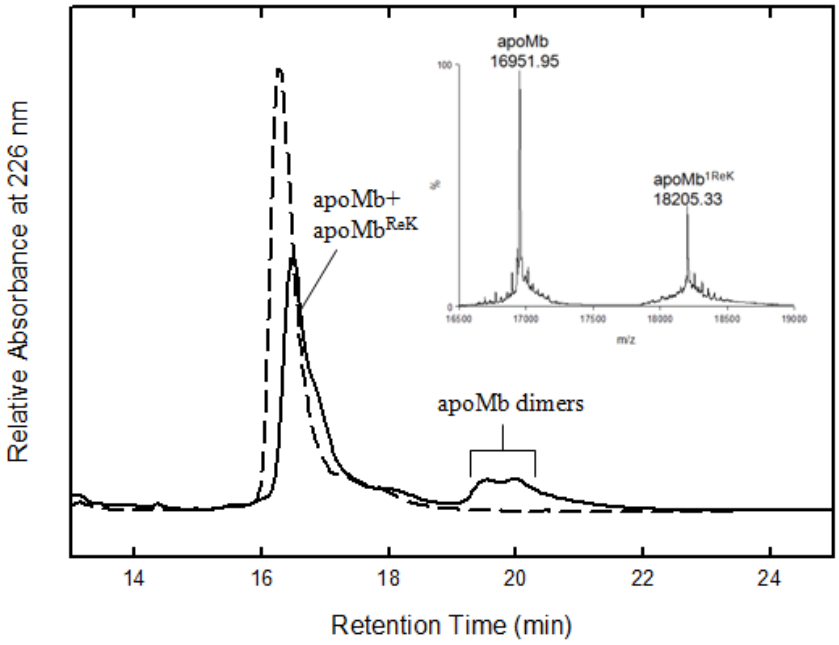
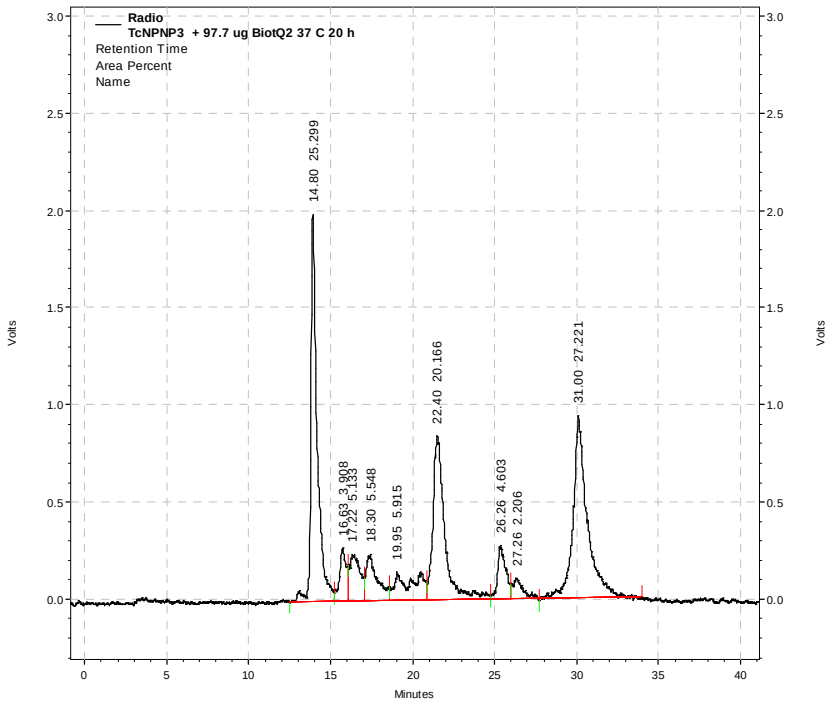


Figure 14. TGase mediated conjugation of apoMb with ReQ**(A)**. RP-HPLC analysis of the reaction after 0 hours (dotted line) and 2 hours (solid line) at 37 °C and with a molar ratio protein-substrate of 1.2:1.. With the new reaction conditions the yield of reaction is considerable decresed. The identities of the eluting protein material were confirmed by MS analysis. **(B)** Chromatographic peaks of apoMb and apoMb^{1ReK} are overlapped. RP-HPLC analysis of the reaction after 0 hours (dotted line) and 2 hours (solid line). To estimate the yield production of apoMb^{1ReK} we consider the relative intensity of deconvoluted MS analysis signal. About ¼ of the total amount of apoMb result to be modified.

A



B

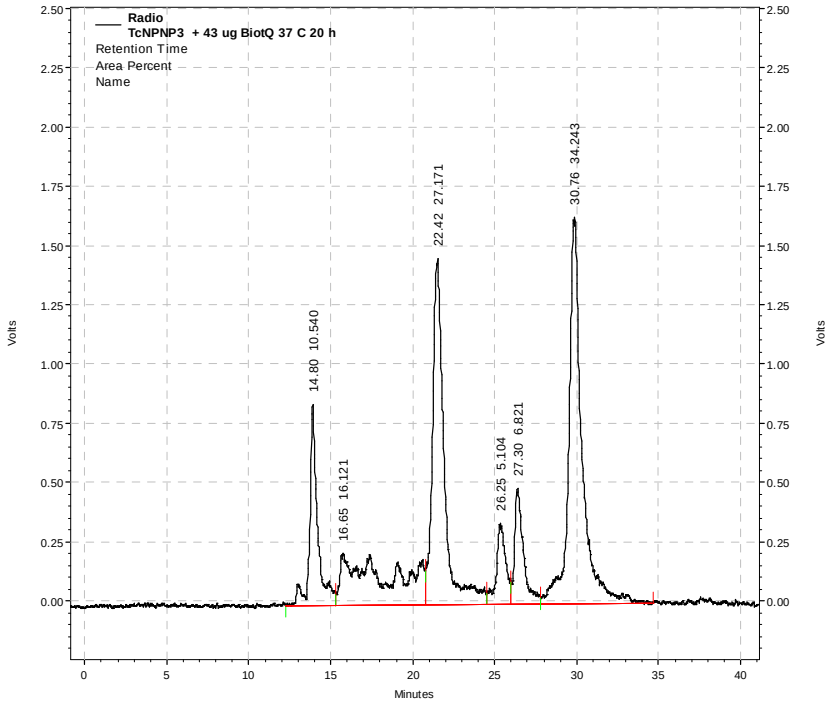


Figure 15. RP-HPLC of the post-labelling reaction of apoMb-ReK conjugated with ^{99m}Tc complex **(A)** and apoMb-ReQ conjugated with ^{99m}Tc complex **(B)** obtained with a radio detector.

3.7. Synthesis and characterisation of H2ATSM/C. Another type of BFCA that we synthesized it is represented by H2ATSM modified with 6-aminocaproic acid (H2ATSM/C). This modification allows to obtain a chelator with a free primary amino group that can act as acyl acceptor in TGase mediated reactions. The chelator H2ATMS was reacted with Fmoc-6-aminocaproic acid (H2ATSM/C) and the product of the reaction was deprotected and analyzed by RP-HPLC and ESI-MS. Chromatogram shows a high purity of the derivative since only one species is present in solution (Figure 16A). The collected peak was then analyzed by MS spectrometry in order to confirm the identity of the synthesized compound (Figure 16B). The mass spectrum of the RP-HPLC collected peak shows m/z signals that are correlated to the modified and unblocked chelator (375.23 m/z and 218.58 m/z) or to the Cu^{+} complex of H2ATSM/C (436.15 m/z). The signal 270.18 m/z seems to be generated by fragmentation of the chelator, since the same signal is observed in the MS/MS spectrum of the signal 375.23 m/z (data not shown), while the signal 323.05 m/z is correlated with Cu^{2+} complex of not modified H2ATSM (obtained by fragmentation during MS analysis) (Table 9).

Table 9. Observed and calculated molecular weight values of H2ATMS/C species purified by RP-HPLC.

Molecular weight (Da)		**species**
observed	**calculated**	
374.23	374.17	H2ATMS/M
435.15	435.07	H2ATMS/M+Cu
322.05	321.99	H2ATMS+Cu

Overall these results indicate that we obtained the H2ATSM/C with good level of purity. The high avidity of the compound toward Cu is demonstrated by MS analysis, since we observed m/z values corresponding to the complex ATSM/C+ Cu^{+} in spite of Cu was not added to the solution. It is possible that the chelator bound the metal from metallic parts of the mass spectrometer or HPLC instrument.

We also studied the stability of H2ATMS/C in phosphate buffer at pH 7. The kinetic of stability was followed by RP-HPLC (Figure 17). The chromatograms of the solution after 1 hour of incubation showed the presence of two different peaks that by MS analysis gave m/z

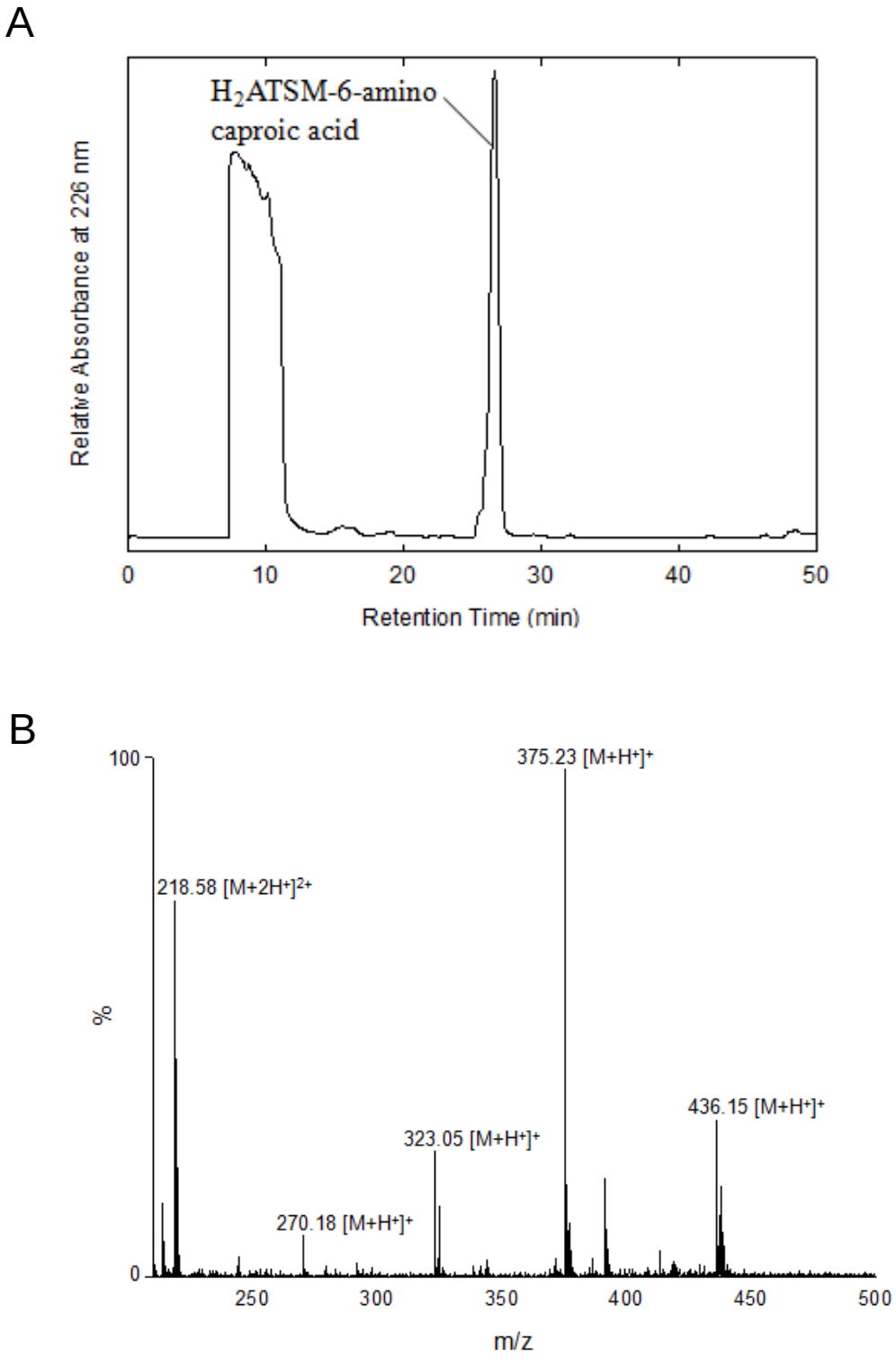


Figure 16. Analysis of H2ATMS-6-aminocaproic acid. In **(A)** it is shown the RP-HPLC chromatogram of the synthesized compound after to deprotetion of the amine group and its purification via precipitation. Loaded compound is dissolved in DMSO. In **(B)** it is shown the MS analysis that confirmed the presence of the modification.

values both correspond to H2ATMS/C. The presence of two different peaks for the same compound is likely due to the formation of isomers of the complex in aqueous solution.

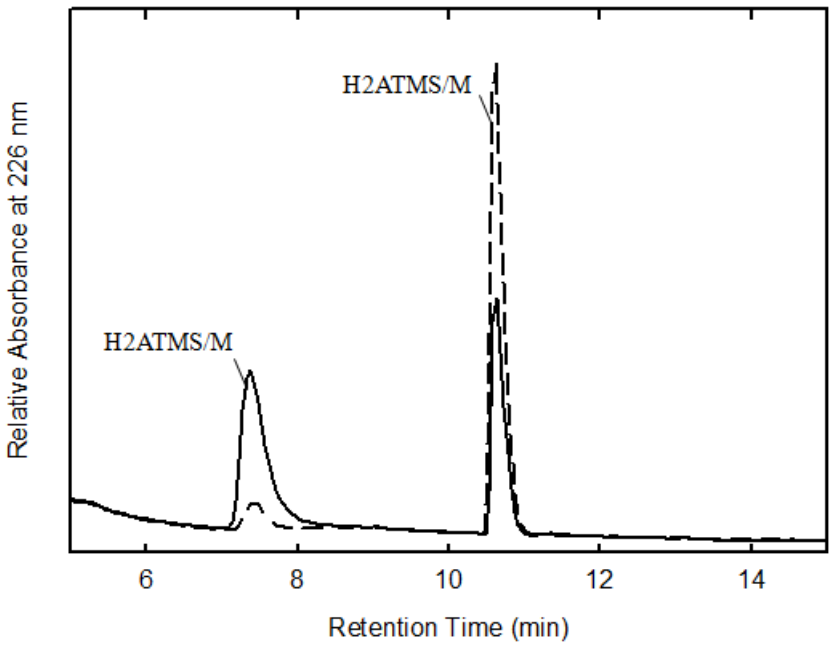


Figure 17. Stability test of H2ATMS/M. Phosphate buffer at pH 7 was added to the DMSO solution of the compound (final concentration was 0.1 mg/ml). The kinetic of stability was followed by RP-HPLC. The RP-HPLC chromatograms after o hours (dotted line) and 1 hour of incubation (solid line) are shown.

3.8. Site-specific modification of apoMb with H2ATSM/C mediated by TGase. In order to modify apoMb at the level of Gln residues with H2ATSM/C, we carried out the TGase mediated reaction using the Cu chelator H2ATSM/C which contains a free amino group that can act as acyl acceptor. The TGase mediated reaction of apoMb with H2ATSM/C was performed at 25 °C with a molar ratio protein/substrate of 1/10, while the weight ratio apoMb/TGase was 20/1. Different time of reaction were analyzed. Since the RP-HPLC chromatograms did not show a nice separation between not reacted apoMb and its modified form, the yield of reaction was extimated on the basis of MS measurements of the eluted peak (Figure 18) .The MS spectrum of the modified apoMb shows mass values corresponding to the native protein and to its derivative with H2ATSM/C both in complexed form with Cu and not (Table 10). Some species originated from the fragmentation of the H2ATSM/C moiety in the mass spectrometer are also present. The yield was calculated on the basis of the relative intensities and it was found to be around the 50%.

Table 10. Molecular masses of apoMb conjugated with H2ATSM/C observed from the MS analysis of the RP-HPLC collected peaks.

Molecula weight (Da)		specie
observed	**calculated**	
16950.93	16951.50	apoMb
17309.63	17308.67	apoMb+ H2ATSM/C
17371.16	17370.67	apoMb+ H2ATSM/C+Cu

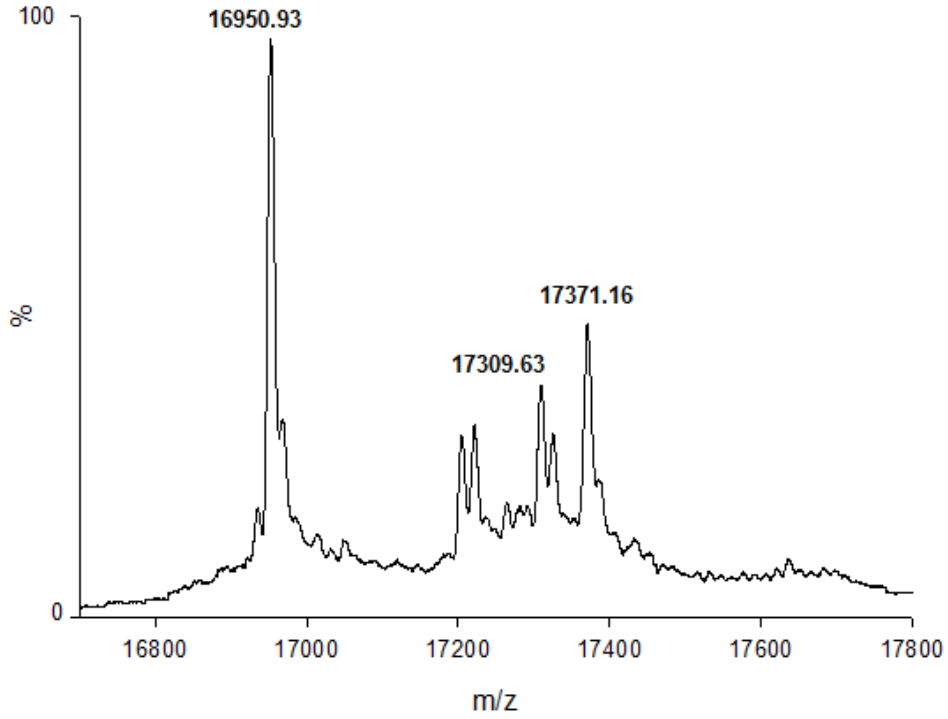


Figure 18. Deconvoluted MS spectra for the RP-HPLC fraction corresponding to the reaction time 5 h of the TGase mediated conjugation of apoMb with H2ATSM/C.

DISCUSSION

The modification of proteins, including antibodies, is nowadays of great interest in biochemical and biomedical research, since modified proteins can display improved physicochemical and functional properties. In particular, labelled proteins with fluorescent probes or radionuclides can be used for the imaging of molecular processes, as well as for diagnostic and therapeutic applications. Today, the modification of proteins is predominantly achieved by using chemical methods, but these methods are often difficult to control and usually can attack different amino acid side chains of a protein. Thus, the risk of

heterogeneous modification and over-labeling is inherent, leading to deleterious effects on the native structure and biological activity of the protein. Therefore, the development of alternative approaches for the controlled modification of proteins is of significant and timely interest.

In this Project, we evaluated TGase-mediated strategies for the radiolabeling of proteins. These procedures require the design and synthesis of suitable and novel metal-chelating agents. To this aim, the reagents previously used for technetium-99m labeling *(17)* were newly synthesized with the view that these reagents should contain a Lys or Gln amino acid to be used in the TGase-mediated reaction. The TGase-mediated derivatization of apoMb using as substrate the $^{185/187}$Re complexes (ReQ and ReK, the Gln and Lys containing complexes respectively) resulted in a very site-specific modification of the protein both at the level of Gln residues (by using ReK) and at level of Lys residues (by using ReQ). The specificity of the modification is very high if we consider that apoMb contains 6 Gln residues and 19 Lys residues but only Gln91 (Gln152 is a secondary site of modification) and Lys96 and Lys98 are modified by TGase. These results are in agreement with the observation that TGase modifies preferentially residues that are located at the level of flexible/disordered regions of proteins *(34)*. Indeed, Gln91, Lys96 and Lys98 are located at the level of the region of helix F, that results more flexible in apoMb as demonstrated by limited proteolysis experiments and NMR measurements *(24-27)*. Moreover, the TGase reaction gave good yields of the ReQ- and ReK-apoMb derivatives (56.5 and 75%, respectively) indicating that it is a promising approach for the radiolabelling of proteins.

This methodology was then transferred to the labeling of apoMb with the ^{99m}Tc containing complexes. Radiolabeling was performed via prelabeled BFCA with subsequent covalent binding to apoMb mediated by TGase (pre-labelling approach) or first binding of BFCA to the protein and then radiolabeling of the protein bioconjugate (post-labelling approach). The post-labelling approach gave the best results even if the yields of radiolabelled protein were low while the pre-labelling approach did not yield any derivative. . These results can be explained with the reaction conditions that were use for the labelling with ^{99m}Tc. Indeed, in the pre-labelling approach the substrate was not in excess over the protein with the consequence that the main product of the TGase mediated reaction was the apoMb dimer. In the case of the post-radiolabelling approach, binding of ^{99m}Tc to the BFCA-apoMb derivative was performed at a temperature much lower that required for an high efficient metal binding. Indeed, the formation of the asymmetrical nitride complexes of ^{99m}Tc gives optimal yields by using a temperature of 80-100°C *(16, 17, 31)*. Proteins generally can not be exposed to these temperature conditions since

they can cause their irreversible denaturation or aggregation. We thus decided to perform metal binding at 37°C but we obtained an yield of ^{99m}Tc-labelled apoMb too low to propose the application of this method for diagnostic purposes.

We also studied another approach of protein radiolabelling *via* TGase using a BFCA that can coordinate ^{64}Cu. This radionuclide has indeed interesting diagnostic and therapeutic applications (REF). The BFCA was designed to contain a bis(thiosemicarbazone) moiety that has an high affinity towards ^{64}Cu and it can bind this radionuclide at room temperature thus allowing to perform a post-labelling approach without heating. The avidity of this chelating system toward Cu is demonstrated by the results of the MS analyses since we identified a Cu-bound form of the BFCA that originates from the binding of environmental Cu. The bis(thiosemicarbazone) was modified with 6-aminocaproic acid (H2ATSM/C) in order to introduce an amino donor for the TGase-mediated reaction. Importantly, a derivative of apoMb modified with H2ATSM/C was produced *via* TGase with good yield (~50%). We feel that this metal-chelating system can be usefully employed for the site-specific labelling of proteins to be used as imaging and therapeutic agents. Experiments are now in progress in order to transfer this approach at tracer level with ^{64}Cu and to extend its application to the preparation of radiolabeled proteins useful for diagnostic or therapeutic purposes .

References

1. Gullberg, G.T., Reutter, B.W., Sitek, A., Maltz, J.S., Budinger, T. (2010) Dynamic single photon emission computed tomography—basic principles and cardiac applications. *Phys. Med. Biol*. 55: R111–R191.

2. Lammertsma, A.A. (2001) PET/SPECT: functional imaging beyond flow. *Vision Research* 41: 1277–1281.

3. Alavi, A., Reivich, M. (2002) Guest editorial: the conception of FDG-PET imaging. *Semin. Nucl. Med.* 32: 2-5.

4. Cherry, S. (2006) Multimodality In Vivo Imaging Systems: Twice the Power or Double the Trouble? *Annu. Rev. Biomed. Eng.* 8:35-62.

5. Kemp, R., Epstein, F., Catana, C., Tsui, B., Ritman, E. (2010) Small-Animal Molecular Imaging Methods. *J. Nucl. Med.*51:18S-32S.

6. Meares, C.F. and Goodwin, D.A. (1984) Linking radiometals to proteins with bifunctional chelating agents. *J. Protein Chem.* 3, 215-228.

7. Brechbiel, M.W. (2008) Bifunctional chelates for metal nuclides. *Q. J. Nucl. Med. Mol. Imaging* 52, 166-173.

8. Huang, L., Gainkam, L.O., Caveliers, V., Vanhove, C., Keyaerts, M., De Baetselier, P., Bossuyt, A, Revets, H., Lahoutte, T. (2008) SPECT imaging with 99mTc-labeled EGFR-specific nanobody for in vivo monitoring of EGFR expression. *Mol Imaging Biol*. 10:167–175.

9. Jeger, S., Zimmermann, K., Blanc, A., Grunberg, J., Honer, M., Hunziker, P., Struthers, H., Schibli, R. (2010) Site-specific and Stoichiometric Modification of Antibodies by Bacterial Transglutaminase. *Angew. Chem.* 49: 1-4.

10. Delmon-Moingeon, L. I.; Mahmood, A.; Davison, A.; Jones, A. G. (1991) Strategies for labeling monoclonal antibodies and antibody-like molecules with technetium-99m. *Nucl Biol Med*. 18: 47-59.

11. Rhodes, B. A. (1991) Direct Labeling of Proteins with Tc-99m. *Nucl Med Biol*. 18: 667-676.

12. Eckelman, W. C.; Steigman, J. (1991) Direct Labeling with Tc-99m. *Nucl Med Biol*. 18: 3-7.

13. Griffiths, G. L.; Goldenberg, D. M.; Jones, A. L.; Hansen, H. J. (1992) Radiolabeling of Monoclonal-Antibodies and Fragments with Technetium and Rhenium. *Bioconjug Chem*. 3: 91-99.

14. Zamora, P. O.; Rhodes, B. A. (1992) Imidazoles as Well as Thiolates in Proteins Bind Tc-99m. *Bioconjug Chem*. 3: 493-498.

15. Hueting, R., Christlieb, M., Dilworth, J.R., Garcìa Garayoa, E., Gouverneur, V., Jones, M.W., Maes, V., Schibli, R., Sun, X., Tourwè, D.A. (2010) Bis(thiosemicarbazones) as bifunctional chelators for the room temperature 64-copper labeling of peptides. *Dalton Trans.* 39: 3620-3632.

16. Bolzati, C., Refosco, F., Marchiani, A., Ruzza, P. (2010) ^{99m}Tc-Radiolabelled Peptides for Tumour Imaging: Present and Future. *Cur.t Med. Chem.* 17: 2656-2683.

17. Bolzati, C., Caporale, A., Agostini, S., Carta, D., Ceccato, M., Refosco, F., Tisato, F., Schievano, E., Bandoli, G. (2006) Avidin–biotin system: a small library of cysteine biotinylated derivatives designed for the $[^{99m}Tc(N)(PNP)]^{2+}$ metal fragment. *Nucl. Med. And Biol.* 34: 511-522.

18. Bayly, S.R, King, R.C., Honess, D.J., Barnard, P.J., Betts, H.M., Holland, J. P., Hueting, R., Bonnitcha, P.D., Dilworth, J.R., Aigbirhio, F.I. and. Christlieb, M. (2008) *J. Nucl. Med.* 49:, 1862–1868.

19. Holland, J.P., Aigbirhio, F.I., Betts, H.M., Bonnitcha, P.D., Burke, P., Christlieb, M., Churchill, G.C., Cowley, A.R., Dilworth, J.R., Donnelly, P.S., Green, J.C., Peach, J.M., Vasudevan S.R. and Warren, J.E. (2007) *Inorg. Chem.* 46: 465–485.

20. J. L. Lacy, in ‘Miniaturized 62Zn/62Cu generator for high concentration clinical delivery of 62Cu kit formulation for the facile preparation of radiolabeled Cu-bis(thiosemicarbazone) compounds’, Application: 2005, WO 2005084168 A2 20050915.

21. Boschi, A.; Duatti, A.; Uccelli, L. (2005) Development of technetium-99m and rhenium-188 radiopharmaceuticals containing a terminal metal-nitrido multiple bond for diagnosis and therapy. *Top Curr Chem*. 252: 85-115.

22. 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.

23. 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.ù

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

25. 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.

26. 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.

27. 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.

28. 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.

29. Satwekar, A.A. Site-specific modification of proteins by transglutaminase (2011) Ph.D. dissertation.

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

31. Bolzati, C., Boschi, A., Uccelli, L., Malagò, E., Bandoli, G., Tisato, F., Refosco, F., Pasqualini, R., Duatti, A. (1999) Synthesis of a Novel Class of Trigonal Bipyramidal Nitrido Tc(V) Complexes with Phosphino –Thiol Ligands. Crystal structure of [^{99g}Tc(N)(L^1)2 [L= 2-(Diphenylphosphino)ethanethiolato] and [^{99g}Tc(N)(L^5)2] [L^5= 2-(Ditolylphosphino)propanethiolato]. *Inorg. Chem.* 38: 4473-4479.

32. Holland, J.P., Aigbirhio, F.I., Betts, H.M., Bonnitcha, P.D., Burke, P., Christlieb, M., Churchill, G.C., Cowley, A.R., Dilworth, J.R., Donnelly, P.S., Green, J.C., Peach, J.M., Vasudevan, S.R., Warren, J.E. (2007) Functionalized Bis(thiosemicarbazonato)

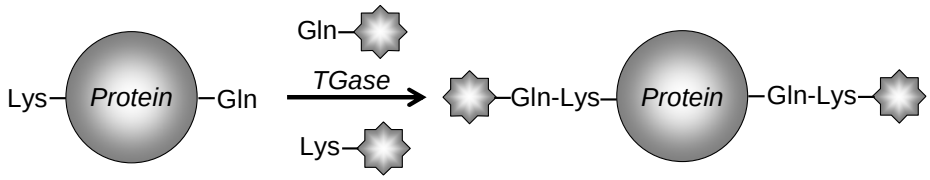
Complexes of Zinc and Copper: Synthetic Platforms Toward Site-Specific Radiopharmaceuticals. *Inorg. Chem.* 46: 465-485.

33. Fields, G. B., Angeletti, R.H.,. Carr, S.A ., Smith, A. J., Stults, J.T., Williams, L.C., and Young, J.D. (1994) Techniques in Protein Chemistry V. *Edited by John W. Crabb.* 501-507.

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

IV. GENERAL DISCUSSION AND OUTLOOK

The main aim of this study was to deduce molecular features of the TGase-mediated reactions using well characterised protein substrates, such as α-lactalbumin, avidin and apomyoglobin (apoMb). At the same time, we were willing to conduct initial experiments to develop suitable procedures for covalently linking to proteins a few peptide moieties able to bind metal ions, including radionuclides. 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 shown below (Fig. 1) illustrates the permutations of substrates of the TGase reactions herewith investigated. As an amino-donor we used dansyl-cadaverine (DC) and as acyl-donor N-carbobenzoxy-glutaminyl-glycine (ZQG). Therefore, 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.



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.

TGases attack proteins at flexible sites. The results presented in this Thesis and previously by our laboratory (*1–3*) have convincingly shown with a number of TGase reactions, performed under different experimental conditions and various protein substrates, that local flexibility or even local unfolding dictates the site-specific modification of a Gln residue in a protein, while a less stringent specificity was observed for Lys residues. In particular, it is here shown that the apo-form of bovine α-lactalbumin in its partly folded or molten globule state at neutral pH reacts primarily at Gln39 with TGase in the presence of DC as the amino-donor, while none of the six Gln residues of the protein reacts when the native, folded and calcium-loaded protein is used as substrate for TGase. Of interest, TGase selectively incorporates DC at the level of Gln residues encompassed by the chain segment which was shown, by means of spectroscopic measurements, to be disordered in the molten globule state of α-lactalbumin (*4*, and references cited therein). Moreover, the same chain region acting as substrate for TGase is also selectively hydrolyzed by several proteases, thus demonstrating that both TGase and proteases act on the flexible/disordered protein regions of α-lactalbumin.

Also the 153-residue chain of apoMb reacts selectively with TGase at the level of a Gln91, a residue encompassed by the chain segment which has been shown to be disordered by NMR and proteolysis experiments (*5*). Instead, in the case of the well-structured and rigid avidin molecule, none of the Gln residues of the protein is reactive towards TGase. That chain flexibility is dictating the site-specific TGase reaction(s) is also demonstrated by the fact that native globular proteins usually are very poor substrates for TGases, while a variety of natively or intrinsically disordered proteins are substrates for TGase *in vivo*, including synuclein, tau protein and huntingtin. With these proteins, as expected, the TGase-mediated reactions are not site-specific and several Gln residues are modified. Since these disordered proteins are involved in severe diseases as those of Alzheimer, Parkinson and Huntington, a role of TGase in these diseases has been proposed (see *3*, for references).

The TGase-mediated reactions involve the initial step of formation of the reactive thioester intermediate at the level of the Cys residue at the enzyme's active site (see Introduction). Considering that Cys64 in microbial TGase is located in a deep cleft of the protein fold (see Introduction), it can be anticipated that the most critical step in the TGase-mediated reaction(s) is the difficult formation of the covalent reactive intermediate involving on one side the buried Cys residue and on the other a protein-bound Gln residue. This intermediate is difficult to be formed with a Gln-protein, while it is formed much more easily with flexible and structureless peptides, including the small ZQG substrate. It can be anticipated that the reactive intermediate can react with a variety of nucleophiles, besides the ε-amino group of a protein-bound Lys-residue. Indeed, it has been found that a large variety of small nucleophiles can react with the 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 thioester.

The requirement of a flexible polypeptide chain for the TGase's reaction is consistent with the fact that, on one side, it can explain the often observed stringent specificity for protein bound Gln residues as acyl-donors and, on the other, that in general there is only a moderate specificity for protein-bound Lys residues as amino-donors (see the results obtained with α-lactalbumin). This can be explained 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, here we have shown that even the reactivity of a Lys residue can be very selective, as shown by the specific modification of a single Lys residue when the native, calcium-loaded α-lactalbumin was used as substrate. The modified Lys is located at the C-terminal end of the protein and thus it is expected to be flexible, since in general the N- and C-terminal segments of globular proteins are significantly more flexible than the protein core.

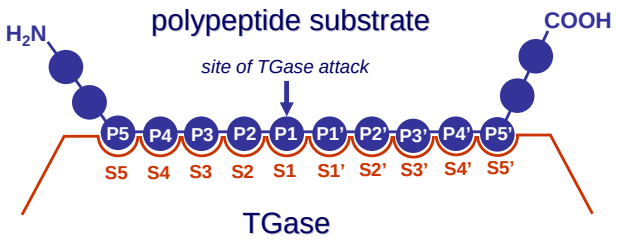


Figure 2. Schematic representation of the binding of a polypeptide substrate at the active site of TGase. A 10-12 residue segment of a polypeptide chain interacts with its side-chain residues (P) at a series of subsites (S) of TGase. The interaction of the substrate at the TGase's active requires a specific 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 (*6*).

TGase acts similarly to a protease. The sites of Gln modification by TGase occur at regions prone to suffer also limited proteolysis phenomena, as demonstrated here for α-lactalbumin and apoMb (see also ref. *3*). This fact prompts us to propose that a mechanism of local unfolding of the site of enzymatic attack is a 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 (*6*) 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 Fig. 2, 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. 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.

Oulook. 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. 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. We have here performed initial experiments for covalently binding metal-chelating moieties to proteins. We have shown that the reaction can be specific, but additional experiments are required to provide a *proof-of-principle* of the technique. Summing up, the results here described using model proteins as TGase substrates are interesting for an understanding of the enzymatic reaction and are also of practical utility. 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 proteins in general, including protein drugs (*7*).

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.M. (2009) The site-specific TGase-mediated PEGylation of proteins occurs at flexible sites. In *PEGylated Protein Drugs: Basic Science and Clinical Applications* (Veronese, F.M., Ed.), pp. 89–112, Birkäuser Verlag, Switzerland.

4. 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.

5. 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.

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

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