Université de Montréal

Illuminating biomolecules: shedding light on the utility of labeling using transglutaminases

par Natalie M. Rachel

Département de chimie
Faculté des arts et sciences

Thèse présentée à la Faculté des études supérieurs et postdoctorales
en vue de l'obtention du grade de philosophiæ doctor (Ph.D.)
en chimie

Avril, 2017

Résumé

Le développement des technologies de recombinaison en biologie moléculaire fut un point tournant pour les sciences biologiques. Depuis cette découverte, diverses avancées extraordinaires qui ont un impact direct sur les humains ont pu être accomplies dans les domaines de recherches qui découlent de cette technologie. L'étude des enzymes produites en utilisant cette technique est le fondement de leurs applications éventuellement accessibles. À cet effet, la biocatalyse est un sous-domaine de l'enzymologie en développement continuel. Les chimistes et ingénieurs utilisent les composantes de systèmes biologiques ou même des systèmes complets afin de complémenter ou remplacer des méthodologies existantes. Cette thèse étudie la famille d'enzymes transglutaminase (TGase) comme biocatalyseur afin d'explorer et d'étendre l'ubiquité et les innovations rendues possibles grâce aux enzymes.

Les TGases sont des enzymes versatiles. Leur homologue bactérien, la transglutaminase bactérienne (MTG), est couramment utilisé à l'échelle industrielle pour la transformation alimentaire. Depuis une dizaines d'années, de nombreux efforts ont été faits afin de trouver de nouvelles applications des TGases. En premier lieu, une revue des accomplissements, progrès et défis reliés au développement des TGases sera décrite.

Les TGases sont intrinsèquement des catalyseurs de la formation de lien isopeptidiques entre une glutamine et une lysine. Par ce fait, elles ont été initialement testées dans cette thèse pour la synthèse de peptides. Une forme de l'enzyme TGase de mammifères fut en mesure de générer les composés dipeptidiques Gly-Xaa et D-Ala-Gly avec une faible conversion.

La MTG possède plusieurs caractéristiques qui font de cette enzyme un candidat intéressant pour le développement de biotechnologies. Elle est stable, non dépendante d'un cofacteur et connait peu de compétition pour sa réaction catalytique inverse. La majeure partie de cette thèse porte exclusivement sur l'utilisation de la MTG. Nous avons développé et caractérisé une réaction chimio-enzymatique en un seul pot pour la conjugaison de peptides et protéines. La présence de glutathion en quantité suffisante permet de contourner l'incompatibilité de la MTG avec le cuivre et ouvre la porte à l'utilisation de la réaction de cycloaddition entre un alcyne et un azoture catalysée par le cuivre, afin d'effectuer le marquage

fluorescent de protéines. L'utilisation d'autres méthodes de chimie « click » sans métaux fut aussi étudiée afin d'incorporer divers substrats protéiques. Le marquage de protéines avec la MTG fut investigué de manière combinatoire. Précisément, la ligation de Staudinger, la cycloaddition azoture-alcyne promue par la tension de cycle, ainsi que la ligation de tetrazine (TL) ont été testées. Différents niveaux de conversion ont été atteints, le plus prometteur étant celui obtenu avec la TL.

Une étude par cristallographie a été effectuée afin d'élucider comment les substrats contenant une glutamine interagissent avec la MTG. Une méthode de purification alternative de la MTG a été développée afin d'atteindre ce but. Une discussion sur les stratégies et défis est présentée.

Finalement, la conjugaison entre un système contenant la MTG comme biocatalyseur de marquage, le domaine B1 de la protéine G (GB1) comme substrat et d'un fluorophore contenant une amine comme sonde fut étudié. Comme deux des constituants de ce système sont des protéines, l'ingénierie d'enzyme peut être entreprise afin d'améliorer leurs propriétés. Une banque de 24 variantes de GB1 fut construite grâce à une approche semi-rationnelle afin d'investiguer quels facteurs sont déterminants pour la sélectivité de la MTG envers la glutamine. Chaque variante étudiée comportait une seule glutamine à une position variable afin d'évaluer l'impact des éléments de structure secondaire où se retrouve la glutamine. L'efficacité pour le marquage a pu être améliorée d'au moins un ordre de grandeur pour huit des substitutions étudiées. Comme chacune des structures secondaires fut marquée, il fut démontré que la MTG n'en préfère pas une en particulier. De plus, la réactivité de la MTG envers la variante I6Q-GB1 fut augmentée en créant des mutations dans son site actif. Ces résultats permettent de comprendre d'avantage la sélectivité de la MTG envers la glutamine, tout en démontrant le potentiel de cette enzyme à être modifiée afin d'être améliorée.

Mots-clés : Biocatalyse, bioconjugation, chimie des clics, ingénierie enzymatique, marquage des protéines, transglutaminase

Abstract

The development of recombinant molecular biology technologies was a turning point for the biological sciences, which has since evolved into dozens upon dozens of different subfields and contributed to extraordinary advances for humans. At the core of many of these advances are the enzymes produced by these techniques, with efforts to understand their form and function laying the groundwork for their application. One of these continuously advancing subfields rooted in enzymology is biocatalysis, in which chemists and engineers embrace biological components and systems to complement, or even replace, existing methodologies. This thesis seeks to further contribute to the advancement and ubiquity of enzymes to be incorporated into future innovations. To this end, transglutaminase (TGase) is the biocatalyst selected for study.

TGases are versatile enzymes, with the bacterial homolog, microbial transglutaminase (MTG) being readily used in industrial processes for years, particularly for food processing. An abundance of efforts seeking to apply TGases to other processes have been made within the last decade. We commence by reviewing the accomplishments, progress, and challenges to developing TGase towards new goals.

TGase naturally catalyzes the formation of isopeptide bonds utilizing a glutamine and lysine substrates, and one of its first unconventional applications we investigated was for peptide synthesis. We determined the ability and specificity of one form of TGase for various amino acid-derived substrates, observing the formation of Gly-Xaa and D-Ala-Gly dipeptide products, albeit at a low conversion.

MTG exhibits several characteristics that make it an appealing candidate for biotechnological development, such as its independence from a cofactor, little competition for its reverse catalytic reaction, and increased stability relative to mammalian TGases. Therefore, the remainder of this thesis pertains exclusively to MTG. We developed and extensively characterized a one-pot chemoenzymatic peptide and protein conjugation scheme. The presence of sufficient glutathione circumvents the incompatibility of the copper-catalyzed azide-alkyne cycloaddition with MTG owing to the presence of copper. We ultimately utilized this chemoenzymatic conjugation scheme for fluorescent protein labeling.

We continue to expand upon combinatorial methods to undertake protein labeling by investigating to what extent metal-free click chemistries can be utilized in combination with MTG. Specifically, the Staudinger ligation, strain-promoted azide-alkyne cycloaddition, and tetrazine ligation (TL) were assayed on protein substrates to reveal varying levels of effective conjugation, with the TL being the most promising of the three.

The details surrounding the manner in which MTG interacts with its glutamine-containing substrate remains unclear. To address this knowledge gap, we sought to pursue crystallography studies, which required the development a modified purification strategy. We discuss the strategies we investigated and the challenges surrounding such efforts.

Finally, we present a conjugation system consisting of MTG as the labeling biocatalyst, the B1 domain of Protein G (GB1) as a substrate, and a small-molecule amine belonging to a recently developed class of fluorophores as a probe. As two components of this system are proteins, enzyme engineering can be applied to further improve their properties. A semi-rational approach was used to generate a 24-member GB1 library to probe the structural determinants of MTG's glutamine selectivity. Each variant contained a single glutamine at varying positions covering all secondary structure elements, and assayed for reactivity. Eight substitutions resulting in an increased labeling efficiency of at least an order of magnitude were distributed throughout all secondary structure elements, indicating that MTG does not favor one preferentially. In addition, introducing point mutations within MTG's active site also resulted in increased reactivity towards variant I6Q-GB1. Our results contribute further to understanding the nature of MTG's glutamine selectivity, while simultaneously demonstrating the potential enzyme engineering has to improve and adjust this system.

Keywords : Biocatalysis, bioconjugation, click chemistry, enzyme engineering, protein labeling, transglutaminase

Table of contents

List of tables

List of figures

List of schemes

List of abbreviations

7HC……...7-hydroxycoumarin

N-AcLysOMe…….*N*-acetyl-L-lysine methyl ester hydrochloride

ACN……...…Acetonitrile

ACP…….Acyl carrier protein

α-LA……...α-Lactalbumin

$AlaNH_2$……L-Alaninamide hydrochloride

AMBER…….Assisted model building with energy refinement

Amp……..Ampicillin

Boc……..*tert*-butyloxycarbonyl

CALB……...……. *Candida antarctica* lipase B

Cbz……...Carbobenzyloxy

Cbz-L-Gln-Gly……...*N*-carbobenzyloxy-L-glutaminylglycine

Cbz-Gly-7HC…….*N*-carbobenzyloxyglycyl-coumarin-7-ylester

Cbz-Gly-GABA-7HC …...4-(*N*-carbobenzyloxyglycylamino)-butyric acid-coumarin-7-yl ester

Cbz-Phe-GABA-7HC…...4-(*N*-carbobenzyloxyphenylalanylamino)-butyric acid-coumarin-7-yl ester

CuAAC……...……Copper-catalzyed azide-alkyne cycloaddition

DMF……... *N,N*-dimethylformamide

DMSO…….Dimethyl sulfoxide

DNA ..... Deoxyribonucleic acid

EDTA……...…..…..Ethylenediaminetetraaceticacid

ESI…….Electrospray ionization

FAB……Fast atom bombardment

FPLC……...…Fast protein liquid chromatography

gTG2…… Guinea pig liver transglutaminase

GB1……...……Domain B1 of Protein G

Gln…….....Glutamine

$GlyNH_2$…….……Glycinamide

gTG2……..……..Guinea pig liver transglutaminase

hDHFR…….Human dihydrofolate reductase

HPLC……..High performance liquid chromatography

HRMS……..…High-resolution mass spectrometry

k_{cat} ……..... Catalytic rate constant

K_i ..... Inhibitory constant

K_M ..... Michaelis constant

LC-MS……..…Liquid chromatography mass spectrometry

LeuOMe……..L-leucine methyl ester

LB…….Lysogeny broth

MBP…….Maltose binding protein

MD…….Molecular dynamics

MeOH…….....Methanol

MOE……... Molecular Operating Environment

MOPS…….…….3-(*N*-morpholino) propanesulfonic acid

MS……Mass spectrometry

MTG……...……Microbial transglutaminase

NAMD……. Nanoscale molecular dynamics

Ni-NTA…….....Nickel-nitrilotriacetic acid

NMR……..Nuclear magnetic resonance

OD……..Optical density

PAGE……. Polyacrylamide gel electrophoresis

PCP…….Peptide carrier protein

PCR…….…….Polymerase chain reaction

PDB……...…Protein Data Bank

PEG…… Polyethylene glycol

pelB…… Periplasmic pectate lyase leader sequence

SDS…….. Sodium dodecyl sulfate

SPAAC……...……..Strain-promoted azide-alkyne cycloaddition

TGase…….....Transglutaminase

TL……...Tetrazine ligation

T_m……..……Melting temperature

WT……Wild-type

ZQG……... *N*-carbobenzyloxy-L-glutaminylglycine

“You can’t really tell how much you can do until you try to do something that’s… more.”

- Gary "Lazarus Lake" Cantrell, creator of the Barkley Marathons

“I remember thinking, some years back, that I KNEW I was gonna be okay, but I wanted to skip ahead to that bit! It's so tiresome knowing you'll be fine but not being at the fine point yet, and you gotta go through all the crap.”

- *Ursula Vernon*

Acknowledgements

For those who know me, it comes as no surprise to see me state that in my nearly 31 years of existence, my Ph.D. – for a variety of reasons – has easily been the hardest thing I`ve ever had to do in my life. To say that I wouldn't have reached this point on my own is such an understatement that it makes me want to launch into a fit of hysterical laughter.

First and foremost, I want to extend deep, sincere gratitude towards my supervisor, Prof. Joelle Pelletier. To this day, I haven't the slightest clue what in the world it is Joelle saw in me. I had an unimpressive undergraduate academic record and no publications on the horizon from my masters, ruling me out as scholarship material. I was insecure and lacked confidence. Yet, she evidently saw some sort of potential and recruited me. Throughout my degree, she was patient, open-minded, thoughtful, as well as understanding and compassionate when medical or personal complications would manifest. She also taught me a plethora about refining my professional soft skills (writing, presenting, teaching, negotiating, etc.), at which she excels. As I am now vehemently opinionated and unapologetically steadfast individual, we've butted heads countless times, but even so, we would always ultimately find a solution. Despite our tribulations, I respect her, and having her as a supervisor has been stimulating, memorable, and inspiring. Thank you, Joelle.

Next, there's my fellow colleague, friend, and former roommate – a unique trifecta! Maximilian Ebert. Oh, Max. I've wanted to kill you more times than I can possibly count. You are loud (fine, so am I), you can be obnoxious (FINE, so can I), and at times too much of a troublemaker even for me. But you are one of the most dedicated (you woke up at 4am to drive me to the train station in Stuttgart, remember? Hopefully not), loyal, passionate, unique, creative, hilarious and brilliant individuals I have ever met. You always were willing to listen to me if I had questions about some aspect of my research, and are such a polyvalent fountain of knowledge, often offering so many ideas that would end up being key to helping me progress. You took me into your home for a couple of months when my finances were particularly dire. I will never, ever forget the look on your face upon seeing me appear in Hamburg secretly to surprise you. Thank you for making those 10 days one of the most memorable experiences of my life – I was one of the first to meet Julie, what an honor that was! I hope we remain in touch

and friends for a very long time – I look forward to following your career and life with great enthusiasm, you magnificent bastard.

Dr. Daniela Quaglia joined the lab reasonably late into my degree, but within that time, she has profoundly influenced me more than anyone I have ever worked with (sorry, Max). She is one of the most professional, competent, practical, and meticulous researchers I have ever met. She is literally now my personal standard for scientific conduct. She has a plethora of experience and remembers all the details of it with that steel trap of a mind she has, and helped me immeasurably with my work. Her sharpness and perfectionism helped me stay on track and avoid pitfalls. She's also an absolutely wonderful, passionate and caring example of humanity, and I found it contagious and motivating. Thank you for all our stimulating conversations and even showing me things within my own city that I had yet to experience (skating in Parc Lafointaine is awesome). You made me realize what a long way I have to go to even come close to being as awesome as you are, but that with a lot of work, I can get there.

Maria Gundersen was the pioneer of the microbial transglutaminase project, which I inherited after she completed her M.Sc. I would like to thank her for all the work she established in that time from which I directly benefited (expression and purification protocols, for example), as well as for her to take the time that she did to train me on instrumentation and other experimental logistics.

I would like to extend my profound thanks to my other colleagues: Dr. Sophie Gobeil, Dr. David Charbonneau, Jacynthe Toulouse, Oliver Rousseau, Sarah M. J. Abraham, and Lorea Alejaldre Ripalda. There's countless times where I needed to know where such a reagent was located, or how to use this instrument, or needed help with this experiment that I hadn't done yet but someone else did, etc, and life would have been much more difficult without their help and time. I would like to particularly thank them for their patience with me when I was dealing with depression (which they may or may not know about) during which I was much more distant, unhappy and troubled. I know that having someone in your workplace with this attitude is a drag, and for whatever its worth, I apologize for that. Luckily, there were also many good times over the years, from conferences to random social events. So, one more time: thank you all.

One of my best homies, Ruby Kulikowska, played an absolute critical factor in keeping what little shreds of sanity I have left. We bonded over kendo and our mutual kendo dojo and you helped me immeasurably with my dojo transition. Over the past few years, we've travelled together, snowboarded together, competed together, and even started and independent projects together. Thank you for being so mature and down to earth, you've kept me grounded many, many times in recent years. I can't wait to snowboard the Rockies with you in the future, duke it out in kendo, and see how much more awesome you become as time goes on.

I've known Dr. Guillaume Beaudoin since undergrad, and I consider Guillaume another one of my best friends, despite that we haven't lived in the same city since 2010. It's been amazing to have a friend that I know is so easily reachable and willing to listen to… anything. Additionally, a number of life changing experiences of mine were with or because of Guillaume. He's also one of the most intelligent people I've met by far (no but seriously, holy crap he's like a Sheldon who understands social queues and isn't an asshole), an incredible scientist, and an AMAZING resource for scientific discussion over the years. This is another case where I'm so thankful to have him as a friend that I can't articulate it effectively. So I'll simply finish with thank you so much, Guillaume.

Speaking of distant friends: Salut Raphaëlle ! It's pretty incredible we're still so close after going through a break up, knowing each other for only 11 months while you were still living in Montréal, and relocating clear across the country. You helped keep me grounded with your realism, having more life experience than I, and being such an open book about everything. Thank you for inviting me to tag along in your pilgrimage across Canada, another trip I won't forget. I'm so insanely happy that you've found your path in Calgary and I'm looking forward to having you be a part of everyday life once again; see you soon.

Lauren Narcross and Alec Sweet are also two of my very best friends. You know what's even better than having lovely friends? Having them live a block away from you for nearly six years. It will definitely be the thing I miss most about living on Queen Mary and Montréal in general. Lauren is one of the most intelligent and talented scientists I know, and was an immeasurably valuable resource. Alec, thank you in particular for discussions about martial arts, helping me decide on the right one for me, and other general life discussions – gaming, running, technology politics, culture, self-care, you name it. I've learned a lot from you and always find

our time spent together engaging. I'm so absolutely honored y'all let me hang around and be your friends, thank y'all so much. Please become really successful so you can afford to visit me often.

Robyn Porterfield is another wonderful human I encountered during undergrad, and she's been a great person to have in my life ever since. She makes me laugh like an idiot and her high level of analytical intelligence is awe-inspiring. Our lives have taken us down different paths as we've gotten older, and it turns out that's okay – because even if we don't talk or see each other awhile, when we do hang out, it's lovely and as if nothing changed. Thank you for being you.

France Desrochers was my first and by far best roommate – four years, yo! France and I shared a lot those four years, which spanned critical development points in both of our lives. We each had our ups and downs with our studies, work, personal lives, you name it – but France was ultimately always super supportive and lovely. Thank you for being you.

Daigo Shima and Akané d'Orangeville were two of my final roommates in Montréal. Daigo, thank you for being one of the kindest and most generous people on this planet, always or having a mate with which to drink beers. Akané, it was wonderful to have a kendoka roommate! Whether it was sending me rodent GIFs when I needed a laugh, listening to me complain, or to simply be present and hold my hand while I'm having a breakdown in the kitchen at 1am on a weeknight, you were very kind and gentle to me in a way that very few (like, almost nobody) have ever been. Maybe you were simply being polite, and that's okay. It really helped me. Thank you, oh so sincerely thank you so much… especially, for Arwin.

To Anne-Catherine (A4) Bédard: oh man you're such a sweetheart. I'm still super indebted to you for all the assistance you provided when I was working on my one-pot chemoenzymatic CuAAC paper. THANK YOU for the last-minute translation of my abstract! You're also incredibly thoughtful; you were literally one of the only people who saw submitting my first 1st author research paper as a cause for celebration! It meant a LOT to me. Thank you.

A shout out to Marija Landekic: I said that A4 was one of the only people who saw my first article submission as a cause for celebration, right? Marija was the other. She not only found a reason why this meant something but convinced (easily) everyone else that we were out

with (thank you McGill Kendo!) of the same thing too, and suddenly everyone was happy for me. Marija, you monster, my liver is still recovering from that one. Thank you.

A massive thank you to Ursula Vernon and Kevin Sonney for being some of the most incredible human beings on the planet. I stand firmly by that statement. Thank you both for being so kind and generous to not only host me for a visit in your home, but pretty much taking care of me the entire time I was there. Your relentless entertainment, passion, thoughtfulness and care will not be forgotten. I especially want to acknowledge Ursula (Kevin too, but let's face it, she's the public figure) for being so open and public about her struggles with severe anxiety and depression. It really helped me when I was dangerously depressed to figure out what was going on and that it wasn't normal.

Thank you to the AMS/Université de Montréal Kendo Club for all your guidance. Additionally, the McGill Kendo Club has been extremely gracious to have be join in on their practices dozens upon dozens of times now – thank you. I've been able to befriend many wonderful people and I'm grateful.

Thank you to the members of Kyabetsu, you're all wonderful.

Thank you to my family for putting up with my shit for 31 years and still supporting me despite all that.

Thank you, everyone. I'm so very glad that you all exist and I'm humbled that you let me hang around.

This is all entirely your fault.

Chapter 1 - Introduction

1.1. Overview

The purpose of this thesis is twofold: in general terms, to demonstrate that biocatalysts are robust entities and should be embraced by synthetic chemists when applicable to complement their research, and specifically, to expand on utility of a selected biological catalyst for amide bond synthesis, with the particular interest of applying the knowledge to protein labeling. To this end, microbial transglutaminase (MTG) is the biocatalyst selected for study. The introduction will present specific highlights through the chronological progression of protein science, demonstrating how the field has evolved and allowed for biocatalysts to be employed and exquisitely refined towards specific purposes. As protein labeling is the major application examined in this thesis, this topic as well as complementary chemical techniques to enhance labeling will also be discussed. In addition, recent applications and mechanistic discoveries concerning transglutaminases are covered extensively in Chapter 2, under the form of a published review article.

1.2. Protein labeling

The common goal shared by all chapters that this thesis strives to achieve is to ultimately contribute towards protein labeling and visualization. There is much to consider to accomplish this task; the sections of this chapter that follow will outline each aspect and their fundamental considerations in detail. This section will introduce the progress researchers have made to label proteins, as well as the challenges that remain.

A complicating aspect of studying biological systems and components is that they cannot be directly visualized at the cellular or molecular level. Because of this, researchers have developed technologies to make this possible, with one of the oldest examples being the microscope. To this day, microscopy is considered a standard technique to observe biological phenomena, including those involving proteins. However, as biological or cellular media are composed of a plethora of different compounds and macromolecules, it is necessary to distinguish a protein of interest from other proteins and components. Methodologies to

accomplish this are discussed at length in Section 1.4; here, we will briefly provide some of the motivations behind protein labeling.

A strong contributor to the impact of microscopy is protein labeling for imaging.[1] Fluorescent proteins were some of the earliest and most commonplace imaging probes conjugated onto proteins, with their implementation serving as indicators of cyclic AMP,[2] cellular tyrosine kinases,[3] and alteration of redox equilibriums in mammalian cells.[4] This progress resulted in designing a rainbow of engineered fluorescent proteins, allowing for visualization over a range of wavelengths. Monitoring the localization of a protein of interest is commonly performed by labeling with dyes, using techniques such as immunofluorescence which requires labeled antibodies for detection.[5-6] A more creative, recent development employing a fusion-based fluorescent protein label was shown to track neuronal proteins with light and electron microscopy.[7] These labeling methodologies allow for researchers to employ imaging techniques to study and understand the cellular processes *in vivo* with high special and temporal resolution.

From a medical perspective, fine-tuning protein therapeutics with covalent modifications has shown much value. Conjugating therapeutic proteins with polyethylene glycol (PEG) chains increases their viability by reducing their susceptibility to proteolysis as well as their immunogenicity.[8] More recently, the production of antibiotics labeled with cytotoxic payloads as therapeutics is creating new progress in the domain of medicine.[9] The high specificity of antibodies for their target can help reduce toxicity that occurs when payloads hit healthy cells rather than the diseased target. Tuning these antibody-drug conjugates with respect to the location and stoichiometry of their payloads also affects their pharmacokinetics and toxicity,[10] and remains an intense area of study.

This brief outline of milestones accomplished by protein labeling describes some of its applications and demonstrates why this topic remains a hot topic of research. As it stands, there is currently no universal labeling methodology that will work for every biological system. Labeling thus requires a broad range of different tools, some being better suited for specific purposes than others. In the next sections, the technical considerations of each component to be considered when developing a labeling system will be examined.

1.3. Biocatalysis: Early discovery and usage of enzymes

The first usage of enzyme-catalyzed chemistry was done without even being aware of the existence of enzymes. Long before enzymes were formally discovered and the knowledge on how to extract and purify them was developed, microorganisms were used to perform enzymatic transformations. Classic examples include fermented foods and alcoholic beverages, the ancient roots of such techniques stemming from Mesopotamia, China, and Japan. Millennia later, an elegant historical demonstration of biological catalysis was performed by Louis Pasteur in 1857. He cultured the mold *Penicillium glaucum* and added it to racemic tartaric acid ammonium salt, to yield the purified (-)-enantiomer; (+)-tartaric acid had been selectively consumed.[11] Concurrently, further conversions using isolated biologically-derived substances were being observed, such as the conversion of starch into sugar by a glutinous component of wheat by a chemist named Gottlieb Kirchhoff.[12] In 1833, Payen and Persoz successfully isolated and studied "diastase", which also hydrolyzed starch to produce dextrin and sugar.[13] A key observation was that small amounts of isolate sufficed to liquefy large quantities of starch. We now recognize that these chemists had discovered the activity of one of the first documented enzymes, amylase. Nonetheless, it wasn't until 1878 that Wilhelm Kühne coined the name "enzyme" for this class of chemically active material – derived from Greek, the term means "in yeast".[14]

Emil Fischer made significant contributions to the understanding of enzymes, including his proposal that enzymes function according to a "lock and key" model, which was published in 1894.[15] He was convinced that enzymes were proteins, although no concrete evidence existed to prove his claim. It was known they were biological in nature, but were considered to be in a class of their own. More hints adding to the poorly understood nature of enzymes came soon after Fischer's work, in 1897, when Eduard Buchner investigated and successfully fermented sugar using cell-free yeast extracts. This is considered a milestone, as he established that intact whole-cell microorganisms were not necessary to achieve conversions catalyzed from a

biological source.[16] It wouldn't be until 1926, upon the successful crystallization of urease, that enzymes were concretely confirmed to be proteins.[17]

With the composition of enzymes being better understood, chemists sought ways to discover more enzymatic activities and to employ them in further chemical transformations. This required a robust methodology to purify enzymes from diverse cellular components – as well as from other enzymes. Early sources of enzymes, other than microorganisms, included egg whites, animal tissues and blood.[18] Electrophoresis was one of the first means to separate proteins, with Picton and Linder reporting the separation of haemoglobin using electric current in 1892,[19] and Hardy doing the same with other globulins.[20] Despite improvements to electrophoresis technology over time, it remained limited as a purification methodology due to low yields of pure protein, which was typically on the milligram scale. This changed dramatically upon the development of cellulose-based ion-exchange chromatographic resins capable of binding proteins, by Peterson and Sobers in 1956.[21] This subsequently lead to the development of numerous substituted chromatographic resins,[22-23] allowing for separation using various chemical or physical properties of the protein.

The bulk purification of enzymes greatly increased their availability, and new possibilities for biocatalysis could be explored. As with any new scientific breakthrough, many synthetic chemists were wary of applying enzymes to their methodologies, and it took the work of a many of researchers to help popularize and validate them. For example, in 1985, Whitesides and Wong published an extensive review describing the characteristics, benefits, and applications of an abundance of different enzymes to the synthesis of fine (chiral) chemicals, pharmaceuticals, and biotechnology.[24] In parallel, Klibanov demonstrated that some enzymes retain some activity in organic solvents.[25] These efforts, along with those of many others,[26] helped convince many academic and industrial chemists to embrace biocatalysis, although many skeptics remain. Ironically, the widespread use of enzymes in organic synthesis inspired the development of synthetic enzyme mimics,[27-29] boasting advantages such as tunable structures and catalytic efficiencies, excellent tolerance to experimental conditions, lower cost, and purely synthetic routes to their preparation.[30] This field continues to be an area of intense interest to

this day, contributing high quality developments to the scientific community, serving sometimes as a complementary school of thought to biocatalysis, and sometimes antagonizing it.

1.3.1 Improving and creating novel biocatalysts by directed evolution

Standardized bulk purification of enzymes increased their availability; however, it limited the availability to naturally-occurring enzymes. The specificities of many of these natural enzymes were extensively probed for, and ultimately led to the successful identification of activity toward non-natural substrates. While this proved to be helpful to the inclusion of enzymes in the repertoire of useful catalysts, the natural activities, specificities, and physical properties of enzymes were often limiting, preventing them from being applied to industrial processes. The conversions performed by enzymes that could accept non-natural substrates were typically suboptimal, those reactions not having benefitted from the tailoring effects of millions of years of evolution. High temperatures, extreme pH and the presence of organic solvents are commonplace in large-scale industrial compound production,[31] all of which are typically incompatible with the use of enzymes. Finding a way to circumvent the limitations of natural enzymes would be an important development for them to be widely used and considered.

The solution to this problem would lay in the realm of molecular biology. As the understanding of DNA and how to effectively manipulate it in the laboratory became evident, researchers attempted to produce recombinant DNA.[32] This slowly progressed to introducing a non-native gene of interest into the DNA of a host organism, and being able to control the expression of the gene of interest using the molecular machinery of the host. One of the first and most successful uses of this groundbreaking technology occurred when the recombinant expression of human insulin in *Escherichia coli* was accomplished in 1979 by Arthur Riggs and Keiichi Itakura .[33-34] This opened the door to new possibilities for synthesizing enzymes, as it was now no longer necessary to extract enzymes from their natural sources and organisms. Despite this development, it did not solve the above-cited limitations: researchers remained at the mercy of the natural properties of the enzymes with which they chose to work.

An attempt to address those limitations made use of UV radiation and chemical mutagens to introduce mutations and observe the changes they bring within organisms, cells, and later, enzymes. However, these changes were non-specific and consequently, damaging to complex

biological systems.[35-36] Improvements in tackling the challenge of introducing targeted, specific mutations into DNA can be attributed to the progress made in molecular cloning and assembly of recombinant DNA. These methods provided researchers a way in which to isolate, manufacture, and modify the source material for protein synthesis from other biological components. The development of PCR in 1983 by Kary Mullis was a huge leap forward, simplifying the laboratory synthesis of genes.[37] Concurrently, researchers had discovered that by designing and synthesizing oligonucleotides to bind a specific site of their gene of interest, the oligonucleotides served as means to introduce targeted mutations within the sequence of the gene, with one of the first enzymes specifically mutated being β-lactamase.[38] Other early enzymes that were mutagenized in such a way include tyrosyl tRNA synthetase,[39] which ultimately resulted in compromised activity due to an increase in K_M, and dihydrofolate reductase,[40] which allowed identifying a catalytic residue as well as two others that perturbed the local structure.

Eventually, by introducing mutations in enzymes, it was revealed that not all amino acid residues will disrupt enzyme function upon substitution. Furthermore, substitution of residues that have been more highly conserved throughout evolution was found to be far more likely to be damaging than substitution of variable residues.[41] Mapping conserved residues thus helped to identify which residues should substituted; knowing that much of the sequence space was not critical to enzyme function also made it possible to introduce multiple mutations into the same gene. By those means, multiple rounds of mutagenesis could be performed while assaying for the desired catalytic property, whether that was the improvement of native catalytic activity or the appearance of another, non-native property. Thus began the development of an approach to artificially generate new, non-natural enzymes *in vitro*, now known as the discipline of enzyme engineering; it will be presented in more detail in Section 1.6.

1.3.2 Incorporation of enzymes into chemical synthesis

One of the earliest examples of using a purified enzyme in a synthetic chemistry context was the work of J.W. Cornforth and colleagues in 1969, in which they described a multi-stage one-pot synthesis of *S*-malate using the three enzymes acetate kinase, phosphotransacetylase, and malate synthase on acetic acid.[42] As bulk enzymatic production and purification technology

became established, the biocatalytic toolbox rapidly became more diverse, as illustrated by the products that were generated in early chemoenzymatic strategies. In the early 1970's, aldolase was exploited for C-C bond formation in synthesizing hexose epoxides from dihydroxyacetone phosphate and various aldehydes. This was made possible due to aldolase having loose specificity for its aldehyde substrate, allowing for it to be fed non-natural aldehydes. These epoxides were used to perform mechanistic studies on sugar isomerases, which were inhibited by the epoxides.[43]

Determining the substrate specificities of enzymes, and in particular which enzymes can accept non-natural substrates as well as the extent of their stereospecificity, was essential knowledge to extract to improve the effectiveness and applicability of biocatalysts. The promiscuity of liver alcohol dehydrogenase was one of those examined for this purpose. Its selectivity was probed, and ultimately successfully used to stereospecifically reduce several aldehydes using its reducing co-factor NADH, described at the time as "the biological equivalent of sodium borohydride".[44] Finally, it was being discovered that compounds that were notoriously difficult and expensive to synthesize by traditional chemical means could be generated by incorporating biocatalytic steps in synthetic schemes. One of these compounds, a pyrrole called porphobilinogen, was labelled with 13C by a dehydratase enzyme. This product was converted enzymatically into labeled porphyrins to enable mechanistic studies.[45]

The chemoenzymatic feats that have been accomplished following these pioneering examples have been broad and varied, with some profoundly impacting synthetic strategies.

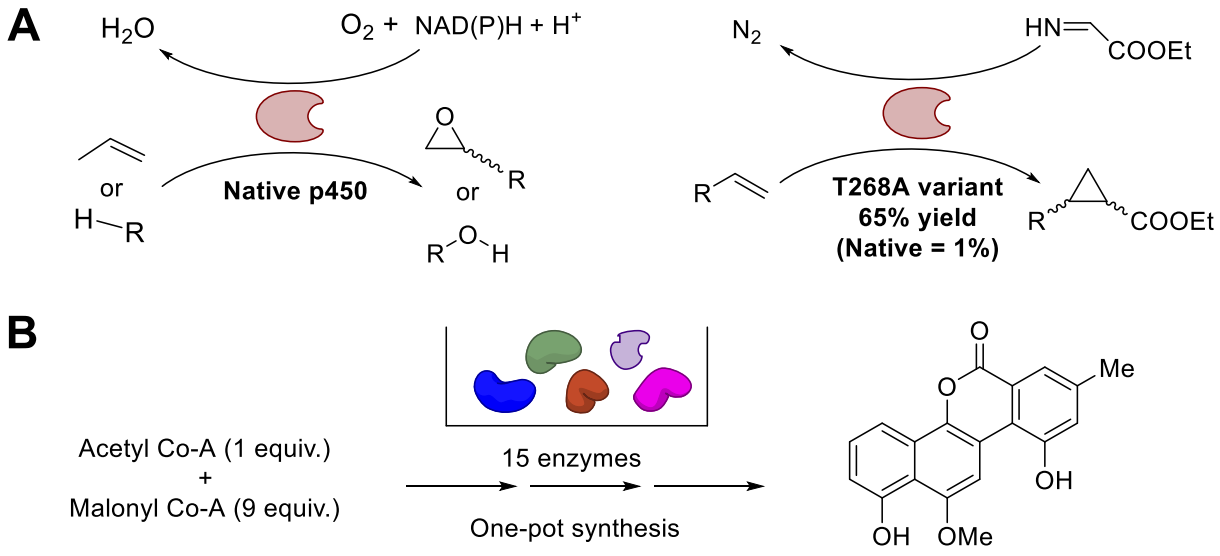


Figure 1-1 Recent examples of biocatalysis engineering.

A) Native cytochrome p450 will oxidize terminal alkynes and hydrocarbons to epoxides and alcohols, respectively. The enzyme was rationally designed to alter the substrate specific to allow for carbene transfer, yielding cyclopropanes.[46] Additional variants produced exhibiting varying diastereo- and enantioslectivity.

B) Scheme for a one-pot cascade synthesis of the core chromophore found in gilvocarcin natural products, produced by using 15 separate enzyme-catalyzed steps.[47]

The advent of enzyme engineering by directed evolution has played an immense role in feasibility of biocatalysis, making it possible to tailor enzymes to a desired substrate specificity, increased reactivity and efficiency, and tolerance to various reaction mediums.[48-51] A well-known example is how directed evolution was employed to successfully design a transaminase capable of asymmetric synthesis of chiral amines, specifically tailored for the large-scale synthesis of a chiral intermediate for sitagliptin (Januvia), the blockbuster antidiabetic.[52] In 2012, the engineered transaminase replaced the previously used rhodium-catalyzed enamine hydrogenation process of Merck & Co, with higher total and space/time yields and enantiomeric excess, and an important reduction in toxic waste.

Taking a more rational approach to biocatalyst engineering, the well-characterized oxidation mechanism of cytochrome p450 monooxygenase was elegantly reworked to engineer

its substrate specificity (Figure 1-1, panel A).[46] The iron atom present in the monooxygenase's heme cofactor is responsible for coordinating the transfer of an oxygen atom (oxene transfer) to its olefin substrate. The authors successfully hypothesized that the electronic properties of a carbene intermediate would be analogous and capable of interacting with the heme, producing a cyclopropane instead of an epoxide. The activity and stereospecificity of the monooxygenase was further tuned by rounds of directed evolution, affording a collection of enzymes with a wide variety of properties. Further progress recently made by applying this mechanism-based substrate engineering strategy to include enzyme-catalyzed enantioselective aziridination[53] and biocatalytic synthesis of a key cyclopropane intermediate to the pharmaceutical, Ticagrelor.[54] These examples highlight the power of a single enzyme harnessed in tandem with other chemical reagents, but there has also been progress made by combining multiple enzymes, often in a one-pot format, to meet a synthetic goal; this is commonly referred to as cascade reactions.[55-56] Taking the term "multiple" to another level, a model compound containing the unique chromophore common to all members of the gilvocarcin group of natural products, known for their anti-tumor activity, was synthesized by employing a one-pot cascade composed of 15 enzymes from various sources (Figure 1-1, panel B).[47] As illustrated in these examples, the strategies of both engineering and applying biocatalysts to synthetic chemistry goals have met considerable progress. Even so, there remains an abundance of enzymes that catalyze reactions that would enhance the synthetic chemist's repertoire, but have yet to be either obtained from their natural source or engineered into a form that can be applied to such a purpose.

1.4. Amide bond synthesis

One of the most common chemical transformations that remains a challenge to this day is the formation of amide bonds. Favorable properties such as high polarity, stability and conformational diversity are the reason for their abundance in nature as well as in synthetic targets, but these properties are also what make them challenging to synthesize. Complex amide bond-containing natural products have intriguing structural and functional properties, and have found utility in fields such as medicine, agriculture and biotechnology (Figure 1-2). Amide-containing natural products are formed within their native organism using cellular machinery, the most abundant being ribosomes to synthesize linear peptides and proteins one

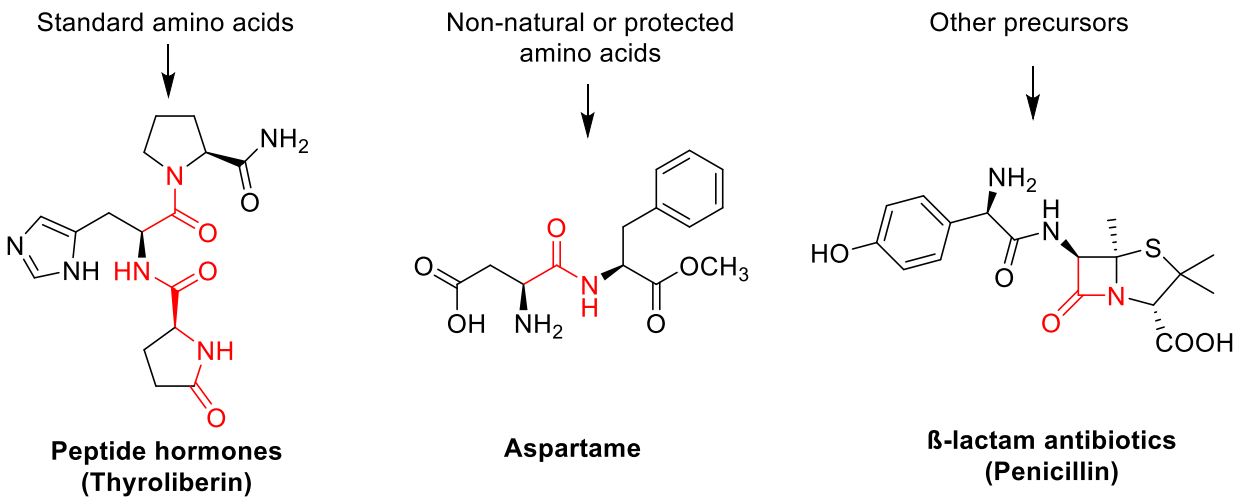


Figure 1-2 Examples of amide-containing compounds.

Amide functional groups are highlighted in red. Aspartame and penicillin are produced annually on the kiloton scale. Peptide hormones are commonly used as therapeutics; aspartame is an artificial sweetener; β-lactam antibiotics are used to treat bacterial infections.

amino acid at a time. Other common amide-bond forming mechanisms include non-ribosomal synthases, which build highly diverse and synthetically challenging compounds. Before researchers became capable of manipulating enzymes, chemists were the sole driving force behind developing and executing molecular synthetic strategies, and were responsible for devising creative ways to reliably and selectively form amides. For the scope of this thesis, only select, highly successful synthetic strategies for amide bond formation will be introduced; for an exhaustive review of other current and upcoming methods, we encourage the reader to consult recent reviews.[57-58]

The pioneering of solid-phase peptide synthesis in 1963 was a breakthrough for routine laboratory synthesis of peptides, and remains the standard methodology in use today.[59] Selectivity proved to be a sizable stumbling block, owning to cross-reactivity within the peptide's side chains, which was circumvented in solid-phase synthesis and other strategies by the use of bulky protecting groups. These methods proved to be effective, but have the disadvantage of employing poor atom economy and many steps, with multiple functional groups often requiring protection for the formation of a single amide bond.[60] For the synthesis of full-

sized proteins, historically thought to be unachievable due to their size and structural complexity, the development of native chemical ligation[61] granted chemists this power in their laboratories. The synthesis of the catalytically-active, 203 residue HIV-1 protease in 2007 is a notable example.[62] Despite this achievement, chemical synthesis of full proteins is seldom practiced; biochemical methods are overwhelmingly employed, though they tend to be limited to inclusion of only the 20 natural amino acids, while chemical synthesis has no such restriction.

Amide bonds are also prevalent in numerous non-peptide active pharmaceutical ingredients (API). The most common way to install these is by the acylation of an amine with an activated carboxylic acid. This transformation accounts for 16% of all reactions in the synthesis of APIs, and 25% of pharmaceuticals in 1999 contained at least one amide bond.[63-64] Coupling reagents are required, otherwise one typically obtains a carboxylate-ammonium salt due to the thermodynamic barrier to amide formation. In the interest of increasing atom economy, boronic acid catalysts were found to be effective in creating amides from amines and carboxylic acids without a coupling reagent, with the first highly-active of such catalysts being reported in 1996.[65] Recent efforts have since yielded improved catalysts, such as halogen-substituted phenylboronic acids, which can proceed at room temperature, are recoverable, and produce no wasteful by-products.[66] The reactions must still be performed in solvent, as density functional theory calculations predict that the elimination of water from a tetrahedral intermediate is the rate-determining step.[67] This highlights a critical obstacle surrounding amide synthesis: it is exceedingly difficult to effectively perform these reactions in the presence of water. Indeed, a recent SciFinder survey revealed that of ~680,000 amidation reactions investigated, dichloromethane or *N*,*N*-dimethylformamide were the most common solvents used, at 36% and 47%, respectively. As both of these solvents face major regulatory issues, efforts have been made into finding alternative reaction media.[68]

It is now possible to employ enzymes in the laboratory with ease. This presents an opportunity to develop diverse biocatalyzed solutions for amide bond formation, with their intrinsic affinity for aqueous reaction conditions being a major advantage. One of the first enzymatic systems used for amide synthesis were proteases.[69-70] The hydrolysis of peptide bonds by proteases proceeds via the nucleophilic attack by the active-site (deprotonated) serine or cysteine residue on the peptide carbonyl, forming a covalent acyl-enzyme intermediate and

eliminating the amine product; an activated nucleophile – a hydroxide anion – repeats those steps to release the enzyme and the carboxylate product. It was observed early on that proteases can hydrolyse esters, broadening their applicability.[71] Hypothetically, any activated nucleophile that can be accommodated into the enzyme active site can react with the acyl-intermediate, thus procuring a route to acyl-group modification. Indeed, by implementing strategies to exclude competing water such as biphasic systems[72] or water mimics,[73] amide formation was observed. However, these methodologies have severe restrictions owing to the more thermodynamically favorable competing hydrolytic reaction, ultimately limiting yields and substrate scopes. Additionally, proteases are unstable in anhydrous reaction media.

The utility of other enzymes in amide-bond formation has since been explored, such as ligases, lipases, and transglutaminases,[74] but competing hydrolysis remained a common obstacle. Each class of enzymes was found to come with their own challenges: ligases utilize a carboxylic acid and amine substrate, and like in traditional synthesis, require activation, which is provided by the hydrolytic coupling of ATP – an expensive step unless one operates inside living cells. Ligases display a remarkably diverse substrate range, there being ligases capable of accepting simple molecules like formic acid (the carboxylic acid substrate) or an ammonium ion (the amine substrate), to complex biological macromolecules like proteins. Lipases naturally catalyze the hydrolysis of fatty acids, and are typically reactive towards long-chained acyl-glyceride substrates. Lipases have been reported to form amides by acylating an amine substrate; once again, the competing aminolysis reaction was also observed.[75] Transglutaminases function similarly to proteases such that they also form an acyl-enzyme intermediate from an ester or amide substrate, and will be discussed at length in Section 1.6 and Chapter 2. It is worth noting that, while mammalian transglutaminases suffer from the competing hydrolysis reaction, little competing hydrolysis is observed with the microbial homolog of transglutaminase, potentially making it a unique means of synthesizing amide bonds.

On the level of chemical synthesis, the number and variety of catalysts for amide bond formation are vast and diverse. Despite this, there is currently no chemical or biological catalyst without notable drawbacks, providing motivation for continued interest and research into addressing these limitations. However, with substrates of some of the systems discussed above

being proteins, and the molecular glue holding proteins together being amides, their formation has applications beyond that of chemical synthesis.

1.5. Site-specific bioconjugation

Biological macromolecules are complex and, as a result, are highly diverse, enabling them to display a variety of different properties and activities. Many of these properties can be harnessed to solve human problems; indeed, in recent years, there has been a steady increase in number of biologics employed for pharmaceutical needs, particularly antibodies.[76] Their properties can be further altered after synthesis. Natural proteins are processed by post-translational modification reactions. This downstream fine-tuning allows for the cell to exert additional control over its machinery, with these modifications modulating enzymatic activities, molecular interactions and recognition, and bestowing functionality that is beyond the chemistry possessed by the standard 20 amino acids.[77]

One of the earliest ways researchers attempted to mimic this natural process was by using *N*-hydroxysuccinimidyl esters, which covalently modify amino groups within proteins, with the ε-amine of lysine being the most reactive. The fundamental limitation of these esters, and the limitation that plagues most bioconjugation efforts to this day, is the poor selectivity it displays for a precise location or residue. Considering that a typical protein contains numerous surface-exposed lysine residues, and if conjugation is designed to occur in a reaction media containing other (lysine-containing) proteins, then the matter of selectivity quickly becomes complicated.

Strategies investigating functional groups and regions of proteins offering distinct reactivity became of high interest in an attempt to improve upon the problem of selectivity. In light of this, covalent modification of cysteine materialized as one strategy; because cysteine is less frequently surface-exposed than lysines, it is an attractive target.[78] The *N*-terminus, particularly if it contained a serine or a cysteine, was another viable target region[79] with techniques including the native chemical ligation discussed in the previous section.

More recently, biochemists have gone back to the natural systems which inspired such synthetic ambitions, and modified the fundamentals of cellular protein synthesis by expanding the genetic code, allowing for the coding and incorporation of abiotic amino acids.[80-81] These

abiotic amino acids are given two possible fates: they can display the desired chemical moiety directly without any further modification, or a unique reactive functional group will be introduced, which can be exclusively modified with a chemoselective reporter in a downstream

Table 1-1 Summary of bioconjugation approaches.

Approach	Requirements	Advantage(s)	Disadvantage(s)
N-hydroxysuccinimidyl esters	○ Primary amino group (lysine) ○ Small molecule or peptide tag	○ High reactivity	○ Inferior selectivity
Nucleophilic modification	○ Accessible cysteine or serine ○ Small molecule tag	○ High reactivity with improved selectivity	○ Poor selectivity
Native chemical ligation	○ Accessible *N*-terminus ○ Peptide-thioester tag	○ Modification made to protein backbone	○ Cumbersome for large proteins ○ Synthetically challenging ○ Poor selectivity
Abiotic amino acid incorporation	○ Engineered gene, specialized expression system ○ Synthetic amino acid ○ (Optional) Small molecule or protein tag	○ Highly selective	○ Mutated gene sequence required
Enzymatic	○ Accessible reactive residue(s) ○ Small molecule or protein tag	○ Highly selective ○ Diversity of reactions available to tailor to a system	○ Encoded recognition sequence typically required

step.[82] The later concept is known as bioorthogonal chemistry, a chemical discipline that develops biocompatible reagents that are orthogonal to the reactive groups encountered in complex biological environments; this topic will be explored in further detail in Section 1.5.

Unnatural amino acid incorporation marked a paradigm shift in site-selective protein modification,[83-85] as it addresses the issue of selectivity directly, as well as proving a means to express proteins with unique chemical properties.[86-88] However, an engineered cellular system containing all the non-canonical components and reagents is required, as well as a mutated gene encoding the protein of interest, limiting its applicability to proteins that are not naturally encoded.

An alternative route to tackling the non-selective nature of existing chemical methodologies takes advantage of the high substrate specificity of enzymes, without the need for a genetically encoded recognition site using non-canonical amino acids. Different classes of enzymes have yielded chemoenzymatic bioconjugation strategies, which can be divided into two broad categories based on the location of conjugation. The first category is the largest, which are enzymes that target either the *N*- or *C*-terminus of the protein substrate: formylglycine generating enzyme, phosphopantetheinyl transferase, farnesyltransferase, biotin ligase, and lipoic acid ligase all fall into this category.[89] The second category corresponds to those that modify a site at any location within the protein substrate, as long as it is accessible to the catalyst. Enzymes capable of this are sortase[90] and transglutaminase.[91]

The strength of the first category of enzyme systems is that they are each highly or exclusively specific for an amino acid recognition sequence, although they must be at a surface-exposed terminus of the protein substrate. A recognition sequence must be encoded within the protein substrate, which typically requires producing a non-native form of the protein. In addition, genetically encoded tags limit their usage to proteins and exclude other biomolecules. This limitation is shared with genetic incorporation of non-canonical amino acids discussed above, although encoding a recognition sequence does not require an engineered cellular expression system. One enzyme that is an exception to the recognition sequence requirement is the bacterial homolog of transglutaminase, commonly referred to as MTG. While recent efforts have revealed an engineered sequence with improved specificity,[92] the enzyme is promiscuous with both its glutamine- and lysine-containing substrates.

Finding a balance between selectivity and applicability remains a delicate challenge: a system that is too selective may not have sufficiently broad applicability to be attractive. Ultimately, the solution for site-specific bioconjugation may lie within fine tuning a

combination of chemical and enzymatic reactions, with the explosion of bioorthogional chemistry techniques within the last 15 years being indicative of this.

Table 1-2 Residue locations and sequences of enzymatic bioconjugation.

Adapted from Rashidan et al.[89] X = One of the canonical amino acids; a = aliphatic amino acid; ACP = acyl carrier protein; PCP; peptide carrier protein.

Enzyme	Conjugation location	Recognition sequence
Formylglycine generating enzyme		CXPXR or 13-mer LCTPSRGSLFTGR
Phosphopantetheinyl transferase	*N*- or *C*-terminus	ACP, PCP or ybbR tags (11 mer: DSLEFIASKLA; 13 mer: VLDSLEFIASKLA; 17 mer: GSQDVLDSLEFIASKLA)
Lipoic acid ligase		GFEIDKVWYDLDA
Biotin ligase		GLNDIFEAQKIEWHE
Farnesyltransferase	*C*-terminus	CaaX
Sortase	Any	LPXTG
Transglutaminase		XXQXX

1.6. Bioorthogonal chemistries

For a chemical transformation to be considered bioorthogonal, it must be mild enough to proceed within a cellular environment without disrupting it; be reactive in water or aqueous media, at physiological temperatures and pH, and exhibit rapid kinetics.[93] Click chemistry is typically bioorthogonal, and serves largely the purpose of covalent *in vitro* or *in vivo* labeling of biomolecules.[94] Another more recent, but growing, application of click chemistry is the production of antibody-drug conjugates, in which a therapeutic compound is introduced onto a specific position of the antibody to procure targeted drug delivery.[9] In both cases, the click transformation covalently links the chemical reporter or payload onto the biomolecule.

As proteins and other complex biomolecules are highly diverse in their physiochemical properties, it's no surprise that a wide range of bioorthogonal reactions have since been developed to best meet biotechnological needs while circumventing the limitations of any selected system. Among the first bioorthogonal reactions to be successfully applied to protein labeling were aldehyde and ketone condensations.[95] Due to interference from endogenous aldehydes and ketones, it may not have been the most practical strategy, but it set the precedent. Utilizing abiotic functional groups would circumvent this non-specific interference, and it is upon this opportunity that the reactive and biologically absent azide group capitalized. The year 2000 marked the first time azides were employed as a reporter, in the Staudinger ligation.[96] It is an adaption of the classic Staudinger reduction of azides with triphenylphosphine.[97] Azides and phosphines are not completely biologically inert, however; cross-reactivity with thiols and disulfides, respectively, have been observed. Nonetheless, the Staudinger ligation has been successfully applied to labeling cell surfaces,[98] validating its potential to work within a living system.

The azide became a functional group celebrity when it was reported that it reacts effectively with a terminal alkyne to yield a 1,2,3-triazole product. The Copper-catalyzed Azide-Alkyne Huisgen Cycloaddition (CuAAC) proceeds at room temperature, in aqueous solution, in the presence of Cu(I), which is typically generated in situ from Cu(II) and a reducing agent such as sodium ascorbate.[99] The CuAAC was adapted from classical chemistry,[100-101] and it was only in 2001 that the bioorthogonal reaction conditions were determined for the transformation. Since then, the reaction has become the quintessential click reaction, and has been developed further to improve reactivity and compatibility, as well as being applied to protein labeling, organic synthesis, medicinal chemistry, and surface chemistry.[102-106]

Despite these successes, the copper catalyst is toxic to live cells,[107] and it can interfere with protein function by chelating the thiol group of cysteines[108] which also limits its application to protein substrates. This limited biocompatibility spawned a second generation of click chemistry reactions, which are simply referred to as copper-free click chemistry. Inspired by the chemistry using ring strain to activate alkynes,[109-110] Carolyn Bertozzi's research group was the first to bring such a generation of reactions to fruition by designing cyclooctyne reagents capable of targeting azides through a strain-promoted [3+2] cycloaddition (SPAAC).[111] The initial

reagents for strain-promoted cycloadditions proved to react relatively slowly, being on the same order as the Staudinger ligation and slower than the CuAAC. Cyclooctynes with improved reactivity have since been synthesized,[112] which has helped improve the applicability of the SPAAC, including to the labeling of live cells and embryos.[113-114]

The success of the SPAAC has since inspired creative repurposing of classical organic chemistry so that those reactions can be employed as bioorthogonal strategies, potentially filling any voids left by the CuAAC and the SPAAC; these diverse reactions have been reviewed recently.[93, 115] For example, some of these reactions sought to improve upon the notoriously low solubility of strained cyclooctynes. Between cross-reactivity, catalyst toxicity and poor solubility, it is not straightforward to determine the conditions that should be employed, or even which bioorthogonal chemistry is best suited for a selected system for study.

Nonetheless, when click chemistry is successfully combined with enzymes, these chemoenzymatic strategies are very powerful, making it one of the most effective ways to specifically perform bioconjugation. In this thesis, we develop the combination of click chemistries with transglutaminase enzymes, described in more detail below.

1.7. Transglutaminase-catalyzed amide bond formation

As mentioned in Sections 1.3 and 1.4, transglutaminases are enzymes that catalyze the formation of amide bonds between two protein substrates. Specifically, they perform an acyl-transfer reaction between the γ-carboxamide of a peptide- or protein-bound glutamine and the ε-amino group of a lysine residue, yielding an isopeptide bond.[116] Transglutaminases have been identified in a wide variety of organisms, although two forms in particular are the focus of biotechnological applications: a calcium-dependent, GTP-activated transglutaminase found in the tissue of animals and humans, referred to as transglutaminase 2 (TG2), and a bacterial form called microbial transglutaminase (MTG; Figure 1-3). We recently reviewed the applications of transglutaminases, presented in Chapter 2. In our review, we address not only recent advances in medical and biotechnological applications of transglutaminases, but also the poorly understood aspects of these enzymes, such as substrate specificity, and the efforts made to clarify them.

A topic that was not discussed in our review concerns the biocatalytic considerations for transglutaminase, which we will briefly presented here. These examples refer exclusively to the microbial enzyme, as it has the advantage of being calcium- and GTP-independent, thermostable and tolerant to organic solvents and various pHs,[117] making it better suited for biocatalytic applications than its mammalian counterpart. Despite these favorable properties, the biocatalytic

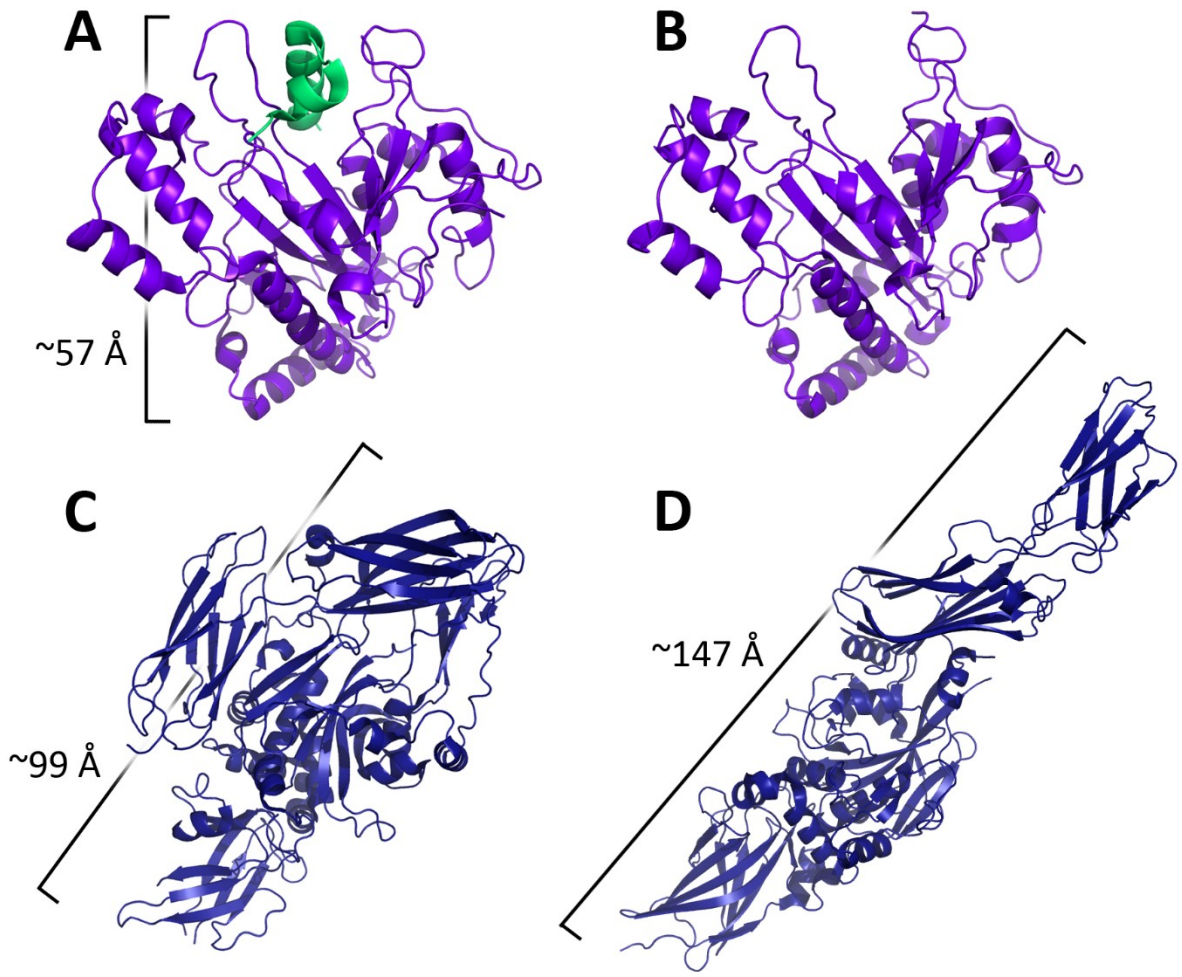


Figure 1-3 Crystal structures of transglutaminase.

A) Zymogenic microbial transglutaminase (MTG); the pro-sequence (green) must be cleaved to expose the active site and render the enzyme functional. PDB ID: 3IU0.

B) Active MTG. PDB ID: 1IU4.

C) GDP-bound human transglutaminase 2 (TG2) in its "closed" confirmation. PDB ID: 1KV3.

D) Inhibitor-bound TG2 in its "activated" confirmation. PDB ID: 2Q3Z.

achievements accomplished with MTG remain modest. For example, MTG has been shown to be effective at PEGylating therapeutic protein substrates, serving to decrease their susceptibility to proteolysis.[118] MTG was also found to catalyze Henry reactions of aliphatic, aromatic, and hetero-aromatic aldehydes with nitroalkanes at room temperature.[119] The yields of the transformation varied heavily depending on the substrates selected, although the highest observed reached 96%. Based on sparse literature, we surmised that there is untapped potential within transglutaminase for further biocatalysis investigations. We start by addressing whether TG2 could be utilized for peptide synthesis, which is the topic of Chapter 3. We progress by developing and extensively characterizing a one-pot chemoenzymatic peptide and protein conjugation scheme using MTG and four of the bioorthogonal chemistries discussed in Section 1.5; Chapters 4 and 5 describe the details of these efforts.

A major reason for restricted biocatalytic applications of MTG is its constrained, but poorly understood, specificity for peptide- or protein-bound glutamine substrates. On the other hand, it has been shown to be promiscuous towards its amine substrate: in addition to peptide- or protein-bound lysines, MTG reacts with a number of small amines. This makes MTG a potential candidate for undertaking site-specific protein modification.[120-121] Indeed, previous works have shown transglutaminase to be used successfully as a tool for bioconjugation,[89, 91] and we also explore this topic at depth in our review in Chapter 2. We hypothesized that MTG's bioconjugation ability could be improved, and that this is the application for which MTG is best suited. This is the underlying theme in most of our research chapters. We apply multiple approaches to work towards this goal: the one-pot chemoenzymatic reactions we developed in Chapters 4 and 5 employ biocatalysis to successfully bioconjugate proteins. Detailed structural information revealing the manner in which MTG interacts with its glutamine-containing substrate would greatly enhance the capacity to engineer MTG towards site-selective protein conjugation; to this effect, Chapter 6 describes our attempts to develop a purification scheme for crystallography trials.

1.8. Protein engineering

As mentioned in Section 1.2.2, protein engineering is a technique which will profoundly impact an enzyme's function, with many approaches available. These approaches are worth

examining to better appreciate the context of Chapter 7, in which we present a conjugation system consisting of MTG as the labeling biocatalyst and the B1 domain of Protein G (GB1) as a substrate.

In a general sense, the design of non-natural enzymes by protein engineering can be broken down into a two-step process: the first is the generation of mutated DNA sequences to encode variants of the enzyme of interest, and the second is the screening of these variants so that the desired property can be evaluated, allowing for the identification of improved variants. The variants exhibiting improvement, which is defined by the screening method, can then be further mutated as required, until a satisfactory variant is found, or further improvements are no longer observed. Protein engineering can be divided into three general categories: rational design, semi-rational design, and random mutagenesis.[122-124]

Rational design is utilized when one can reasonably hypothesize, or know with certainty, the identity of residues responsible for conferring a selected property, such as catalytic activity, substrate preference, or stability. Crystal structures revealing the intricacies of key structural elements and the active site are generally required for rational design, where knowing detailed information such the immediate molecular environment of a selected residue is essential. The effect of point mutations and sequence alignments of homologous enzymes are also valuable tools to reveal functionally significant and conserved residues, respectively. Specific mutants will be produced by site-directed mutagenesis and tested for the desired effect on the enzyme property targeted. Nonetheless, even a crystal structure obtained with the highest resolution does not immediately reveal the intricacies of catalytic mechanism; the complexity and size of enzymes make it easy to miss residues that may have profound effects, but are not obvious targets. Additionally, the effects of combinations of residue mutations are extremely difficult to predict. Furthermore, enzymes are dynamic molecules such that crystal structures only partly capture their physical properties. These facts account for the low success rate of rational design in improving catalytic properties.

Random mutagenesis removes any control the researcher has over selecting the location of the mutation. The methods used to introduce mutations into the gene of interest are, in principle, unbiased. They can be adjusted to tune the average number of mutations included in the gene, but their location cannot be dictated. Structural information is not required, which can

be advantageous. However, statistically speaking, the majority of variants generated randomly will either not be functional or interesting. This will require that a large number of mutants are produced (which can be on the scale from 10^5 to 10^{10}) to best guarantee the odds of identifying useful variants, and a high-throughput screen is necessary in this case.[125] Screening millions or billions of mutants requires an observable output such as cell survival linked to the desired enzyme activity, which is not amenable to most enzyme systems as few enzymes are essential

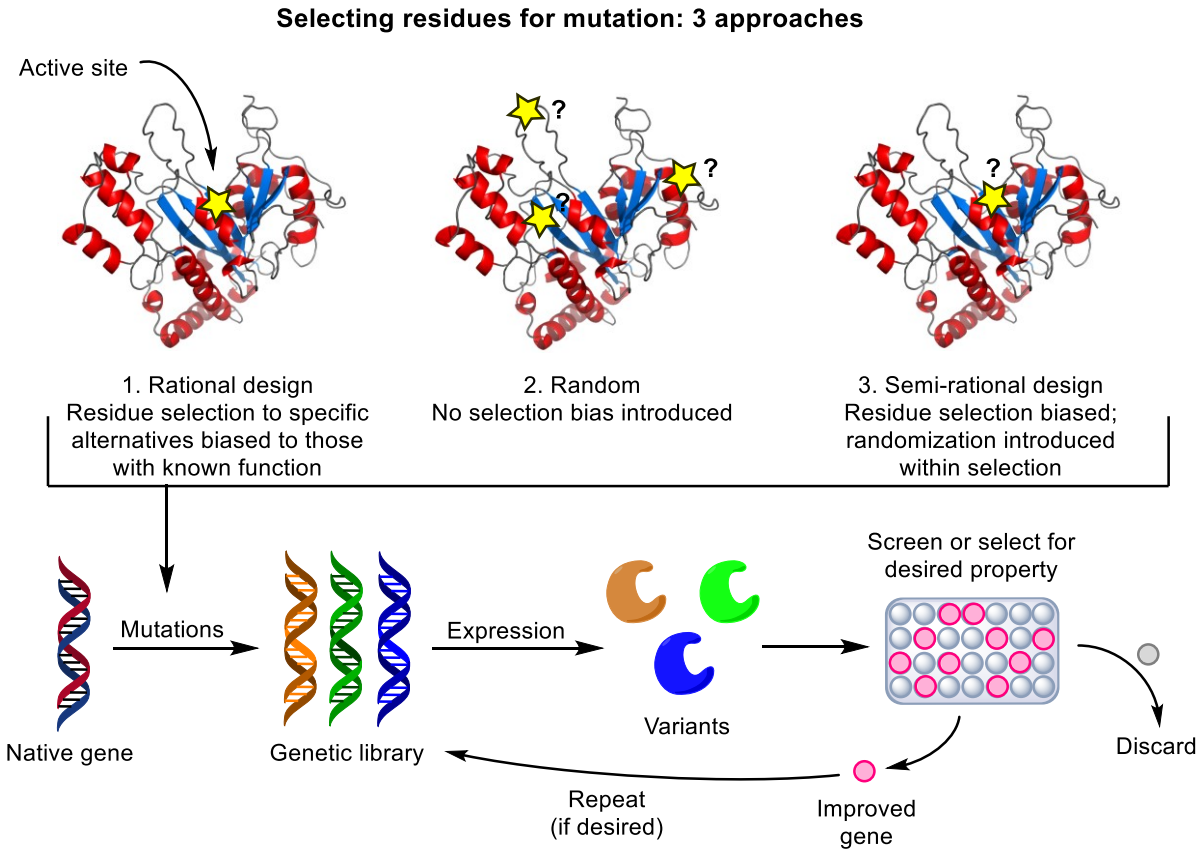


Figure 1-4 Engineering enzyme variants.

First, one of three approaches must be chosen as to how mutations (yellow stars) will be introduced. Then, a library of mutant genes can be generated, which will express variants with altered function and must be evaluated by a selection or screening assay. If further rounds of mutagenesis are desired, a mutant gene (typically exhibiting the greatest improvement) will become the parent for these additional rounds.

to cell survival. Screening many thousands of mutants requires an observable such as a colorimetric or fluorimetric signal associated with the desired enzyme activity, which is, once

again, limiting (FACS sorting increases this to millions of variants). Screening by analytical methods coupled to autosampling, such as NMR or MS, is limited to screening only hundreds of mutants, such that there is statistically little chance of identifying an improved variant resulting from a randomly inserted mutation.

Semi-rational design, as inferred from its name, combines the advantageous attributes of the previous two approaches, and is arguably the most effective strategy for improving enzyme properties[126-128] although the degree of success is heavily dependent on the enzyme system. In this approach, select residues are targeted for mutagenesis as they are in rational design, but there is variability in the substitutions made. Site-saturation mutagenesis, for example, is the practice of replacing a residue with all 19 other naturally-encoded amino acid possibilities, to exhaustively evaluate all possible effects that can be made within this single location. Combinatorial mutagenesis takes this a step further by mutating multiple residues simultaneously, so that effects of multiple mutations can be evaluated. Rationally selecting residues helps keep library sizes modest and focused, while introducing a degree of randomness (as by performing site-saturation mutagenesis) increases the diversity of mutants generated and reduces the bias imposed by the researcher.

Depending on the approach selected to introduce mutations, high-throughput screens are often necessary, as the number of variants that can be generated in a single round can as high as millions, making manual evaluation utterly impossible. In the 1990's, strategies for directed evolution began to materialize, with the number of works describing enzymes with altered properties occurring increasingly frequently. Creating enzymes for improving synthetic chemistry methodologies was one of the early goals, with the evolution of an enantioselective lipase being a notable example.[129] Other applications included the degradation of polychlorinated biphenyls,[130] increased thermostability,[131] and artificially expanding the chemical diversity of the genetic code.[132] Over the past 30 years, a plethora of both general and specific screening methodologies and strategies for approaching directed evolution for a multitude of enzymes have been developed, and reviewed extensively.[49, 133-136]

Based on structural information inferred from the crystal structure, results reported in the literature, as well as the absence of an effective high-throughput screen, we ultimately hypothesized that a semi-rational approach was best suited to improve MTG as a biocatalyst for

protein labeling. This work is explained in detail in Chapter 7. Briefly, as we did not know which amino acid position would be optimal for reactivity within the protein substrate to be conjugated by MTG, we generated a library of variant protein substrates. To this effect, we used the B1 domain of Protein G (GB1) that naturally contains a single glutamine residue; each variant expressed a single glutamine at various locations on its structure, probing all secondary structure elements, and was assayed for reactivity. In addition, point mutations within MTG's active site were also evaluated for improved activity towards native GB1 and its variants.

1.9. Summary

Ever since technology has been developed to manipulate enzymes in the laboratory with ease, their applications have become widespread and diverse. Enzymes are biocatalysts, and some of the earliest work performed with them was to aid in synthetic chemistry methodologies. These efforts continue today, with many enzymes being successfully implemented in large-scale industrial processes. They are capable of performing even longstanding difficult, but important and common transformations, such as amide bond synthesis. These transformations can be applied not only to synthetic applications, but to biological macromolecules as well, resulting in site-specific covalent modification. These modifications can be fine-tuned with the help of biologically compatible chemistry, and in the following chapters, transglutaminases will be the focus of combining these chemistries to further explore the application for which they appear to be best suited, site-specific protein labeling.

The research component of this thesis begins in Chapter 3, in which we expand the biocatalytic capability of transglutaminase for peptide synthesis. It is followed, in Chapter 4, by a one-pot biocatalysis protocol in which we successfully combined the conjugation reaction of transglutaminase with the chemical fine-tuning capability of the CuAAC, in which the enzyme had been previously observed to be inactive. Chapter 5 builds upon the concepts described in Chapter 4 by expanding the utility of our one-pot chemoenzymatic bioconjugation strategy to other bioorthogonal chemistries and protein substrates.

Despite the efforts of many researchers, the specificity that transglutaminase shows for its glutamine-containing substrate remains poorly understood, ultimately limiting its utility for widespread specific protein labeling. In Chapter 6, we address our attempts to construct, express,

and purify a form of microbial transglutaminase that could be subjected to crystallographic experiments in the presence of an inhibitor. As no crystal structure of MTG in the presence of a ligand exists, this would provide crucial information to understand and engineering the enzyme to improve its activity with respect to its glutamine substrate. Finally, in Chapter 7, we probe the determinants for specificity of MTG for its glutamine substrate, in which we employ semi-rational design to create multiple versions of a model protein substrate and evaluate MTG's reactivity as a function of secondary and tertiary protein structure.

1.10. References

1. Lippincott-Schwartz, J.; Patterson, G. H., Development and use of fluorescent protein markers in living cells. *Science* **2003,** *300*, 87-91.
2. Zaccolo, M.; De Giorgi, F.; Cho, C. Y.; Feng, L.; Knapp, T.; Negulescu, P. A.; Taylor, S. S.; Tsien, R. Y.; Pozzan, T., A genetically encoded, fluorescent indicator for cyclic AMP in living cells. *Nat. Cell Biol.* **2000,** *2*, 25-29.
3. Ting, A. Y.; Kain, K. H.; Klemke, R. L.; Tsien, R. Y., Genetically encoded fluorescent reporters of protein tyrosine kinase activities in living cells. *Proc. Natl. Acad. Sci. U.S.A.* **2001,** *98*, 15003-15008.
4. Dooley, C. T.; Dore, T. M.; Hanson, G. T.; Jackson, W. C.; Remington, S. J.; Tsien, R. Y., Imaging dynamic redox changes in mammalian cells with green fluorescent protein indicators. *J. Biol. Chem.* **2004,** *279*, 22284-22293.
5. Weller, T. H.; Coons, A. H., Fluorescent antibody studies with agents of varicella and herpes zoster propagated *in vitro*. *Proc. Soc. Exp. Biol. Med.* **1954,** *86*, 789-794.
6. Fink, M. A.; Malmgren, R. A.; Rauscher, F. J.; Orr, H. C.; Karon, M., Application of Immunofluorescence to the Study of Human Leukemia. *J. Natl. Cancer Inst.* **1964,** *33*, 581-588.
7. Butko, M. T.; Yang, J.; Geng, Y.; Kim, H. J.; Jeon, N. L.; Shu, X.; Mackey, M. R.; Ellisman, M. H.; Tsien, R. Y.; Lin, M. Z., Fluorescent and photo-oxidizing TimeSTAMP tags track protein fates in light and electron microscopy. *Nat. Neurosci.* **2012,** *15*, 1742-1751.
8. Veronese, F. M.; Pasut, G., PEGylation, successful approach to drug delivery. *Drug Discov. Today* **2005,** *10*, 1451-1458.
9. Agarwal, P.; Bertozzi, C. R., Site-specific antibody-drug conjugates: the nexus of bioorthogonal chemistry, protein engineering, and drug development. *Bioconjug Chem* **2015,** *26*, 176-192.
10. Junutula, J. R.; Raab, H.; Clark, S.; Bhakta, S.; Leipold, D. D.; Weir, S.; Chen, Y.; Simpson, M.; Tsai, S. P.; Dennis, M. S.; Lu, Y.; Meng, Y. G.; Ng, C.; Yang, J.; Lee, C. C.; Duenas, E.; Gorrell, J.; Katta, V.; Kim, A.; McDorman, K.; Flagella, K.; Venook, R.; Ross, S.; Spencer, S. D.; Lee Wong, W.; Lowman, H. B.; Vandlen, R.; Sliwkowski, M. X.; Scheller, R. H.; Polakis, P.; Mallet, W., Site-specific conjugation of a cytotoxic drug to an antibody improves the therapeutic index. *Nat. Biotechnol.* **2008,** *26*, 925-932.
11. Pasteur, L., Mémoire sur la fermentation de l'acide tartrique. *C. R. Séances Acad. Sci.* **1858,** *46*, 615-618.

12. Sumner, J. B. S., G. F., *Chemistry and Methods of Enzymes*. Academic Press, Inc.: New York, NY, 1953.
13. Payen, A. P., J.F., Mémoire sur la diastase, les principaux produits de ses réactions et leurs applications aux arts industriels. *Annales de chimie et de physique, 2me. série* **1833,** *53*, 73-92.
14. Kühne, W., Über das Verhalten verschiedener organisirter und sog. ungeformter Fermente. *Verhandlungen des naturhistorisch-medicinischen Vereins zu Heidelberg.* **1877,** *1*, 190-193.
15. Fischer, E., Einfluss der Configuration auf die Wirkung der Enzyme. *Ber. Dtsch. Chem. Ges.* **1894,** *27*, 2985-2993.
16. Buchner, E., Alkoholische Gährung ohne Hefezellen (Vorläufige Mitteilung). *Ber. Dtsch. Chem. Ges.* **1897,** *30*, 117-124.
17. Sumner, J. B., The isolation and crystallization of the enzyme urease. *J. Biol. Chem.* **1926,** *69*, 435-441.
18. Perrett, D., From 'protein' to the beginnings of clinical proteomics. *Proteomics Clin Appl.* **2007,** *1*, 720-738.
19. Picton, H. L., S. E., Solution and pseudo-solution, Part 1. *J. Chem. Soc.* **1892,** *61*, 148-172.
20. Hardy, W. B. W., W. C. D., On the coagulation of proteid by electricity. *J. Physiol.* **1896,** *24*, 288-304.
21. Peterson, E. A. S., H. A., Chromatography of proteins. I. Cellulose ion-exchange adsorbents. *J. Am. Chem. Soc.* **1954,** *76*, 751-755.
22. Porath, J.; Flodin, P., Gel filtration: a method for desalting and group separation. *Nature* **1959,** *183*, 1657-1659.
23. Hjertén, S., The Preparation of Agarose Spheres for Chromatography of Molecules and Particles. *Biochimica et biophysica acta* **1964,** *79*, 393-398.
24. Whitesides, G. M.; Wong, C.-W., Enzymes as Catalysts in Synthetic Organic Chemistry. *Angew. Chem., Int. Ed.* **1985,** *24*, 617-638.
25. Klibanov, A. M., Improving enzymes by using them in organic solvents. *Nature* **2001,** *409*, 241-246.
26. Drauz, K.; Gröger, H.; May, O., *Enzyme Catalysis in Organic Synthesis, 3rd ed.* Wiley-VCH: Weinheim, 2012.
27. Joshi, T.; Graham, B.; Spiccia, L., Macrocyclic metal complexes for metalloenzyme mimicry and sensor development. *Acc. Chem. Res.* **2015,** *48*, 2366-2379.
28. Sallmann, M.; Limberg, C., Utilizing the Trispyrazolyl Borate Ligand for the Mimicking of O_2-Activating Mononuclear Nonheme Iron Enzymes. *Acc. Chem. Res.* **2015,** *48*, 2734-2743.
29. Raynal, M.; Ballester, P.; Vidal-Ferran, A.; van Leeuwen, P. W., Supramolecular catalysis. Part 2: artificial enzyme mimics. *Chem. Soc. Rev.* **2014,** *43*, 1734-1787.
30. Kuah, E.; Toh, S.; Yee, J.; Ma, Q.; Gao, Z., Enzyme Mimics: Advances and Applications. *Chemistry* **2016,** *22* (25), 8404-8430.
31. Arnold, F. H., Directed evolution: Creating biocatalysts for the future. *Chem. Eng. Sci.* **1996,** *51*, 5091-5102.
32. Jackson, D. A.; Symons, R. H.; Berg, P., Biochemical method for inserting new genetic information into DNA of Simian Virus 40: circular SV40 DNA molecules containing lambda phage genes and the galactose operon of *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* **1972,** *69*, 2904-2909.

33. Johnson, I. S., Human insulin from recombinant DNA technology. *Science* **1983,** *219*, 632-637.
34. Goeddel, D. V.; Kleid, D. G.; Bolivar, F.; Heyneker, H. L.; Yansura, D. G.; Crea, R.; Hirose, T.; Kraszewski, A.; Itakura, K.; Riggs, A. D., Expression in Escherichia coli of chemically synthesized genes for human insulin. *Proc. Natl. Acad. Sci. U.S.A.* **1979,** *76*, 106-110.
35. Auerbach, C., The chemical production of mutations. The effect of chemical mutagens on cells and their genetic material is discussed. *Science* **1967,** *158*, 1141-1147.
36. Brandenburger, A.; Godson, G. N.; Radman, M.; Glickman, B. W.; van Sluis, C. A.; Doubleday, O. P., Radiation-induced base substitution mutagenesis in single-stranded DNA phage M13. *Nature* **1981,** *294*, 180-182.
37. Bartlett, J. M.; Stirling, D., A short history of the polymerase chain reaction. *Methods Mol. Biol.* **2003,** *226*, 3-6.
38. Dalbadie-McFarland, G.; Cohen, L. W.; Riggs, A. D.; Morin, C.; Itakura, K.; Richards, J. H., Oligonucleotide-directed mutagenesis as a general and powerful method for studies of protein function. *Proc. Natl. Acad. Sci. U.S.A.* **1982,** *79*, 6409-13.
39. Winter, G.; Fersht, A. R.; Wilkinson, A. J.; Zoller, M.; Smith, M., Redesigning enzyme structure by site-directed mutagenesis: tyrosyl tRNA synthetase and ATP binding. *Nature* **1982,** *299*, 756-758.
40. Villafranca, J. E.; Howell, E. E.; Voet, D. H.; Strobel, M. S.; Ogden, R. C.; Abelson, J. N.; Kraut, J., Directed mutagenesis of dihydrofolate reductase. *Science* **1983,** *222*, 782-788.
41. Poteete, A. R.; Rennell, D.; Bouvier, S. E., Functional significance of conserved amino acid residues. *Proteins* **1992,** *13*, 38-40.
42. Cornforth, J. W.; Redmond, J. W.; Eggerer, H.; Buckel, W.; Gutschow, C., Asymmetric methyl groups, and the mechanism of malate synthase. *Nature* **1969,** *221*, 1212-1213.
43. Schray, K. J.; O'Connell, E. L.; Rose, I. A., Inactivation of muscle triose phosphate isomerase by D- and L-glycidol phosphate. *J. Biol. Chem.* **1973,** *248*, 2214-2218.
44. Battersby, A. R.; Kelsey, J. E.; Staunton, J.; Suckling, K. E., Studies of enzyme-mediated reactions. 3. Stereoselective labelling at C-2 of tyramine: stereochemistry of hydroxylation at saturated carbon. *J. Chem. Soc., Perkin Trans. 1* **1973,** *15*, 1609-1615.
45. Battersby, A. R. H., E.
McDonald, E., Biosynthesis of type-III porphyrins: nature of the rearrangement process. *J. S. C. Chem. Comm.* **1973,** (13), 442-443.
46. Coelho, P. S.; Brustad, E. M.; Kannan, A.; Arnold, F. H., Olefin cyclopropanation via carbene transfer catalyzed by engineered cytochrome P450 enzymes. *Science* **2013,** *339*, 307-310.
47. Pahari, P.; Kharel, M. K.; Shepherd, M. D.; van Lanen, S. G.; Rohr, J., Enzymatic total synthesis of defucogilvocarcin M and its implications for gilvocarcin biosynthesis. *Angew. Chem., Int .Ed.* **2012,** *51*, 1216-1220.
48. Sheldon, R. A.; Pereira, P. C., Biocatalysis engineering: the big picture. *Chem. Soc. Rev.* **2017,** *46*, 2678-2691.
49. Zhao, H.; Chockalingam, K.; Chen, Z., Directed evolution of enzymes and pathways for industrial biocatalysis. *Curr. Opin. Biotechnol.* **2002,** *13*, 104-110.
50. Reetz, M. T., Biocatalysis in organic chemistry and biotechnology: past, present, and future. *J. Am. Chem. Soc.* **2013,** *135*, 12480-12496.

51. Koeller, K. M.; Wong, C. H., Enzymes for chemical synthesis. *Nature* **2001,** *409*, 232-240.
52. Savile, C. K.; Janey, J. M.; Mundorff, E. C.; Moore, J. C.; Tam, S.; Jarvis, W. R.; Colbeck, J. C.; Krebber, A.; Fleitz, F. J.; Brands, J.; Devine, P. N.; Huisman, G. W.; Hughes, G. J., Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture. *Science* **2010,** *329*, 305-309.
53. Farwell, C. C.; Zhang, R. K.; McIntosh, J. A.; Hyster, T. K.; Arnold, F. H., Enantioselective Enzyme-Catalyzed Aziridination Enabled by Active-Site Evolution of a Cytochrome P450. *ACS Cent. Sci.* **2015,** *1*, 89-93.
54. Hernandez, K. E.; Renata, H.; Lewis, R. D.; Jennifer Kan, S. B.; Zhang, C.; Forte, J.; Rozzell, D.; McIntosh, J. A.; Arnold, F. H., Highly Stereoselective Biocatalytic Synthesis of Key Cyclopropane Intermediate to Ticagrelor. *ACS Catal.* **2016,** *6*, 7810-7813.
55. Ricca, E. B., B.; Schrittwieser, J. H., Multi-Enzymatic Cascade Reactions: Overview and Perspectives. *Adv. Synth. Catal.* **2011,** *353*, 2239-2262.
56. France, S. P.; Hepworth, L. J.; Turner, N. J.; Flitsch, S. L., Constructing Biocatalytic Cascades: In Vitro and in Vivo Approaches to de Novo Multi-Enzyme Pathways. *ACS Catal.* **2017,** *7*, 710-724.
57. Pattabiraman, V. R.; Bode, J. W., Rethinking amide bond synthesis. *Nature* **2011,** *480*, 471-479.
58. de Figueiredo, R. M.; Suppo, J. S.; Campagne, J. M., Nonclassical Routes for Amide Bond Formation. *Chem. Rev.* **2016,** *116*, 12029-12122.
59. Merrifield, R. B., Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. *J. Am. Chem. Soc.* **1963,** *85*, 2149-2154.
60. Coin, I.; Beyermann, M.; Bienert, M., Solid-phase peptide synthesis: from standard procedures to the synthesis of difficult sequences. *Nat. Protoc.* **2007,** *2*, 3247-3256.
61. Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B., Synthesis of proteins by native chemical ligation. *Science* **1994,** *266*, 776-779.
62. Torbeev, V. Y.; Kent, S. B., Convergent chemical synthesis and crystal structure of a 203 amino acid "covalent dimer" HIV-1 protease enzyme molecule. *Angew. Chem., Int. Ed.* **2007,** *46*, 1667-1670.
63. Ghose, A. K.; Viswanadhan, V. N.; Wendoloski, J. J., A knowledge-based approach in designing combinatorial or medicinal chemistry libraries for drug discovery. 1. A qualitative and quantitative characterization of known drug databases. *J. Comb. Chem.* **1999,** *1*, 55-68.
64. Roughley, S. D.; Jordan, A. M., The medicinal chemist's toolbox: an analysis of reactions used in the pursuit of drug candidates. *J. Med. Chem.* **2011,** *54*, 3451-3479.
65. Ishihara, K.; Ohara, S.; Yamamoto, H., 3,4,5-Trifluorobenzeneboronic Acid as an Extremely Active Amidation Catalyst. *J. Org. Chem.* **1996,** *61*, 4196-4197.
66. Al-Zoubi, R. M.; Marion, O.; Hall, D. G., Direct and waste-free amidations and cycloadditions by organocatalytic activation of carboxylic acids at room temperature. *Angew. Chem.. Int. Ed.* **2008,** *47*, 2876-2879.
67. Marcelli, T., Mechanistic insights into direct amide bond formation catalyzed by boronic acids: halogens as Lewis bases. *Angew. Chem., Int. Ed.* **2010,** *49*, 6840-6843.
68. MacMillan, D. S.; Murray, J.; Sneddon, H. F.; Jamieson, C.; Watson, A. J. B., Evaluation of alternative solvents in common amide coupling reactions: replacement of dichloromethane and *N*,*N*-dimethylformamide. *Green Chem.* **2013,** *15*, 596-600.

69. Jakubke, H.; Kuhl, P.; Könnecke, A., Basic Principles of Protease-Catalyzed Peptide Bond Formation. *Angew. Chem., Int. Ed.* **1985,** *24*, 85-93.
70. Bordusa, F., Nonconventional amide bond formation catalysis: programming enzyme specificity with substrate mimetics. *Braz. J. Med. Biol. Res.* **2000,** *33*, 469-485.
71. Kaufman, S.; Neurath, H.; Schwert, G. W., The specific peptidase and esterase activities of chymotrypsin. *J. Biol. Chem.* **1949,** *177*, 793-814.
72. Martinek, K.; Semenov, A. N.; Berezin, V., Enzymatic synthesis in biphasic aqueous-organic systems. I. Chemical equilibrium shift. *Biophys. Acta, Enzymol.* **1981,** *658*, 76-89.
73. Kitaguchi, H.; Klibanov, A. M., Enzymic peptide synthesis via segment condensation in the presence of water mimics. *J. Am. Chem. Soc.* **1989,** *111*, 9272-9273.
74. Goswami, A.; Van Lanen, S. G., Enzymatic strategies and biocatalysts for amide bond formation: tricks of the trade outside of the ribosome. *Mol. Biosyst.* **2015,** *11*, 338-353.
75. Liljeblad, A.; Kallio, P.; Vainio, M.; Niemi, J.; Kanerva, L. T., Formation and hydrolysis of amide bonds by lipase A from *Candida antarctica*; exceptional features. *Org. Biomol. Chem.* **2010,** *8*, 886-95.
76. Cockburn, I.; Henderson, R.; Orsenigo, L.; Pisano, G., *US Industry in 2000, Studies inCompetitive Performance.* The National Academy Press.: Washington, DC, 1999.
77. Walsh, C. T.; Garneau-Tsodikova, S.; Gatto, G. J., Jr., Protein posttranslational modifications: the chemistry of proteome diversifications. *Angew. Chem., Int. Ed.* **2005,** *44*, 7342-7372.
78. Chalker, J. M.; Bernardes, G. J.; Lin, Y. A.; Davis, B. G., Chemical modification of proteins at cysteine: opportunities in chemistry and biology. *Chem. Asian J.* **2009,** *4*, 630-640.
79. Gaertner, H. F.; Offord, R. E., Site-specific attachment of functionalized poly(ethylene glycol) to the amino terminus of proteins. *Bioconjug. Chem.* **1996,** *7*, 38-44.
80. Wang, L.; Schultz, P. G., Expanding the genetic code. *Angew. Chem., Int. Ed.* **2004,** *44*, 34-66.
81. Xie, J.; Schultz, P. G., A chemical toolkit for proteins--an expanded genetic code. *Nat. Rev. Mol. Cell Biol.* **2006,** *7*, 775-782.
82. Lang, K.; Chin, J. W., Cellular incorporation of unnatural amino acids and bioorthogonal labeling of proteins. *Chem. Rev.* **2014,** *114*, 4764-4806.
83. Liu, W.; Brock, A.; Chen, S.; Chen, S.; Schultz, P. G., Genetic incorporation of unnatural amino acids into proteins in mammalian cells. *Nat. Methods* **2007,** *4*, 239-244.
84. Neumann, H.; Wang, K.; Davis, L.; Garcia-Alai, M.; Chin, J. W., Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. *Nature* **2010,** *464*, 441-444.
85. Hohsaka, T.; Sisido, M., Incorporation of non-natural amino acids into proteins. *Curr. Opin. Chem. Biol.* **2002,** *6*, 809-815.
86. Mendel, D.; Ellman, J. A.; Chang, Z.; Veenstra, D. L.; Kollman, P. A.; Schultz, P. G., Probing protein stability with unnatural amino acids. *Science* **1992,** *256*, 1798-1802.
87. Wang, L.; Xie, J.; Deniz, A. A.; Schultz, P. G., Unnatural amino acid mutagenesis of green fluorescent protein. *J. Org. Chem.* **2003,** *68*, 174-176.
88. Wang, J.; Xie, J.; Schultz, P. G., A genetically encoded fluorescent amino acid. *J. Am. Chem. Soc.* **2006,** *128*, 8738-8739.
89. Rashidian, M.; Dozier, J. K.; Distefano, M. D., Enzymatic labeling of proteins: techniques and approaches. *Bioconjug. Chem.* **2013,** *24*, 1277-1294.

90. Popp, M. W.; Antos, J. M.; Grotenbreg, G. M.; Spooner, E.; Ploegh, H. L., Sortagging: a versatile method for protein labeling. *Nat. Chem. Biol.* **2007,** *3*, 707-708.
91. Lin, C. W.; Ting, A. Y., Transglutaminase-catalyzed site-specific conjugation of small-molecule probes to proteins in vitro and on the surface of living cells. *J. Am. Chem. Soc.* **2006,** *128*, 4542-4543.
92. Oteng-Pabi, S. K.; Pardin, C.; Stoica, M.; Keillor, J. W., Site-specific protein labelling and immobilization mediated by microbial transglutaminase. *Chem. Commun.* **2014,** *50*, 6604-6606.
93. Sletten, E. M.; Bertozzi, C. R., Bioorthogonal chemistry: fishing for selectivity in a sea of functionality. *Angew. Chem., Int. Ed.* **2009,** *48*, 6974-6998.
94. Lang, K.; Chin, J. W., Bioorthogonal reactions for labeling proteins. *ACS Chem. Biol.* **2014,** *9*, 16-20.
95. Cornish, V. W.; Hahn, K. M.; Schultz, P. G., Site-specific protein modification using a ketone handle. *J. Am. Chem. Soc.* **1996,** *118*, 8150-8151.
96. Saxon, E.; Bertozzi, C. R., Cell surface engineering by a modified Staudinger reaction. *Science* **2000,** *287*, 2007-2010.
97. Staudinger, H.; Meyer, J., Über neue organische Phosphorverbindungen III. Phosphinmethylenderivate und Phosphinimine. *Helv. Chim. Acta.* **1919,** *2*, 635-646.
98. Saxon, E.; Luchansky, S. J.; Hang, H. C.; Yu, C.; Lee, S. C.; Bertozzi, C. R., Investigating cellular metabolism of synthetic azidosugars with the Staudinger ligation. *J. Am. Chem. Soc.* **2002,** *124*, 14893-14902.
99. Kolb, H. C.; Finn, M. G.; Sharpless, K. B., Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angew. Chem., Int. Ed.* **2001,** *40*, 2004-2021.
100. Michael, A., Uber die Einwirkung von Diazobenzolimid auf Acetylendicarbonsauremethylester. *J. Prakt. Chem.* **1893,** *48*, 94.
101. Huisgen, R., 1,3-Dipolar cycloadditions. Past and future. *Angew. Chem., Int. Ed.* **1963,** *2*, 565-598.
102. Fokin, V. V., Click imaging of biochemical processes in living systems. *ACS Chem. Biol.* **2007,** *2*, 775-778.
103. Moses, J. E.; Moorhouse, A. D., The growing applications of click chemistry. *Chem. Soc. Rev.* **2007,** *36*, 1249-1262.
104. Tron, G. C.; Pirali, T.; Billington, R. A.; Canonico, P. L.; Sorba, G.; Genazzani, A. A., Click chemistry reactions in medicinal chemistry: Applications of the 1,3-dipolar cycloaddition between azides and alkynes. *Med. Res. Rev.* **2007,** *28*, 278-308.
105. Lutz, J. F.; Zarafshani, Z., Efficient construction of therapeutics, bioconjugates, biomaterials and bioactive surfaces using azide-alkyne "click" chemistry. *Adv. Drug Deliv. Rev.* **2008,** *60*, 958-970.
106. Meldal, M.; Tornøe, C. W., Cu-catalyzed azide-alkyne cycloaddition. *Chem. Rev.* **2008,** *108*, 2952-3015.
107. Wolbers, F.; ter Braak, P.; Le Gac, S.; Luttge, R.; Andersson, H.; Vermes, I.; van den Berg, A., Viability study of HL60 cells in contact with commonly used microchip materials. *Electrophoresis* **2006,** *27*, 5073-5080.
108. Wu, Z.; Fernandez-Lima, F. A.; Russell, D. H., Amino acid influence on copper binding to peptides: cysteine versus arginine. *J. Am. Soc. Mass. Spectrom.* **2010,** *21*, 522-533.
109. Alder, K.; Stein, G., Über das abgestufte Additionsvermögen von ungesättigten Ringsystemen. II. *Justus Liebigs Ann. Chem.* **1933,** *501*, 1-48.

110. Wittig, G.; Krebs, A., Zur Existenz niedergliedriger Cycloalkine, I. *Chem. Ber.* **1961,** *94*, 3260-3275.
111. Agard, N. J.; Prescher, J. A.; Bertozzi, C. R., A strain-promoted [3 + 2] azide-alkyne cycloaddition for covalent modification of biomolecules in living systems. *J. Am. Chem. Soc.* **2004,** *126*, 15046-15047.
112. Jewett, J. C.; Sletten, E. M.; Bertozzi, C. R., Rapid Cu-free click chemistry with readily synthesized biarylazacyclooctynones. *J. Am. Chem. Soc.* **2010,** *132*, 3688-3690.
113. Baskin, J. M.; Prescher, J. A.; Laughlin, S. T.; Agard, N. J.; Chang, P. V.; Miller, I. A.; Lo, A.; Codelli, J. A.; Bertozzi, C. R., Copper-free click chemistry for dynamic in vivo imaging. *Proc. Natl. Acad. Sci. U.S.A.* **2007,** *104*, 16793-16797.
114. Laughlin, S. T.; Baskin, J. M.; Amacher, S. L.; Bertozzi, C. R., In vivo imaging of membrane-associated glycans in developing zebrafish. *Science* **2008,** *320*, 664-667.
115. Patterson, D. M.; Nazarova, L. A.; Prescher, J. A., Finding the right (bioorthogonal) chemistry. *ACS Chem. Biol.* **2014,** *9*, 592-605.
116. Griffin, M.; Casadio, R.; Bergamini, C. M., Transglutaminases: nature's biological glues. *Biochem. J.* **2002,** *368*, 377-396.
117. Yokoyama, K.; Nio, N.; Kikuchi, Y., Properties and applications of microbial transglutaminase. *Appl. Microbiol. Biotechnol.* **2004,** *64*, 447-454.
118. Fontana, A.; Spolaore, B.; Mero, A.; Veronese, F. M., Site-specific modification and PEGylation of pharmaceutical proteins mediated by transglutaminase. *Adv. Drug. Deliv. Rev.* **2008,** *60*, 13-28.
119. Tang, R.; Guan, Z.; He, Y.; Zhu, W., Enzyme-catalyzed Henry (nitroaldol) reaction. *J. Mol. Cat. B.* **2010,** *63*, 62-67.
120. Ohtsuka, T.; Sawa, A.; Kawabata, R.; Nio, N.; Motoki, M., Substrate specificities of microbial transglutaminase for primary amines. *J. Agric. Food. Chem.* **2000,** *48*, 6230-6233.
121. Gundersen, M. T.; Keillor, J. W.; Pelletier, J. N., Microbial transglutaminase displays broad acyl-acceptor substrate specificity. *Appl. Microbiol. Biotechnol.* **2014,** *98*, 219-230.
122. Arnold, F. H., When blind is better: protein design by evolution. *Nat. Biotechnol.* **1998,** *16*, 617-8.
123. Arnold, F. H., Design by Directed Evolution. *Acc. Chem. Res.* **1998,** *31*, 125-131.
124. Bornscheuer, U. T.; Pohl, M., Improved biocatalysts by directed evolution and rational protein design. *Curr. Opin. Chem. Biol.* **2001,** *5*, 137-143.
125. Xiao, H.; Bao, Z.; Zhao, H., High Throughput Screening and Selection Methods for Directed Enzyme Evolution. *Ind. Eng. Chem. Res.* **2015,** *54*, 4011-4020.
126. Chica, R. A.; Doucet, N.; Pelletier, J. N., Semi-rational approaches to engineering enzyme activity: combining the benefits of directed evolution and rational design. *Curr. Opin. Biotechnol.* **2005,** *16*, 378-384.
127. Lutz, S., Beyond directed evolution--semi-rational protein engineering and design. *Curr. Opin. Biotechnol.* **2010,** *21*, 734-743.
128. Davids, T.; Schmidt, M.; Bottcher, D.; Bornscheuer, U. T., Strategies for the discovery and engineering of enzymes for biocatalysis. *Curr. Opin. Chem. Biol.* **2013,** *17*, 215-220.
129. Reetz, M. T.; Zonta, A.; Schimossek, K.; Liebeton, K.; Jaeger, K. E., Creation of enantioselective biocatalysts for organic chemistry by in vitro evolution. *Angew. Chem. Int. Ed. Engl.* **1997,** *36*, 2830-2832.

130. Kumamaru, T.; Suenaga, H.; Mitsuoka, M.; Watanabe, T.; Furukawa, K., Enhanced degradation of polychlorinated biphenyls by directed evolution of biphenyl dioxygenase. *Nat. Biotechnol.* **1998,** *16* (7), 663-666.
131. Giver, L.; Gershenson, A.; Freskgard, P. O.; Arnold, F. H., Directed evolution of a thermostable esterase. *Proc. Natl. Acad. Sci. U.S.A.* **1998,** *95*, 12809-12813.
132. Liu, D. R.; Magliery, T. J.; Pastrnak, M.; Schultz, P. G., Engineering a tRNA and aminoacyl-tRNA synthetase for the site-specific incorporation of unnatural amino acids into proteins in vivo. *Proc. Natl. Acad. Sci. U.S.A.* **1997,** *94*, 10092-10097.
133. Lutz, S.; Patrick, W. M., Novel methods for directed evolution of enzymes: quality, not quantity. *Curr. Opin. Biotechnol.* **2004,** *15*, 291-297.
134. Packer, M. S.; Liu, D. R., Methods for the directed evolution of proteins. *Nat. Rev. Genet.* **2015,** *16*, 379-394.
135. Currin, A.; Swainston, N.; Day, P. J.; Kell, D. B., Synthetic biology for the directed evolution of protein biocatalysts: navigating sequence space intelligently. *Chem. Soc. Rev.* **2015,** *44*, 1172-1239.
136. Ebert, M. C.; Pelletier, J. N., Computational tools for enzyme improvement: why everyone can - and should - use them. *Curr. Opin. Chem. Biol.* **2017,** *37*, 89-96.

Chapter 2 - Biotechnological applications of transglutaminase

2.1 Context

This thesis is dedicated to presenting the diverse applications of enzymes, with a specific focus on one enzyme class in particular, transglutaminases. As described in Section 1.6, transglutaminases naturally catalyze amide bond formation using glutamine and lysine side chains. To introduce the intricacies about transglutaminases, we present the following chapter, which is a published review for the journal *Biomolecules*, accepted for publication in October 2013. Entitled *Biotechnological Applications of Transglutaminases*, it covers recent advances in the applications of transglutaminases outside of the food processing industry. Beyond biotechnological applications, we also describe in detail recent progress made in being able to implement and manipulate transglutaminases with ease, such as different expression systems, assays, and investigations into their substrate specifies. We illustrate that transglutaminases, despite their current stumbling blocks, are of high interest owing to their abundant potential for new solutions to scientific challenges.

My contribution to this review was the conceptualization, literature search and writing, which was supported by Prof. Joelle Pelletier.

Biotechnological applications of transglutamianses

Natalie M. Rachel[1,2,3] and Joelle N. Pelletier[1,2,3,4]

[1] Département de Chimie, Université de Montréal, 2900 Boulevard Edouard-Montpetit, Montréal, Québec, H3T 1J4, Canada

[2] CGCC, the Center in Green Chemistry and Catalysis, Montréal, H3A 0B8, Canada

[3] PROTEO, the Québec Network for Protein Function, Structure and Engineering, Québec, G1V 0A6, Canada

[4] Départment de Biochimie, Université de Montréal, 2900 Boulevard Edouard-Montpetit, Montréal, Québec, H3T 1J4, Canada

Biomolecules **2013**, *3*, 870-888.

Corresponding Author: Joelle N. Pelletier <joelle.pelletier@umontreal.ca>

2.2 Abstract

In nature, transglutaminases catalyze the formation of amide bonds between proteins to form insoluble protein aggregates. This specific function has long been exploited in the food and textile industries as a protein cross-linking agent to alter the texture of meat, wool, and leather. In recent years, biotechnological applications of transglutaminases have come to light in areas ranging from material sciences to medicine. There has also been a substantial effort to further investigate the fundamentals of transglutaminases, as many of their characteristics remain poorly understood. Those studies also work towards the goal of developing transglutaminases as more efficient catalysts. Progress in this area includes structural information and novel chemical and biological assays. Here, we review recent achievements in this area in order to illustrate the versatility of transglutaminases.

2.3 Introduction

Harnessing the catalytic properties of enzymes is a field of research that continues to receive increasing attention. One of the most attractive characteristics of biocatalysts is that they are often highly chemo-, regio-, and stereo-selective. This provides potential for highly specific chemical transformations of complex, functionalized molecules. Additionally, biocatalysts are non-toxic, degradable, and functional in aqueous media at moderate temperatures and pressure, making them of high interest in the development of environmentally respectful synthetic methodologies. Due to these desirable properties, chemists are increasingly incorporating enzymes into their reaction schemes.

The synthesis of amide bonds has the potential to benefit greatly from biocatalysis. The high stability of the amide functionality makes it one of the most favorable and commonly used in organic synthesis.[1] Some examples of compounds containing biocatalyzed amide bonds are found in the large-scale production of Atorvastatin (commercialized as Lipitor™), Nylon, penicillin, and aspartame. The high activation barrier to amide-bond formation is synthetically challenging; further development of biocatalysts for formation of a broad range of compounds remains of interest. Transglutaminases (TGases) are a family of enzymes (EC 2.3.2.13) that catalyze an acyl-transfer reaction between the γ-carboxamide group of a protein- or peptide-bound glutamine and the ε-amino group of a lysine residue, resulting in the formation of a relatively protease-resistant isopeptide bond (Figure 2-1).[2] TGases, having evolved to catalyze the formation of amide bonds with little competition from the reverse hydrolytic reaction, are a promising biocatalytic alternative to classical organic chemistry for amide bond synthesis.

TGases have been identified in many different of taxonomic groups, including microorganisms, plants, invertebrates, and mammals.[3] With respect to application, the vast majority of research has been done on two forms of the enzyme: the first is a calcium-dependant TGase found in tissues of animals and humans, referred to as transglutaminase 2 (TG2). TG2 is implicated in a number of physiological roles including endocytosis, cell-matrix assembly, apoptosis, and cellular adhesive processes.[4-6] There is much interest in studying TG2 from a medical standpoint to better understand its role in disease, including

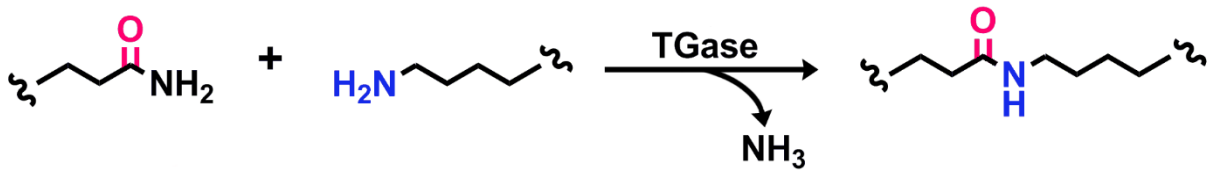


Figure 2-1 Amide bond formation catalyzed by TGase.

Peptide- or protein-bound glutamines and lysines serve as substrates, releasing ammonia in the process.

cataract formation,[7] celiac sprue,[8] and psoriasis.[9] The second enzyme is a calcium-independent, microbial transglutaminase (MTG), which was first isolated from *Streptomyces mobaraense*[10] and has since been isolated from other microbial strains, including, but not limited to, *S. griseocarneum*, *S. hygroscopicus*, and *B. subtilis*.[11-12] Both types of TGases have been studied extensively in academia and industry. Mechanisms for the reaction catalyzed by both TGase types have been proposed. The catalytic triad characteristic to cysteine proteases is present in the human factor XIII TGase (Cys314, His373, and Asp396).[13] These residues correspond to Cys276, His334, and Asp358 in the highly conserved active site of guinea pig TG2.[14] In the proposed mechanism, the cysteine and the histidine residues are principally involved in the acyl transfer reaction, where the aspartic acid residue hydrogen bonds with the histidine, maintaining a catalytically-competent orientation. The crystal structure of MTG revealed that this triad is not conserved; rather, it was proposed that MTG uses a cysteine protease-like mechanism in which Asp255 plays the role of the histidine residue in factor XIII-like TGases.[15]

Of the two, MTG is more robust, and is commonly employed as a tool in the food industry to catalyze the cross-linking of meat, soy, and wheat proteins to improve and modify their texture and tensile properties.[11, 16] Despite the medical importance of TG2 and widespread industrial use of MTG, many properties such as ligand binding, catalytic mechanism, and function in health and disease remain poorly understood, ultimately hindering further successful integration of these enzymes into novel applications and processes. Nonetheless, researchers are continually looking for ways to exploit the cross-linking activity of TGases for novel applications outside of the fields of human physiology and the food industry. Examples include tissue engineering,[17] as well as textile and leather processing.[18] These applications generally utilize TGase to serve the same purpose it does in the food industry: non-specific protein cross-

linking to provide improved physical and textural properties. A recent example involved increasing the mechanical strength of amniotic membrane, for applications in regenerative medicine.[19] The advances made in these fields have been covered in recent reviews,[20-21] and will not be discussed in detail here. This review focuses on recent advances made in studying TGases in the scope of biotechnology and characterization, including advances in assay development, site-specific modification of biomacromolecules, and protein labeling.

2.4 Production and engineering of TGases

2.4.1 Transglutaminase expression and purification

Both TG2 and MTG are readily recombinantly expressed and purified in bacterial hosts.[22-23] Using these methods, the production of TG2 in a hexa-histidine labeled form has become routine,[22, 24-25] although other forms of TG2 can remain a challenge to obtain in good yield. A complementary technique for the purification of hTG2 was recently reported, in which hTG2 was expressed as a fusion with glutathione S-transferase (GST) and followed by a one-step affinity chromatography purification.[26] Unlike TG2, the purification of the most widely used MTG (from *S. mobaraensis* and homologs) is complicated by the fact that the native enzyme is expressed as a zymogen (pro-MTG); a 46-residue *N*-terminal pro-sequence must be proteolytically cleaved in order for MTG to be rendered functional. There are reports of other MTGs that can be directly expressed as recombinant, active enzymes,[27-28] however these are not as well characterized. Three solutions to this problem have been reported: (1) expression of pro-MTG followed by in vitro activation using a protease[29-30]; (2) direct expression of insoluble MTG lacking its *N*-terminal pro-sequence (mature MTG) followed by refolding,[23] or (3) co-expression of pro-MTG with the activating protease in *Streptomyces*[31] or *E. coli* .[32] Each of these strategies has limitations: the first strategy can achieve high yields and activity, but involves lengthy activation methodologies (N.M. Rachel and J.N. Pelletier, unpublished observations). The second often leads to a low expression or insoluble protein, while the third strategy can result in protein degradation, affecting the yield.[33]

Recently, MTG from *S. hygroscopicus* was successfully produced in its active form in *E. coli* by simultaneously expressing the pro-sequence and mature MTG as separate

polypeptides under the control of a single T7 promoter.[34] Expression of the pro-sequence prior to the mature MTG polypeptide was found to be essential for activity, as well as an *N*-terminal pelB sequence for periplasmic localization. This supports the hypothesis that the pro-sequence is required for proper folding and soluble expression of MTG. Improved efficiency of MTG maturation in *Streptomyces* was also recently reported, by engineering more protease-labile linkers into the pro-propeptide.[35] The structural basis for this requirement can be understood upon observing the crystal structure of pro-MTG, which was determined at 1.9-Å resolution (Figure 2-2).[36] The pro-sequence folds into an α-helix, covering the putative active site cleft by adopting an L-shaped conformation. The active site cleft is predominantly composed of two flexible loop regions, explaining how the presence of this ordered helix stimulates proper folding, in a fashion similar to that of the pro-sequences for subtilisin BPN' and other proteases.[37]

Two biophysical studies focusing on the detailed mechanism of unfolding and refolding of MTG were reported by Suzuki and colleagues.[38-39] In the first, a two-step refolding process of acid-denatured MTG was proposed after probing the effect of pH and salt concentration. The authors then applied this protocol to pro-MTG in the second report, such that by partially unfolding the enzyme, the internal residues would be exposed when in the presence of a deuterated solvent. This solvent exposure is often necessary so that hydrogen back-exchange occurs for all residues in the protein, allowing for accurate measurements using nuclear magnetic resonance (NMR) spectroscopy to be taken. Complete back-exchanges for internal residues of pro-MTG were observed by NMR spectroscopy, and the authors were able to recover the properly folded form of both pro-MTG and mature MTG, reporting refolding yields of 84% and 40%, respectively.

2.4.2 Engineering TGases for altered function and properties

The design of enzymes with improved or non-native properties has become a common approach.[40-42] Engineering TGases may provide solutions to increase their applicability in biocatalytic contexts. TG2 has been engineered towards catalyzing amide bond formation

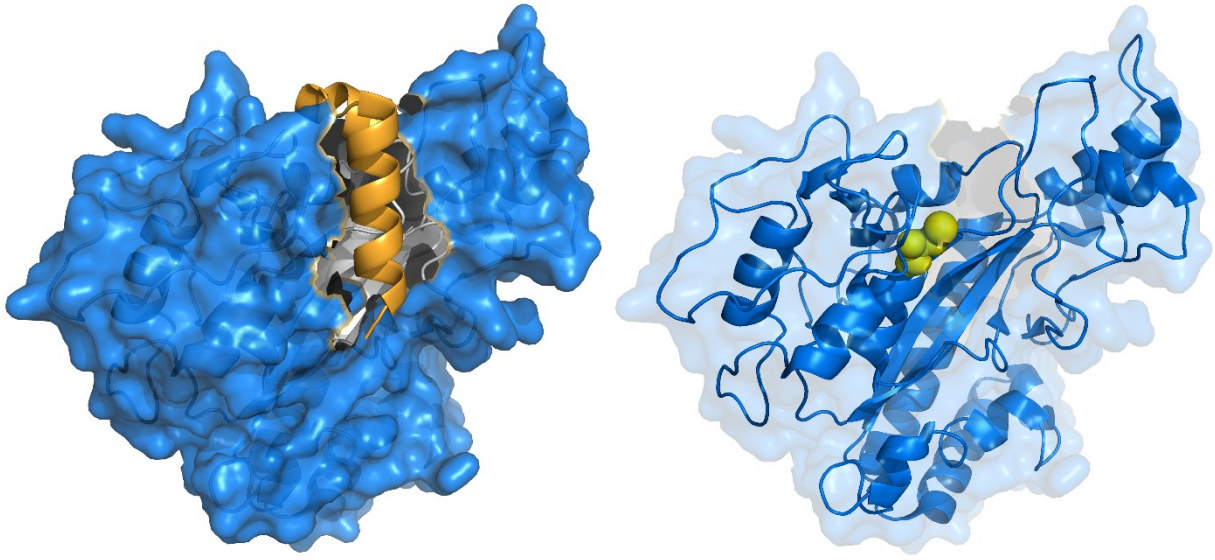


Figure 2-2 Crystal structure of MTG (PDB ID: 3IU0).

The active site of the zymogen is covered (left) by an α-helix (gold), which is cleaved upon activation, exposing the active site cysteine residue (right, yellow spheres) that is critical for activity.

between various synthetic substrates, by altering its substrate specificity.[43] A model peptide substrate, benzyloxycarbonyl-L-glutaminylglycine (Z-QG), was modified to yield a fluorescent umbelliferyl ester derivative (Z-GU) in order to screen for variants of TG2 with altered transpeptidase activity. Two separate point mutations were identified, which broaden the substrate scope of TG2, resulting in variants that can accept threonine methyl ester. To the best of our knowledge, this remains the only study focused on evolving TG2, and so the efforts in this field remain largely conservative.

With respect to MTG, logistical complications of expressing the mature enzyme and the lack of a simple, high-throughput screening assay remain major challenges for engineering. Nonetheless, enhancing the activity and thermostability of MTG has been probed by two different studies. Pietzsch and colleagues[44] performed random mutagenesis using a microtiter plate-based screening method adapted to the standard hydroxamate assay[45] to measure activity. A library of 5500 clones generated randomly by error-prone PCR was initially screened, 70 of which showed higher activity following incubation at 60°C. Following another round of mutagenesis, the nine clones with the highest residual activity were further characterized. The

single-residue variant Ser2Pro was found to have an optimal functioning temperature of 55°C, an improvement of 5°C compared to the native enzyme. More recent efforts using saturation mutagenesis and DNA-shuffling by the same group yielded a triply substituted variant of MTG exhibiting a 12-fold and 10-fold higher half-life at 60°C and 50°C, respectively,[46] although the Ser2Pro variant remained the most active at 55°C. Chen and colleagues also evolved thermostable variants of MTG by combining saturation mutagenesis and the deletion of various N-terminal residues.[47] The variant Del 1-4E5D, which lacks the first four *N*-terminal residues and substitutes the fifth residue, exhibits a modest 1.85-fold higher specific activity and a 2.7-fold higher half-life at 50°C compared to the wild-type enzyme.

Determining what residues to be the focus of mutagenesis is key to the success of any protein engineering initiative. In order to probe which residues may be necessary for MTG activity, an alanine screen of 29 residues that are either located in proximity to, or constitute the putative active site, was performed.[48] Docking and molecular dynamics simulations were also performed in order to propose the manner in which the model peptide substrate Z-QG binds to the enzyme, and the mutagenesis results were interpreted in the context of the docking results. The results suggest that an extended surface along the active site cleft is involved in binding of a protein substrate. Furthermore, it appears that a number of hydrophobic and aromatic residues are important for interacting with Z-QG, which is summarized in Figure 2-3. Despite this data, further evolution of TGases has yet to be reported.

2.5 Substrate specificity

While the acyl-transfer reaction catalyzed by TGase between the peptide- or protein-bound glutamine and lysine substrates is well characterized, the preference the enzymes display towards a specific peptide sequence is not obvious. Most glutamine and lysine residues will serve as a substrate, with varying degrees of reactivity, as long as they are accessible to TGase.[49] This limits the application scope of TGases where reactivity towards a specific substrate is required, such as protein labeling. Ten years ago, highly-reactive glutamine-containing substrates for TG2 were reported, which in some cases are related to physiologically-relevant targets,[25] and in other cases were empirically designed and contain

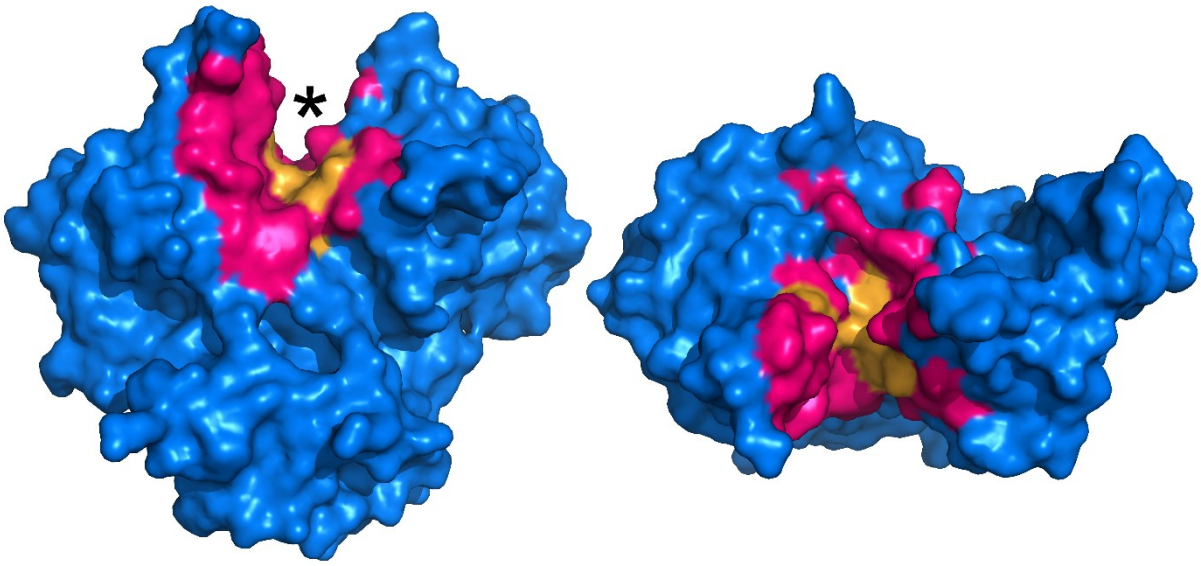


Figure 2-3 Surface representation of MTG (PDB ID: 1UI4).

This illustrates active site residues investigated by mutagenesis (pink and orange regions).[48] *The active site cleft is indicated by an asterisk. Residues in orange, upon substitution to alanine, resulted in activity of 5% or less than the wild type, revealing their importance.*

more than one glutamine for increased reactivity.[50] The secondary structure surrounding the glutamine appears to be important in defining reactivity.[25] With respect to MTG, the native substrates and physiological function of the enzyme are not known. This has led researchers to approach the question of TGase's poorly understood substrate preferences from two different perspectives. The first is to probe the specificity of the enzyme towards specific peptide or protein substrates of interest by analyzing which glutamine or lysine residues are reactive and to what degree. The second is to screen libraries of peptide sequences or other compounds with the goal of either identifying a preferred sequence pattern, or to identify highly reactive substrates. Recent advances with both of these approaches TGase substrate specificity offer further insight into the utility as well as the remaining limitations of these enzymes toward their biotechnological application.

The reactivity of MTG towards glutamine residues on several different proteins has been recently investigated. Using the sensitivity of mass spectrometry (MS), the identification of the glutamine residues most reactive towards MTG-catalyzed PEGylation was described.[51] In that study, a monodisperse Boc-PEG-NH_2 was used as the amine substrate on three model proteins:

granulocyte colony stimulating factor (GCSF), human growth hormone (hGH), and apomyoglobin (apoMb). The former two proteins were selected for their importance as therapeutic proteins, and apoMb for being a model protein regarding the investigation of protein structure, folding, and stability. Despite the fact that GCSF, hGH, and apoMb have 17, 13, and 6 glutamine residues, respectively, only one or two per protein were modified by MTG. All effectively PEGylated glutamines were within disordered regions, suggesting that a flexible polypeptide substrate facilitates binding of MTG to target glutamines. A similar study used type I collagen as a protein substrate.[52] The resulting intermolecular collagen cross-links were quantified by digesting the collagen sample and separating of the fragments by HPLC. No more than five cross-links were formed out of a maximum of 27 possible. At least half of the cross-links were located within the triple helical region of the collagen molecule; however, the specific residues that were modified by MTG were not identified. Importantly, the cross-links were introduced by MTG only after the collagen had been at least partially heat-denatured, supporting the correlation between structural disorder of the target and recognition by MTG. To further investigate the importance of secondary structure and MTG's apparent preference for flexible polypeptide regions, the reactivity of MTG towards apoMb, α-lactalbumin (α-LA) and fragment 205-316 of thermolysin was analyzed.[53] These extensively studied proteins are models of α-helices, β-sheets and unstructured regions, respectively. Once more, despite many glutamine residues being present, few were substrates, with flexible or unstructured regions experiencing the highest reactivity. MTG discriminated notably less against protein-bound lysine as substrates, although those located in disordered regions were indeed more reactive. While this is by no means an exhaustive study of MTG's substrate reactivity with respect to secondary structure, MTG's reactivity towards flexible or unfolded regions for both glutamine and lysine protein substrates is further enforced.

Notwithstanding those advances, searching for superior glutamine recognition sequences that can be grafted onto a desired labeling target (often referred to as a Q-tag) requires a high-throughput methodology in order to screen varying glutamine-containing sequences in an efficient manner. This had been previously done by phage display,[54-55] in which phage-displayed dodecapeptide libraries on the order of 10^{11} members were screened for reactivity toward TG2 and MTG. Regarding MTG, a preference for an aromatic amino acid *N*-terminal to

the glutamine was observed, as well as for an arginine and a hydrophobic amino acid at the +1 or +2 positions. However, no clear preferred amino acid pattern was obvious among the results. Building on this data, sequences determined to be the most reactive were synthesized and tested as penta- and heptapeptide substrates.[56] The pentapeptides' affinity for MTG were as low as Z-QG (in the range of 50 mM); however two heptapeptides, 7M42 (Ac-YELQRPY-NH_2) and 7M48 (Ac-WALQRPH-NH_2), were found to have a 4.5 and 19-fold decrease in K_M, indicating that the identity of surrounding amino acids affect K_M. Using a complementary approach, the search for a Q-tag was expanded by recently employing mRNA display as a high-throughput screen.[57] Peptides that served as substrates became covalently bound via MTG reaction with hexa-lysine conjugated beads. Two pentapeptide sequences in particular were reported to have considerably higher reactivity and affinity for MTG (RLQQP and RTQPA), which vary considerably from the results obtained via phage display. In light of these results, valuable insight into the sequence and structural preferences for efficient TGase recognition of glutamine has been obtained. However, they do not yet converge onto a single, high-affinity Q-tag. The identification of a peptide sequence that is highly specific for MTG has also yet to be demonstrated, and so the precise requirements for selective glutamine binding to TGases remain under investigation.

The structural requirement of MTG's amine (lysine) substrate has previously been suggested to be considerably less strict than that of its amide (glutamine) substrate.[58-60] Along the same line of thought, as with the glutamine substrate, a recent study used an in vivo Förster resonance energy transfer (FRET) quenching assay in order to screen for highly reactive lysine recognition sequences (K-tag) in *E. coli*.[61] The sequences screened were limited to pentapeptides with a lysine fixed at the center position. Although there was no repeated or consensus sequence determined by the screen, the pentapeptide KTKTN was found to be of reactivity comparable to a hexa-lysine tag. Synthetic amide and amine substrates were also previously tested for activity in order to determine if MTG could utilize non-natural substrates.[62] This was investigated in greater detail recently by screening amine compounds with increased diversity of chemical substituents and functional groups.[63] Overall, MTG was found to be highly promiscuous for its primary amine substrate, and amines attached to a less hindered carbon as well as amines with a longer hydrocarbon linker exhibited increased reactivity. Aromatic and

small, polar amine-bearing compounds were observed to be excellent substrates as well. These studies help broaden the scope for modification of glutamine-containing peptides and proteins by TGases.

2.6 Assays

Assay development is key to the advancement of medicine, cell biology, and biotechnology. With respect to TGase, some goals for novel or improved assays include: the identification of highly specific substrates or inhibitors, higher sensitivity, cellular visualization in order to better understand the role of TGase in disease, and facilitation of TGase engineering by high-throughput screens. The detection of TGase activity is not immediately obvious due to the fact that none of its reactants or products absorb strongly at a distinctive wavelength, nor are they fluorescent. A standard end-point, colorimetric assay was developed early on (Figure 2-4A). The assay uses Z-QG as a model glutamine substrate and hydroxylamine as the amine substrate. The addition of TGase catalyzes the formation of an isopeptide bond and a hydroxamate group, and upon the addition of a concentrated ferric chloride solution, results in the development of a yellow color.[45] The hydroxamate assay remains in use to this day in order to determine kinetic constants, but its discontinuous nature and low molar absorptivity limit its applicability. As a result, a number of novel TGase assays have since been developed for use not only in vitro, but in vivo as well. Some colorimetric and fluorometric examples include sensitive assays involving the enzymatic release of *p*-nitrophenol, 7-hydroxycoumarin, and the production of chromophoric anilide.[64-66]

An alternative approach has been to label a protein substrate of interest in a reaction mediated by TG2 with a biotinylated fluorophore and subsequently isolate the newly biotinylated protein with streptavidin beads, allowing for immobilization and separation of the product.[67] The sensitivity of this assay allows for detection of 0.6 mU purified TG2, and can also be applied to crude lysates, making it possible to screen for low transpeptidase activities. However, the sensitivity is less than that of assays using dansylcadaverine to detect product formation, which have been reported to detect as little as 60 μU[68] and 10.8 μU[69] of TG2. This fluorescent alkylamine is commonly used as a substrate for TGases to fluorescently label proteins, and removal of unreacted dansyl cadaverine may reduce background. To address this

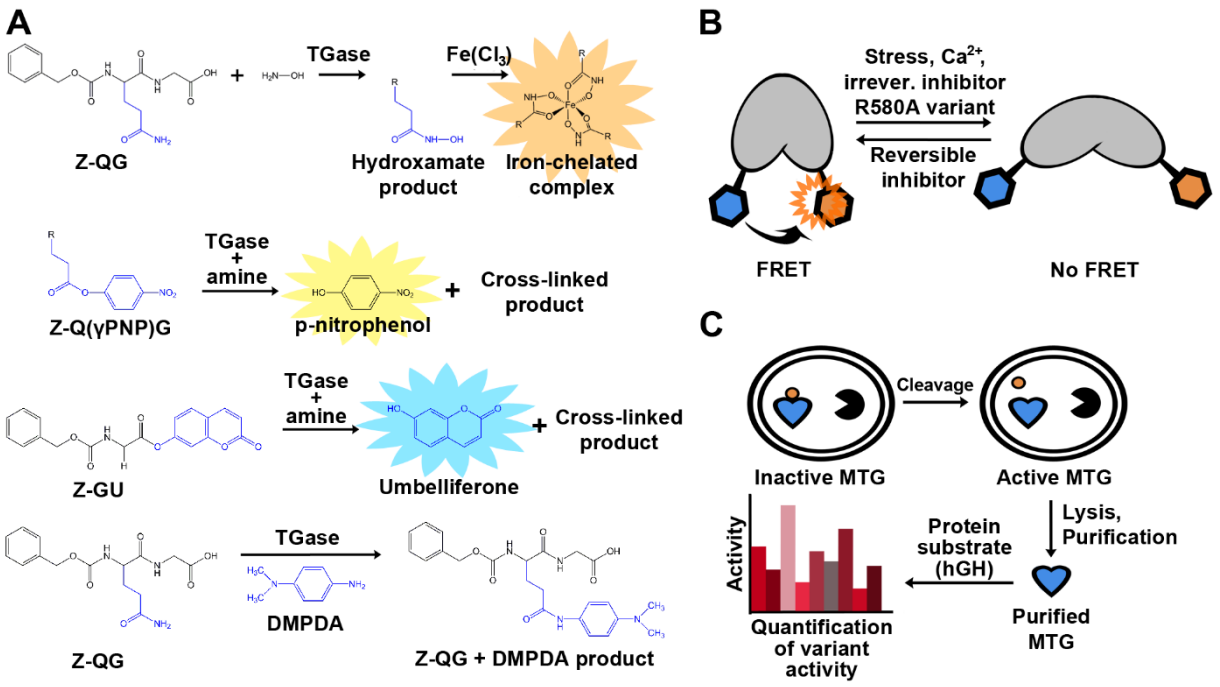


Figure 2-4 Examples of assays used for detection of TGase activity.

A) Colorimetric and fluorescent product release activity assays. The hydroxamate assay (top) remains the standard method to determine and compare TGase activity. TG2 activity can also be quantified by the release of p-nitrophenol (PNP; λ_{max} = 405 nm), umbelliferone (λ_{em} = 465 nm), or by the formation of an anilide product (λ_{max} = 278 nm) following conjugation with N,N-*dimethyl-1,4-phenylenediamine (DMPDA).*

B) Cartoon representation of the TG2 conformational FRET sensor.

C) In vivo *activation of MTG allowing for in-cell assaying.*

issue, magnetic dextran coated charcoal has been used to capture and magnetically sediment unreacted dansyl cadaverine, in a method readily adapted to 96-well plate format.[69] The first assay monitoring the change in fluorescence anisotropy has been recently described.[70] A fluorescein-labeled substrate peptide is monitored for an increase in fluorescence anisotropy as it is cross-linked to a significantly larger substrate, bovine serum albumin (BSA). The assay allows for detection of TG2 as low as 300 pM. The assay also detects product formation; however, a large difference in mass between substrates and product is required in order for detection to occur.

Crystal structures of TG2 reveal that the enzyme undergoes a sizeable conformational change upon substrate binding.[71] In the presence of GDP/GTP, TG2 adopts a "closed" conformation that is inactive.[72] When bound to a substrate-mimicking inhibitor, TG2 was found to be in an "open" conformation, suggesting that the open conformation is the catalytically active form of the enzyme.[72] These conformational changes were recently used as a basis for novel activity assays of TG2. In the first assay, TG2 is used as a biosensor that allows for quantitative assessment in live cells using FRET, as measured by fluorescence lifetime imaging microscopy (FLIM) (Figure 2-4B).[73] This concept was further developed to monitor the real-time, ligand-induced conformational changes of TG2 using kinetic capillary electrophoresis, making this a rapid detection method.[74] As mentioned above, Kim and coworkers recently reported a FRET quenching assay to screen MTG activity in *E. coli*.[61] Each of the two peptide substrates is genetically fused to a fluorescent protein; if the peptide substrates are cross-linked upon exposure to TGase, a FRET quenching results. This approach is highly flexible in that it will allow library screening for either peptide substrate.

Previously, interest has been expressed to engineer TGases towards novel applications.[43-44] With regard to MTG, its requirement for activation complicates the development of a high-throughput screening assay. In effort to circumvent this obstacle, Zhao and co-workers demonstrated an *in vivo* selection assay for MTG (Figure 2-4C).[32] MTG was co-expressed with the 3C protease in order to activate the enzyme. The authors performed site-saturation mutagenesis on two different residues, Y62 and Y75, and used the assay to identify a variant that favors the conjugation of PEG to a specific glutamine (Q141) of human growth hormone. Two variants were found to be exclusively specific for Q141, even after 30 hours of reaction time. In order to determine activity, a previously established scintillation proximity assay was used,[75] complexifying the methodology. A simple, continuous, colorimetric TGase assay was recently adapted in order to easily determine kinetic parameters of MTG with different substrates. Glutamate dehydrogenase activity was coupled to ammonia release upon deamination of the glutamine substrate for MTG, resulting in a decrease in NADH readily observed at λ_{max} = 340 nm.[56]

2.7 TGases as biocatalysts for the production of novel biomaterials

The earliest biocatalytic use of TGases was in the food industry,[11, 16] which continues on a large scale to this day. Novel biotechnological applications have since been fostered to expand the biocatalytic utility of TGases outside of the food industry. Progress in this field has hastened in conjunction with recognition of their flexibility with respect to the primary amine substrate. This has helped open the door of possibilities with regard to covalently modifying protein- or peptide-bound glutamines with a wide array of compounds. The increasing diversity is welcomed: as previously discussed, a number of polymer-protein conjugates have been prepared with TGase using PEG to tailor the properties of the substrate protein to towards a more favorable therapeutic profile, such as enhanced stability and decreased toxicity. Recently, the polymer repertoire was expanded by synthesizing conjugates using hydroxyethyl starch.[76] It is a biodegradable alternative to PEG for commercial use as a blood plasma volume expander, potentially making it a more suitable polymer for protein conjugation. Taking this concept a step further, protein lipidation was demonstrated using MTG, with the goal of altering the behavior of the conjugated protein by controlling its localization via increased amphiphilicity.[77] Proteins can be regarded as biopolymers themselves, and can thus be assembled into larger biomolecular complexes in order to achieve altered functionality and properties. However, such a complex is only of use if its assembly can be controlled. A supramolecular protein complex, composed of *E. coli* alkaline phosphatase (AP) and streptavidin, was constructed with the aid of MTG.[78] The strong avidin-biotin interaction was exploited to direct the assembly of these two protein building blocks into a larger complex, by having AP site-specifically conjugated with biotin using MTG. The location of biotin conjugation on AP was crucial to create large structures and retain AP activity. Finally, MTG has also been found to be effective at modifying the structure of peptides containing a glutamine and lysine residue by cyclization.[79]

Proteins and peptides are not the only biological molecules that have been modified using TGases; MTG has been recently used to site-specifically attach diverse compounds, at multiple positions, onto antibodies.[80-81] Glycosylation normally prevents TGase from effectively modifying antibodies, but the glycosylation pattern was modified such that MTG

was able to react at specific locations. The resulting antibody-drug conjugates (ADCs) are of interest as potential therapeutic solutions, and tweaking their pharmokinetic properties by conjugation with different compounds may yield new therapeutic avenues that were previously unfeasible.

2.8 Protein labeling

A specific application of TGases that is gaining importance is their use as a tool to site-specifically label proteins with the goal of visualization within complex biological systems, such as in living cells. The typical strategy is to introduce an amide- or amine-containing fluorophore substrate into the system, along with TGase, to form an isopeptide bond with a specific lysine or glutamine, respectively, on the target protein (Figure 2-5).

A fluorescent analog of the conventional model glutamine substrate, Z-QG, has been synthesized. Fluorescein-4-isothiocyanate-β-Ala-QG was shown to be an effective glutamine substrate for MTG for reaction with a lysine-containing peptide tag (dubbed as a "K-tag"), genetically encoded at the *N*-terminus of the peptide or protein of interest.[82-83] This K-tag was six amino acids in length, and both the second and fourth residues were lysines (MKHKGS). Mass spectrometry revealed that MTG displayed a high preference for the second lysine. The same group later developed two 13-mer peptidyl loop K-tags, each containing a single lysine, specifically recognized by MTG;[84] no direct comparison of the reactivity of the 6-mer and 13-mer tags was conducted. The 13-mer tags were encoded into bacterial alkaline phosphatase (BAP), which had been selected because MTG does not recognize any of its native glutamine or lysine residues as substrates. High labeling yields (>94%) were obtained when the 13-mer tags were inserted in vicinity of the active site, or at a location distal from the active site (Figure 2-6A). However, insertion distal from the active site provided higher reactivity. The reactivity of the two 13-mer tags was comparable. Using a different approach, incorporation of a fluorescent substrate was observed by an intramolecular FRET between two fluorescent substrate proteins, allowing an evaluation of transamidation activity of TG2.[85] With this assay, propargylamine was found to be an excellent substrate for TG2. Following propargylation of a glutamine residue in casein, the resulting alkyne-modified residue was fluorescently labeled

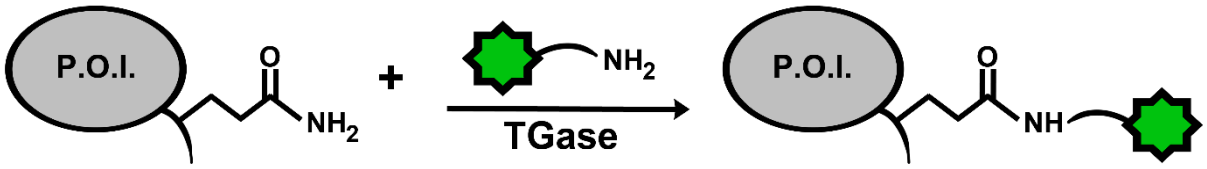


Figure 2-5 General scheme for protein labeling using TGase.

The protein of interest (P.O.I.) carries an accessible glutamine residue, for TGase-catalysed reaction with an amine-substituted fluorophore; alternatively, the P.O.I. carries a reactive lysine residue for reaction with a glutamine-modified fluorophore.

through a copper-catalyzed Huigsen cycloaddition with an azido-fluorescein conjugate (click chemistry),[86] thus providing a general route for labeling with a variety of azido-containing compounds. MTG was also found to be capable of using propargylamine as a substrate; additionally, it can use amino azides as substrates, to allow ulterior click chemistry with a variety of alkyne-containing compounds.[63] The techniques above offer high reactivity *in vitro*; however, they have not yet been tested in the context of cellular visualization.

TG2 is associated with tumor growth and drug resistance, but attempts to detect TG2 in tissues can often be plagued by false positives. Magnetic resonance imaging is a powerful diagnostic tool, and TGase may in the future be detected in tumor cells by using a new contrast agent[87] containing a primary amine, designed so that it would serve as a substrate for MTG (Figure 2-6B). Upon cross-linking the agent onto a tumor, a MRI signal is created. Called chemical exchange saturation transfer (CEST), a particular proton signal associated with the CEST agent is selectively saturated, and the proton remains in exchange with surrounding water molecules. As a result, the MRI signal from the water surrounding the CEST agent is reduced, allowing for its location to be determined. The signal generated before and after cross-linking of the contrast agent differs, allowing for easy differentiation between the two species. Once again, this work remains at the level of *in vitro* experimentation in a model system and has yet to be tested *in vivo*. TGase-mediated labeling has also been further expanded to label biological macromolecules other than proteins, such as DNA and RNA[88-89] (Figure 2-6C). Nucleic acid hybridization techniques make it possible to detect the expression pattern of a particular gene, which may be indicative of a disease. In situ hybridization (ISH) requires binding of a target DNA sequence to a probe, followed by detection with radioisotopes, fluorophores, or antibodies.

In a new hybridization procedure dubbed transglutaminase-mediated in situ hybridization (TransISH), a Z-QG-labeled DNA-peptide conjugate was synthesized using DNA primers containing Z-QG-dUTP. The labeled DNA can then be denatured and cross-linked to alkaline phosphatase (AP) containing a K-tag in a process mediated by MTG. The DNA-linked AP will then dephosphorylate 5-bromo-4-chloro-3-indolyl phosphate, leading to the development of a blue chromophore. The same concept was also applied to mRNA.[90] As additional detection is not required with TransISH, it simplifies common ISH protocols by bypassing these steps, allowing direct staining after washing the unhybridized probe.

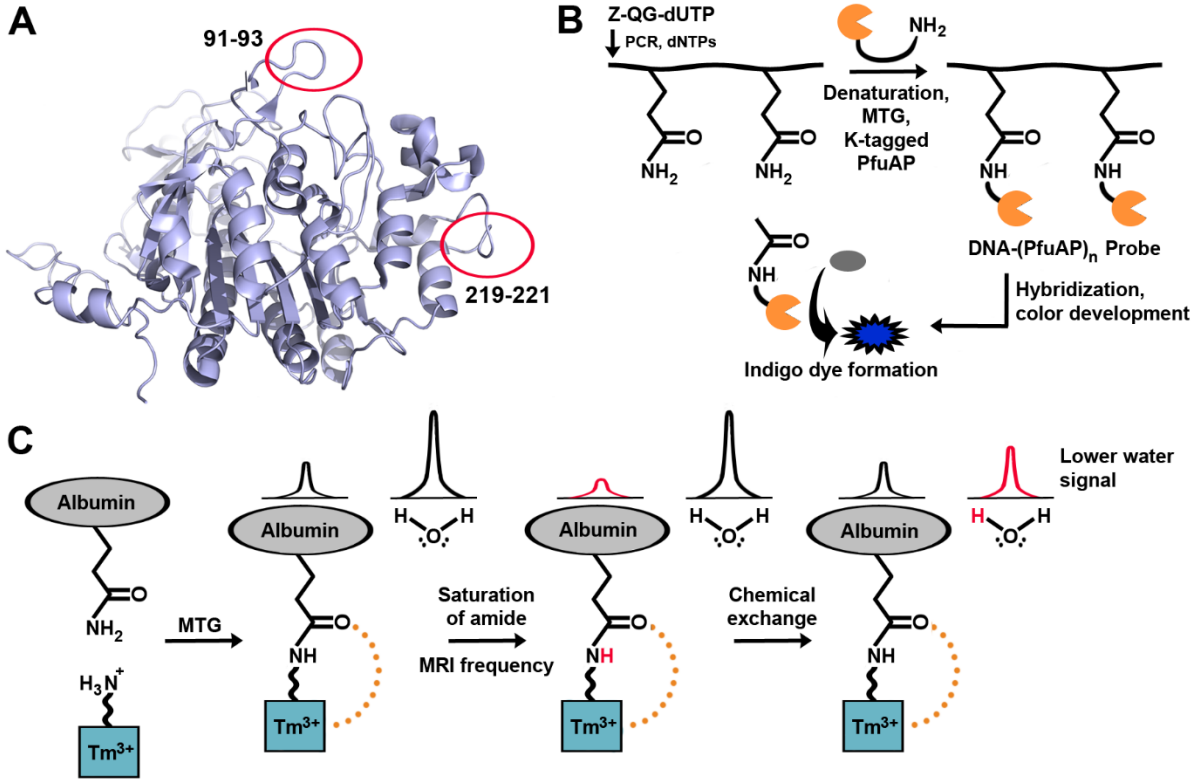


Figure 2-6 Examples of TGases applied for visualization of biomacromolecules.

A) Locations of independently encoded 13-mer peptidyl loop K-tags on bacterial alkaline phosphatase.

B) MTG-aided enzymatic detection of nucleic acids.

C) The paramagnetic agent is cross-linked to a glutamine, generating the CEST effect. Magnetic resonance saturation is transferred to water following saturation of the amide proton.

Fluorescent tagging has also been performed using TG2 activity in order to monitor cellular processes as well as the implication of TGases themselves in disease. Click chemistry was employed in a clinical context to monitor native TG2-mediated protein serotonylation (TPS). With little discrimination with regard to its protein substrate, this process involves TG2 cross-linking of serotonin to glutamine residues, and is implicated in necessary biological processes as well as disease.[91-92] A modified analog of serotonin, propargylserotonin, was synthesized so that it could react with azide-functionalized substrates and enhance the understanding of Ras and its role in previously unknown processes.[93] In addition, of clinical relevance, TG2 is known to play a role in fibrosis and vascular calcification. In order to probe this further, mechanism-based fluorescent inhibitors were designed to covalently label TG2, to investigate how its activity may relate to stiffening of arterial tissues.[94]

2.9 Conclusions

Notable progress has been made in both fundamental and applied research of TGases, although many challenges remain. New efforts in engineering their production have been made, with recent biophysical studies supplementing the knowledge base on the enzymes. However, despite recent work with respect to engineering TGase towards new and different capacities, the goals and results remains largely conservative. Better understanding and characterizing the substrate specificity remains a prime interest so that TGase can be effectively applied in existing and for novel applications. The enzymes have also increasingly become a tool to accomplish new feats in biotechnology. New methods have been developed for detecting and quantifying TGase activity, allowing for increased sensitivity and even in vivo assessment. TGases' natural ability to use protein and peptide substrates gives them potential to label target proteins or peptides, but is limited by its specificity. Some of the techniques discussed in this review have found ways to work around this limitation, however, many remain at the level of proof-of-concept, leaving room for further development and optimization.

2.10 Acknowledgements

The authors would like to acknowledge Dr. Alexis Vallée-Bélisle (Université de Montréal) for helpful discussion and comments.

2.11 Conflict of interest

The authors declare no conflict of interest.

2.12 References

1. Pattabiraman, V. R.; Bode, J. W., Rethinking amide bond synthesis. *Nature* **2011,** *480*, 471-479.
2. Griffin, M.; Casadio, R.; Bergamini, C. M., Transglutaminases: nature's biological glues. *Biochem. J.* **2002,** *368*, 377-396.
3. Shleikin, A. G.; Danilov, N. P., [Evolutionary-biological peculiarities of transglutaminase. Structure, physiological functions, application]. *Zh Evol Biokhim Fiziol* **2011,** *47*, 13-14.
4. Autuori, F.; Farrace, M. G.; Oliverio, S.; Piredda, L.; Piacentini, M., "Tissue" transglutaminase and apoptosis. *Adv Biochem Eng Biotechnol* **1998,** *62*, 129-136.
5. Abe, S.; Yamashita, K.; Kohno, H.; Ohkubo, Y., Involvement of transglutaminase in the receptor-mediated endocytosis of mouse peritoneal macrophages. *Biol. Pharm. Bull.* **2000,** *23*, 1511-1513.
6. Chen, J. S.; Mehta, K., Tissue transglutaminase: an enzyme with a split personality. *Int. J. Biochem. Cell Biol.* **1999,** *31*, 817-836.
7. Shridas, P.; Sharma, Y.; Balasubramanian, D., Transglutaminase-mediated cross-linking of alpha-crystallin: structural and functional consequences. *FEBS Lett.* **2001,** *499*, 245-250.
8. Shan, L.; Molberg, O.; Parrot, I.; Hausch, F.; Filiz, F.; Gray, G. M.; Sollid, L. M.; Khosla, C., Structural basis for gluten intolerance in celiac sprue. *Science* **2002,** *297*, 2275-2279.
9. Schroeder, W. T.; Thacher, S. M.; Stewart-Galetka, S.; Annarella, M.; Chema, D.; Siciliano, M. J.; Davies, P. J.; Tang, H. Y.; Sowa, B. A.; Duvic, M., Type I keratinocyte transglutaminase: expression in human skin and psoriasis. *J. Invest. Dermatol.* **1992,** *99*, 27-34.
10. Ando, H.; Adachi, M.; Umeda, K.; Matsuura, A.; Nonaka, M.; Uchio, R.; Tanaka, H.; Motoki, M., Purification and characteristics of a novel transglutaminase derived from microorganisms. *Agric. Biol. Chem.* **1989,** *53*, 2613-2617.
11. Zhu, Y.; Rinzema, A.; Tramper, J.; Bol, J., Microbial transglutaminase - a review of its production and application in food processing. *Appl. Microbiol. Biotechnol.* **1995,** *44*, 277-282.
12. Suzuki, S.; Izawa, Y.; Kobayashi, K.; Eto, Y.; Yamanaka, S.; Kubota, K.; Yokozeki, K., Purification and characterization of novel transglutaminase from *Bacillus subtilis* spores. *Biosci. Biotechnol. Biochem.* **2000,** *64*, 2344-2351.
13. Pedersen, L. C.; Yee, V. C.; Bishop, P. D.; Le Trong, I.; Teller, D. C.; Stenkamp, R. E., Transglutaminase factor XIII uses proteinase-like catalytic triad to crosslink macromolecules. *Protein Sci.* **1994,** *3*, 1131-1135.
14. Ikura, K.; Nasu, T.; Yokota, H.; Tsuchiya, Y.; Sasaki, R.; Chiba, H., Amino acid sequence of guinea pig liver transglutaminase from its cDNA sequence. *Biochemistry* **1988,** *27*, 2898-2905.
15. Kashiwagi, T.; Yokoyama, K.; Ishikawa, K.; Ono, K.; Ejima, D.; Matsui, H.; Suzuki, E., Crystal structure of microbial transglutaminase from Streptoverticillium mobaraense. *J. Biol. Chem.* **2002,** *277*, 44252-44260.

16. Motoki, M.; Seguro, K., Transglutaminase and its use for food processing. *Trends Food Sci. Technol.* **1998,** *9*, 204-210.
17. Zeugolis, D. I.; Panengad, P. P.; Yew, E. S.; Sheppard, C.; Phan, T. T.; Raghunath, M., An in situ and in vitro investigation for the transglutaminase potential in tissue engineering. *J. Biomed. Mater. Res. A* **2010,** *92*, 1310-1320.
18. Cortez, J.; Bonner, P. L.; Griffin, M., Application of transglutaminases in the modification of wool textiles. *Enzyme Microb. Technol.* **2004,** *34*, 64-72.
19. Chau, D. Y.; Brown, S. V.; Mather, M. L.; Hutter, V.; Tint, N. L.; Dua, H. S.; Rose, F. R.; Ghaemmaghami, A. M., Tissue transglutaminase (TG-2) modified amniotic membrane: a novel scaffold for biomedical applications. *Biomed. Mater.* **2012,** *7*, 045011.
20. Zhu, Y.; Tramper, J., Novel applications for microbial transglutaminase beyond food processing. *Trends Biotechnol.* **2008,** *26*, 559-565.
21. Teixeira, L. S.; Feijen, J.; van Blitterswijk, C. A.; Dijkstra, P. J.; Karperien, M., Enzyme-catalyzed crosslinkable hydrogels: emerging strategies for tissue engineering. *Biomaterials* **2012,** *33*, 1281-1290.
22. Gillet, S. M.; Chica, R. A.; Keillor, J. W.; Pelletier, J. N., Expression and rapid purification of highly active hexahistidine-tagged guinea pig liver transglutaminase. *Protein Expr. Purif.* **2004,** *33*, 256-264.
23. Yokoyama, K. I.; Nakamura, N.; Seguro, K.; Kubota, K., Overproduction of microbial transglutaminase in *Escherichia coli*, in vitro refolding, and characterization of the refolded form. *Biosci. Biotechnol. Biochem.* **2000,** *64*, 1263-1270.
24. Shi, Q.; Kim, S. Y.; Blass, J. P.; Cooper, A. J., Expression in *Escherichia coli* and purification of hexahistidine-tagged human tissue transglutaminase. *Protein Expr. Purif.* **2002,** *24*, 366-373.
25. Piper, J. L.; Gray, G. M.; Khosla, C., High selectivity of human tissue transglutaminase for immunoactive gliadin peptides: implications for celiac sprue. *Biochemistry* **2002,** *41*, 386-393.
26. Roy, I.; Smith, O.; Clouthier, C. M.; Keillor, J. W., Expression, purification and kinetic characterisation of human tissue transglutaminase. *Protein Expr. Purif.* **2013,** *87*, 41-46.
27. Kobayashi, K.; Hashiguchi, K.; Yokozeki, K.; Yamanaka, S., Molecular cloning of the transglutaminase gene from *Bacillus subtilis* and its expression in *Escherichia coli*. *Biosci. Biotechnol. Biochem.* **1998,** *62*, 1109-1114.
28. Placido, D.; Fernandes, C. G.; Isidro, A.; Carrondo, M. A.; Henriques, A. O.; Archer, M., Auto-induction and purification of a *Bacillus subtilis* transglutaminase (Tgl) and its preliminary crystallographic characterization. *Protein Expr. Purif.* **2008,** *59*, 1-8.
29. Marx, C. K.; Hertel, T. C.; Pietzsch, M., Soluble expression of a pro-transglutaminase from Streptomyces mobaraensis in Escherichia coli. *Enzyme Microb. Technol.* **2007,** *40*, 1543-1550.
30. Sommer, C.; Volk, N.; Pietzsch, M., Model based optimization of the fed-batch production of a highly active transglutaminase variant in *Escherichia coli*. *Protein Expr. Purif.* **2011,** *77*, 9-19.
31. Zhang, D.; Zhu, Y.; Chen, J., Microbial transglutaminase production: understanding the mechanism. *Biotechnol. Genet. Eng. Rev.* **2009,** *26*, 205-222.
32. Zhao, X.; Shaw, A. C.; Wang, J.; Chang, C. C.; Deng, J.; Su, J., A novel high-throughput screening method for microbial transglutaminases with high specificity toward Gln141 of human growth hormone. *J. Biomol. Screen.* **2010,** *15*, 206-212.

33. Kikuchi, Y.; Date, M.; Yokoyama, K.; Umezawa, Y.; Matsui, H., Secretion of active-form *Streptoverticillium mobaraense* transglutaminase by *Corynebacterium glutamicum*: processing of the pro-transglutaminase by a cosecreted subtilisin-Like protease from *Streptomyces albogriseolus*. *Appl. Environ. Microbiol.* **2003,** *69*, 358-366.
34. Liu, S.; Zhang, D.; Wang, M.; Cui, W.; Chen, K.; Du, G.; Chen, J.; Zhou, Z., The order of expression is a key factor in the production of active transglutaminase in *Escherichia coli* by co-expression with its pro-peptide. *Microb. Cell Fact.* **2011,** *10*, 112.
35. Chen, K.; Liu, S.; Wang, G.; Zhang, D.; Du, G.; Chen, J.; Shi, Z., Enhancement of *Streptomyces* transglutaminase activity and pro-peptide cleavage efficiency by introducing linker peptide in the C-terminus of the pro-peptide. *J. Ind. Microbiol. Biotechnol.* **2013,** *40*, 317-325.
36. Yang, M. T.; Chang, C. H.; Wang, J. M.; Wu, T. K.; Wang, Y. K.; Chang, C. Y.; Li, T. T., Crystal structure and inhibition studies of transglutaminase from *Streptomyces mobaraense*. *J. Biol. Chem.* **2011,** *286*, 7301-7307.
37. Eder, J.; Fersht, A. R., Pro-sequence-assisted protein folding. *Mol. Microbiol.* **1995,** *16*, 609-614.
38. Suzuki, M.; Sakurai, K.; Lee, Y. H.; Ikegami, T.; Yokoyama, K.; Goto, Y., A back hydrogen exchange procedure via the acid-unfolded state for a large protein. *Biochemistry* **2012,** *51*, 5564-5570.
39. Suzuki, M.; Yokoyama, K.; Lee, Y. H.; Goto, Y., A two-step refolding of acid-denatured microbial transglutaminase escaping from the aggregation-prone intermediate. *Biochemistry* **2011,** *50* (47), 10390-10398.
40. Savile, C. K.; Janey, J. M.; Mundorff, E. C.; Moore, J. C.; Tam, S.; Jarvis, W. R.; Colbeck, J. C.; Krebber, A.; Fleitz, F. J.; Brands, J.; Devine, P. N.; Huisman, G. W.; Hughes, G. J., Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture. *Science* **2010,** *329*, 305-309.
41. Clouthier, C. M.; Pelletier, J. N., Expanding the organic toolbox: a guide to integrating biocatalysis in synthesis. *Chem. Soc. Rev.* **2012,** *41*, 1585-1605.
42. Brustad, E. M.; Arnold, F. H., Optimizing non-natural protein function with directed evolution. *Curr. Opin. Chem. Biol.* **2011,** *15*, 201-210.
43. Keillor, J. W.; Chica, R. A.; Chabot, N.; Vinci, V.; Pardin, C.; Fortin, E.; Gillet, S. M. F. G.; Nakano, Y.; Kaartinen, M. T.; Pelletier, J. N.; Lubell, W. D., The bioorganic chemistry of transglutaminase—from mechanism to inhibition and engineering. *Can. J. Chem.* **2008,** *276*, 271-276.
44. Marx, C. K.; Hertel, T. C.; Pietzsch, M., Random mutagenesis of a recombinant microbial transglutaminase for the generation of thermostable and heat-sensitive variants. *J. Biotechnol.* **2008,** *136*, 156-162.
45. Folk, J. E.; Cole, P. W., Mechanism of action of guinea pig liver transglutaminase. I. Purification and properties of the enzyme: identification of a functional cysteine essential for activity. *J. Biol. Chem.* **1966,** *241*, 5518-5525.
46. Buettner, K.; Hertel, T. C.; Pietzsch, M., Increased thermostability of microbial transglutaminase by combination of several hot spots evolved by random and saturation mutagenesis. *Amino Acids* **2012,** *42*, 987-996.
47. Chen, K.; Liu, S.; Ma, J.; Zhang, D.; Shi, Z.; Du, G.; Chen, J., Deletion combined with saturation mutagenesis of N-terminal residues in transglutaminase from Streptomyces

hygroscopicus results in enhanced activity and thermostability. *Process Biochem.* **2012,** *47*, 2329-2334.
48. Tagami, U.; Shimba, N.; Nakamura, M.; Yokoyama, K.; Suzuki, E.; Hirokawa, T., Substrate specificity of microbial transglutaminase as revealed by three-dimensional docking simulation and mutagenesis. *Protein Eng. Des. Sel.* **2009,** *22*, 747-752.
49. Coussons, P. J.; Price, N. C.; Kelly, S. M.; Smith, B.; Sawyer, L., Factors that govern the specificity of transglutaminase-catalysed modification of proteins and peptides. *Biochem. J.* **1992,** *282 (Pt 3)*, 929-930.
50. Hu, B. H.; Messersmith, P. B., Rational design of transglutaminase substrate peptides for rapid enzymatic formation of hydrogels. *J. Am. Chem. Soc.* **2003,** *125*, 14298-14299.
51. Mero, A.; Spolaore, B.; Veronese, F. M.; Fontana, A., Transglutaminase-mediated PEGylation of proteins: direct identification of the sites of protein modification by mass spectrometry using a novel monodisperse PEG. *Bioconjug. Chem.* **2009,** *20*, 384-389.
52. Stachel, I.; Schwarzenbolz, U.; Henle, T.; Meyer, M., Cross-linking of type I collagen with microbial transglutaminase: identification of cross-linking sites. *Biomacromolecules* **2010,** *11*, 698-705.
53. Spolaore, B.; Raboni, S.; Ramos Molina, A.; Satwekar, A.; Damiano, N.; Fontana, A., Local unfolding is required for the site-specific protein modification by transglutaminase. *Biochemistry* **2012,** *51*, 8679-8689.
54. Sugimura, Y.; Hosono, M.; Wada, F.; Yoshimura, T.; Maki, M.; Hitomi, K., Screening for the preferred substrate sequence of transglutaminase using a phage-displayed peptide library: identification of peptide substrates for TGASE 2 and Factor XIIIA. *J. Biol. Chem.* **2006,** *281*, 17699-17706.
55. Sugimura, Y.; Yokoyama, K.; Nio, N.; Maki, M.; Hitomi, K., Identification of preferred substrate sequences of microbial transglutaminase from *Streptomyces mobaraensis* using a phage-displayed peptide library. *Arch. Biochem. Biophys.* **2008,** *477*, 379-383.
56. Oteng-Pabi, S. K.; Keillor, J. W., Continuous enzyme-coupled assay for microbial transglutaminase activity. *Anal. Biochem.* **2013,** *441*, 169-173.
57. Lee, J. H.; Song, C.; Kim, D. H.; Park, I. H.; Lee, S. G.; Lee, Y. S.; Kim, B. G., Glutamine (Q)-peptide screening for transglutaminase reaction using mRNA display. *Biotechnol. Bioeng.* **2013,** *110*, 353-362.
58. Ohtsuka, T.; Sawa, A.; Kawabata, R.; Nio, N.; Motoki, M., Substrate specificities of microbial transglutaminase for primary amines. *J. Agric. Food. Chem.* **2000,** *48*, 6230-6233.
59. Nonaka, M.; Matsuura, Y.; Motoki, M., Incorporation of a lysine- and lysine dipeptides into as1-Caesin by Ca^{2+}- independent microbial transglutaminase. *Biosci. Biotech. Biochem.* **1996,** *60*, 131-133.
60. Ikura, K.; Sasaki, R.; Motoki, M., Use of transglutaminase in quality-improvement and processing of food proteins. *Agric. Food. Chem.* **1992,** *2*, 389-407.
61. Lee, J. H.; Song, E.; Lee, S. G.; Kim, B. G., High-throughput screening for transglutaminase activities using recombinant fluorescent proteins. *Biotechnol. Bioeng.* **2013,** *110*, 2865-2873.
62. Kulik, C.; Heine, E.; Weichold, O.; Möller, M., Synthetic substrates as amine donors and acceptors in microbial transglutaminase-catalysed reactions. *J. Mol. Catal. B Enzym.* **2009,** *57*, 237-241.
63. Gundersen, M. T.; Keillor, J. W.; Pelletier, J. N., Microbial transglutaminase displays broad acyl-acceptor substrate specificity. *Appl. Microbiol. Biotechnol.* **2014,** *98*, 219-230.

64. de Macedo, P.; Marrano, C.; Keillor, J. W., A direct continuous spectrophotometric assay for transglutaminase activity. *Anal. Biochem.* **2000,** *285*, 16-20.
65. Leblanc, A.; Gravel, C.; Labelle, J.; Keillor, J. W., Kinetic studies of guinea pig liver transglutaminase reveal a general-base-catalyzed deacylation mechanism. *Biochemistry* **2001,** *40*, 8335-8342.
66. Gillet, S. M.; Pelletier, J. N.; Keillor, J. W., A direct fluorometric assay for tissue transglutaminase. *Anal. Biochem.* **2005,** *347*, 221-226.
67. Gnaccarini, C.; Ben-Tahar, W.; Lubell, W. D.; Pelletier, J. N.; Keillor, J. W., Fluorometric assay for tissue transglutaminase-mediated transamidation activity. *Bioorg. Med. Chem.* **2009,** *17*, 6354-63549.
68. Jeitner, T. M.; Fuchsbauer, H. L.; Blass, J. P.; Cooper, A. J., A sensitive fluorometric assay for tissue transglutaminase. *Anal. Biochem.* **2001,** *292*, 198-206.
69. Wu, Y. W.; Tsai, Y. H., A rapid transglutaminase assay for high-throughput screening applications. *J. Biomol. Screen.* **2006,** *11*, 836-843.
70. Kenniston, J. A.; Conley, G. P.; Sexton, D. J.; Nixon, A. E., A homogeneous fluorescence anisotropy assay for measuring transglutaminase 2 activity. *Anal. Biochem.* **2013,** *436*, 13-15.
71. Begg, G. E.; Holman, S. R.; Stokes, P. H.; Matthews, J. M.; Graham, R. M.; Iismaa, S. E., Mutation of a critical arginine in the GTP-binding site of transglutaminase 2 disinhibits intracellular cross-linking activity. *J. Biol. Chem.* **2006,** *281*, 12603-12609.
72. Pinkas, D. M.; Strop, P.; Brunger, A. T.; Khosla, C., Transglutaminase 2 undergoes a large conformational change upon activation. *PLoS Biol.* **2007,** *5*, e327.
73. Caron, N. S.; Munsie, L. N.; Keillor, J. W.; Truant, R., Using FLIM-FRET to measure conformational changes of transglutaminase type 2 in live cells. *PLoS One* **2012,** *7*, e44159.
74. Clouthier, C. M.; Mironov, G. G.; Okhonin, V.; Berezovski, M. V.; Keillor, J. W., Real-time monitoring of protein conformational dynamics in solution using kinetic capillary electrophoresis. *Angew. Chem., Int. Ed.* **2012,** (51), 12464-12468.
75. Madi, A.; Karpati, L.; Kovacs, A.; Muszbek, L.; Fesus, L., High-throughput scintillation proximity assay for transglutaminase activity measurement. *Anal. Biochem.* **2005,** *343*, 256-262.
76. Besheer, A.; Hertel, T. C.; Kressler, J.; Mader, K.; Pietzsch, M., Enzymatically catalyzed HES conjugation using microbial transglutaminase: Proof of feasibility. *J. Pharm. Sci.* **2009,** *98*, 4420-4428.
77. Abe, H.; Goto, M.; Kamiya, N., Protein lipidation catalyzed by microbial transglutaminase. *Chemistry* **2011,** *17*, 14004-14008.
78. Mori, Y.; Wakabayashi, R.; Goto, M.; Kamiya, N., Protein supramolecular complex formation by site-specific avidin-biotin interactions. *Org. Biomol. Chem.* **2013,** *11*, 914-922.
79. Touati, J.; Angelini, A.; Hinner, M. J.; Heinis, C., Enzymatic cyclisation of peptides with a transglutaminase. *Chembiochem* **2011,** *12*, 38-42.
80. Strop, P.; Liu, S. H.; Dorywalska, M.; Delaria, K.; Dushin, R. G.; Tran, T. T.; Ho, W. H.; Farias, S.; Casas, M. G.; Abdiche, Y.; Zhou, D.; Chandrasekaran, R.; Samain, C.; Loo, C.; Rossi, A.; Rickert, M.; Krimm, S.; Wong, T.; Chin, S. M.; Yu, J.; Dilley, J.; Chaparro-Riggers, J.; Filzen, G. F.; O'Donnell, C. J.; Wang, F.; Myers, J. S.; Pons, J.; Shelton, D. L.; Rajpal, A., Location matters: site of conjugation modulates stability and pharmacokinetics of antibody drug conjugates. *Chem. Biol.* **2013,** *20*, 161-167.

81. Jeger, S.; Zimmermann, K.; Blanc, A.; Grunberg, J.; Honer, M.; Hunziker, P.; Struthers, H.; Schibli, R., Site-specific and stoichiometric modification of antibodies by bacterial transglutaminase. *Angew. Chem., Int. Ed.* **2010,** *49*, 9995-9997.
82. Kamiya, N.; Abe, H., New fluorescent substrates of microbial transglutaminase and its application to peptide tag-directed covalent protein labeling. *Methods Mol. Biol.* **2011,** *751*, 81-94.
83. Kamiya, N.; Abe, H.; Goto, M.; Tsuji, Y.; Jikuya, H., Fluorescent substrates for covalent protein labeling catalyzed by microbial transglutaminase. *Org. Biomol. Chem.* **2009,** *7*, 3407-3412.
84. Mori, Y.; Goto, M.; Kamiya, N., Transglutaminase-mediated internal protein labeling with a designed peptide loop. *Biochem. Biophys. Res. Commun.* **2011,** *410*, 829-833.
85. Gnaccarini, C.; Ben-Tahar, W.; Mulani, A.; Roy, I.; Lubell, W. D.; Pelletier, J. N.; Keillor, J. W., Site-specific protein propargylation using tissue transglutaminase. *Org. Biomol. Chem.* **2012,** *10*, 5258-5265.
86. Kolb, H. C.; Finn, M. G.; Sharpless, K. B., Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angew. Chem., Int. Ed.* **2001,** *40*, 2004-2021.
87. Hingorani, D. V.; Randtke, E. A.; Pagel, M. D., A catalyCEST MRI contrast agent that detects the enzyme-catalyzed creation of a covalent bond. *J. Am. Chem. Soc.* **2013,** *135*, 6396-6398.
88. Kitaoka, M.; Tsuruda, Y.; Tanaka, Y.; Goto, M.; Mitsumori, M.; Hayashi, K.; Hiraishi, Y.; Miyawaki, K.; Noji, S.; Kamiya, N., Transglutaminase-mediated synthesis of a DNA-(enzyme)n probe for highly sensitive DNA detection. *Chemistry* **2011,** *17*, 5387-5392.
89. Takahara, M.; Hayashi, K.; Goto, M.; Kamiya, N., Tailing DNA aptamers with a functional protein by two-step enzymatic reaction. *J. Biosci. Bioeng.* **2013,** *116*, 660-665.
90. Kitaoka, M.; Mitsumori, M.; Hayashi, K.; Hiraishi, Y.; Yoshinaga, H.; Nakano, K.; Miyawaki, K.; Noji, S.; Goto, M.; Kamiya, N., Transglutaminase-mediated in situ hybridization (TransISH) system: a new methodology for simplified mRNA detection. *Anal. Chem.* **2012,** *84*, 5885-5891.
91. Watts, S. W.; Priestley, J. R.; Thompson, J. M., Serotonylation of vascular proteins important to contraction. *PLoS One* **2009,** *4*, e5682.
92. Walther, D. J.; Stahlberg, S.; Vowinckel, J., Novel roles for biogenic monoamines: from monoamines in transglutaminase-mediated post-translational protein modification to monoaminylation deregulation diseases. *FEBS J.* **2011,** *278*, 4740-4755.
93. Lin, J. C.; Chou, C. C.; Gao, S.; Wu, S. C.; Khoo, K. H.; Lin, C. H., An in vivo tagging method reveals that Ras undergoes sustained activation upon transglutaminase-mediated protein serotonylation. *Chembiochem* **2013,** *14*, 813-817.
94. Chabot, N.; Moreau, S.; Mulani, A.; Moreau, P.; Keillor, J. W., Fluorescent probes of tissue transglutaminase reveal its association with arterial stiffening. *Chem. Biol.* **2010,** *17*, 1143-1150.

Chapter 3 - Specificity of transglutaminase-catalyzed peptide synthesis

3.1 Context

Research on transglutaminases within the Pelletier research group was established by her former doctoral student, Prof. Roberto Chica, working in collaboration with Prof. Jeffrey Keillor (University of Ottawa), his co-supervisor. With Prof. Keillor's mechanistic expertise of transglutaminases, and Prof. Joelle Pelletier's experience in enzyme engineering, the functionalities and applications of mammalian transglutaminase were investigated. At that time, mammalian transglutaminases had been extensively characterized, and the poorly understood microbial transglutaminase was not a subject of study in either research group. Prof. Chica's doctoral research was devoted biocatalytic investigations of the guinea pig liver transglutaminase (gTG2). He collected preliminary data indicating that gTG2 could synthesize peptide bonds within the main chain of a peptide analog, rather than an isopeptide bond between side chains. This valuable biocatalytic reaction was demonstrated using activated ester substrates rather than its native amide (glutamine) substrate. While this work was described as a chapter in his thesis, it did not reach its experimental conclusion at that time and remained in limbo. When I started my doctoral research, I was offered by Profs Chica and Pelletier to participate in completing the investigations into transglutaminase's capacity to synthesize peptide bonds. Prof. Chica's advances led us to speculate that if unactivated substrates could be used for peptide synthesis catalyzed by gTG2, the reaction would occur with a low efficiency. The fluorometric assay then in use would not likely be sufficiently sensitive to detect this activity. Hence, my contribution in employing a highly sensitive LC-MS assay to detect substrate consumption and product formation was welcomed.

This chapter is a reproduction of the contents of a published article in the *Journal of Molecular Catalysis B: Enzymatic,* entitled: *Specificity of Transglutaminase-Catalyzed Peptide Synthesis*. My contribution to this research paper was the development and employment of the LC-MS assays. Antony St-Jacques performed the kinetic assays and computational simulations. Dan Curry assisted in transglutaminase expression, purification, and kinetic assays. Dr. Steve

Gillet developed the 7-hydroxycoumarine kinetic assay. Dr. Christopher Clouthier synthesized and characterized all non-commercial peptide reagents. All authors contributed to the conceptualization, although Prof. Roberto Chica pioneered the key hypotheses. Research was performed in the laboratories of both Profs Roberto Chica and Joelle Pelletier. The manuscript was drafted mainly by Prof. Roberto Chica, with all authors writing contributions covering their respective experimental responsibilities, and assistance from Profs Jeffrey Keillor and Joelle Pelletier. Supplemental information associated with this manuscript can be consulted in Annex 1 of this thesis.

Specificity of transglutaminase-catalyzed peptide synthesis

Antony D. St-Jacques[1,2], Natalie M. Rachel[3,4,5], Dan R. Curry[1,2], Steve M.F.G. Gillet[3], Christopher M. Clouthier[1,2], Jeffrey W. Keillor[1,2], Joelle N. Pelletier[3,4,5], and Roberto A. Chica[1,2,4]

[1] Department of Chemistry and Biomolecular Sciences, University of Ottawa, Ottawa, Ontario, K1N 6N5, Canada

[2] Centre for Catalysis Research and Innovation, University of Ottawa, Ottawa, Ontario, K1N 6N5, Canada

[3] Département de Chimie, Université de Montréal, Montréal, Québec, H3T 1J4, Canada

[4] PROTEO Network, Université Laval, Québec, Québec, G1V 0A6, Canada

[5] Center for Green Chemistry and Catalysis, Montréal, Québec, H3A 0B8, Canada

J. Mol. Catal. B: Enzymatic **2016**, *123*, 53–61.

Corresponding author: Roberto A. Chica <rchica@uottawa.ca>

3.2 Abstract

Biocatalytic methods for peptide synthesis are of high value due to the rapidly increasing approval of peptide-based therapeutics and the need to develop new analogs. Guinea pig liver transglutaminase (gTG2) catalyzes the cross-linking of peptides and proteins via the formation of γ-glutamyl-ε-lysylisopeptide bonds. In this study, we investigate gTG2-catalyzed peptide bond formation between various amino acid-derived donor and acceptor substrates. Using LC–MS analysis, we demonstrate that gTG2 forms Gly-Xaa and d-Ala-Gly dipeptide products, confirming that its natural transamidation activity can be co-opted for peptide synthesis. An aromatic ester of Gly was the most efficient acyl-donor substrate tested; aromatic esters of D-Ala and L-Ala showed 50-fold lower reactivity or no reactivity, respectively. A computational strategy combining computational protein design algorithms and molecular dynamics simulations was developed to model the binding modes of donor substrates in the gTG2 active site. We show that the inability of gTG2 to efficiently catalyze peptide synthesis from donors containing alanine results from the narrow substrate binding tunnel, which prevents bulkier donors from adopting a catalytically productive binding mode. Our observations pave the way to future protein engineering efforts to expand the substrate scope of gTG2 in peptide synthesis, which may lead to useful biocatalysts for the synthesis of desirable bioactive molecules.

3.3 Introduction

The amide bond is among the most versatile functional groups in synthetic organic chemistry due to its high polarity, stability, and well-characterized conformational preference.[1] In particular, facile peptide bond formation – whether between natural or unnatural amino acids – is of extremely high value due to the rapidly increasing approval of peptide-based therapeutics and the need to develop new analogues. Conventional chemical approaches to peptide bond synthesis require chemical activation, protection, and deprotection steps for each bond formed as well as orthogonal protection of reactive substituents. As a result, peptide bond synthesis remains an important challenge in chemistry.[2] Enzymatic approaches have attempted to alleviate these limitations. This is generally performed by running proteases "backward", toward bond synthesis rather than hydrolysis (recently reviewed[3]). Despite engineering of proteases and optimization of reaction conditions, hydrolysis of existing peptide bonds reduces yield. Using an enzyme that has evolved to synthesize an amide bond, rather than hydrolyze it, could prove advantageous in enzyme-catalyzed peptide bond synthesis.

One such enzyme is tissue transglutaminase (TG2), which catalyzes the Ca^{2+}-dependent cross-linking of peptides and proteins via the formation of γ-glutamyl-ε-lysyl isopeptide bonds.[4-6] The catalytic reaction follows a modified ping-pong mechanism in which a glutamine-containing protein or peptide, the acyl-donor substrate, reacts with the catalytic cysteine residue to form a thioester bond. The resulting covalent acyl-enzyme intermediate then reacts with a second substrate, the acyl-acceptor, to yield the isopeptide-containing product and free enzyme in a transamidation reaction. In the absence of an amine acyl-acceptor, the acyl-enzyme intermediate can be hydrolyzed, transforming the acyl-donor glutamine residue into glutamate and regenerating the free enzyme.[7]

TG2 enzymes exhibit broad specificity towards the acyl-acceptor substrate.[8] Although the native acyl-acceptor substrate is generally a lysine-containing protein or peptide, many non-natural primary amines, such as glycinamide,[9-10] and anilines, such as *N*,*N*-dimethyl-1,4-phenylenediamine,[11] can also react. However, amines containing free carboxylic acid groups, such as free amino acids, do not act as substrates.[10] On the other hand, TG2 displays narrow specificity for its acyl-donor substrates. The side chain of a protein or peptide-bound L-Gln

residue is the native substrate while the side chain of the similar amino acid L-Asn is not reactive.[9] In addition to amides, γ-glutamyl aromatic ester derivatives of L-Glu, such as *N*-carbobenzyloxy-L-glutamyl(γ-*p*-nitrophenyl ester)glycine (Fig. 3-1A), have also been shown to be acyl-donor substrates of TG2 and are used to measure the enzyme's activity.[12] However, secondary amide derivatives of L-Gln, such as *N*-γ-methyl-L-glutamine or anilides, are not substrates of TG2:[13] the γ-carboxamide group of L-Gln is the only known amide that is an acyl-donor substrate of TG2.

We and others previously demonstrated that TG2 could use a novel class of acyl-donor substrates that are neither L-Gln nor L-Glu derivatives.[14-15] Namely, 4-(*N*-carbobenzyloxyglycylamino)-butyric acid-coumarin-7-yl ester (Cbz-Gly-GABA-7HC) and 4-(*N*-carbobenzyloxyphenylalanylamino)-butyric acid-coumarin-7-yl ester (Cbz-Phe-GABA-7HC) (Fig. 3-1B) can react with TG2 to release 7-hydroxycoumarin (7HC), resulting in a fluorescence increase that makes these compounds useful for quantifying TG2 reaction rates. The scaffolds of these substrates, based on known irreversible inhibitors of TG2,[16-17] differ from L-Glu aromatic ester acyl-donor substrates of TG2 in that the reactive ester function is located on the main chain of the peptide analogue, rather than on the side chain. As a result, they give rise to products that do not contain a γ-glutamyl-ε-lysyl isopeptide bond. An analogue in which the 7HC leaving group is attached directly to the glycine residue carboxylate group, *N*-carbobenzyloxyglycyl-coumarin-7-yl ester (Cbz-Gly-7HC, Fig. 3-1C), is also a donor substrate of TG2.[18] Significantly, the reaction of this substrate with an acceptor amine substrate would result in the formation of a peptide-like α-amide bond (Scheme 3-1). These results illustrate that specificity for acyl-donor substrates with aromatic ester functions is broader than had previously been supposed and demonstrate that the enzyme can generate products with novel scaffolds.

In this study, we investigate guinea pig liver TG2 (gTG2)-catalyzed peptide bond formation between the Cbz-Gly-7HC donor substrate in combination with various amino acid-derived acceptors. Using LC-MS analysis of the reaction products, we demonstrate that the enzyme is able to react directly with the α-carboxyl group of Cbz-Gly-7HC to form Gly-Xaa dipeptide products, confirming that its natural transamidation activity can be co-opted for peptide synthesis. Additionally, we explore the substrate specificity of the enzyme in peptide synthesis by measuring its reactivity toward a variety of potential acyl-donor substrates having

an aromatic ester function on the α-carboxyl group of various amino acids. We observed that the aromatic ester of Gly is an efficient acyl-donor substrate; the aromatic ester of D-Ala is also reactive though to a lesser extent, and that of L-Ala showed no detectable reactivity.

To elucidate how the stereochemical configuration of the side-chain of alanine-containing donor substrates affects gTG2 catalytic efficiency, we used a computational strategy combining computational protein design and molecular dynamics simulations to model the binding modes of donors in the gTG2 active site. We show that the inability of gTG2 to efficiently catalyze peptide synthesis from donors other than Cbz-Gly-7HC results from the narrow substrate binding tunnel, which prevents bulkier donors to adopt a catalytically productive binding mode. Our observations pave the way to future protein engineering efforts to expand the substrate scope of gTG2 in peptide synthesis, which may lead to useful biocatalysts for the synthesis of desirable bioactive molecules.

3.4 Materials and methods

3.4.1 Materials

All reagents used were of the highest available purity. Lysozyme, 7HC, *N*-acetyl-L-lysine methyl ester hydrochloride (*N*-AcLysOMe), *N*-carbobenzyloxy-L-glutaminylglycine (Cbz-L-Gln-Gly), glycinamide ($GlyNH_2$) and L-leucine methyl ester (LeuOMe) hydrochlorides were purchased from Sigma-Aldrich. L-Alaninamide hydrochloride ($AlaNH_2$) was purchased from Novabiochem (Mississauga, ON). Ni-NTA agarose resin was purchased from Qiagen (Mississauga, ON). Restriction enzymes and DNA-modifying enzymes were from New England Biolabs. All aqueous solutions were prepared using water purified with a Millipore BioCell system.

3.4.2 Synthesis of donor substrates

3.4.2.1 Synthesis of Cbz-Gly-7HC and CBZ-L-Ala-7HC

The synthesis of Cbz-Gly-7HC was based on a previously reported protocol.[14] Namely, 0.2 g (1 mmol) of Cbz-Gly and 0.4 g (2.5 mmol) of 7HC were dissolved in 10 mL of ethyl acetate. Then, 0.22 mL (0.2 g, 2 mmol) of *N*-methylmorpholine and 0.8 mL (0.63 g, 5 mmol) of

N,*N*-diisopropylcarbodiimide were added with stirring at room temperature. Stirring was continued until the complete disappearance of Cbz-Gly, as followed by thin layer chromatography (ethyl acetate). The reaction mixture was then washed once with 1 M NaOH, three times with 0.1 M NaOH, 3 times with 0.1 M HCl, once with saturated $NaHCO_3$, and once with brine. The organic phase was then dried over $MgSO_4$, filtered and evaporated under reduced pressure. The resulting residue was purified by silica gel chromatography (ethyl acetate) to remove traces of diisopropylurea, giving the desired ester in 70 % yield (0.25 g). Cbz-L-Ala-7HC was synthesized according to a similar protocol.

Cbz-Gly-7HC. ^{1}H NMR (300 MHz, $CDCl_3$): δ 4.30 (2H, d), 5.18 (2H, s), 5.33 (1H, s), 6.44 (1H, d), 7.08 (1H, d), 7.10 (1H, s), 7.37 (5H, m), 7.51 (1H, d), 7.70 (1H, d). ^{13}C NMR (75 MHz, $CDCl_3$): δ 168.5, 160.6, 156.7, 154.9, 152.9, 143.1, 136.3, 129.0, 128.9, 128.6, 128.5, 118.4, 117.2, 116.6, 110.5, 67.7, 43.2. HRMS (FAB) calculated for $C_{19}H_{16}NO_6$ ($[M^+H]^+$): 354.0972, found 354.0968.

Cbz-L-Ala-7HC. ^{1}H NMR (300 MHz, $CDCl_3$): δ 1.63 (3H, d), 4.62 (1H, m), 5.18 (2H, s), 5.33 (1H, d), 6.42 (1H, d), 7.07 (1H, d), 7.09 (1H, s), 7.39 (5H, m), 7.51 (1H, d), 7.70 (1H, d). ^{13}C NMR (75 MHz, CDCl3): δ 171.5, 160.5, 156.0, 154.9, 153.1, 143.1, 136.4, 129.0, 128.9, 128.6, 128.5, 118.4, 117.2, 116.6, 110.5, 67.5, 50.2, 18.5. HRMS (FAB) calculated for $C_{20}H_{18}NO_6$ ($[M^+H]^+$): 368.1129, found 368.1118.

3.4.2.2 Synthesis of Cbz-D-Ala-7HC

The synthesis of Cbz-D-Ala-7HC followed the protocol employed for Cbz-L-Ala-7HC. Equimolar amounts of Cbz-D-Ala (4 mmol, 0.89 g) and 7HC (4 mmol, 0.65 g) were dissolved in 15 mL of dichloromethane at room temperature. To the stirring solution, 0.38 mL (4.4 mmol) of N-methylmorpholine and 0.82 mL (8 mmol) of *N*,*N*-diisopropylcarbodiimide were added. The consumption of Cbz-D-Ala was monitored by thin-layer chromatography. Upon completion, the reaction mixture was washed successively with 0.1 M NaOH, 0.1 M HCl, saturated sodium bicarbonate, and brine. The organic phase was dried over $MgSO_4$, filtered, and concentrated under reduced pressure. The crude product was purified via flash column using chloroform/methanol (9:1), affording the titular compound in 66 % yield (0.59 g).

Cbz-D-Ala-7HC. ^{1}H NMR ($CDCl_3$, 400 MHz): δ 1.41 (d, 3H), 4.25 (q, 1H), 5.10 (s, 2H), 6.63 (d, 1H), 7.37 (m, 5H), 7.55 (br s, 1H), 7.71 (d, 1H), 7.84 (d, 1H). 13C NMR ($CDCl_3$, 100 MHz): δ 17.6, 53.5, 67.4, 110.1, 115.4, 116.1, 118.7, 127.9, 128.1, 129.7, 136.5, 146.3, 156.6, 156.9, 157.1, 161.8, 169.3. HRMS (ESI) Calculated for $C_{20}H_{17}NO_6$: 367.1056. Found: 367.1060.

3.4.2.3 Synthesis of Cbz-GlyNH$_2$

The synthesis of Cbz-GlyNH$_2$ was adapted from a previously reported protocol.[19] Glycinamide hydrochloride (18 mmol, 2.00 g) was dissolved in water (60 mL) and acetone (8 mL), prior to the addition of Na_2CO_3 (54 mmol, 5.7 g) and $NaHCO_3$ (18 mmol, 1.5 g). Benzyl chloroformate (22 mmol, 3.20 mL) was added dropwise to the stirring solution over the course of 30 minutes. The resulting mixture was stirred for 3 hours at room temperature, after which the products were isolated by washing with diethyl ether (50 mL). The protected product was precipitated out of solution by the slow addition of 0.1 M HCl. The precipitate was filtered and subsequently dried *in vacuo* to afford a white solid in 86% yield (3.24 g).

^{1}H NMR ($CDCl_3$, 400 MHz): δ 3.87 (s, 2H), 5.07 (s, 2H), 7.26 (m, 5H), 7.35 (s, 2H), 7.95 (s, 1H). ^{13}C NMR ($CDCl_3$, 100 MHz): δ 45.1, 67.3, 126.2, 126.9, 128.9, 136.3, 156.8, 170.1. HRMS (ESI) calculated for $C_{10}H_{12}N_2O_3$: 208.0848. Found: 208.0851.

3.4.3 Overexpression and purification of gTG2

Recombinant gTG2 was overexpressed and purified from *Escherichia coli* according to a protocol developed in our laboratory[20] with the following modifications. After Ni-NTA purification, the eluant was transferred to a 15-mL Amicon Ultra tube (Millipore) with a molecular weight cut-off of 30 kDa and the gTG2 solution was desalted by centrifugation with 25 mM Tris-acetate buffer (pH 7.0) containing 0.5 mM EDTA. The samples were aliquoted, snap-frozen on dry ice and stored at -80 °C. Typical yields were 1.5-10 mg/L of approximately 85 % pure protein, as estimated from Coomassie Blue staining following SDS-PAGE, in agreement with previous results.[20]

3.4.4 Specific activity

The hydroxamate assay[9] was used to quantify gTG2 activity. Briefly, gTG2 was incubated at 37 °C for 10 minutes with 30 mM Cbz-L-Gln-Gly and 100 mM hydroxylamine in 200 mM Tris-acetate buffer (pH 7.0) containing 5 mM $CaCl_2$ and 1 mM EDTA. The reaction was quenched with a solution containing 2.0 M ferric chloride hexahydrate, 0.3 M trichloroacetic acid, and 0.8 M HCl. The mixture was vortexed and left at room temperature for 10 minutes before measuring absorbance at 525 nm. One unit (U) of gTG2 produces 1 μmol of L-glutamic acid γ-monohydroxamate per minute at 37 °C.

3.4.5 Kinetic assays

All assays were performed in triplicate. The following solutions were prepared: a standard stock buffer solution (100 mM MOPS buffer pH 7.0, 5 mM $CaCl_2$, and 0.05 mM EDTA), a 2- mM (Cbz-Gly-7HC, Cbz-L-Ala-7HC) or 100- mM (Cbz-D-Ala-7HC) solution of acyl-donor substrate in *N*,*N*-dimethylformamide (DMF), and a 200- mM solution of acyl-acceptor substrate *N*-AcLysOMe in water. Prior to performing the assays, a "fluorescence coefficient" was determined daily by measuring the arbitrary fluorescence intensities corresponding to five concentrations of 7HC at concentrations ranging from 0.05 to 12.5 μM in 5 % DMF in the stock buffer solution at 25 °C. The value of this "fluorescence coefficient" varied only slightly (<5 %) each day. For the hydrolysis reaction, activity was measured by adding 10 mU of purified gTG2 to each well of a TCT Luminescence 96-well microtiter plate (Thermo Electron) containing a 0.5-100 μM solution of the acyl-donor substrate in stock buffer. For the transamidation reaction, the same amount of the purified enzyme was added to 0.5-80 mM of the acyl-acceptor substrate in stock buffer solution containing 100 μM of acyl-donor substrate Cbz-Gly-7HC. The acyl-acceptor substrate was replaced by water in the blank. DMF was present at 5 % in the final reaction mixtures. The increase in fluorescence due to the release of 7HC was followed at 25 °C against a blank at λ_{ex} 340 nm and λ_{em} 465 nm in a FluoStar Optima microtiter plate reader (BMG Labtech). Linear slopes of fluorescence versus time were measured over the first <10 % conversion of substrate to product and were converted into initial rates using the fluorescence coefficient.

3.4.6 LC-MS

Reaction mixtures containing 150 μM of ester donor substrate (Cbz-Gly-7HC, Cbz-L-Ala-7HC, or Cbz-D-Ala-7HC) or 20 mM of amide donor substrate (Cbz-GlyNH_2) and 50 mM of acceptor substrate (GlyNH_2, AlaNH_2, or LeuOMe) were prepared in a buffer composed of 100 mM MOPS pH 7.0, 5 mM $CaCl_2$, 0.05 mM EDTA, and 5 % DMF. The pH of each substrate mixture was verified with indicator paper prior to the addition of enzyme. The reaction was initiated upon the addition of 0.1 U/mL of gTG2 (or an equivalent volume of buffer for reactions run in the absence of enzyme) in a final volume of 2 mL. Reactions were incubated at 37 °C for up to 20 min. Control reactions without gTG2 or without amine acceptor were run for each combination of substrate mixtures. Experiments and controls were performed in triplicate.

Disappearance of substrates and appearance of dipeptide products were monitored by ESI LC-MS. Aliquots of reaction mixture were taken immediately after the addition of enzyme (0 min) and after 2 min, 10 min, and 20 min of reaction time. Formic acid (98 %, 10 μL) was added to each aliquot (480 μL) and the mixture was vortexed to quench the reaction. The change in pH from 7.0 to < 2 was verified with indicator paper. An internal standard solution (10 μL of 33.1 mM 4-methoxybenzamide in neat DMSO) was added to the quenched reaction, which was then filtered using 0.2- μm polytetrafluoroethylene filters (Corning) to remove particulates. The filtered sample (20 μL, or 10 μL for the reactions containing Cbz-GlyNH_2) was injected onto a Synergi 4- μm, polar reverse phase, 80- Å, 50 × 2- mm liquid chromatography column (Phenomenex) on a Waters 2545 HPLC apparatus. Elution was achieved with a 5-70 % MeOH/H_2O gradient. Masses were detected under positive ionization mode with a Waters 3100 single quadrupole mass detector.

3.4.7 Homology modeling

A homology model of gTG2 was prepared as described previously.[21] Briefly, an alignment of the human and guinea pig liver TG2 sequences (83% identity) was performed using ClustalW[22] with default parameters. Atomic coordinates for human TG2 in complex with a covalent inhibitor were retrieved from the Protein Data Bank (PDB ID: 2Q3Z[23]). Using the sequence alignment and the crystal structure with all non-protein atoms removed, ten models were generated by Modeller 9.15[24] with default parameters. All models had regions with

unfavorable residue interactions and had to be further refined using the following procedure. Following the addition of hydrogens, the Protonate 3D utility[25], available in the Molecular Operating Environment (MOE) software package[26], was used to solvate the ten models with water in rectangular boxes under periodic boundary conditions with a box cut-off of 6 Å, and to add counter-ions (Na^+ and Cl^-). Then, each structure was energy minimized by conjugate gradient minimization to a root-mean-square gradient below 0.01 kcal mol-1 Å-1 using the AMBER99 force field[27] with a combined explicit and implicit reaction field solvent model set up using MOE. Following analysis of all-atom contacts and geometry using MolProbity[28], the best homology model was selected for further experiments.

3.4.8 Construction of acyl-enzyme intermediates

Using MOE, the catalytic Cys residue (Cys277) on the gTG2 homology model was acetylated. The carbonyl moiety of the acetyl group was then manually oriented via dihedral angle rotation to form a hydrogen bond with the indole nitrogen of Trp241, which has been shown to be essential for catalytic activity, presumably by stabilizing the transition state.[29] Following energy minimization as described above (root-mean-square gradient below 0.1 kcal mol^{-1} $Å^{-1}$) to optimize H-bonding between the acetyl group and Trp241, the acetylated Cys277 residue was extracted from the gTG2 structure and used as a template to build a Cys residue acylated with the Cbz-glycyl moiety in MOE. The added Cbz and Gly atoms were then energy minimized *in vacuo* by conjugate gradient minimization to a root-mean-square gradient below 0.1 kcal mol^{-1} $Å^{-1}$ in order to refine bond lengths and angles. The resulting minimized acylated Cys residue was then used to generate rotamers via the introduction of the following dihedrals: O1-C1-C2-N1, 90 ± 20° and -90 ± 20°; C1-C2-N1-C3, 180 ± 20°; C2-N1-C3-O3, 180°; N1-C3-O3-C4, 120 ± 20°, 180 ± 20°, and -120 ± 20°; C3-O3-C4-C5, 60 ± 20°, 180 ± 20°, and -60 ± 20°; O3-C4-C5-C6A, 0 ± 20°, 60 ± 20°, and 120 ± 20° (see Fig. A 1-1 for atom names). The internal energy of the resulting 13,122 rotamers was evaluated *in vacuo* using MOE with the previously described force field, and only rotamers whose energy was within 10 kcal/mol from the lowest energy rotamer were included in the final rotamer library. A similar procedure was utilized to prepare a rotamer library for Cys277 acylated with the Cbz-D-alanyl moiety.

To build the gTG2 acyl-enzyme intermediates, computational protein design was performed using the fast and accurate side-chain topology and energy refinement (FASTER) algorithm as implemented in PHOENIX[30-32]. Rotamers for acylated Cys277 and surrounding residues (positions Gln169, Trp241, Asn243, Tyr245, Met252, Gln276, Trp278, Phe316, Arg317, Met330, Trp332, Asn333, Phe334, His335, and Cys336) were optimized on the fixed backbone of the gTG2 homology model. The backbone independent Dunbrack rotamer library with expansions of ± 1 standard deviation around $\chi 1$ and $\chi 2$[33] was used to model side-chain conformations. A four-term potential energy function consisting of a van der Waals term from the Dreiding II force field with atomic radii scaled by 0.9[34], a direction sensitive hydrogen-bond term with well depth at 8.0 kcal/mol[35], an electrostatic energy term modelled using Coulomb's law with a distance dependent dielectric of 10, and an occlusion-based solvation penalty term[31] were used to evaluate rotamer combinations. The lowest energy acyl-enzyme intermediate structure obtained from each donor was retained for further analysis.

3.4.9 Molecular dynamics

For generation of molecular dynamics (MD) trajectories, structures of the gTG2 acyl-enzyme intermediates prepared as described in Section 2.8 were used as templates. The thioester bond between Cys277 and the acyl groups was hydrolyzed *in silico*, resulting in noncovalent complexes (gTG2 bound with Cbz-Gly or Cbz-D-Ala) that were energy minimized to alleviate steric clashes following the procedure described in Section 2.7, with the exception that minimization was conducted until a root-mean-square gradient below 0.1 kcal mol^{-1} $Å^{-1}$ was achieved. For the complex with Cbz-L-Ala, a methyl group was added with MOE to the Cα of Gly prior to minimization. The minimized and solvated noncovalent complexes were used as input to NPT (constant number, pressure, and temperature) MD simulation at 300 K. MD trajectories were heated over 500 picoseconds and equilibrated for an additional nanosecond. This was followed by a 1.5-nanosecond production run sampled at 10-picosecond increments. All MD simulations were performed using the AMBER99 and extended Hückel theory[36] force fields in NAMD.[37]

3.5 Results

3.5.1 Wild-type gTG2 can form peptide bonds

Previously, we showed that recombinant gTG2 can react with Cbz-Gly-7HC as a donor substrate in conjunction with a variety of amino acid derivatives as acceptor substrates.[18] These observations were based on the increased rate of 7HC release during the enzymatic reaction in the presence of acceptor substrates, relative to the rate of 7HC release in their absence. While we had not monitored the appearance of the final reaction products, our observations were consistent with gTG2 having an intrinsic peptide synthase activity (Scheme 3-1). In the current work, we applied a sensitive LC-MS assay to monitor dipeptide product formation directly, and thus confirm that gTG2 catalyzes peptide bond synthesis.

In the first step in the development of this LC-MS assay, we synthesized Cbz-Gly-7HC and determined the kinetic parameters of its gTG2-mediated hydrolysis using spectrofluorometric analysis. Michaelis-Menten kinetics demonstrated that Cbz-Gly-7HC is an acyl-donor substrate of wild-type gTG2, having an apparent K_M of 15 ± 12 μM and an apparent k_{cat} of 0.128 ± 0.007 s^{-1} (Table 3-1). This K_M is similar to that measured for the gTG2-catalyzed hydrolysis of Cbz-Gly-GABA-7HC (9 ± 2 μM) while the k_{cat} is approximately 10-fold lower (1.25 ± 0.08 s^{-1}) [14]. The lower k_{cat} results from the absence of the γ-aminobutyric acid linker in Cbz-Gly-7HC relative to Cbz-Gly-GABA-7HC (Fig. 3-1), the lack of which may decrease accessibility of the substrate's reactive carbonyl group for nucleophilic attack by the catalytic thiol.

We also determined the kinetic parameters for the gTG2-catalyzed transamidation reaction of acyl-donor substrate Cbz-Gly-7HC with the widely used acyl-acceptor substrate *N*-AcLysOMe. As previously observed with other amino acid derivatives,[18] we confirmed that the rate of release of 7HC from Cbz-Gly-7HC increased in the presence of *N*-AcLysOMe

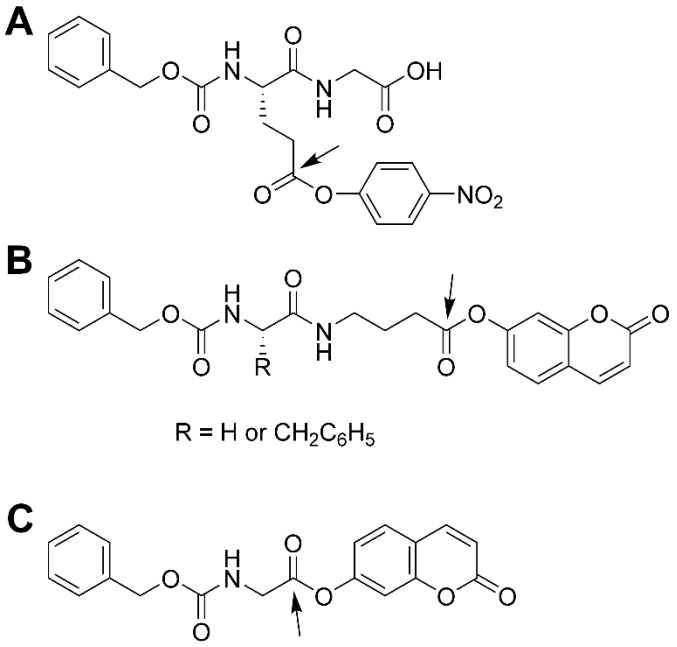
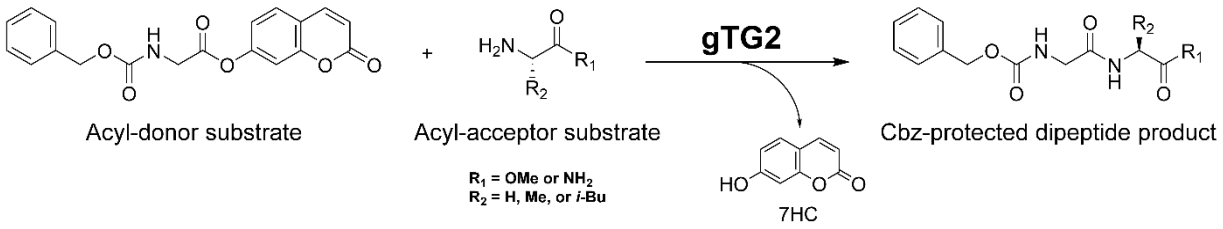


Figure 3-1 Aromatic ester donor substrates of gTG2.

The reactive carbonyl group of these compounds is indicated by an arrow. (A) N-*Carbobenzyloxy-* L*-glutamyl(γ-p-nitrophenyl ester) glycine; (B) Cbz-Gly-GABA-7HC and Cbz-Phe-GABA-7HC; (C) Cbz-Gly-7HC.*

relative to the rate of hydrolysis. This result suggests that gTG2 catalyzes the formation of a covalent bond between the α-carboxyl group of Gly in Cbz-Gly-7HC and the ε-amino group of Lys in *N*-AcLysOMe, whose apparent K_M and k_{cat} values were determined to be 4 ± 1 mM and 0.20 ± 0.02 s^{-1}, respectively.

Next, we confirmed gTG2-mediated peptide synthesis by LC-MS analysis to identify the reaction products. We assayed derivatives of three different amino acids previously shown to act as acyl-acceptor substrates of gTG2[18] in which the negatively-charged carboxylate is neutralized under the form of a primary amide or a methyl ester. Chromatograms of the reaction time-course of Cbz-Gly-7HC with acceptors $GlyNH_2$ (Fig. 3-2A), $AlaNH_2$ (Fig. 3-2B), and LeuOMe (Fig. 3-2C) unequivocally demonstrate the time-dependent increase in concentration of the corresponding dipeptide product. Since the Cbz-Gly-7HC donor substrate is an activated ester, it is highly reactive with nucleophilic amines such as the amino acid derivatives tested herein. This is illustrated by the fact that the LC-MS chromatograms also show significant



Scheme 3-1 Peptide synthesis reaction catalyzed by gTG2.

amounts of dipeptide product being formed in the absence of gTG2 (2.4- to 4-fold catalyzed/uncatalyzed product ratio at the 20 min time-point). However, dipeptide products are formed more rapidly in the presence of enzyme, confirming the intrinsic peptide synthase activity of wild-type gTG2. This is particularly clear at the earlier time points, where the catalyzed/uncatalyzed product ratio is 4- to 8-fold after 10 min and greater than 10-fold for $GlyNH_2$ at the 2 min time point.

3.5.2 Donor substrate specificity of gTG2-catalyzed peptide synthesis

Having confirmed that gTG2 catalyzes the synthesis of Cbz-Gly-Xaa dipeptides, we next investigated whether alternate donor substrates could be utilized. Because the natural substrate of gTG2 is the γ-carboxamide group of an L-Gln residue, we first tested the amide analog of Cbz-Gly-7HC, Cbz-$GlyNH_2$, using $GlyNH_2$ as the acceptor substrate. Cbz-$GlyNH_2$ does not react with gTG2 at concentrations up to 50 mM (Fig. A 1-2), its solubility limit. It has previously been observed that aromatic ester acyl-donor substrates of gTG2 have a lower K_M value than the corresponding amide: Cbz-L-Gln-Gly has an apparent K_M of 4.1 mM in the hydrolysis reaction[38] whereas its aromatic ester analog *N*-carbobenzyloxy-L-glutamyl(γ-*p*-nitrophenyl ester)glycine has an apparent K_M of 0.02 mM.[12] The two orders of magnitude lower K_M of the aromatic ester could be due to improved binding conferred by the *p*-nitrophenol aromatic leaving group. This improved binding could also occur in Cbz-Gly-7HC relative to Cbz-$GlyNH_2$ through beneficial π-stacking interactions between the aromatic leaving group and the aromatic side chains of the tunnel-wall residues of gTG2 (Trp241 and Trp332).[23, 39]

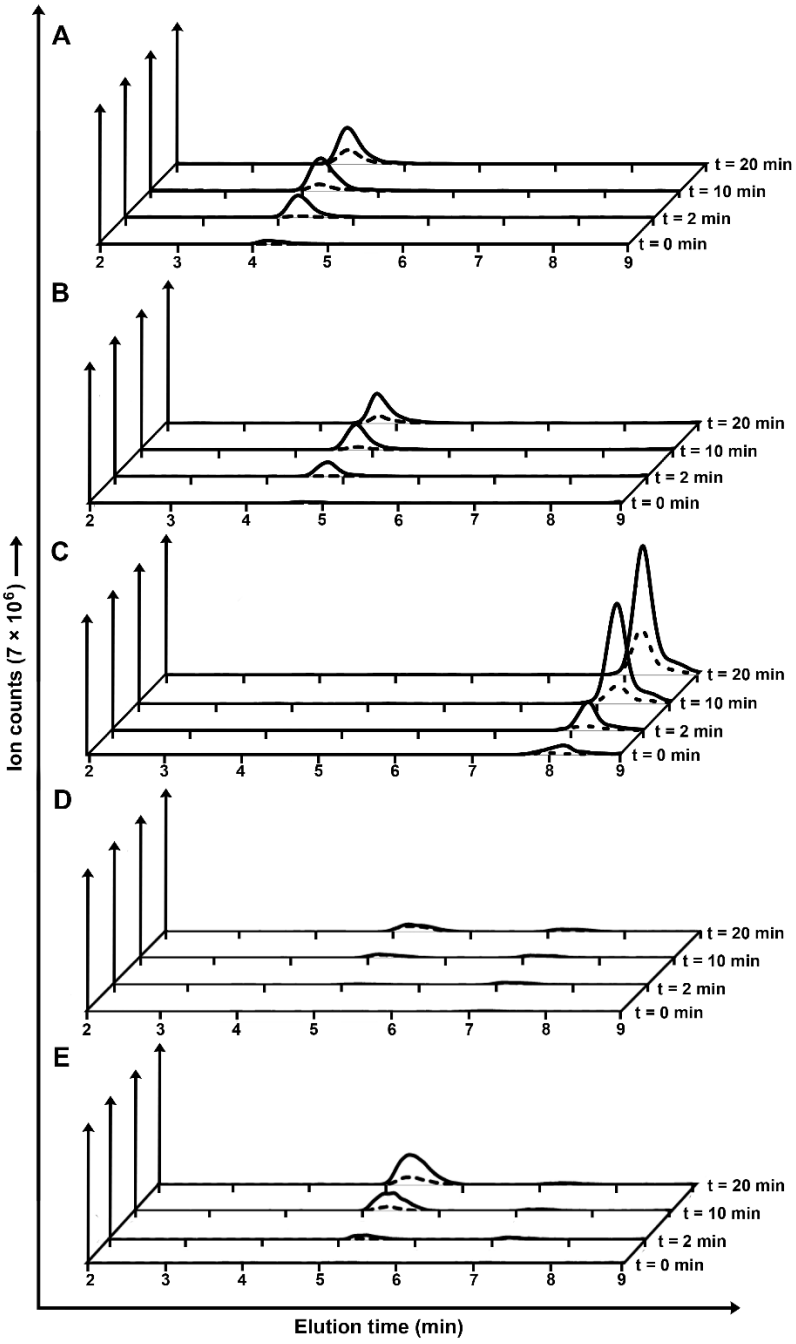


Figure 3-2 LC–MS traces of gTG2-catalyzed peptide synthesis reaction mixtures.

gTG2-catalyzed synthesis of dipeptide products (solid lines) and uncatalyzed control reactions (dashed lines). Various combination of aromatic ester donors and amino acid derivative acceptors were tested: (A) Cbz-Gly-7HC + $GlyNH_2$; (B) Cbz-Gly-7HC + $AlaNH_2$; (C) Cbz-Gly-7HC + LeuOMe; (D) Cbz-L-Ala-7HC + $GlyNH_2$; (E) Cbz-D-Ala-7HC + $GlyNH_2$.

Table 3-1 Apparent kinetic parameters for acyl-donor substrates of gTG2 in hydrolysis reactions.

Errors indicated are standard errors of best-fit parameters.

[a] *Data from Ref 14.*

[b] *No detectable activity.*

[c] *Saturation could not be achieved within solubility limit of donor substrate.*

Substrate	K_M (μM)	k_{cat} (s^{-1})	k_{cat}/K_M (x 10^4 M^{-1} s^{-1})
Cbz-Gly-7HC	15 ± 2	0.128 ± 0.007	0.85
Cbz-L-Ala-7HC	N.D.[b]	N.D.	N.D.
Cbz-D-Ala-7HC	-[c]	-	0.018
Cbz-Gly-GABA-7HC[a]	9 ± 2	1.25 ± 0.08	14

To investigate whether the donor-substrate specificity of gTG2 includes compounds with a substituted α-carbon, we verified whether activated ester donors containing an amino acid other than Gly would react with gTG2. Thus, we synthesized coumarin-7-yl esters of Cbz-protected L-Ala and D-Ala, which contain small methyl-group side chains. Cbz-L-Ala-7HC did not react with gTG2 (Fig. 3-2D) suggesting that the methyl side chain of the L-alanine residue observed with Cbz-D-Ala-7HC (Fig. 3-2E). We thus measured the kinetic parameters for thisdonor substrate with gTG2 using a fluorometric assay. Although we could not saturate the may hinder productive binding at the active site. However, a clearly detectable activity was enzyme with this compound at its solubility limit (50 μM in 5 % DMF), we were able to measure its k_{cat}/K_M, which is approximately 50-fold lower than that of Cbz-Gly-7HC (Table 3-1). These results support observations that substituents, even small ones such as methyl groups, located in close proximity to the reactive carbonyl group of the donor cannot be accommodated readily in the gTG2 active site and are detrimental to activity.[40-41]

3.5.3 Structural basis for donor substrate specificity of gTG2

To elucidate the structural basis for the observed acyl-donor specificity of gTG2, we generated models of the acyl-enzyme intermediates formed during the gTG2-catalyzed hydrolysis of the Cbz-Gly-7HC and Cbz-D-Ala-7HC substrates. We did not generate an acyl-enzyme intermediate structure for hydrolysis of Cbz-L-Ala-7HC as this compound is not a gTG2 substrate. In the acyl-enzyme intermediates, the catalytic Cys277 residue of gTG2 is covalently bound to the Cbz-glycyl or Cbz-D-alanyl moiety through a thioester bond. In the acyl-enzyme intermediate for gTG2-catalyzed hydrolysis of Cbz-Gly-7HC, the glycyl group fits, with no steric clashes, into a tunnel formed by residues Trp241, Gln276, Trp278, Trp332, and Phe334 (Fig. A 1-3), while the Cbz phenyl ring is positioned outside the tunnel and lies in a cleft on the surface of the enzyme. Closer inspection of this acyl-enzyme intermediate suggests that the presence of a methyl side chain on the Cα atom resulting in an L or D configuration would be detrimental to binding as it would clash with either residue Phe334 or Trp332, respectively (Fig. 3-3A). This is indeed what is observed in the acyl-enzyme intermediate structure of gTG2 with a Cbz-D-alanyl moiety (Fig. 3-3B). In this model structure, the side chain of Trp332 adopts an alternate conformation, presumably to alleviate unfavorable steric interactions with the methyl side chain of D-Ala.

Based on these observations, we hypothesized that the methyl side chain of alanine is detrimental to activity because it decreases productive binding for Cbz-D-Ala-7HC and abolishes binding for Cbz-L-Ala-7HC. To test these hypotheses, we generated noncovalent complexes of gTG2 bound to the hydrolysis products Cbz-Gly, Cbz-L-Ala, and Cbz-D-Ala from the acyl-enzyme structures, and used these complexes as input structures for molecular dynamics simulations. The goal of these simulations was to evaluate the binding modes of products in the gTG2 active site.

Carboxylic acid products of hydrolysis were selected as ligands because gTG2 should be able to bind these compounds due to microscopic reversibility and because we could not unambiguously specify where the 7HC group would bind. To evaluate the efficiency with which these compounds are bound in the gTG2 active site, we measured the distance between the nucleophilic sulfur atom of the catalytic residue Cys277 and the electrophilic carbonyl carbon of the Cbz-Gly, Cbz-D-Ala, or Cbz-L-Ala products during the course of a 1.5-nanosecond MD

simulation. As shown in Figure 3-4A, sulfur-carbon distances for Cbz-Gly and Cbz-D-Ala are much lower than those obtained for Cbz-L-Ala. Specifically, the sulfur-

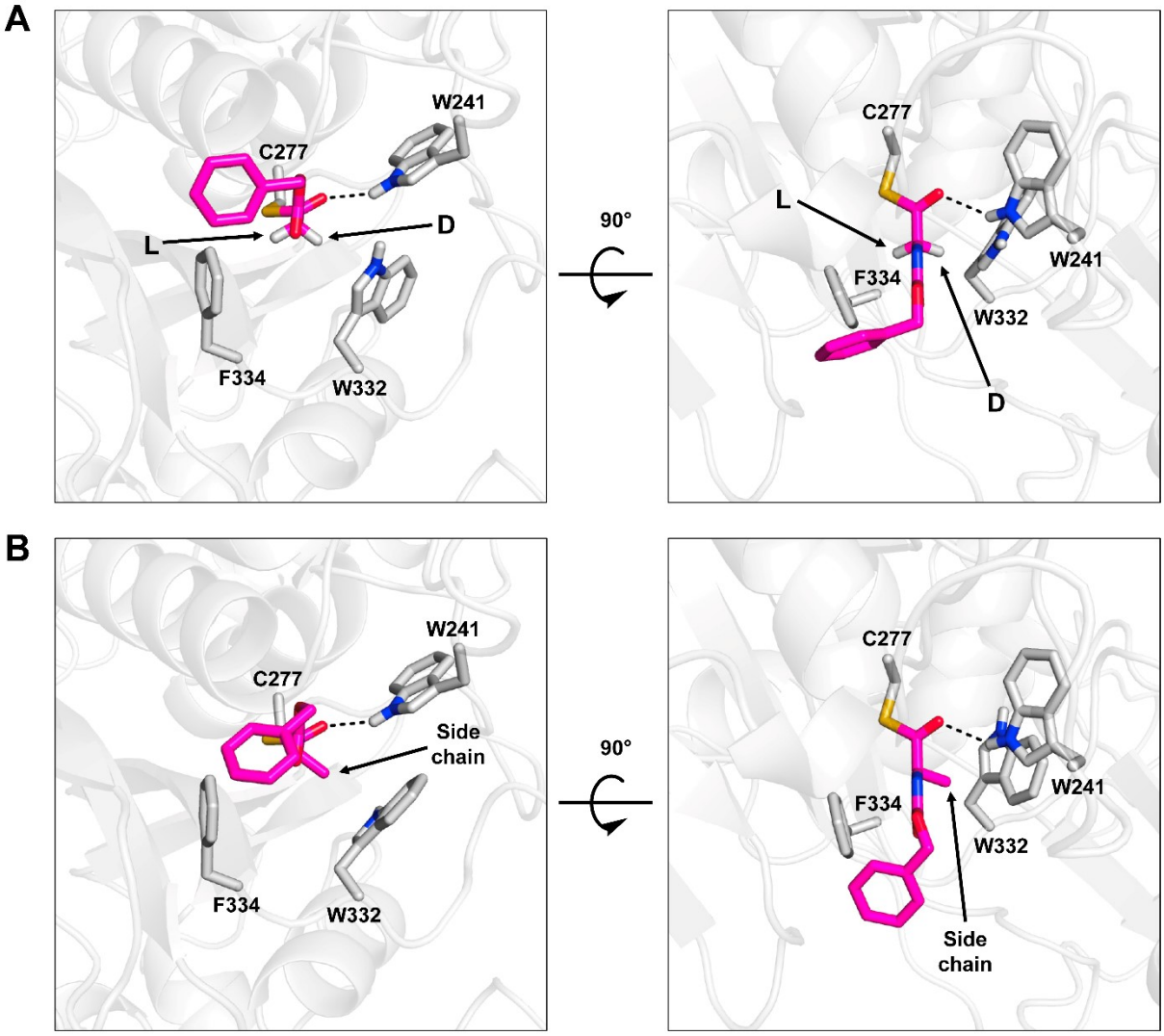


Figure 3-3 gTG2 acyl-enzyme intermediate models.

The active site of gTG2 (white) with the catalytic Cys277 residue acylated by the Cbz-Gly (A) or Cbz-d-Ala (B) moiety (magenta) is shown. An H-bond between Trp241 and the carbonyl oxygen of the acyl group is indicated by a dashed line. Cα hydrogens of Cbz-Gly are shown as sticks, and the resulting configuration of Cα following replacement of each hydrogen by a methyl group side chain is indicated. The methyl side chain of Cbz-D-Ala is indicated by an arrow. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

carbon distance for Cbz-Gly is centered at approximately 5 Å for the duration of the MD simulation while for Cbz-D-Ala, this distance increases to approximately 7 Å after 0.5 nanosecond, suggesting a second distinct binding mode (Fig. 3-4B).On the other hand, the sulfur-carbon distance for Cbz-L-Ala remains centered at approximately 15 Å throughout the simulation. This significantly higher distance results from the fact that the Cbz-L-Ala molecule exits rapidly the gTG2 active site, suggesting that it cannot be bound by the enzyme, in agreement with our kinetic data.

In light of our results, we propose that Cbz-D-Ala-7HC is a poor substrate and that Cbz-L-Ala-7HC is not a substrate of gTG2 because their methyl side chain clashes with tunnel wall residue Trp332 or Phe334, respectively. We postulate that the clash between the methyl

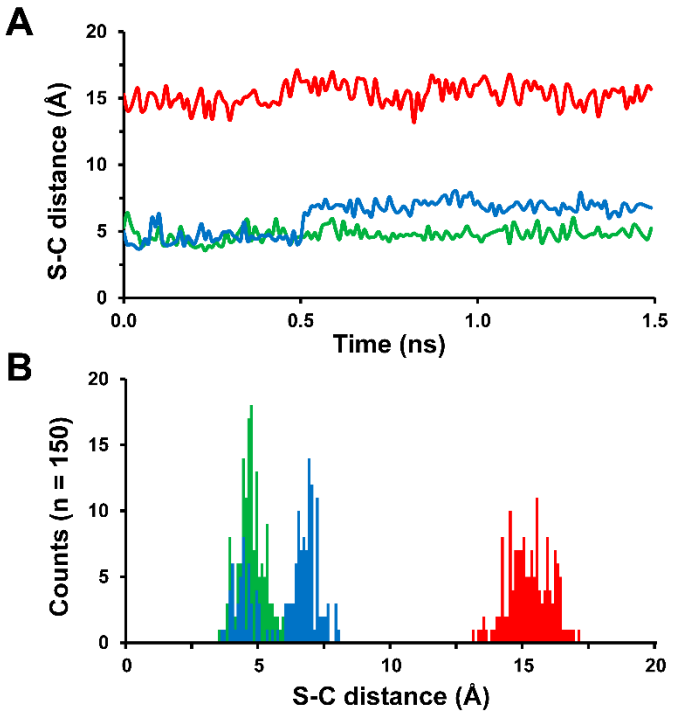


Figure 3-4 Molecular dynamics simulations results.

(A) Distance between the nucleophilic sulfur atom of catalytic residue Cys277 and the electrophilic carbonyl carbon atom of the Cbz-Gly (green), Cbz-d-Ala (blue), or Cbz-l-Ala (red) products during the course of a 1.5-ns MD simulation. (B) Distributions depict the number of MD snapshots with sulfur-carbon distances grouped in incremental bins of 0.1 Å. Each snapshot was taken at 10-ps intervals. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

side chain of D-Ala and Trp332 can be more easily accommodated in the active site because Trp332 is located on a loop formed by residues Asn318-Asn333. This loop is highly flexible, as illustrated by the fact that no electronic density is present for most of the residues comprising it in the human TG2 crystal structure.[23] The higher mobility of this loop would enable Trp332 to move out of the way from the methyl side chain of D-Ala, allowing retention of catalytic activity, albeit at a lower level. On the other hand, Phe334 is part of a β-sheet formed by residues Thr295-Phe301 and Phe334-Trp341, and its phenyl side chain is stacked against the backbone of residues Gln169 and Gly170. These interactions would make Phe334 more rigid, preventing it from moving away from the methyl side chain of L-Ala. These observations raise the possibility that the acyl-donor substrate scope of gTG2 may be expanded by mutating Phe334 and Trp332 in order to increase the space available for substrates containing alternate side chains.

3.6 Discussion

The peptide synthase activity of gTG2 described here results from its transferase activity: the enzyme can transfer a Gly or a D-Ala moiety onto the α-amino group of various amino acid derivatives, thus forming peptide bonds. The transamidase activity of gTG2 relies on its capacity to exclude water from the active site.[42] If water had free access to the thioester bond of the covalent acyl-enzyme intermediate, amine acyl-acceptor substrates would not be able to compete with it for the acyl transfer reaction, as water is much more abundant. Thus, the intermediate thioester must be sequestered in the active site long enough for amines to enter and act as acyl-acceptor substrates. The exclusion of water may result from the hydrophobicity of residues that form the tunnel leading to the catalytic residues, namely Trp241, Trp332, and Phe334. This ability to exclude water from the active site differentiates TG2 from the cysteine proteases, such as papain, that share a similar segment of α-helix and β-sheet containing the catalytic triad.[43]

Protease-catalyzed peptide synthesis is the topic of much current research.[3] Serine and cysteine proteases can catalyze peptide synthesis through a kinetically controlled process in which the protease (hydrolase) acts as a transamidase.[44-45] This process requires a protease that can form a covalent acyl-enzyme intermediate, as is the case with gTG2. Competition between hydrolysis and aminolysis is always present during the degradation of this acyl-enzyme

intermediate, resulting in lower yields for the synthesis of peptides, since proteases are not efficient at excluding water from their active site. Furthermore, proteases hydrolyze the peptide products, further lowering the overall yield of peptide synthesis.

Proteases used in kinetically controlled peptide synthesis have reported transamidation/hydrolysis ratios in the range of 102-104[46-47] whereas gTG2 has a similar 102 increase in rate of transamidation relative to hydrolysis when the acceptor substrate is hydroxylamine.[12] Further, the catalytic efficiency of wild-type gTG2 for the synthesis of various Cbz-Gly-L-Xaa dipeptides ranges from 12 to 141 $M^{-1}s^{-1}$.[18] Papain, a cysteine protease that has been used in peptide synthesis, has catalytic efficiencies of 5 and 49 $M^{-1}s^{-1}$ for the synthesis of the Boc-Gly-L-Phe-N_2H_2Ph dipeptide and the Boc-L-Tyr(Bzl)-Gly-Gly-L-Phe-L-Leu-N_2H_2Ph pentapeptide, respectively.[48] These comparisons suggest that gTG2 could also be used as a catalyst for the synthesis of peptide bonds.

An advantage of gTG2-catalyzed peptide synthesis is that it requires no organic co-solvent. Indeed, the 5 % DMF used in the transamidation assay of Cbz-Gly-7HC and amino acid derivatives by gTG2 is required only to help solubilize the acyl-donor substrate. This is not the case with papain, with which the synthesis of peptides must be carried out in a mixture containing 40 % ethanol[48-49] in order to decrease the activity of water. A further advantage of gTG2 for the synthesis of peptide bonds is that the enzyme cannot recognize secondary amides as acyl-donor substrates, thus limiting hydrolysis of the dipeptide product and potentially increasing yields. However, gTG2 suffers from its apparent need of an aromatic leaving group in acyl-donor substrates, a limitation for peptide synthesis. In addition, the narrow specificity for the amino acid residue found at the *C*-terminus of acyl-donors hinders the general applicability of gTG2 for peptide synthesis. Nevertheless, it may be possible to expand the specificity of gTG2 for additional donor substrates by mutating active site residues Trp332 and Phe334 that form part of the substrate binding tunnel.

3.7 Conclusion

Herein, we confirmed the peptide synthase activity of wild-type gTG2 using LC-MS. This enzyme can form peptide bonds between Cbz-protected Gly or D-Ala, and a variety of polar or hydrophobic amino acid derivatives with a catalytic efficiency similar to the cysteine protease

papain. Although the specificity of gTG2 for peptide bond formation is limited, future engineering efforts based on our computational models to expand its donor substrate specificity may lead to the development of a new tool for the enzymatic synthesis of peptides and complement the known specificities of other proteases.

3.8 Acknowledgements

R. A. C., J. N. P and J. W. K. acknowledge grants from the Natural Sciences and Engineering Research Council of Canada (NSERC), and the Canada Foundation for Innovation. R. A. C. also acknowledges a grant from the Ontario Research Fund; J. N. P and J. W. K. acknowledge a grant from FRQNT. A. D. S. is the recipient of an Ontario Graduate Scholarship.

3.9 References

1. Pattabiraman, V. R.; Bode, J. W., Rethinking amide bond synthesis. *Nature* **2011,** *480*, 471-479.
2. Constable, D. J. C.; Dunn, P. J.; Hayler, J. D.; Humphrey, G. R.; Leazer, J. J. L.; Linderman, R. J.; Lorenz, K.; Manley, J.; Pearlman, B. A.; Wells, A.; Zaks, A.; Zhang, T. Y., Key green chemistry research areas-a perspective from pharmaceutical manufacturers. *Green Chem.* **2007,** *9*, 411-420.
3. Yazawa, K.; Numata, K., Recent advances in chemoenzymatic peptide syntheses. *Molecules* **2014,** *19*, 13755-13774.
4. Fesus, L.; Piacentini, M., Transglutaminase 2: an enigmatic enzyme with diverse functions. *Trends Biochem. Sci.* **2002,** *27*, 534-539.
5. Griffin, M.; Casadio, R.; Bergamini, C. M., Transglutaminases: nature's biological glues. *Biochem. J.* **2002,** *368*, 377-396.
6. Lorand, L.; Graham, R. M., Transglutaminases: crosslinking enzymes with pleiotropic functions. *Nat. Rev. Mol. Cell Biol.* **2003,** *4*, 140-156.
7. Keillor, J. W.; Clouthier, C. M.; Apperley, K. Y.; Akbar, A.; Mulani, A., Acyl transfer mechanisms of tissue transglutaminase. *Bioorg. Chem.* **2014,** *57*, 186-197.
8. Aeschlimann, D.; Paulsson, M., Transglutaminases: protein cross-linking enzymes in tissues and body fluids. *Thromb Haemost* **1994,** *71*, 402-415.
9. Folk, J. E.; Chung, S. I., Transglutaminases. *Methods Enzymol.* **1985,** *113*, 358-375.
10. Clarke, D. D.; Mycek, M. J.; Neidle, A.; Waelsch, H., The incorporation of amines into protein. *Arch. Biochem. Biophys.* **1959,** *79*, 338-354.
11. de Macedo, P.; Marrano, C.; Keillor, J. W., A direct continuous spectrophotometric assay for transglutaminase activity. *Anal. Biochem.* **2000,** *285*, 16-20.
12. Leblanc, A.; Gravel, C.; Labelle, J.; Keillor, J. W., Kinetic studies of guinea pig liver transglutaminase reveal a general-base-catalyzed deacylation mechanism. *Biochemistry* **2001,** *40*, 8335-8342.

13. Halim, D.; Caron, K.; Keillor, J. W., Synthesis and evaluation of peptidic maleimides as transglutaminase inhibitors. *Bioorg. Med. Chem. Lett.* **2007,** *17*, 305-308.
14. Gillet, S. M.; Pelletier, J. N.; Keillor, J. W., A direct fluorometric assay for tissue transglutaminase. *Anal. Biochem.* **2005,** *347*, 221-226.
15. Klock, C.; Jin, X.; Choi, K.; Khosla, C.; Madrid, P. B.; Spencer, A.; Raimundo, B. C.; Boardman, P.; Lanza, G.; Griffin, J. H., Acylideneoxoindoles: a new class of reversible inhibitors of human transglutaminase 2. *Bioorg. Med. Chem. Lett.* **2011,** *21*, 2692-2696.
16. Choi, K.; Siegel, M.; Piper, J. L.; Yuan, L.; Cho, E.; Strnad, P.; Omary, B.; Rich, K. M.; Khosla, C., Chemistry and biology of dihydroisoxazole derivatives: selective inhibitors of human transglutaminase 2. *Chem. Biol.* **2005,** *12*, 469-475.
17. Pardin, C.; Gillet, S. M.; Keillor, J. W., Synthesis and evaluation of peptidic irreversible inhibitors of tissue transglutaminase. *Bioorg. Med. Chem.* **2006,** *14*, 8379-8385.
18. Keillor, J. W.; Chica, R. A.; Chabot, N.; Vinci, V.; Pardin, C.; Fortin, E.; Gillet, S. M. F. G.; Nakano, Y.; Kaartinen, M. T.; Pelletier, J. N.; Lubell, W. D., The bioorganic chemistry of transglutaminase—from mechanism to inhibition and engineering. *Can. J. Chem.* **2008,** *276*, 271-276.
19. Pehere, A. D.; Abell, A. D., An improved large scale procedure for the preparation of *N*-Cbz amino acids. *Tetrahedron Lett.* **2011,** *52*, 1493-1494.
20. Gillet, S. M.; Chica, R. A.; Keillor, J. W.; Pelletier, J. N., Expression and rapid purification of highly active hexahistidine-tagged guinea pig liver transglutaminase. *Protein Expr. Purif.* **2004,** *33*, 256-264.
21. Pardin, C.; Roy, I.; Chica, R. A.; Bonneil, E.; Thibault, P.; Lubell, W. D.; Pelletier, J. N.; Keillor, J. W., Photolabeling of tissue transglutaminase reveals the binding mode of potent cinnamoyl inhibitors. *Biochemistry* **2009,** *48*, 3346-3353.
22. Larkin, M. A.; Blackshields, G.; Brown, N. P.; Chenna, R.; McGettigan, P. A.; McWilliam, H.; Valentin, F.; Wallace, I. M.; Wilm, A.; Lopez, R.; Thompson, J. D.; Gibson, T. J.; Higgins, D. G., Clustal W and Clustal X version 2.0. *Bioinformatics* **2007,** *23*, 2947-2948.
23. Pinkas, D. M.; Strop, P.; Brunger, A. T.; Khosla, C., Transglutaminase 2 undergoes a large conformational change upon activation. *PLoS Biol.* **2007,** *5*, e327.
24. Eswar, N.; Webb, B.; Marti-Renom, M. A.; Madhusudhan, M. S.; Eramian, D.; Shen, M. Y.; Pieper, U.; Sali, A., Current protocols in bioinformatics/editoral board,Andreas D. Baxevanis. [et al.], Unit 5 6 (2006) (Chapter 5).
25. Labute, P., Protonate3D: assignment of ionization states and hydrogen coordinates to macromolecular structures. *Proteins* **2009,** *75*, 187-205.
26. Chemical Computing Group Inc., i., Chemical Computing Group Inc., 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2012.
27. Wang, J.; Cieplak, P.; Kollman, P. A., How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* **2000,** *21*, 1049-1074.
28. Davis, I. W.; Leaver-Fay, A.; Chen, V. B.; Block, J. N.; Kapral, G. J.; Wang, X.; Murray, L. W.; Arendall, W. B., 3rd; Snoeyink, J.; Richardson, J. S.; Richardson, D. C., MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **2007,** *35* (Web Server issue), W375-83.
29. Murthy, S. N.; Iismaa, S.; Begg, G.; Freymann, D. M.; Graham, R. M.; Lorand, L., Conserved tryptophan in the core domain of transglutaminase is essential for catalytic activity. *Proc. Natl. Acad. Sci. U.S.A.* **2002,** *99*, 2738-2742.

30. Allen, B. D.; Nisthal, A.; Mayo, S. L., Experimental library screening demonstrates the successful application of computational protein design to large structural ensembles. *Proc. Natl. Acad. Sci. U.S.A.* **2010,** *107*, 19838-19843.
31. Chica, R. A.; Moore, M. M.; Allen, B. D.; Mayo, S. L., Generation of longer emission wavelength red fluorescent proteins using computationally designed libraries. *Proc. Natl. Acad. Sci. U.S.A.* **2010,** *107*, 20257-20262.
32. Privett, H. K.; Kiss, G.; Lee, T. M.; Blomberg, R.; Chica, R. A.; Thomas, L. M.; Hilvert, D.; Houk, K. N.; Mayo, S. L., Iterative approach to computational enzyme design. *Proc. Natl. Acad. Sci. U.S.A.* **2012,** *109*, 3790-3795.
33. Dunbrack, R. L., Jr.; Cohen, F. E., Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Sci.* **1997,** *6*, 1661-1681.
34. Mayo, S. L.; Olafson, B. D.; Goddard, W. A., DREIDING: a generic force field for molecular simulations. *J. Phys. Chem.* **1990,** *94*, 8897–8909.
35. Dahiyat, B. I.; Mayo, S. L., Probing the role of packing specificity in protein design. *Proc. Natl. Acad. Sci. U.S.A.* **1997,** *94*, 10172-10177.
36. Gerber, P. R.; Muller, K., MAB, a generally applicable molecular force field for structure modelling in medicinal chemistry. *J. Comput. Aided Mol. Des.* **1995,** *9*, 251-268.
37. Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K., Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005,** *26*, 1781-1802.
38. Day, N.; Keillor, J. W., A continuous spectrophotometric linked enzyme assay for transglutaminase activity. *Anal. Biochem.* **1999,** *274*, 141-144.
39. Chica, R. A.; Gagnon, P.; Keillor, J. W.; Pelletier, J. N., Tissue transglutaminase acylation: Proposed role of conserved active site Tyr and Trp residues revealed by molecular modeling of peptide substrate binding. *Protein Sci.* **2004,** *13*, 979-991.
40. Gross, M.; Folk, J. E., Mapping of the active sites of transglutaminases. I. Activity of the guinea pig liver enzyme toward aliphatic amides. *J. Biol. Chem.* **1973,** *248*, 1301-1306.
41. Gross, M.; Folk, J. E., Mapping of the active sites of transglutaminases. II. Activity of the guinea pig liver enzyme toward methylglutamine peptide derivatives. *J. Biol. Chem.* **1973,** *248*, 6534-6542.
42. Nemes, Z.; Petrovski, G.; Csosz, E.; Fesus, L., Structure-function relationships of transglutaminases--a contemporary view. *Prog. Exp. Tumor Res.* **2005,** *38*, 19-36.
43. Kashiwagi, T.; Yokoyama, K.; Ishikawa, K.; Ono, K.; Ejima, D.; Matsui, H.; Suzuki, E., Crystal structure of microbial transglutaminase from Streptoverticillium mobaraense. *J. Biol. Chem.* **2002,** *277*, 44252-44260.
44. Lombard, C.; Saulnier, J.; Wallach, J. M., Recent trends in protease-catalyzed peptide synthesis. *Protein Pept. Lett.* **2005,** *12*, 621-629.
45. Mitsuhashi, J.; Nakayama, T.; Narai-Kanayama, A., Mechanism of papain-catalyzed synthesis of oligo-tyrosine peptides. *Enzyme Microb. Technol.* **2015,** *75-76*, 10-17.
46. Kumar, D.; Bhalla, T. C., Microbial proteases in peptide synthesis: approaches and applications. *Appl. Microbiol. Biotechnol.* **2005,** *68*, 726-736.
47. Riechmann, L.; Kasche, V., Peptide synthesis catalyzed by the serine proteinases chymotrypsin and trypsin. *Biochim. Biophys. Acta* **1985,** *830*, 164-172.
48. Kullmann, W., Kinetics of chymotrypsin- and papain-catalysed synthesis of [leucine]enkephalin and [methionine]enkephalin. *Biochem. J.* **1984,** *220*, 405-416.

49. Kullmann, W., Proteases as catalysts for enzymic syntheses of opioid peptides. *J. Biol. Chem.* **1980,** *255*, 8234-8238.

Chapter 4 - One-pot peptide and protein conjugation: combination of enzymatic transamidation and click chemistry

4.1 Context

Our investigations into the biocatalytic utility of mammalian transglutaminase revealed that it was severely limited by a narrow substrate scope, sensitivity to reaction media, and its propensity to accept water as a nucleophilic substrate in lieu of the desired nucleophile. In light of these results, we focused our efforts to investigating microbial transglutaminase. The fact that it was not as well characterized as mammalian transglutaminase was initially intimidating, but ultimately provided an opportunity to shed light onto an alternative way to achieve biocatalytic amide bond formation. Research on MTG by our group and others revealed specific characteristics that set the stage for the investigations that ultimately led to the following article. Firstly, MTG is fairly robust, being able to tolerate reaction media beyond the gentile requirements for biological systems. Secondly, it can accept a broad range of amine substrates. Finally, it requires peptide- or protein-bound glutamines as its amide substrate. Taken together, we hypothesized that the enzyme's capacity to covalently modify peptides and proteins could be expanded, specifically in the context of its combination with other chemical reactions.

This chapter is a reproduction of an article published in the journal *Chemical Communications*, entitled: *One-Pot Peptide and Protein Conjugation: a Combination of Enzymatic Transamidation and Click Chemistry*. My contribution was the conceptualization and realization of laboratory experiments, performed in the laboratory of Prof. Joelle Pelletier. The manuscript was drafted by myself with assistance from Prof. Joelle Pelletier. Supporting information associated with this manuscript can be consulted in Annex 2 of this thesis.

This version varies slightly from the published article, to address the comments and questions of the evaluation committee: specifically, additional sentences were added to clarify certain topics.

One-pot peptide and protein conjugation: a combination of enzymatic transamidation and click chemistry

Natalie M. Rachel[1,2,3] and Joelle N. Pelletier[1,2,3,4]

[1] Département de Chimie, Université de Montréal, 2900 Boulevard Edouard-Montpetit, Montréal, Québec, H3T 1J4, Canada

[2] CGCC, the Center in Green Chemistry and Catalysis, Montréal, H3A 0B8, Canada

[3] PROTEO, the Québec Network for Protein Function, Structure and Engineering, Québec, G1V 0A6, Canada

[4] Départment de Biochimie, Université de Montréal, 2900 Boulevard Edouard-Montpetit, Montréal, Québec, H3T 1J4, Canada

Chem. Comm. **2016**, *52*, 2541–2544.

Corresponding Author: Joelle N. Pelletier <joelle.pelletier@umontreal.ca>

4.2 Abstract

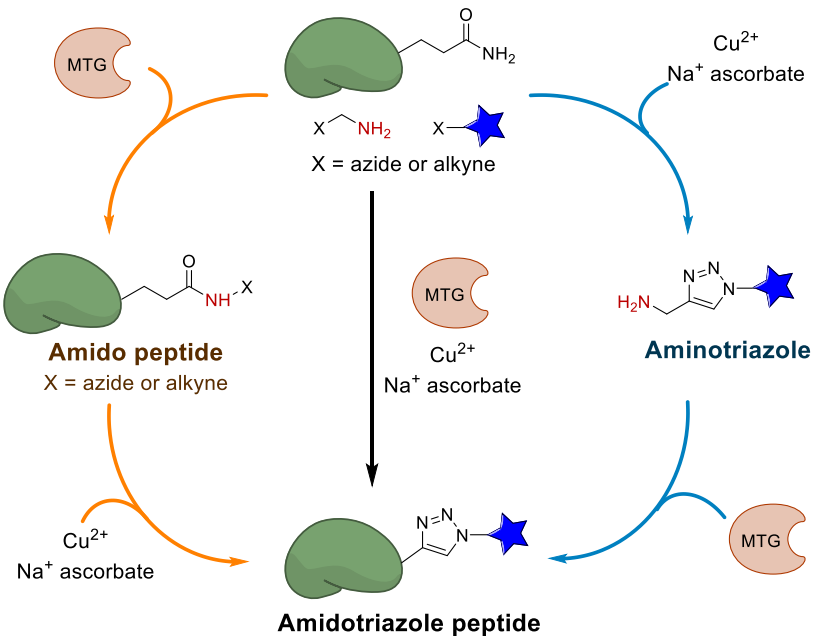
Enzymatic transamidation and copper-catalyzed azide–alkyne cycloaddition (CuAAC) were combined to yield covalently conjugated peptides and proteins. The addition of glutathione preserved enzymatic activity in the presence of copper. Tuning the reaction kinetics was key to success, providing up to 95% conversion. This one-pot reaction allowed for targeted fluorescent protein labeling.

4.3 Article content

Site-specific modification of peptides and proteins allows us to control their chemical, structural, and functional properties.[1-2] Biocatalyzed conjugation is a promising alternative to traditional metal-catalyzed conjugation, with enzymes offering high specificity for the biological target and working under mild reaction conditions.[3-4] Enzymes successfully employed to this end (recently reviewed[5]) include the targeted conjugation of an azidebearing compound to a protein using an engineered lipoic acid ligase, allowing for a downstream cycloaddition reaction,[6] and use of phosphopantetheinyl transferase to conjugate target proteins with chemically modified Coenzyme A analogues carrying chemical moieties such as biotin[7] and fluorophores.[8]

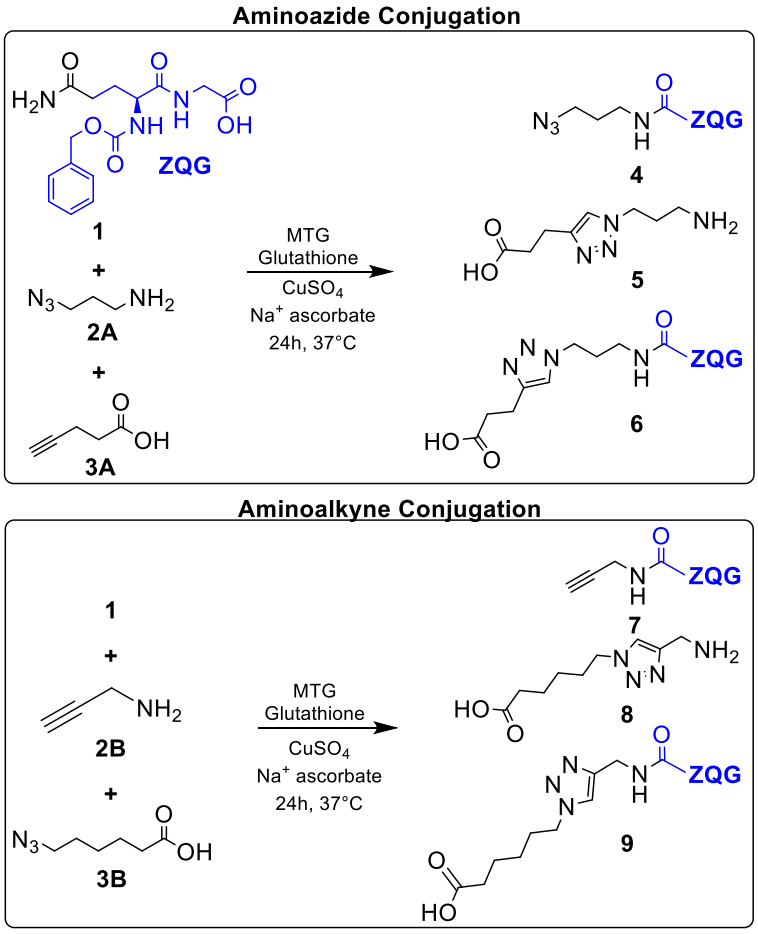
Performing bioconjugation in conjunction with chemical transformations can provide high reaction control and specificity while accessing great chemical diversity. Combining reactions and eliminating tedious purification steps generally streamlines procedures and increase the yield, making one-pot strategies for protein modification highly desirable.[9-10] An important barrier to conducting simultaneous chemical and enzymatic reactions is the incompatibility of most biocatalysts with chemical catalysts, particularly metals. Copper-catalyzed Huisgen azide–alkyne cycloaddition (CuAAC) has been previously used with enzymatic systems.[11-14] In those efforts, CuAAC followed the enzyme reaction as very few enzymes have been demonstrated to be compatible with CuAAC, illustrating the magnitude of this challenge. To the best of our knowledge, Candida antarctica lipase B (CALB) is the only enzyme that has successfully been employed in a simultaneous chemoenzymatic reaction.[15-16] Despite the appeal of one-pot conjugation strategies, their success is often counter-intuitive due to the sensitivity of biomolecules to conditions employed in traditional organic synthesis, and conversely, to the incompatibility of organic reagents with biological reaction media. As a result, the development of one-pot chemoenzymatic conjugation strategies remains limited.

Microbial transglutaminase (MTG) crosslinks peptide or protein substrates by catalyzing the formation of isopeptide bonds.[17] It accelerates acyl-transfer between the γ-carboxamide group of a glutamine-containing substrate, and the ε-amino group of a lysine-



Scheme 4-1 Simultaneous and subsequent chemoenzymatic one-pot protein labeling reactions.

containing substrate. Its robustness towards moderately high temperatures and water-miscible organic solvents further broadens its range of applications.[18] Additionally, while still not fully understood, MTG displays selectivity towards its glutamine substrate, with reports of its reactivity being successfully directed towards engineered targets.[19-20] Previously, we demonstrated the high flexibility of MTG towards chemically diverse amine-bearing substrates as substitutes for lysine.[21] In the same work, we demonstrated the successful conjugation of the model peptide *N*-benzyloxycarbonyl-L-glutaminylglycine (ZQG) with propargylamine using MTG. The propargylated peptide was purified, and the CuAAC reaction with an azide group was undertaken. This chemoenzymatic transformation was conducted in a stepwise fashion because it had previously been observed that the enzymatic activity of MTG was incompatible with the Cu^{2+} used in the CuAAC reaction.[22] This limitation was circumvented in a recent report in which the product of MTG-catalyzed conjugation successfully underwent the subsequent CuAAC reaction without requiring purification.[23] Building on these advancements and combining all reagents simultaneously in a one-pot fashion would streamline the process (Scheme 4-1, black arrow). Determining conditions in which MTG remains functional in the presence of CuAAC reagents, in particular copper, is



Scheme 4-2 Summary of reactions.

crucial. Furthermore, MTG and the peptide or protein substrates must not interfere with the CuAAC reaction. In addition, while MTG reacts effectively with the azido- or alkyne-bearing amines such as propargylamine and azidopropylamine,[21] it would be ideal if it could also react with the aminotriazole product of the CuAAC reaction so that all possible reaction pathways could be productive (Scheme 4-1, see blue arrows). While the CuAAC reaction has been demonstrated to be functional in the presence of MTG and substrates,[23] herein we address the remaining points. To this end, we present a one-pot strategy in which MTG is active in the presence of copper and sodium ascorbate. Considering the flexibility of MTG for its amine substrate and the wide variety of commercially-available CuAAC reagents, this methodology promises to be a powerful tool to produce diversely conjugated peptides and proteins.

Table 4-1 Subsequent one-pot reactions.

		Substrate conversion (%)					
		4 or **7**[b]		**5** or **8**[b]		**6** or **9**[b]	
Reagents	Conditions[a]	1 h	24 h	1 h	24 h	1 h	24 h
1, 2A, 3A	CuAAC first	4.6 ± 0.1	5.4 ± 0.2	92 ± 1.7	98 ± 3.6	5.4 ± 0.2	7.3 ± 0.1
	MTG first	58 ± 4.7	27 ± 12	< LOD	26 ± 0.6	< LOD	53 ± 11
1, 2B, 3B	CuAAC first	< LOD	< LOD	83 ± 1.9	82 ± 1.7	18 ± 0.9	22 ± 0.8
	MTG first	> 99	1.5 ± 0.5	< LOD	11 ± 1.0	< LOD	76 ± 6.5

[a] *Unless otherwise indicated, ZQG, alkyne and azide were present at 30 mM; details of reaction conditions are provided in the ESI. CuAAC first: CuAAC reaction performed first such that MTG was added after 24 h incubation at 37 °C with all other reagents; MTG first: $CuSO_4$ and Na^+ ascorbate was added after 2 h incubation at 37 °C with all other reagents, including MTG. Mean values and standard deviations are calculated from triplicate measurements.* [b] ***4–6*** *are the products corresponding to the azide conjugation reaction, and* ***7–9*** *to the alkyne conjugation reaction. LOD refers to limit of detection.*

First, upon quantifying MTG activity by standard assay procedures in the presence of 2.5 mM $CuSO_4$ and ascorbate, product formation was low (Table A 2-1). The proposed mechanism of MTG involves a nucleophilic attack of the thiolate ion of Cys64 on the acyl donor, the amide side-chain of a glutamine residue.[24] Cu^{1+} chelates thiols, including cysteine,[25] and may thus inactivate the catalytic thiolate. We hypothesized that the reduced form of glutathione could maintain Cys64 reactive as a free thiol.[21,26] The addition of glutathione restored activity of MTG in the presence of $CuSO_4$ and ascorbate for at least an hour but not more than an hour and a half, providing working conditions for MTG in the presence of CuAAC reagents. Additionally, we found that the presence of glutathione did not inhibit the CuAAC reaction (Table A 2-2).

Second, as the two reactions should occur in the same container simultaneously, two amine species may co-exist: the amine substrates azidopropylamine (**2A**) and propargylamine (**2B**), as well as their respective aminotriazole products, **5** and **8** resulting from the CuAAC (Scheme 4-2). To test both possibilities, we characterized the effectiveness of the reactions

Table 4-2 Simultaneous one-pot reactions.

		Substrate conversion (%)					
		4 or **7**[b]		**5** or **8**[b]		**6** or **9**[b]	
Reagents	Conditions[a]	1 h	24 h	1 h	24 h	1 h	24 h
1, 2A, 3A	A	< LOD	< LOD	4.9 ± 0.4	88 ± 1.7	< LOD	< LOD
	B	82 ± 4.9	37 ± 1.9	10 ± 0.6	37 ± 2.0	< LOD	52 ± 2.6
	C	82 ± 3.1	36 ± 1.0	7.0 ± 0.1	58 ± 2.3	< LOD	63 ± 2.6
	D	93 ± 2.6	33 ± 0.8	< LOD	30 ± 0.5	< LOD	59 ± 0.8
	E	74 ± 2.7	71 ± 8.8	< LOD	78 ± 7.3	< LOD	5.2 ± 0.4
	F	70 ± 2.8	26 ± 0.1	0.9 ± 0.1	55 ± 1.3	< LOD	50 ± 0.8
1, 2B, 3B	A	< LOD	< LOD	5.2 ± 0.3	91 ± 1.2	< LOD	< LOD
	B	96 ± 3.5	< LOD	9.8 ± 0.3	12 ± 0.5	< LOD	82 ± 1.8
	C	86 ± 4.2	0.5 ± 0.1	8.2 ± 1.5	51 ± 0.5	< LOD	95 ± 0.7
	D	> 99	0.9 ± 0.1	< LOD	9.4 ± 0.1	< LOD	80 ± 1.3
	E	87 ± 2.4	84 ± 7.6	3.7 ± 1.4	84 ± 3.4	< LOD	3.7 ± 0.2
	F	85 ± 5.3	< LOD	0.5 ± 0.1	8.2 ± 0.5	< LOD	86 ± 4.8

[a] *Unless otherwise indicated, ZQG, alkyne and azide were present at 30 mM; details of reaction conditions are provided in the ESI. A: No glutathione; B:* ***1****,* ***2A****,* ***3A****,* ***2A****,* ***2B*** *are present in equimolar concentration; C:* ***2A****,* ***2B****,* ***3A****,* ***3B*** *in 2-fold molar excess to* ***1****; D: Glutathione in 2-fold molar excess (10 mM); E:* ***3A*** *and* ***2B*** *in 2-fold molar excess to* ***1****,* ***2A*** *and* ***3B****; F: Concentrations of* ***1****,* ***2A****,* ***3A****,* ***2A****,* ***2B*** *doubled. Mean values and standard deviations are calculated from triplicate measurements.* [b] ***4****–****6*** *are the products corresponding to the azide conjugation reaction, and* ***7****–****9*** *to the alkyne conjugation reaction. LOD refers to the limit of detection.*

performed subsequently in one pot, in either reaction order (Scheme 4-1, orange and blue arrows) using amine **2A** or **2B** (Table 4-1). Substrate conversion to the amidotriazole peptides **6** and **9** was moderate, at 53% for the azide and 76% for the alkyne conjugation after 24 h (Table

4-1, MTG first). These results confirm that MTG is more reactive towards propargylamine than azidopropylamine, a trend observed for all reactions.

To evaluate the reactivity of aminotriazoles **5** and **8** as substrates for MTG, reactions containing all substrates and reagents, except MTG, were incubated for 24 h to maximize the formation of **5** and **8**. MTG was subsequently added. Little formation of **6** or **9** was observed (Table 4-1, CuAAC first), revealing that MTG does not react with aminotriazoles effectively. Our group previously observed a correlation between substrate conversion and the number of methylenes separating the primary amine from the rest of the substrate.[21] Considering the alkyl chain lengths of **5** and **8**, we expected them to react well. Their low reactivity is consistent with the lower reactivity of aryl containing, short-chain amines,[21] suggesting that aromaticity decreases MTG amine substrate reactivity to an extent that is not recovered by elongation of the intervening alkyl chain. MTG's unequal substrate reactivity demonstrates that the kinetics of product formation should be considered. The reaction conditions, specifically the order of addition of the reagents, are not trivial and should be tuned accordingly. If the CuAAC rate of formation of aminotriazole were to outpace the rate of MTG conjugation, this would render the simultaneous addition of all reagents unfeasible. However, as MTG reacts significantly faster under these conditions (Table 4-1), such a methodology is potentially viable. To this end, we performed a simultaneous one-pot chemoenzymatic scheme with all reagents in equimolar concentrations (Table 4-2, conditions A and B). Products **4**, **7**, **6** or **9** were observed only in the presence of glutathione, confirming that glutathione is essential for product formation under these conditions. The extent of substrate conversion to amidotriazole peptides **6** and **9** were 52% and 82%, respectively, after 24 h on par with results of the subsequent reaction in which the MTG conjugation step was performed first (Table 4-1). The MTG-catalyzed conjugation occurs significantly faster than the CuAAC (Table 4-2, condition B), where 82% and 96% of MTG-conjugated products **4** and **7**, respectively, were detected after 1 h, compared to 10% of the triazole products **5** and **8**. Taken with our observations of the subsequent reactions, this suggests that the MTG conjugation with **2A** or **2B** is sufficiently efficient to preclude significant interference by the aminotriazole product **5** or **8**.

In an attempt to increase the overall reaction yield, we performed the reaction with amine **2A** or **2B** and CuAAC partner **3A** or **3B** in 2-fold excess relative to ZQG, to promote reactivity

of MTG and fully conjugate ZQG. These conditions provided the highest conversions to amidotriazole peptides **6** and **9** (Table 4-1, condition C). We noted that the

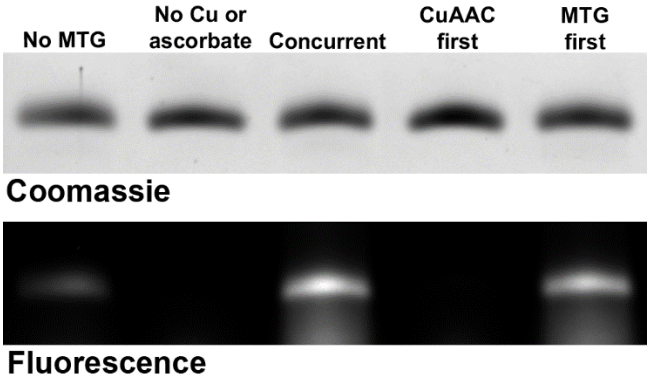


Figure 4-1 Incorporation of Cy5 into α-lactalbumin by MTG and CuAAC.

conversion of **4** to **6** is consistently lower than that of the complementary conversion of **7** to **9**, suggesting that the CuAAC reaction is less efficient with **4** than with **7**. Based on this observation, we hypothesized that if the glutathione concentration was doubled, the formation of products **4** and **7** may be optimized due to increased MTG activity and thus increase the conversions to **6** and **9**. While modest increases of **4** and **7** were observed (Table 4-1, condition D), there was no corresponding increase in the formation of the amidotriazole peptides **6** and **9**. Doubling the concentration of the amine reagent **2A** or **2B** relative to ZQG and respective CuAAC partner **3A** or **3B** resulted in poor formation of the conjugated triazole peptides (Table 4-1 condition E). The correspondingly increased yields of **5** and **8** indicate that the complementary alkyne or azide reacts preferentially with the unconjugated amine **2A** or **2B** rather than the conjugated peptide **4** or **7**, as the concentrations of the amine and the conjugated peptide should approximately be equivalent by the time the CuAAC reaction has gone to completion. We note that decreasing the copper concentration 10-fold produced no triazole-containing product (results not shown). Furthermore, the addition of supplementary MTG to the reaction after the initial aliquot of enzyme had lost activity resulted only in a marginal increase of product formation (results not shown).

In light of these results, we investigated the effect of doubling the concentration of all reagents simultaneously. We hypothesized that the high K_m of MTG for ZQG (27 mM)[27] limits conversion efficiency; doubling all concentrations increases the saturation of MTG and could

increase substrate conversion to amidotriazole peptide. While the efficiency of conversion to the product remained the same (Table 4-1: compare conditions B and F), it was possible to double the volumetric conversion by doubling reagent concentrations. Overall, the simultaneous one-pot reaction is fairly robust and can provide conversion of up to 95%. Key to its success is that the MTG-catalyzed conjugation must occur significantly faster than the CuAAC. To demonstrate the applicability of this simultaneous one-pot chemoenzymatic modification strategy toward protein labeling, we modified α-lactalbumin (α-LA). The surface Gln residues Q39 and Q54 of α-LA are known to react rapidly and preferentially in MTG-catalyzed reactions at micromolar concentrations, which is considerably lower than the millimolar concentrations required for ZQG.[28] Cy5-alkyne or Cy5- azide were used to incorporate a fluorophore onto α-LA in the presence of **2A** or **2B**, respectively. As with the ZQG peptide reactions, the MTG-mediated transamidation of α-LA and CuAAC reactions were performed simultaneously or subsequently to each other. Fluorescence incorporation was observed to be successful relative to the control reactions, except when the CuAAC was performed prior to the transamidation step (Fig. 4-1). This is consistent with our findings that MTG reacts poorly with the aminotriazole substrate (Table 4-1, CuAAC first). We note that a-LA exhibits trace amounts of fluorescence in the presence of Cu^{2+} and ascorbate, independent of MTG. This interaction is likely a result of the Cu^{2+}-binding activity of α-LA.[29]

In summary, we have identified conditions of an effective one-pot reaction for specific covalent modification of glutamine containing peptides and proteins by combining enzyme-catalyzed transamidation and click chemistry. The reaction occurs within 24 h and the highest conversions to amidotriazole peptides were obtained when the amine reagent was in a 2-fold excess relative to the peptide concentration. Aminotriazoles are not good substrates for MTG, and a one-pot simultaneous setup is feasible owing to the fact that under these reaction conditions, MTG's conjugation activity is faster than the CuAAC. The applicability of the system was demonstrated by the conjugation of α-LA with clickable fluorescent dyes. In combination with the controlled Gln selectivity of MTG, this one-pot chemoenzymatic reaction is a straightforward and versatile peptide and protein conjugation methodology that can be used for specific protein tagging.

4.4 References

1. Krishna, O. D.; Kiick, K. L., Protein- and peptide-modified synthetic polymeric biomaterials. *Biopolymers* **2010,** *94*, 32-48.
2. Kochendoerfer, G. G., Site-specific polymer modification of therapeutic proteins. *Curr. Opin. Chem. Biol.* **2005,** *9*, 555-560.
3. Clouthier, C. M.; Pelletier, J. N., Expanding the organic toolbox: a guide to integrating biocatalysis in synthesis. *Chem. Soc. Rev.* **2012,** *41*, 1585-1605.
4. Reetz, M. T., Biocatalysis in organic chemistry and biotechnology: past, present, and future. *J. Am. Chem. Soc.* **2013,** *135*, 12480-12496.
5. Rashidian, M.; Dozier, J. K.; Distefano, M. D., Enzymatic labeling of proteins: techniques and approaches. *Bioconjug. Chem.* **2013,** *24*, 1277-1294.
6. Fernandez-Suarez, M.; Baruah, H.; Martinez-Hernandez, L.; Xie, K. T.; Baskin, J. M.; Bertozzi, C. R.; Ting, A. Y., Redirecting lipoic acid ligase for cell surface protein labeling with small-molecule probes. *Nat. Biotechnol.* **2007,** *25*, 1483-1487.
7. Yin, J.; Liu, F.; Li, X.; Walsh, C. T., Labeling proteins with small molecules by site-specific posttranslational modification. *J. Am. Chem. Soc.* **2004,** *126*, 7754-7755.
8. Yin, J.; Lin, A. J.; Golan, D. E.; Walsh, C. T., Site-specific protein labeling by Sfp phosphopantetheinyl transferase. *Nat. Protoc.* **2006,** *1*, 280-285.
9. Ramachary, D. B.; Jain, S., Sequential one-pot combination of multi-component and multi-catalysis cascade reactions: an emerging technology in organic synthesis. *Org. Biomol. Chem.* **2011,** *9*, 1277-1300.
10. Denard, C. A.; Hartwig, J. F.; Zhao, H., Multistep One-Pot Reactions Combining Biocatalysts and Chemical Catalysts for Asymmetric Synthesis. *ACS Catal.* **2013,** *3*, 2856-2864.
11. Brusa, C.; Ochs, M.; Rémond, C.; Murielle, M.; Plantier-Royon, R., Chemoenzymatic synthesis of “click” xylosides and xylobiosides from lignocellulosic biomass. *RCS Adv.* **2014,** *4*, 9330-9338.
12. Schulz, D.; Holstein, J. M.; Rentmeister, A., A chemo-enzymatic approach for site-specific modification of the RNA cap. *Angew. Chem., Int. Ed.* **2013,** *52*, 7874-7878.
13. Rostovtsev, V. V.; Green, L. G.; Fokin, V. V.; Sharpless, K. B., A stepwise huisgen cycloaddition process: copper(I)-catalyzed regioselective "ligation" of azides and terminal alkynes. *Angew. Chem., Int. Ed.* **2002,** *41*, 2596-2599.
14. Kulkarni, C.; Lo, M.; Fraseur, J. G.; Tirrell, D. A.; Kinzer-Ursem, T. L., Bioorthogonal Chemoenzymatic Functionalization of Calmodulin for Bioconjugation Applications. *Bioconjug. Chem.* **2015,** *26*, 2153-2160.
15. Hassan, S.; Tschersich, R.; Müller, T. J. J., Three-component Chemoenzymatic Synthesis of Amide Ligated 1,2,3-Triazoles. *Tetrahedron Lett.* **2013,** *54*, 4641-4644.
16. Hassan, S.; Ullrich, A.; Muller, T. J., Consecutive three-component synthesis of (hetero)arylated propargyl amides by chemoenzymatic aminolysis-Sonogashira coupling sequence. *Org. Biomol. Chem.* **2015,** *13*, 1571-1576.
17. Ando, H.; Adachi, M.; Umeda, K.; Matsuura, A.; Nonaka, M.; Uchio, R.; Tanaka, H.; Motoki, M., Purification and characteristics of a novel transglutaminase derived from microorganisms. *Agric. Biol. Chem.* **1989,** *53*, 2613-2617.
18. Yokoyama, K.; Nio, N.; Kikuchi, Y., Properties and applications of microbial transglutaminase. *Appl. Microbiol. Biotechnol.* **2004,** *64*, 447-454.

19. Oteng-Pabi, S. K.; Keillor, J. W., Continuous enzyme-coupled assay for microbial transglutaminase activity. *Anal. Biochem.* **2013,** *441*, 169-173.
20. Strop, P.; Liu, S. H.; Dorywalska, M.; Delaria, K.; Dushin, R. G.; Tran, T. T.; Ho, W. H.; Farias, S.; Casas, M. G.; Abdiche, Y.; Zhou, D.; Chandrasekaran, R.; Samain, C.; Loo, C.; Rossi, A.; Rickert, M.; Krimm, S.; Wong, T.; Chin, S. M.; Yu, J.; Dilley, J.; Chaparro-Riggers, J.; Filzen, G. F.; O'Donnell, C. J.; Wang, F.; Myers, J. S.; Pons, J.; Shelton, D. L.; Rajpal, A., Location matters: site of conjugation modulates stability and pharmacokinetics of antibody drug conjugates. *Chem. Biol.* **2013,** *20*, 161-167.
21. Gundersen, M. T.; Keillor, J. W.; Pelletier, J. N., Microbial transglutaminase displays broad acyl-acceptor substrate specificity. *Appl. Microbiol. Biotechnol.* **2014,** *98*, 219-230.
22. Seguro, K. N., N.; Motoki, M., Some characteristics of a microbial protein cross-linking enzyme: Transglutaminase. . *ACS Symp. Ser.* **1996,** *650*, 271-280.
23. Oteng-Pabi, S. K.; Pardin, C.; Stoica, M.; Keillor, J. W., Site-specific protein labelling and immobilization mediated by microbial transglutaminase. *Chem. Commun.* **2014,** *50*, 6604-6606.
24. Kashiwagi, T.; Yokoyama, K.; Ishikawa, K.; Ono, K.; Ejima, D.; Matsui, H.; Suzuki, E., Crystal structure of microbial transglutaminase from Streptoverticillium mobaraense. *J. Biol. Chem.* **2002,** *277*, 44252-44260.
25. Wu, Z.; Fernandez-Lima, F. A.; Russell, D. H., Amino acid influence on copper binding to peptides: cysteine versus arginine. *J. Am. Soc. Mass. Spectrom.* **2010,** *21*, 522-533.
26. Dixon, M.; Quastel, J. H., A new type of reduction–oxidation system. Part I. Cysteine and glutathione. *J. Chem. Soc.* **1923,** (123), 2943-2953.
27. Ohtsuka, T.; Umezawa, Y.; Nio, N.; Kubota, K., Comparison of Deamidation Activity of Transglutaminases. *J. Food Sci.* **2001,** *66*, 26-29.
28. Spolaore, B.; Raboni, S.; Ramos Molina, A.; Satwekar, A.; Damiano, N.; Fontana, A., Local unfolding is required for the site-specific protein modification by transglutaminase. *Biochemistry* **2012,** *51*, 8679-8689.
29. Van Dael, H.; Tieghem, E.; Haezebrouck, P.; Van Cauwelaert, F., Conformational aspects of the Cu^{2+} binding to alpha-lactalbumin. Characterization and stability of the Cu-bound state. *Biophys. Chem.* **1992,** *42*, 235-242.

Chapter 5 - Transglutaminase-catalyzed bioconjugation using one-pot metal-free bioorthogonal chemistries

5.1 Context

The field of bioorthogonal chemistry is of high interest and was accelerated with the advent of the CuAAC. A number of alternative metal-free conjugation reactions have been reported, each with their own advantages, limitations, and characteristics. As our one-pot investigations with microbial transglutaminase and the CuAAC came to a conclusion, we were intrigued by the variation we observed in substrate conversion as a result of reaction conditions, and wondered if the complementary chemical reaction could also play a role in affecting conjugation efficiency. While MTG accepts a variety of amines as substrates, its stringent substrate specificity for glutamine-bound peptides and proteins led to us to hypothesize that it would be best suited as a biocatalyst for protein conjugation, rather than for small-molecule conjugation. Therefore, this work varies from the previous chapter by the investigated glutamine substrates, which are globular proteins rather than a simple protected dipeptide. This chapter is thus focussed on evaluating reactivity amongst the different metal-free conjugation reactions.

This chapter is a reproduction of a manuscript submitted to the journal *ACS Bioconjugate Chemistry*, entitled: *Transglutaminase-Catalyzed Bioconjugation using One-Pot Metal-Free Bioorthogonal Chemistry*. My contribution was the conceptualization and realization of laboratory experiments, performed in the laboratory of Prof. Joelle Pelletier. Jacynthe Toulouse purified one of the protein substrates, hDHFR. The manuscript was drafted by myself with assistance from Prof. Pelletier and revision by Prof. Andreea Schmitzer. Supporting information associated with this manuscript is available for consultation in Annex 3 of this thesis.

Transglutaminase-catalyzed Bioconjugation using One-Pot Metal-Free Bioorthogonal Chemistry

Natalie M. Rachel,[1,2,3] Jacynthe L. Toulouse,[3,4] and Joelle N. Pelletier[1,2,3,4]

[1] Département de Chimie, Université de Montréal, 2900 Boulevard Edouard-Montpetit, Montréal, Québec, H3T 1J4, Canada

[2] CGCC, the Center in Green Chemistry and Catalysis, Montréal, H3A 0B8, Canada

[3] PROTEO, the Québec Network for Protein Function, Structure and Engineering, Québec, G1V 0A6, Canada

[4] Départment de Biochimie, Université de Montréal, 2900 Boulevard Edouard-Montpetit, Montréal, Québec, H3T 1J4, Canada

Manuscript submitted for review.

Corresponding Author: Joelle N. Pelletier <joelle.pelletier@umontreal.ca>

5.2 Abstract

General approaches for controlled protein modification are increasingly sought-after in the arena of chemical biology. Here, using bioorthogonal reactions, we present combinatorial chemoenzymatic strategies to undertake protein labeling. Three metal-free conjugations were simultaneously or sequentially incorporated in a one-pot format with microbial transglutaminase (MTG) to effectuate protein labeling. MTG offers the particularity of crosslinking within a protein sequence, rather than at its extremities, providing a route to internal protein labeling. The reactions are rapid and circumvent the incompatibility posed by metal catalysts. We identify the tetrazine ligation as most reactive for this purpose, as demonstrated by the fluorescent labeling of two proteins. The Staudinger ligation and strain-promoted azide-alkyne cycloaddition are alternatives. Owing to the breadth of labels that MTG can use as a substrate, our results demonstrate the versatility of this system, with the researcher being able to combine specific protein substrates with a variety of labels.

5.3 Article content

Site-selective protein modification is one of the most complex challenges of chemical biology, with applications ranging from visualizing cellular components,[1] modification of complex small molecules[2] and synthesizing drug-antibody conjugates.[3-4] Enzyme-catalyzed bioconjugation offers the advantage of improved selectivity and compatibility with sensitive biological systems relative to traditional chemical methodologies.[1, 5-6] Combining the selectivity of enzyme-catalyzed bioconjugation with versatile bio-orthogonal chemistries will ultimately allow for the specific incorporation of a range of abiotic chemical structures into proteins.[7-8] Click chemistry has transformed the way we approach applying traditional chemical catalysis to sensitive biological systems.[9-10] The copper-catalyzed azide-alkyne cycloaddition (CuAAC) was the first of such reactions[11] and has since been characterized extensively both fundamentally and in applied methodologies.[12-13] Despite these successes, the CuAAC has limitations that curtail its generality. The most noteworthy limitation in the context of biological applications is its reliance on copper, which exerts a significant biological toxicity and is therefore incompatible with many biomolecules,[14] particularly enzymes. Copper toxicity was observed and ultimately circumvented in our previous work, in which we characterized the compatibility of the CuAAC with microbial transglutaminase (MTG) bioconjugation in a one-pot chemoenzymatic system.[15] MTG catalyzes the formation of a stable isopeptide bond between a protein- or peptide-bound glutamine residue and a primary amine,[16] establishing it as a tool for covalent protein modification.[17] Contrary to other enzymatic conjugation methods that typically target protein termini, MTG's labeling sites can be located at an accessible site within the protein. As a result, MTG provides a means to internally incorporate labels into proteins. MTG reacts with a broad scope of amine substrates, and although it is more effective with long-chain amines, shorter ones can be accepted as well.[18] The protected dipeptide, *N*-benzyloxycarbonyl-L-glutaminylglycine (ZQG), was the model substrate used for the characterization of reaction conditions which were ultimately extended to labeling a protein with a fluorophore.[15] Despite these achievements, conjugation by that one-pot scheme will be incompatible with protein substrates that are inactivated, or precipitate, in the presence of copper.

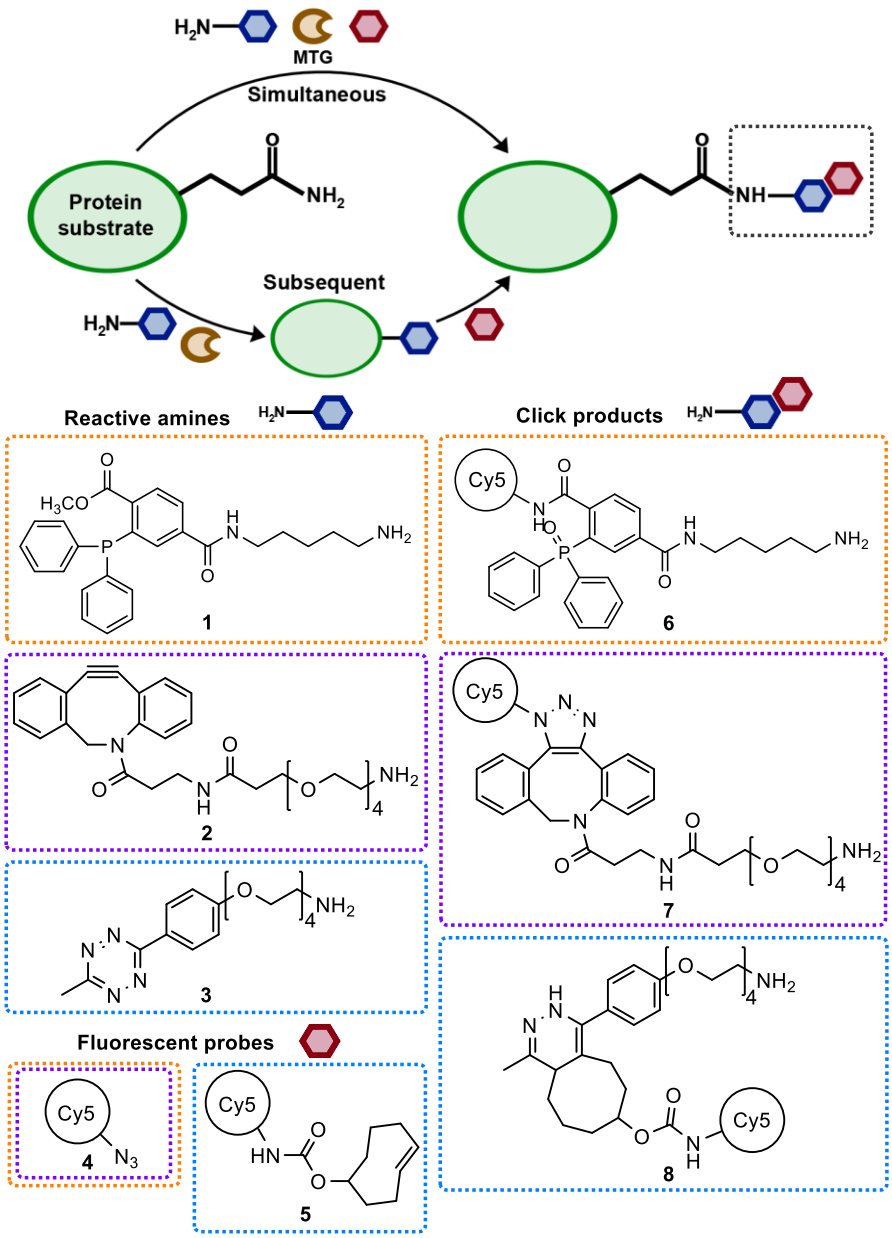


Figure 5-1 Simultaneous and subsequent chemoenzymatic one-pot protein labeling reactions.

The various reactive amines, fluorescent probes, and corresponding bioconjugated product are shown for the Staudinger ligation (orange boxes), SPAAC (purple boxes), and TL (blue boxes).

The frequently-encountered limitation of reagent toxicity has prompted the development of metal-free click chemistry approaches. One of the first was the Staudinger ligation where a triarylphosphine bearing an *o*-ester group reacts with an azide to form an amide bond.[19-21] The

sensitivity of the phosphines to oxidation and their modest aqueous solubility promoted research into other chemistries. The strain-promoted azide-alkyne cycloaddition (SPAAC) was the first metal-free alternative to the CuAAC, allowing catalyst-free [3 + 2] cycloaddition.[22-23] The SPAAC utilizes activated, strained cyclooctynes to initiate spontaneous cycloaddition with a terminal azide and is often a few orders of magnitude more rapid than the Staudinger ligation.[24] In light of this improvement, the interest in extremely rapid bioconjugation reactions produced the tetrazine ligation (TL),[25] which varies from the SPAAC and Staudinger ligation by not relying on azides or alkynes. Instead, *trans*-cyclooctene and *s*-tetrazine react in an inverse-electron-demand Diels-Alder reaction. With a second-order rate constant ranging between 210 and 2.8 x 10^6 $M^{-1}s^{-1}$ depending on solvent composition, the TL has the advantage of being highly reactive at low concentrations.[26]

Metal-free bio-orthogonal alternatives to the CuAAC may thus be advantageous partners to the MTG-catalyzed formation of amide linkages to further its applicability for one-pot protein labeling (Figure 5-1). Here, we compare simultaneous and subsequent metal-free, chemoenzymatic one-pot conjugation schemes to examine whether MTG function is compatible with the various click reagents.

Contrary to our previous one-pot MTG-CuAAC conjugation work,[15] only one of the two patterns of subsequent reagent addition was assessed. Indeed, the metal-free click reactions investigated here all occur orders of magnitude faster than the enzymatic conjugation step.[24, 27] As they are not rate-limiting, the simultaneous reaction scheme approximates a reaction starting with the metal-free click reactions and following with the slower MTG-catalyzed bioconjugation. Reactive amines **2** and **3** differ from **1** both in the main reactive moiety as well as by their linker (Figure 5-1). Similarly to the alkyl linker of **1**,[28] the PEG4 linker of **2** and **3** has been demonstrated to be compatible with MTG[29] and has the benefit of increasing the solubility of the hydrophobic dibenzocyclooctyne moiety of **2**.

With the goal of extending these chemoenzymatic methodologies to protein labeling, we performed the labeling reactions on the substrates bovine α-lactalbumin (α-LA) and human

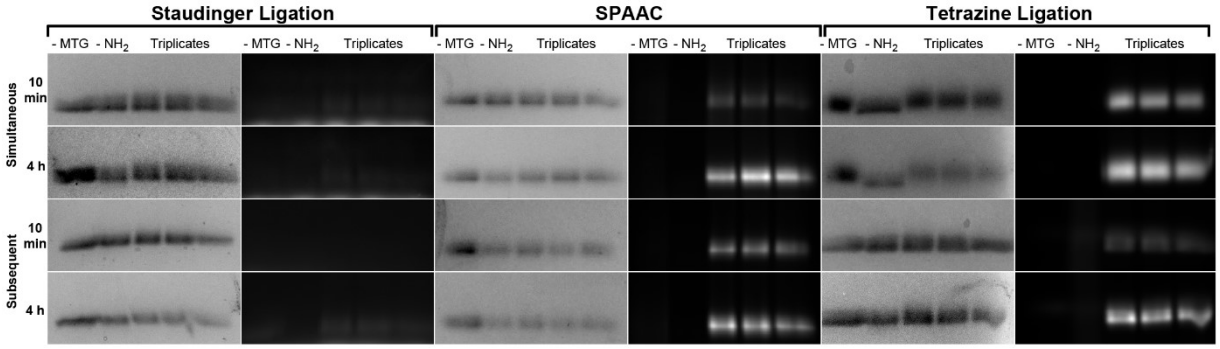


Figure 5-2 One-pot chemoenzymatic protein labeling of α-LA catalyzed by MTG.

The Staudinger ligation, SPAAC and TL were performed at 37°C. Aliquots were taken after 10 min and 4 hours of reaction time, quenched by the addition of formic acid and resolved by tricine SDS-PAGE. The negative controls lacked either MTG (- MTG) or the amine substrate (- NH_2) and were run in parallel. Left-hand panels: Coomassie brilliant blue staining, performed after the gel was excited with a Cy5 filter to detect fluorescence (right-hand panels). The top two and bottom two rows correspond to the simultaneous and subsequent reaction format, respectively.

Table 5-1 Fluorescence intensities for reaction with α-LA[a].

Reaction	Order of addition	Controls		Samples[b]
		No MTG ($\times 10^4$)	No NH_2 ($\times 10^4$)	($\times 10^4$)
Staudinger	Simultaneous	3.4	4.1	34 ± 7
	Subsequent	320	220	200 ± 52
SPAAC	Simultaneous	110	350	1900 ± 150
	Subsequent	70	200	3600 ± 190
Tetrazine	Simultaneous	28	75	5900 ± 480
	Subsequent	160	490	2800 ± 67

[a] Fluorescence intensities were quantified by Image Lab™ on the SDS-PAGE gels of samples taken after 10 min of reaction; [b] Average of triplicate samples.

Table 5-2 Fluorescence intensities for reaction with α-LA[a]

Reaction	Order of addition	Controls		Samples[b]
		No MTG ($\times 10^4$)	No NH_2 ($\times 10^4$)	($\times 10^4$)
Staudinger	Simultaneous	2.9	34	210 ± 17
	Subsequent	450	331	1400 ± 100
SPAAC	Simultaneous	850	543	6800 ± 910
	Subsequent	120	177	5700 ± 250
Tetrazine	Simultaneous	45	29	9800 ± 850
	Subsequent	670	407	7100 ± 400

[a] Fluorescence intensities were quantified by Image Lab™ on the SDS-PAGE gels of samples taken after 10 min of reaction; [b] Average of triplicate samples.

dihydrofolate reductase (hDHFR) rather than on the previously studied model peptide ZQG.

Proteins allow use of significantly lower reagent concentrations because MTG-reactive proteins display higher affinity for MTG than does the ZQG peptide. α-LA was selected as it exhibits excellent reactivity with MTG,[30] and hDHFR was selected because of its therapeutic interest. Bioconjugation on hDHFR was previously performed to characterize the thermodynamics and kinetics of its interaction with methotrexate,[31] and to monitor catalytic conformational transitions.[32] At the outset of this study, hDHFR's reactivity with MTG had not been previously established; it cannot be predicted because the understanding of MTG's preference for protein substrates remains superficial.[30, 33-34]

Glutathione is typically included in protein preparations that require cysteine residues to remain in their reduced form, as is the case for MTG because its catalytic nucleophile is a cysteine. Indeed, we previously found that glutathione was essential in the presence of Cu(I) for the CuAAC-mediated chemoenzymatic labeling reaction, presumably to prevent chelation of the catalytic thiolate by Cu(I).[15] In contrast, exclusion of glutathione in these metal-free reactions led to results that were indistinguishable from those in which glutathione was present (Figure A 3-1), immediately illustrating the benefit of removing the metal.

Beginning with the Staudinger ligation, we combined the α-LA protein substrate with amine **1** and Cy5-azide to execute the chemoenzymatic labeling (Figure 5-1). In all reactions, we maintained a 2-fold excess of the amine relative to the protein substrate, as it was previously shown to give the highest yields for the MTG-catalyzed step.[15] In both a one-pot simultaneous or subsequent format, only low fluorescence was observed on α-LA, even after 4 hours of reaction (Figure 5-2, Tables 5-1 and 5-2). The formation of the desired chemoenzymatic conjugated α-LA product was confirmed by high-resolution mass spectrometry (HRMS; see the Supporting Information). In sharp contrast, the labeling performed with amine **2** via the SPAAC reaction showed significant fluorescence incorporation after 10 min: 55- or 18-fold more than the corresponding Staudinger ligation, its intensity increasing 2- to 4-fold after 4 hours, for the simultaneous and subsequent reactions, respectively. Reaction specificity was confirmed by the absence of labeling when either MTG or amine **2** were absent.

While the MTG-conjugated α-LA species was clearly observed upon resolution by denaturing gel electrophoresis, it was not observed upon performing exact mass analysis using HRMS (Supporting Information). This suggests that the SPAAC reaction proceeded at a yield below the MS detection limit, yet clearly above the threshold for fluorescence detection. We verified whether the low yield could result from poor solubility of **2** by performing the reaction in buffer containing 30% and 50% DMSO, where both MTG and α-LA are stable.[35] The addition of DMSO increased the solubility of **2** by 10-fold and 20-fold, respectively. While fluorescence incorporation was observed on resolving gel after 24 hours (Figure A 3-2), the conjugated α-LA species was not observed by MS. Thus, solubility of **2** is not a significant factor in low conjugation yield, and these results demonstrate that **2** is a poor substrate for MTG.

We investigated the TL using amine **3** which, similar to **2**, has a PEG4 linker separating the amine from the reactive tetrazine site. The fluorescence intensity was 3-fold higher yet than the SPAAC reaction, suggesting that amine **3**, or the click product **8**, were the most reactive amines with MTG. As noted above, the metal-free click reactions investigated here all occur orders of magnitude faster than the enzymatic conjugation step and thus are not rate- limiting.[24, 27] Having all the reagents present in a simultaneous format procured greater reactivity than in a subsequent format, as evidenced by the 2.5-fold increase in fluorescence after 10 min. This suggests that 8 is more reactive with MTG than **3**; the subsequent format favors direct

conjugation of **3** because its click probe, **5**, is not present in the solution while MTG is active. We confirmed that **3** is a better substrate for MTG than **1** or **2** by observing significant formation of the conjugated α-LA species using high resolution liquid chromatography mass spectrometry with detection at 214 nm (Figure 5-3). After 4 hours, the conjugated species were clearly visible, the identity of which was confirmed by MS. The combined peak areas of native and conjugated α-LA were less than the unreacted control α-LA, consistent with formation of a side-product. α-LA is known to multimerize as MTG crosslinks its multiple reactive glutamine and lysine residues.[36] The appearance of an extremely broad peak spanning from 7.5-10 min retention time after 4 hours and 24 hours support this observation.

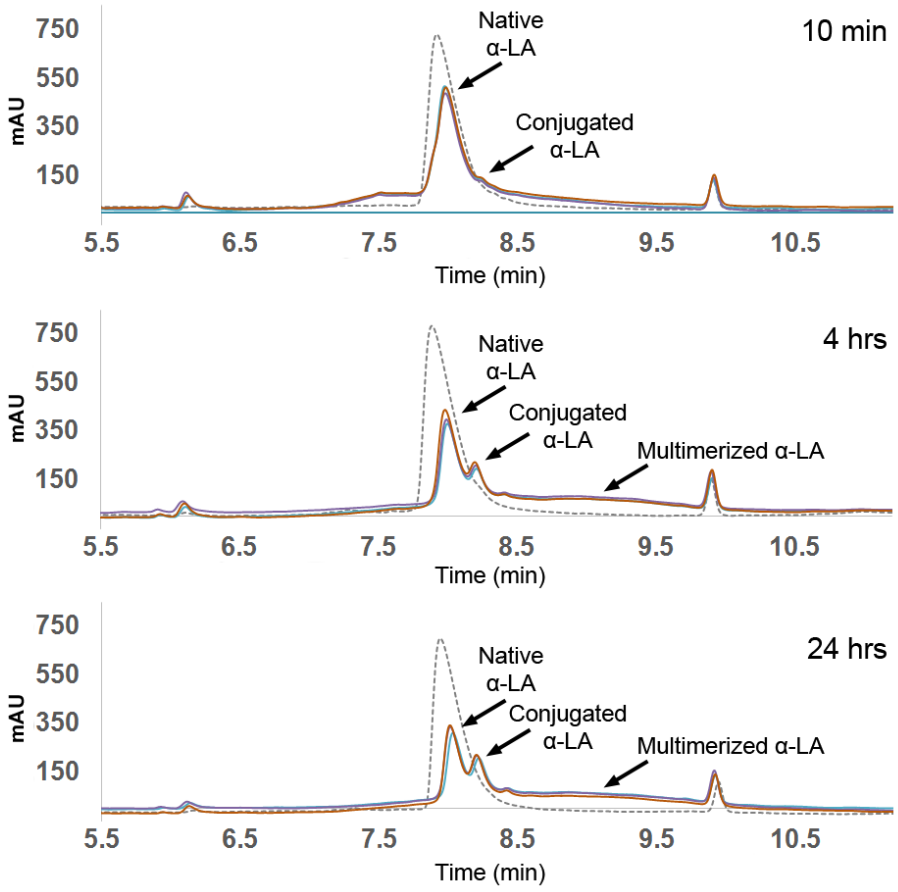


Figure 5-3 LC chromatograms of α-LA upon conjugation with 3 by MTG.

Detection was performed at 214 nm. Grey dashed traces and colored traces correspond to a negative control lacking MTG and reaction triplicates, respectively.

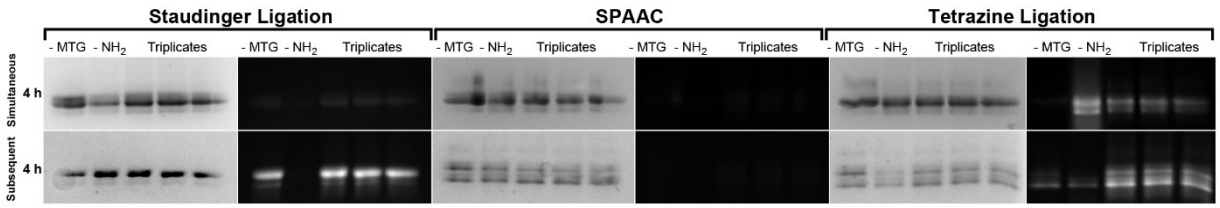


Figure 5-4 One-pot chemoenzymatic protein labeling of hDHFR catalyzed by MTG.
The reactions were performed as previously described for α-LA, above. The simultaneous and subsequent reaction formats are shown.

Table 5-3 Fluorescence intensities for reaction with hDHFR.[a]

Reaction	Order of addition	Controls		Samples[b]
		No MTG ($\times 10^4$)	No NH_2 ($\times 10^4$)	($\times 10^4$)
Staudinger	Simultaneous	310	96	450 ± 84
	Subsequent	3400	59	4200 ± 390
SPAAC	Simultaneous	460	95	410 ± 50
	Subsequent	52	16	130 ± 10
Tetrazine	Simultaneous	510	2700	2700 ± 200
	Subsequent	1000	620	3600 ± 270

[a] Fluorescence intensities were quantified by Image Lab™ on the SDS-PAGE gels of samples taken after 10 min of reaction; [b] Average of triplicate samples.

To investigate the chemoenzymatic labeling of a therapeutically relevant protein, we performed the same reactions with hDHFR. While MTG can conjugate a variety of proteins, it will not exhaustively react with all glutamine or lysine residues.[15] Previously, we also attempted MTG-catalyzed bioconjugation of two other clinically-relevant enzymes, *E. coli* L-asparaginase and TEM-1 β-lactamase, but neither displayed any reactivity with MTG (data not shown). We were gratified to observe that hDHFR served as a substrate for MTG, which to the best of our knowledge, has not yet been reported.

Upon performing the Staudinger ligation on hDHFR, non-specific fluorescence was observed, where the fluorescent labeling was independent of the presence of MTG (Figure 4, Table 5-3); no conjugated product was detected using high-resolution mass spectrometry,

contrary to that observed for α-LA. The same phenomenon was observed with the SPAAC, with no fluorescence being observed when the SPAAC was performed. While non-specific fluorescence incorporation was observed with the TL, particularly in the simultaneous reaction format, the subsequent reaction format is advantageous. These findings demonstrate that hDHFR is less reactive than α-LA yet it can be successfully labeled. Because MTG displays some selectivity towards its protein substrates, even low reactivity can produce strong signal-to-background fluorescence ratio.

We thus demonstrate that the metal-free Staudinger ligation, SPAAC and TL can each be combined with MTG-catalyzed bioconjugation in a one-pot format, although their reactivity varies significantly. Among the reactions, the TL was the most effective in the extent of labeling and in its applicability to both protein substrates tested. We were surprised that the SPAAC was significantly less effective, as the amine **2** and **3** both have the same long-chained linker which should result in similar amine substrate recognition by MTG. It appears that a long spacer between the MTG-reactive amine group and the clickable moiety is not enough to guarantee MTG reactivity. This illustrates how the choice of amine influences not only the reactivity, but the selectivity as well. Indeed, reactivity was not equal for all protein substrates; the Staudinger and the SPAAC reactions were both detected on α-LA but not on hDHFR. The TL is sufficiently reactive to label both α-LA and hDHFR, whereas the SPAAC only functioned with α-LA. This differential reactivity provides the researcher with a degree of control over the labeling methodology. Despite low conversion, fluorescently-labelled hDHFR was successfully visualized, demonstrating the system's potential for labeling medically relevant proteins.

In summary, our results show that metal-free combinatorial chemoenzymatic strategies can be utilized for internal fluorescent protein labeling. Fluorescence detection can be observed within minutes, with the TL ligation being most effective. If the TL is not possible, the SPAAC remains a feasible alternative, with the Staudinger ligation being the least reactive toward these protein targets. Coupled with the reactivity of MTG, we demonstrate the alternatives and their considerations to the CuAAC to synthesize chemically diverse covalently modified proteins, particularly for those that are sensitive to metals.

5.4 Acknowledgements

The authors thank Marie-Christine Tang and Alexandra Furtos of the Regional Mass Spectrometry Centre (Université de Montréal) for their technical aid, and Prof. Andreea Schmitzer (Université de Montréal) for thoughtful revision of the manuscript. This work was supported by Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant RGPIN 227853.

5.5 References

1. Krall, N.; da Cruz, F. P.; Boutureira, O.; Bernardes, G. J., Site-selective protein-modification chemistry for basic biology and drug development. *Nat. Chem.* **2016,** *8* (2), 103-113.
2. Lichtor, P. A.; Miller, S. J., Combinatorial evolution of site- and enantioselective catalysts for polyene epoxidation. *Nat. Chem.* **2012,** *4*, 990-995.
3. Agarwal, P.; Bertozzi, C. R., Site-specific antibody-drug conjugates: the nexus of bioorthogonal chemistry, protein engineering, and drug development. *Bioconjug Chem* **2015,** *26*, 176-192.
4. Akkapeddi, P.; Azizi, S.; Freedy, A.; Cal, P. M. S. D.; Gois, P. M. P.; Bernardes, G. J., Construction of homogeneous antibody–drug conjugates using site-selective protein chemistry. *Chem. Sci.* **2016,** *7*, 2954-2963.
5. Miller, L. W.; Cornish, V. W., Selective chemical labeling of proteins in living cells. *Curr. Opin. Chem. Biol.* **2005,** *9*, 56-61.
6. Rashidian, M.; Dozier, J. K.; Distefano, M. D., Enzymatic labeling of proteins: techniques and approaches. *Bioconjug. Chem.* **2013,** *24*, 1277-1294.
7. Lin, S.; Yang, X.; Jia, S.; Weeks, A. M.; Hornsby, M.; Lee, P. S.; Nichiporuk, R. V.; Iavarone, A. T.; Wells, J. A.; Toste, F. D.; Chang, C. J., Redox-based reagents for chemoselective methionine bioconjugation. *Science* **2017,** *355*, 597-602.
8. Wang, H.; Wang, R.; Cai, K.; He, H.; Liu, Y.; Yen, J.; Wang, Z.; Xu, M.; Sun, Y.; Zhou, X.; Yin, Q.; Tang, L.; Dobrucki, I. T.; Dobrucki, L. W.; Chaney, E. J.; Boppart, S. A.; Fan, T. M.; Lezmi, S.; Chen, X.; Yin, L.; Cheng, J., Selective in vivo metabolic cell-labeling-mediated cancer targeting. *Nat. Chem. Biol.* **2017,** *13*, 415-424.
9. Thirumurugan, P.; Matosiuk, D.; Jozwiak, K., Click chemistry for drug development and diverse chemical-biology applications. *Chem. Rev.* **2013,** *113*, 4905-4979.
10. McKay, C. S.; Finn, M. G., Click chemistry in complex mixtures: bioorthogonal bioconjugation. *Chem. Biol.* **2014,** *21*, 1075-1101.
11. Kolb, H. C.; Finn, M. G.; Sharpless, K. B., Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angew. Chem., Int. Ed.* **2001,** *40*, 2004-2021.

12. Liang, L.; Astruc, D., The copper(I)-catalyzed alkyne-azide cycloaddition (CuAAC) "click" reaction and its applications. An overview. *Coordin. Chem. Rev.* **2011,** *255*, 2933-2945.
13. Uttamapinant, C.; Tangpeerachaikul, A.; Grecian, S.; Clarke, S.; Singh, U.; Slade, P.; Gee, K. R.; Ting, A. Y., Fast, cell-compatible click chemistry with copper-chelating azides for biomolecular labeling. *Angew. Chem., Int. Ed.* **2012,** *51*, 5852-5856.
14. Anderson, C. T.; Wallace, I. S.; Somerville, C. R., Metabolic click-labeling with a fucose analog reveals pectin delivery, architecture, and dynamics in Arabidopsis cell walls. *Proc. Natl. Acad. Sci. U.S.A.* **2012,** *109*, 1329-34.
15. Rachel, N. M.; Pelletier, J. N., One-pot peptide and protein conjugation: a combination of enzymatic transamidation and click chemistry. *Chem. Commun.* **2016,** *52*, 2541-2544.
16. Ando, H.; Adachi, M.; Umeda, K.; Matsuura, A.; Nonaka, M.; Uchio, R.; Tanaka, H.; Motoki, M., Purification and characteristics of a novel transglutaminase derived from microorganisms. *Agric. Biol. Chem.* **1989,** *53*, 2613-2617.
17. Strop, P., Versatility of microbial transglutaminase. *Bioconjug. Chem.* **2014,** *25*, 855-62.
18. Gundersen, M. T.; Keillor, J. W.; Pelletier, J. N., Microbial transglutaminase displays broad acyl-acceptor substrate specificity. *Appl. Microbiol. Biotechnol.* **2014,** *98*, 219-230.
19. Saxon, E.; Armstrong, J. I.; Bertozzi, C. R., A "traceless" Staudinger ligation for the chemoselective synthesis of amide bonds. *Org. Lett.* **2000,** *2*, 2141-2143.
20. Saxon, E.; Bertozzi, C. R., Cell surface engineering by a modified Staudinger reaction. *Science* **2000,** *287*, 2007-2010.
21. Kiick, K. L.; Saxon, E.; Tirrell, D. A.; Bertozzi, C. R., Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation. *Proc. Natl. Acad. Sci. U.S.A.* **2002,** *99*, 19-24.
22. Agard, N. J.; Prescher, J. A.; Bertozzi, C. R., A strain-promoted [3 + 2] azide-alkyne cycloaddition for covalent modification of biomolecules in living systems. *J. Am. Chem. Soc.* **2004,** *126*, 15046-15047.
23. Jewett, J. C.; Sletten, E. M.; Bertozzi, C. R., Rapid Cu-free click chemistry with readily synthesized biarylazacyclooctynones. *J. Am. Chem. Soc.* **2010,** *132*, 3688-3690.
24. Patterson, D. M.; Nazarova, L. A.; Prescher, J. A., Finding the right (bioorthogonal) chemistry. *ACS Chem. Biol.* **2014,** *9*, 592-605.
25. Blackman, M. L.; Royzen, M.; Fox, J. M., Tetrazine ligation: fast bioconjugation based on inverse-electron-demand Diels-Alder reactivity. *J. Am. Chem. Soc.* **2008,** *130*, 13518-13519.
26. Taylor, M. T.; Blackman, M. L.; Dmitrenko, O.; Fox, J. M., Design and synthesis of highly reactive dienophiles for the tetrazine-trans-cyclooctene ligation. *J. Am. Chem. Soc.* **2011,** *133*, 9646-9649.
27. Yang, M. T.; Chang, C. H.; Wang, J. M.; Wu, T. K.; Wang, Y. K.; Chang, C. Y.; Li, T. T., Crystal structure and inhibition studies of transglutaminase from *Streptomyces mobaraense*. *J. Biol. Chem.* **2011,** *286*, 7301-7307.
28. Ohtsuka, T.; Sawa, A.; Kawabata, R.; Nio, N.; Motoki, M., Substrate specificities of microbial transglutaminase for primary amines. *J. Agric. Food. Chem.* **2000,** *48*, 6230-6233.
29. Mero, A.; Spolaore, B.; Veronese, F. M.; Fontana, A., Transglutaminase-mediated PEGylation of proteins: direct identification of the sites of protein modification by mass spectrometry using a novel monodisperse PEG. *Bioconjug. Chem.* **2009,** *20*, 384-389.
30. Spolaore, B.; Raboni, S.; Ramos Molina, A.; Satwekar, A.; Damiano, N.; Fontana, A., Local unfolding is required for the site-specific protein modification by transglutaminase. *Biochemistry* **2012,** *51*, 8679-8689.

31. Rajagopalan, P. T.; Zhang, Z.; McCourt, L.; Dwyer, M.; Benkovic, S. J.; Hammes, G. G., Interaction of dihydrofolate reductase with methotrexate: ensemble and single-molecule kinetics. *Proc. Natl. Acad. Sci. U.S.A.* **2002,** *99*, 13481-13486.
32. Antikainen, N. M.; Smiley, R. D.; Benkovic, S. J.; Hammes, G. G., Conformation coupled enzyme catalysis: single-molecule and transient kinetics investigation of dihydrofolate reductase. *Biochemistry* **2005,** *44*, 16835-16843.
33. Sugimura, Y.; Yokoyama, K.; Nio, N.; Maki, M.; Hitomi, K., Identification of preferred substrate sequences of microbial transglutaminase from *Streptomyces mobaraensis* using a phage-displayed peptide library. *Arch. Biochem. Biophys.* **2008,** *477*, 379-383.
34. Tagami, U.; Shimba, N.; Nakamura, M.; Yokoyama, K.; Suzuki, E.; Hirokawa, T., Substrate specificity of microbial transglutaminase as revealed by three-dimensional docking simulation and mutagenesis. *Protein Eng. Des. Sel.* **2009,** *22*, 747-752.
35. Batista, A. N.; Batista, J. M., Jr.; Bolzani, V. S.; Furlan, M.; Blanch, E. W., Selective DMSO-induced conformational changes in proteins from Raman optical activity. *Phys. Chem. Chem. Phys.* **2013,** *15*, 20147-20152.
36. Nieuwenhuizen, W. F.; Dekker, H. L.; de Koning, L. J.; Groneveld, T.; de Koster, C. G.; de Jong, G. A., Modification of glutamine and lysine residues in holo and apo alpha-lactalbumin with microbial transglutaminase. *J. Agric. Food Chem.* **2003,** *51*, 7132-7139.

Chapter 6 - Microbial transglutaminase purification strategies for structural studies

6.1 Preface

As discussed in Chapter 2, the selectivity MTG displays for its glutamine substrate is only superficially understood. Indeed, MTG discriminates among its protein substrates, as we observed first-hand with a difference in reactivity between α-LA and hDHFR, described in Chapter 5. However, it is not understood how MTG discriminates among different protein substrates; this topic will be specifically addressed in Chapter 7. In addition to the observation of reactivity toward different protein-bound glutamines, apoenzyme crystal structures and point mutations are the primary source of information to better understand this topic. However, neither provides sufficient data to concretely form a hypothesis as to the sequence or structural context that favors the reactivity of a glutamine more than another. To this effect, we sought to gain more information on glutamine-substrate binding by a combination of computational docking, and ligand-bound crystallization.

Two crystal structures of MTG exist: in the first, MTG was crystallized in its active, catalytic form, following cleavage of its protective pro-sequence[1] by two endogenous proteases.[2] The second, published nearly a decade after the first, is the zymogen, revealing how the pro-sequence folds over the active site, preventing potential substrates from binding.[3] These high-resolution structures revealed key information such as the location and nature of MTG's active site and catalytic mechanism.[1] In addition, another smaller TGase from *Bacillus subtilis* was crystallized more recently. Although bearing no amino acid sequence similarity to MTG, this TGase has structural similarities, including the same catalytic dyad mechanism proposed for MTG.[4] However, many unanswered questions remain regarding the manner in which MTG interacts with its substrates.

Hypothesizing the binding mode of MTG is not an obvious task, as it displays a broad, shallow tertiary arrangement flanking the active site; the crevice housing the active-site cysteine itself is fairly narrow, allowing reactivity with the amide of glutamine but not of the shorter

asparagine. The protected dipeptide substrate ZQG was small enough to be docked into the active site,[5] but the question as to how MTG could accommodate a large, bulky protein substrate remains unanswered. In these docking results, ZQG is observed to thread itself through the active site crevice. We performed our own docking experiments using ZQG as well as a

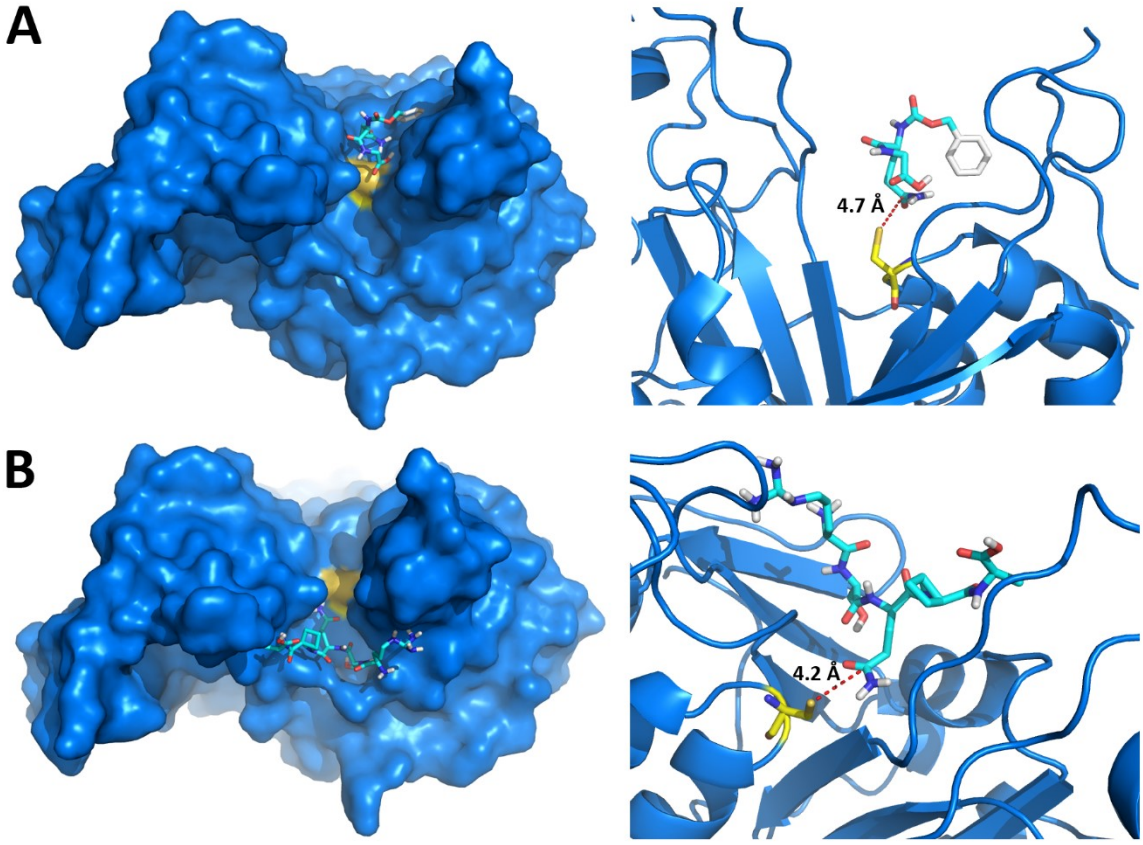


Figure 6-1 Docking of peptide substrates with MTG.

The catalytic cysteine residue, Cys64, is colored in yellow. Both peptide substrates, ZQG (A) and RTQPA (B) are shown. PDB ID: 1IU4.

pentapeptide reported to be a substrate[6] to see if similar results would be obtained. Using AutoDock Vina,[7] the top 10 lowest-energy poses for each substrate were evaluated; the validity of the poses was judged on the distance separating the δ-carbon of glutamine from the γ-sulfur atom of Cys64, the catalytic nucleophile required for activity. If this carbon atom is too far from the sulfur, nucleophilic attack cannot occur, as proposed in the mechanism. Among these poses, only one pose for each substrate was observed to be near 4 Å of Cys64. These docked structures did not thread themselves through the active site; instead, they rested along the outer surface of

the active site, with the glutamine side-chain inserting itself within proximity of Cys64 (Figure 6-1). This offered a potential explanation as to how large proteins can be substrates for MTG. Docking results must be met with scrutiny, however; they are helpful for providing insights and concepts into substrate binding, but without complementary experimental data, they are not sufficiently reliable to stand on their own If a structural rearrangement occurs to better accommodate substrates, such an event will not be captured by docking.

In addition to seeking to better understand the fundamentals of substrate binding, knowing the binding mode of MTG's glutamine substrate would help us address our goal of further developing MTG as a tool for site-specific peptide and protein labeling. We were keen on engineering MTG to increase its effectiveness towards a selected glutamine-containing sequence. Indeed, crystal structures are often of immeasurable aid in guiding the researcher to decide which residues to mutagenize. Lacking a high-throughput assay to screen thousands of MTG variants, it became clear that we would have to apply a semi-rational approach to selecting residues to mutate. Owning to MTG's broad surface and active site crevice, along with conflicting docking results, it was difficult to hypothesize which residues would be best to focus on, and this ultimately thwarted our engineering attempts for years.

While other studies reported in this thesis were being performed, we concurrently pursued efforts to attempt MTG crystallography trials in the presence of a binding partner. The most fundamental item required is a means to obtain an ultra-pure, active preparation of MTG; it is essential that it be pure in order to crystallize, and active in order that it be fit to bind substrate. The reproduction of the methodology that produced ultra-pure MTG by other groups was attempted previously in our laboratory, including denaturing/refolding, but was not successful. The purification strategy we use routinely for MTG produces enzyme that is suitable for biocatalytic studies, but contaminants remain that may impede crystallography trials.

To this end, we sought to develop an alternate purification strategy. Ultimately, the fruits of our labor was in vain, and ultra-pure MTG was not produced. This chapter will present our approach, the limitations we sought to mitigate, examine and discuss our results, as well as the factors contributing to the eventual undoing of our efforts.

6.2 General strategy and considerations

The major complication surrounding the successful expression and purification of active MTG is that it must be handled as a zymogen. It is well documented that expression of

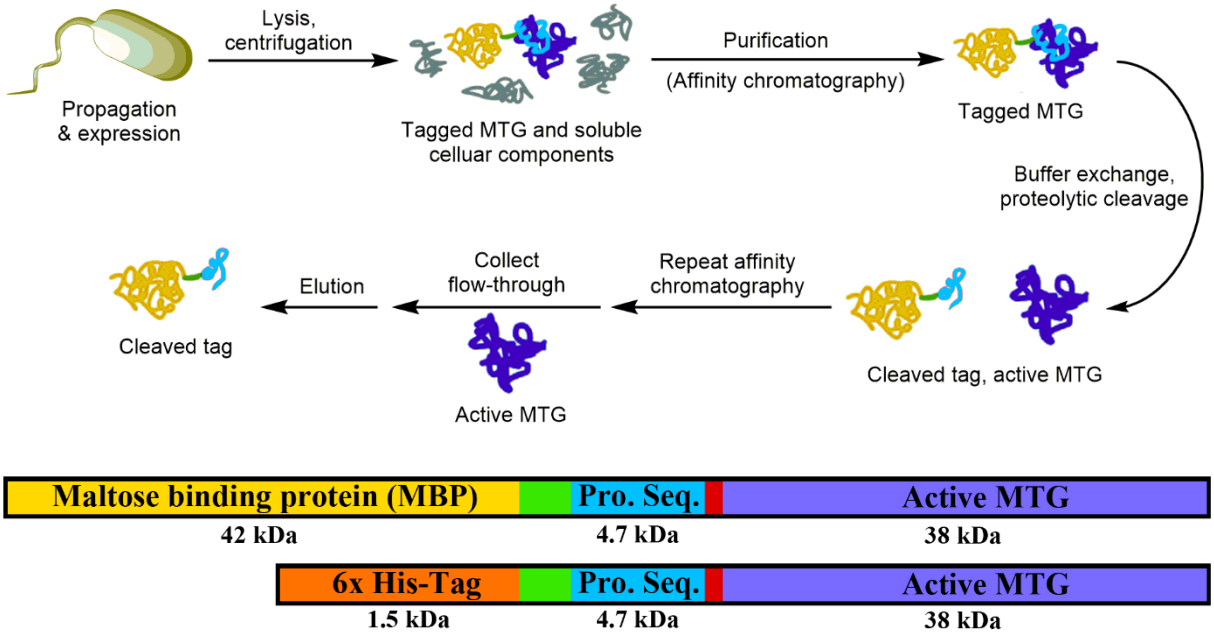


Figure 6-2 : Purification strategy and sequences of tagged MTGs.

Yellow and orange correspond to the affinity tags used to purify MTG, MBP and 6×His-Tag, respectively. Green corresponds to the linker between the affinity tag and the target protein; blue corresponds to the pro-sequence that keeps MTG in its inactive form; red corresponds to the protease cleavage site to remove the pro-sequence; purple corresponds to active MTG.

the active form of MTG results in inclusion bodies, or death of the cells; in simple terms, it likely crosslinks cellular components.[8-9] We discussed several work-around methodologies that have been developed in recent years in Chapter 2. Indeed, when we attempted the direct expression of active MTG, cell density of the *E. coli* cultures dropped catastrophically, and yielded no viable protein (data not shown). Our routine purification strategy involves expression of the zymogen with auto-inducing media (IPTG induction was unsuccessful), treating the clarified cellular lysate with trypsin to cleave the pro-sequence, and purification using affinity chromatography. Specifically, our expression vector containing the ORF for MTG also includes a *C*-terminal 6-histidine tag, which binds to Ni-NTA resin. Peptide tags such as the 6-histidine

tag can be problematic in crystallography trials;[10] indeed, the two crystal structures of MTG do not contain a poly-histidine tag. We noted that if an affinity tag were expressed at the *N*-terminus instead, the active form of MTG would ultimately be absent of the tag, as the *N*-terminus gets cleaved during proteolytic activation. We devised a purification strategy utilizing this concept (Figure 6-2).

Two purification steps would be required: first, the zymogen is purified from all other cellular components with affinity chromatography following overexpression and lysis.

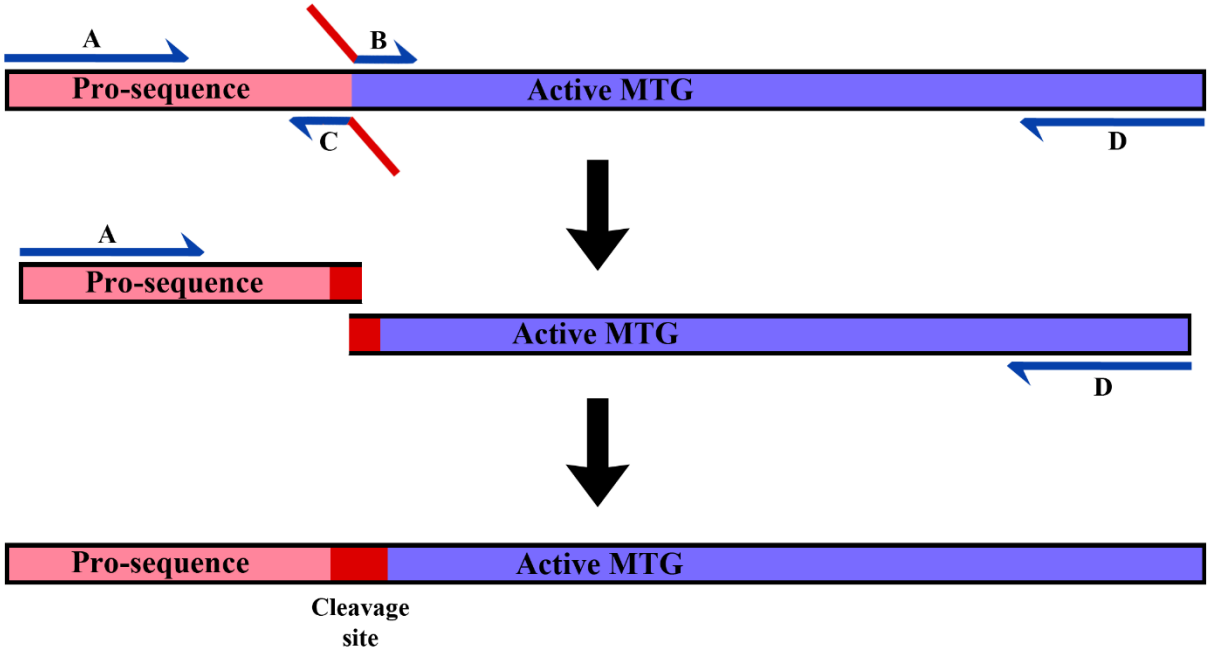


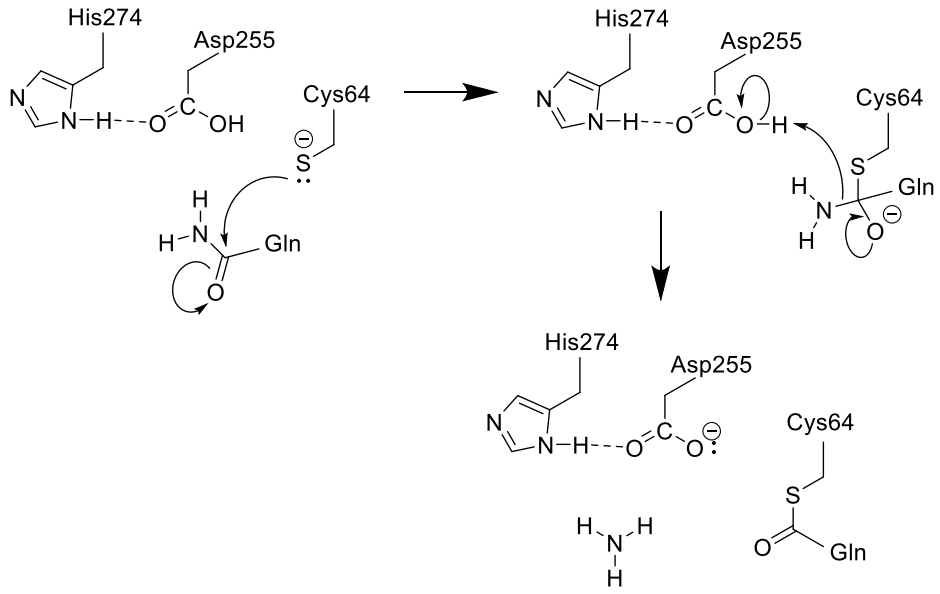
Figure 6-3 Site overlap extension PCR for the creation of cleavable MTG zymogens. *Two internal mutagenic primers, B and C, will bind to the region of the double-stranded DNA to be mutated. During the first round of PCR, fragments AC and BD (here specifically, the pro-sequence and active MTG, respectively) are synthesized with the addition of flanking primers A and D. The fragments are mixed for the second round of PCR using primers A and D exclusively to create the final product, which now contains the mutation.*

Proteolytic cleavage of the pro-sequence follows, upon which activated MTG must be separated from the cleaved *N*-terminus exhibiting the affinity tag. For this second purification step, we hypothesized that by repeating the affinity chromatography separation, active MTG

could be collected in the flow-through as it should not bind to the column. Additionally, any impurities remaining after initial separation via affinity chromatography should re-bind along with the cleaved *N*-terminus, theoretically yielding ultrapure MTG.

To this end, we constructed two MTG clones, each one encoding a distinct *N*-terminal affinity tag and specific proteolytic cleavage sequence (Figure 6-2). A poly-histidine was one tag selected, which is used routinely for protein purification; histidine interacts with nickel ions, and so resins functionalized with nickel ions (Ni-NTA) will bind histidine-rich proteins. Poly-hisitidine tags are typically the go-to methodology for protein purification owing to their ease of manipulation and small, non-disruptive size. It's not uncommon for non-specific binding to the Ni-NTA matrix to occur, yielding contaminants during elution, as histidine is naturally present in proteins. Considering this, we selected maltose binding protein (MBP) as an alternative tag. MBP interacts with amylose-functionalized resin, and will elute in the presence of maltose. We expected to obtain tagged-MTG of higher purity owing to higher specificity of this resin than Ni-NTA; theoretically, only MBP should interact with amylose resin.

After purification of tagged-MTG, pooled elution fractions were subjected to buffer exchange so that proteolytic cleavage would proceed effectively. Proteases are sensitive to buffer composition, and may not function well in the presence of components of the elution buffers. For selective proteolysis of the protective *N*-terminal pro-sequence, the cleavage sequence needs to be introduced in between this sequence and active enzyme. Additionally, MTG must not natively contain any cut sites recognized and accessible by the protease. To achieve the first requirement, we introduced cleavage sites within the native sequence of MTG by site-overlap PCR (Figure 6-3). As with our selection of affinity tags, we chose two proteases that do not cut within MTG: enterokinase and thrombin. Enterokinase was preferred, as it recognizes its cleavage site in such a way that will leave no additional residues on MTG, where thrombin will leave Gly-Ser at the newly formed *N*-terminus. In the event that enterokinase does not cut effectively, a second MTG clone was generated containing a cleavage site for thrombin; this is the clone exhibiting the poly-histidine tag. Indeed, between



Scheme 6-1 Proposed mechanism for the formation of the covalent thioester intermediate within MTG.

Adapted from Kashiwagi et al.[1]

the two clones, MBP-enterokinase-MTG was constructed as an ideal candidate, hypothetically exhibiting a superior tag and cleavage site, with 6-His-thrombin-MTG as a back-up plan.

Concurrently, a suitable binding partner for MTG must be determined for crystallization. The model dipeptide substrate, ZQG, displays micromolar affinity towards MTG, with its K_M being 53 mM.[3] For both co-crystallization and soaking approaches to obtaining crystals of ligand-bound enzyme complex, an excess of ligand is required relative to enzyme concentration to ensure that all active site cavities are occupied. However, ZQG's aqueous solubility is limited to 70-80 mM, making such approaches unfeasible with ZQG. On the other hand, MTG's proposed catalytic mechanism requires the formation of a covalent enzyme-ZQG intermediate after the nucleophilic attack of the substrate glutamine residue by the thiolate of Cys64, and proton transfer to the oxyanion intermediate occurs (Scheme 6-1). If this covalent enzyme-ZQG intermediate is sufficiently stable, it could serve as the species of MTG used for crystallography trials. In the event that it is not, an irreversible peptide inhibitor bound at MTG's active site could serve as a reasonable alternative. While an irreversible inhibitor may exhibit different interactions with the active site than a substrate due to the

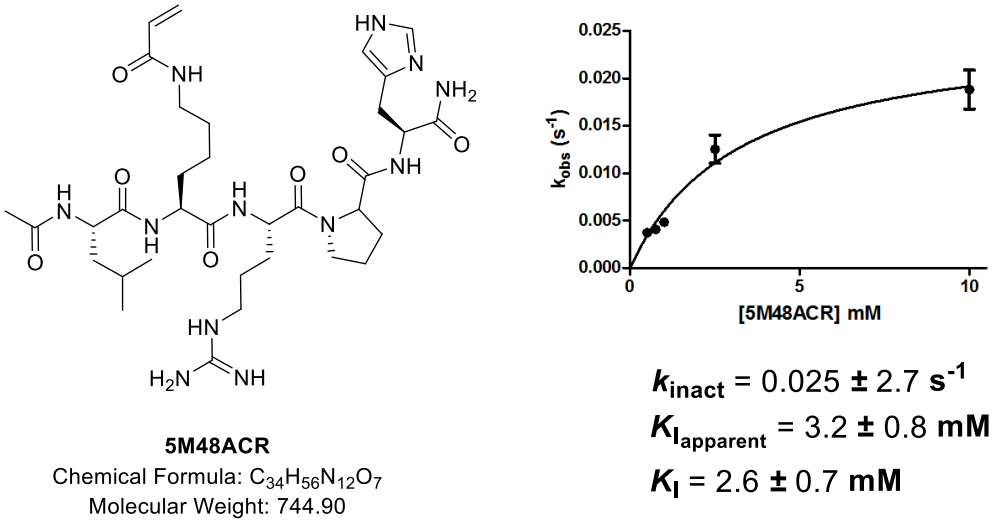


Figure 6-6-4 Structure of 5M48ACR, a covalent inhibitor for MTG.

Inhibition was determined using a GDH-coupled assay.[11] Time-dependent inactivation was fit by non-linear regression to a mono-exponential model, providing first-order rate constants of inactivation, ultimately providing the kinetic parameters k_{inact} and K_I. K_I corresponds to the concentration of inhibitor required so that 50% of the enzyme population is inhibited; $K_{Iapparent}$ is the measurement of K_I under conditions that could compromise the determination of its true value.

forced covalent interaction, it would nevertheless provide insight into the MTG's binding mode, especially if it resembles a natural substrate. Such an inhibitor was developed and synthesized in the laboratory of our collaborator, Prof. Jeffrey Keillor (University of Ottawa, Figure 6-4). The modified pentapeptide was observed to display micromolar affinity for MTG, with an apparent K_I of 3.2 mM being reported by our collaborators. Combined with our expression and purification strategy, we proceeded to execute the experimental steps so that we could ultimately obtain ligand-bound crystals of MTG.

6.3 6.3. Purification and digestion of tagged and cleavable MTG

6.3.1 Stability of MTG-ZQG complex

Three separate samples were prepared: a buffered solution containing MTG only, and two samples of MTG and excess ZQG. One of these two samples was subjected to buffer

exchange after incubation for 2 hours at 37°C, to remove excess ZQG from the solution, whereas this excess was maintained in the other. Once the exchange was complete, the samples were analyzed by LC-MS (Figure 6-5). If the MTG-ZQG complex were stable, this would be reflected in the mass detected after buffer exchange. It was quickly determined that such a complex was not stable, as only the mass corresponding to the native MTG was detected after buffer exchange; indeed, excess ZQG needed to remain present for the complex to exist. In light of these results, we made arrangements with our collaborator to focus efforts towards the usage of an irreversible inhibitor rather than a substrate.

6.3.2 Purification and digestion tests

After construction of an expression vector containing the open reading frame for MTG *N*-terminally tagged with MBP (MBP-MTG), we proceeded to expression and purification tests. MBP-MTG proved to express well (Figure 6-6, Panel A); the purification was scaled up and elution fractions were collected and pooled. To our surprise, a significant amount of non-specific binding was observed, resulting in contaminants in the elution fraction. We decided to proceed buffer exchange and digestion, as we hypothesized that these contaminants would again bind to the amylose resin during the second purification step. An aliquot of the purified, buffer exchanged MBP-MTG was subjected to a cleavage test with enterokinase. Cleavage

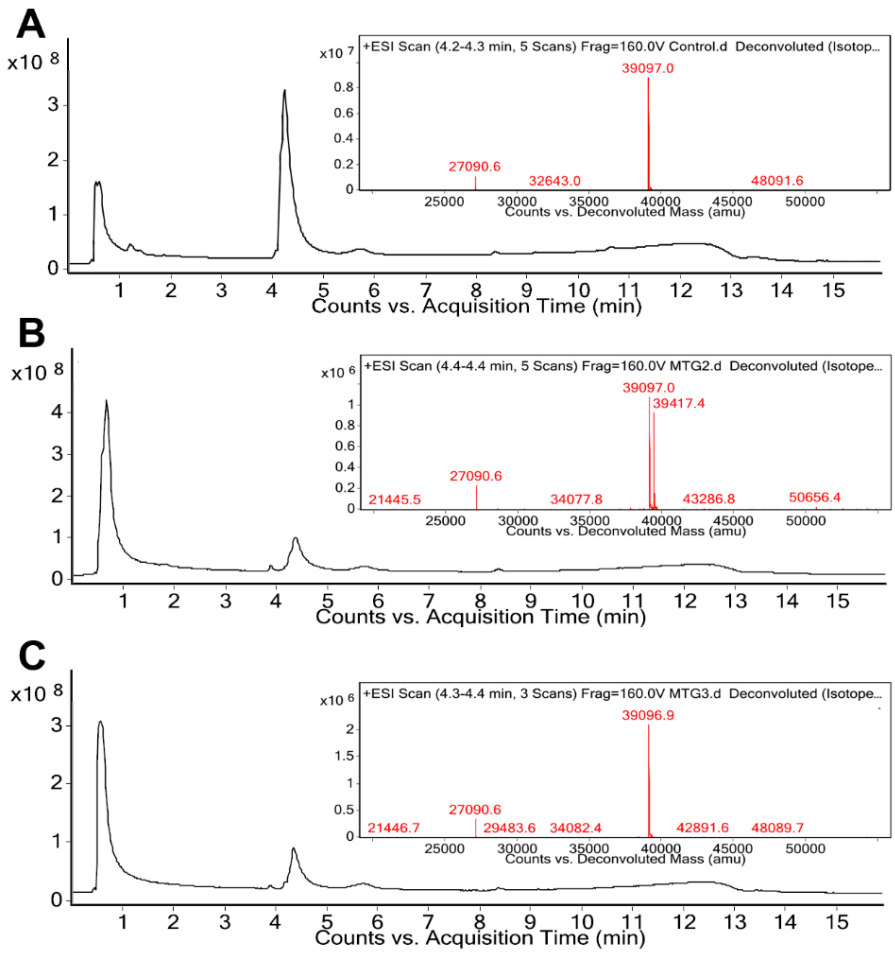


Figure 6-5 LC-MS analysis of the ZQG-MTG complex.

Native MTG (A) with no other treatment was analyzed as a control (39097 amu). While the exact mass of the ZQG-MTG complex was observed (B, 39417.4 amu), native MTG remained. In addition, after ZQG was no longer present in excess, only native MTG was observed (C).

performed overnight, with aliquots taken at various time points (Figure 6-6, Panel B). If the reaction had gone to completion, two bands were expected to be observed on SDS-PAGE: MBP and the pro-sequence, and active MTG at approximately 47 kDa and 38 kDa, respectively. Inexplicably, additional intermediate bands were observed, particularly after 2 hours of digestion, but disappeared and stabilized after further incubation. Cleavage was complete after 14 hours, with the fragments being stable for a minimum of 6 additional hours. Cleaved MBP-MTG was re-loaded onto amylose resin pre-conditioned with buffer. Gratifyingly, over numerous fractions, active MTG was recovered in the flow-through (Figure 6-6, Panel C),

although activity was approximately 10-fold less than a usual preparation, indicating the need for optimization. Even so, we had confirmation that the enterokinase cleavage site is accessible by the protease.

An unexpected complication arose when our supplier of enterokinase ceased production, forcing us to switch manufacturers. Repeating the same cleavage and

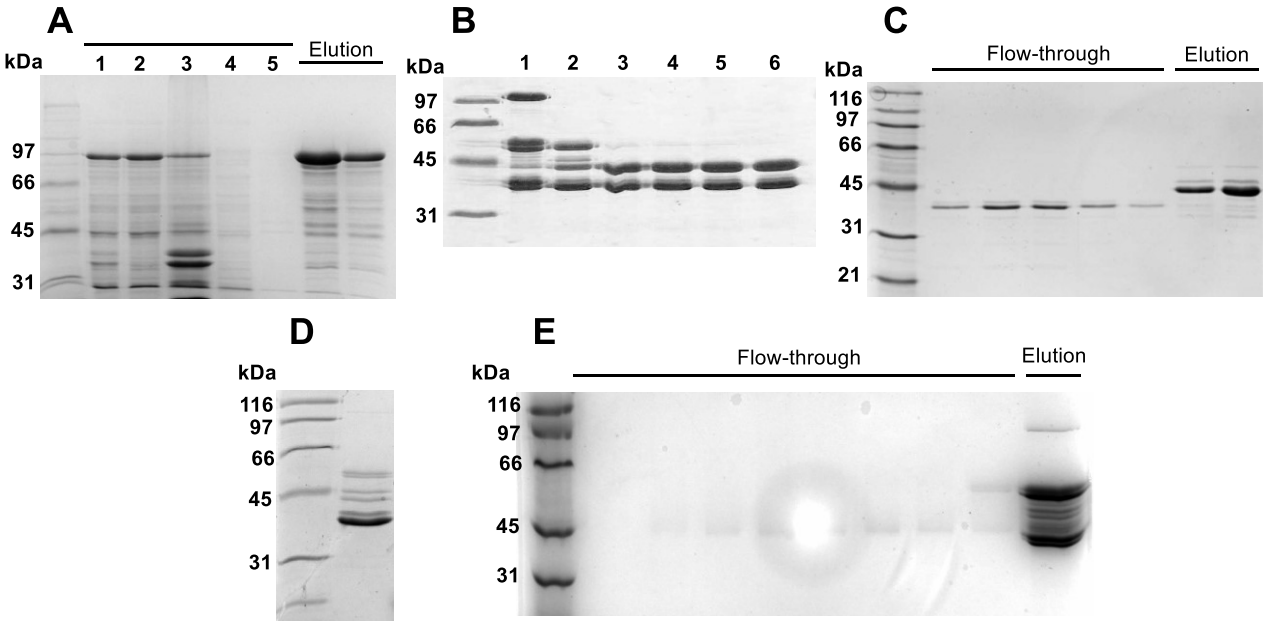


Figure 6-6 SDS-PAGE analysis of MBP-MTG purification, digestions, and re-purification of active MTG.

Expected masses: MBP-MTG = 84.7 kDa; cleaved MBP = 44.2 kDa; cleaved MTG = 38 kDa,

- *A) Purification of MBP-MTG. Lane 1: lysate; lane 2: soluble fraction; lane 3: insoluble fraction; lane 4: flow-through; lane 5: wash.*
- *B) Cleavage test of MBP-MTG with enterokinase. Lane 1: t = 30 min; lane 2: t = 2 h; lane 3: t = 14 h; lane 4: t = 16 h; lane 5: t = 18 h; lane 6: t = 20 h.*
- *C) Re-purification of active MTG on amylose resin after cleavage.*
- *D) Pooled re-purification flow-through fractions after cleavage with enterokinase after switching supplier to Diamed.*

E) Flow-through and pooled elution fractions after cleavage with enterokinase purchased from ProSpec.

re-purification protocols with two different suppliers consistently yielded unsatisfactory results (Figure 6-6, Panels D and E) during both cleavage and re-purification steps. During the former, additional, unidentified bands were observed; even more troubling was that after re-purification, active MTG would co-elute with the cleaved MBP tag. In light of these impediments, we explored the feasibility of our back-up expression clone, 6-His-Thr-MTG.

Expression of 6-His-Thr-MTG proceeded without difficulty, and we moved onto the initial purification step utilizing routine Ni-NTA resin. A surprising but welcome observation was that purification with Ni-NTA yielded protein of higher purity than the amylose-MBP

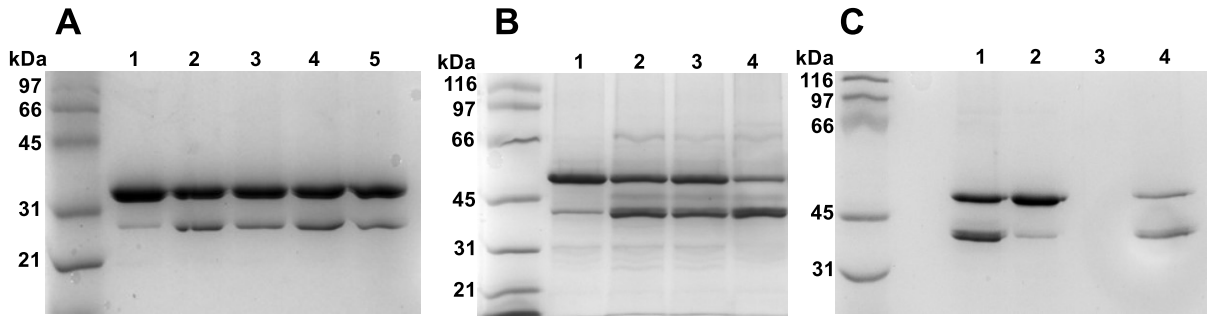


Figure 6-7 SDS-PAGE analysis of 6-His-thrombin-MTG digestions and re-purification of active MTG.

A) Thrombin digestion of 6-His-thrombin-MTG. Lane 1: 6-His-thrombin-MTG control (No thrombin); Lane 2: 6-His-thrombin-MTG, 30°C; Lane 3: 6-His-thrombin-MTG, 37°C; Lane 4: 6-His-thrombin-MTG, 30°C, Triton X-100; Lane 5: 6-His-thrombin-MTG, 37°C, Triton X-100.

B) Further digestion conditions. Lane 1: 6-His-thrombin-MTG after buffer exchange, no thrombin; Lane 2: 6-His-thrombin-MTG + 1U/50 ug of thrombin, 25°C; Lane 3: 6-His-thrombin-MTG + 1U/100 ug of thrombin, 25°C; Lane 4: 6-His-thrombin-MTG + 1U/100 ug of thrombin, 37°C.

C) Digestion and re-purification. Lane 1: 6-His-thrombin-MTG after buffer exchange and digestion, 1U/100 µg thrombin, 37°C; Lane 2: 6-His-thrombin-MTG after buffer exchange and digestion, 1U/200 µg thrombin, 37°C; Lane 3: Flow-through of 6-His-

thrombin-MTG post-digestion; Lane 4: Elution of 6-His-thrombin-MTG post-digestion.

system used previously. Continuing to cleavage with thrombin, results were less fruitful, (Figure 6-7, Panel A) with cleavage proceeding ineffectively, suggesting that thrombin may not exhibit the same accessibility to the cut site than was observed with enterokinase. Addition of a mild detergent can help circumvent by promoting local denaturation, although this proved not to be the case. Fortunately, by varying the temperature and ratio of thrombin to 6-His-thrombin-MTG, we were able to improve the cleavage efficiency, resulting in a 3-fold improvement in digestion efficiency (Figure 6-7, Panel B). However, when separation of active MTG was attempted, we once again observed co-elution of active MTG with the tagged, uncleaved enzyme (Figure 6-7, Panel C). Cleavage and purifications were re-attempted on many occasions, but as these results persisted, coupled with the inherent high-risk nature of crystallography studies, we decided to devote our efforts to other aspects of the project.

6.3.3 Discussion and conclusions

Initially, cleavage and purification with enterokinase was promising. It demonstrated that our initial plan could work, despite that activity was low. However, this likely would have been fairly simple to address, as the initial purifications did not have DTT present in the buffers. Reducing agents such as DTT or glutathione are typically included in MTG purifications as its catalytic residue is a cysteine. Most importantly, it was highly pure. After switching enterokinase suppliers, we were perplexed by how significantly the digestion efficiencies varied. Not only this, but the chromatography lost its reliability, with cleaved, active MTG co-eluting with species exhibiting a tag rather than coming out in the flow-through. If it weren't for this complication, even if tagged MTG had not reached 100 % cleavage, the active MTG could still be recovered from uncut, tagged MTG. As such was not the case, it is difficult to explain these results.

Regarding the co-elution of active MTG with its tagged counterpart, one explanation we formulated was that after cleavage, active MTG would continue to form non-covalent interactions with its cleaved pro-sequence, allowing for it to bind to the resin indirectly. This does not explain how active MTG was successfully recovered initially, as the same clone was

used for these expressions and purifications. In hopes of disrupting any potential interactions, we also attempted purifications under denaturing conditions by including up to 3 M of urea in the purification buffers, although the co-elution of both protein species continued to be observed (data not shown).

Inconsistent and poor digestion efficiencies, as well as potential interactions with the pro-sequence resulting in co-elution lead us to conclude that obtaining ultrapure active MTG may benefit from alternate expression methodologies to obtain folded, active MTG. As mentioned, recent approaches were covered in Chapter 2. However, since that review was published, more work has been done investigating soluble expression of active MTG. In particular, a study published last year described a genetically modified version of MTG that can be expressed in high yields within the cytoplasm of *E. coli*.[12] Mutations within the pro-sequence revealed variants that maintain its chaperone function but destabilize the cleaved pro-sequence/MTG interaction in a temperature dependent fashion. A protease is still required (3C protease was used in the study) to cut the sequence. In addition, such purifications would most likely require a *C*-terminally poly-histidine tagged version so that MTG may be separated from other soluble cellular components, which isn't always favorable for crystallography. It's possible that MTG is extremely sensitive to experimental conditions, such as temperature, buffer composition, and protease selection, making them critical factors that will determine the effectiveness of the purification. Therefore, once a working methodology is found within the laboratory, it should become standard and never be modified.

6.4 Materials and methods

6.4.1 Materials

The plasmid pDJ1-3 was kindly provided by Professor M. Pietzsch (Martin-Luther-Universität, Halle-Wittenberg, Germany). pDJ1-3 encodes the proenzyme of MTG from *S. mobaraensis* inserted between the *Nde*I and XhoI restriction sites of the vector pET20b.(REF) The sequence and served as a template for amplifying the MTG coding sequence containing either entrokinase or thrombin cut sites. The plasmid pMAL-c2X, which is the *N*-terminal MBP expression vector, was kindly provided by Prof. Stephen Michnick (Université de Montréal).

Deionized water (18Ω) was used for all experiments. Products used for the expression and purification of MTG were of biological grade.

Other chemicals used were purchased from the suppliers listed below. Carboxybenzyl–L-glutaminyl–glycine (Z-Gln-Gly, or ZQG) was from Peptide Institute (Osaka, Japan). Glutathione (reduced) and thiamine were from Bioshop (Burlington, Canada). FastDigest *Nco*I, *Bam*HI, *Hind*III, Phusion® High-Fidelity Polymerase and Fast AP Thermosensitive Alkaline Phosphatase were purchased from ThermoFisher Scientific (Waltham, MA, USA). Takara T4 DNA Ligase was purchased from Clontech (Mountain View, CA, USA). FastBreak™ Cell Lysis Reagent was purchased from Promega (Madison, WI, USA). Oligonucleotides were purchased from Sigma Aldrich (St. Louis, USA). Enterokinase was purchased from Feldan (Québec, QC), Diamed (Mississauga, Ontario), and ProSpec (East Brunswick, NJ, USA). Thrombin from bovine plasma was purchased from Sigma Aldrich (St. Louis, USA).

6.4.2 MBP-MTG cloning

The pDJ1-3 plasmid encoding the open reading frame for MTG was used as a template for mutagenesis. An enterokinase cut site was introduced into MTG using site overlap extension PCR,[13] with the primers listed in Table 6-1. Following amplification with Phusion® High Fidelity polymerase, the PCR product was treated with were digested with FastDigest *Bam*HI and *Hind*III restriction enzymes, and religated into pMAL-c2X, yielding the plasmid pMAL-MBP-MTG. which had been cut with the same enzymes and also dephosphoylated, and transformed in *E. coli* DH5α. Ampicillin (Amp) was used at 100 µg/mL for plasmid maintenance. Sequences were confirmed by DNA sequencing (ABI 3730 DNA sequencer, IRIC Genomic Platform at Université de Montréal).

6.4.3 MBP-MTG expression, purification, digestion, and active MTG re-purification

A 2-mL starter culture of *E. coli* DH5α containing the plasmid pMAL-MBP-MTG was propagated overnight at 37°C in LB medium and shaking at 240 rpm. Ampicillin (Amp) was used at 100 µg/mL for plasmid maintenance in all cultures using LB medium. This starter was used to inoculate 200 mL of LB medium at a 1:200 dilution. After 3h of incubation at 37°C

and240 rpm, when an OD_{600} was reached, IPTG was added to a final concentration of 0.2 mM. The temperature was reduced to 20°C and incubated for 18 hours at 240 rpm. Cells were

Table 6-1 Primers for the construction of pMAL-MBP-MTG.

Primer Identity	Direction	Oligonucleotide sequence (5' → 3')
N-terminal MTG	Forward	AAAGGATCCATGGACAATGGCGCGGG
C-terminal MTG	Reverse	TTTCCCAAGCTTTTACGGCCAGCCCTGCTTTAC
Ent cut site	Forward	GTCCTTGTCATCGTCATCGGGGCCCGGAACGAC
	Reverse	CCCGATGACGATGACAAGGACTCCGACGACAGGGTCAC

collected by centrifugation and resuspended in 20 mM Tris base, 150 mM NaCl, 1 mM EDTA, pH 7.5. The cells were lysed using a Constant Systems cell disruptor set at 37 kPSI and cooled to 4°C. After further centrifugation to remove insoluble cellular matter, the supernatant was filtered with 0.2 μM PES syringe filters (Corning) and purified using a 10-mL amylose column (New England Biolabs) equilibrated in 20 mM Tris base, 150 mM NaCl, 1 mM EDTA, pH 7.5. Eluted was carried out with buffer containing 10 mM maltose, using an Åtka FPLC (GE Healthcare). After purification, MBP-MTG was dialyzed against the digestion buffer, 50 mM Tris-HCl, 1 mM $CaCl_2$, pH 8.0. The average yield was ~6 mg of MBP-MTG per litre of culture, with ~ 85% purity as estimated by SDS-PAGE and revelation with Coomassie blue stain. Aliquots were snap frozen and stored at -80°C in 15% glycerol.

Digestion was performed as suggested by the manufacturer: Feldan, Diamed, or ProSpec. For example, with the enterokinase supplied by Feldan, freshly dialyzed MBP-MTG at a concentration of 1 mg/mL with 0.5 U enterokinase added for every 0.2 mg of MBP-MTG: One unit of enterokinase is the amount of enzyme required to digest 0.5 mg of thioredoxin-NP-27 fusion protein to 90% completion in 16 hours at 37°C. Digestions were carried out at 23°C and 37°C, with both temperatures yielding equivalent results, and incubated for up to 22h with aliquots being taken for analysis by SDS-PAGE throughout the digestion.

Attempts to re-purify cleaved MTG were undertaken in the same manner as the purification of MBP, with the flow-through collected as well as elution fractions.

6.4.4 6-His-thrombin-MTG cloning and expression

The pDJ1-3 plasmid encoding the open reading frame for MTG was used as a template for mutagenesis. A thrombin cut site was introduced into MTG using site overlap extension PCR,[13] with the primers listed in Table 6-2. Complementary oligonucleotides encoding the poly-histidine tag to be introduced at the *N*-terminus were combined to a concentration of 500 ng/µL, heated to 90°C and left to cool back down to room temperature so that annealing could occur. The fragment was then digested with FastDigest *Nco*I and *Nde*I. Following amplification with Phusion® High Fidelity polymerase, the PCR product was treated with were digested with FastDigest *Nde*I and *Bam*HI restriction enzymes, and all fragments were combined and religated into pET15b, which had been cut with *Nco*I and *Bam*HI and also dephosphoylated, yielding the plasmid pET15-6-His-thrombin-MTG. It was transformed into *E. coli* BL21 (DE3), with ampicillin (Amp) used at 100 µg/mL for plasmid maintenance. Sequences were confirmed by DNA sequencing (ABI 3730 DNA sequencer, IRIC Genomic Platform at Université de Montréal).

Table 6-2 Primers for the construction of pET15-6-His-thrombin-MTG**.**

Primer Identity	Direction	Oligonucleotide sequence (5' → 3')
N-terminal MTG	Forward	AAAACATATGGACAATGGCGCGGGGG
C-terminal MTG	Reverse	AAAAGGATCCTTACGGCCAGCCCTGCTTTACC
N-terminal His-tag	Forward	AAAAAAAACCATGGGGCAGCAGCCATCATCATCATCATCACAG CAGCGGCCATATGAAAAAAAA
N-terminal His-tag	Reverse	TTTTTTTTCATATGGCCGCTGCTGTGATGATGATGATGATGGC TGCTGCCCCATGGTTTTTTTT
Thrombin cut site	Forward	GTCGCTGCCGCGCGGCACCAGGGGGGCCCGGAACGAC
	Reverse	CCCCTGGTGCCGCGCGGCAGCGACTCCGACGACAGGGTCACC

6.4.5 6-His-thrombin-MTG purification, digestion, and active MTG re-purification

A 2-mL starter culture of *E. coli* BL21 (DE3) containing the plasmid pET15-6-His-thrombin-MTG, which expresses a *C*-terminally 6-His-tagged version of MTG, was propagated overnight at 37°C in ZYP-0.8G medium and shaking at 240 rpm. It was used to inoculate 200 mL of auto-inducing ZYP-5052 medium. After 2h of incubation at 37°C and 240 rpm, the temperature was reduced to 22°C overnight. Cells were collected by centrifugation and resuspended in 50 mM sodium phosphate buffer, 300 mM NaCl, pH 7.5. The cells were lysed using a Constant Systems cell disruptor set at 37 kPSI and cooled to 4°C. After further centrifugation to remove insoluble cellular matter, the supernatant was filtered with 0.2 μM PES syringe filters (Corning) amd 6-His-thrombin-MTG was purified using a 5-mL His-trap nickel-nitrilotriacetic acid (Ni-NTA) column (GE Healthcare) equilibrated in 50 mM phosphate buffer, pH 7.5, with 300 mM NaCl. It was eluted with an imidazole gradient (0 – 250 mM) using an Åtka FPLC (GE Healthcare). After purification, active MTG was dialyzed against the digestion buffer, PBS (140 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.3). The average yield was ~40 mg 6-His-thrombin-MTG per litre of culture, with ~ 90% purity as estimated by SDS-PAGE and revelation with Coomassie blue stain. Aliquots were snap frozen and stored at -80°C in 15% glycerol.

Digestion was performed as suggested by the manufacturer, Sigma. Freshly dialyzed 6-His-thrombin-MTG at a concentration of 1 mg/mL with 1 U thrombin added for every mg of 6-His-thrombin-MTG: In PBS, 1U of thrombin digests 100 μg of test protein to >90% completion, at 22°C after 16h. Digestions were incubated for up to 20h with aliquots being taken for analysis by SDS-PAGE throughout the digestion.

Attempts to re-purify cleaved MTG were undertaken in the same manner as the purification of His-thrombin-MTG, with the flow-through collected as well as elution fractions.

6.4.6 Molecular docking

ZQG and RTQPA pentapeptide were individually docked into the crystal structure of MTG (PDB ID: 1IU4) using Autodock Vina, version 1.1.2.[7] First, a PDBQT file was prepared

by uploading the crystal structure into AutoDock Tools (Version 4, Molecular Graphics Laboratory), removing all non-polar hydrogen atoms and saving the file as PDBQT. Next, the search space was defined in AutoDockTools. The search space restricts where the movable atoms, including those in the flexible side chains, should lie. The spacing of the grid points was set to 1.000 Å and 38 grid points in all three directions were used. Therefore, the search space was cubic with a volume of in which each site measured 38 Å. The PDB file of ZQG and RTQPA was then converted into a PDBQT file by loading the PDB file into AutoDockTools, defining all bonds as rotable bonds and saving the file as PDBQT file. Docking was performed using the default paramaters of AutoDock Vina with an exhaustiveness of 30 and the resulting structures were visualized using PyMOL.[14]

6.5 References

1. Kashiwagi, T.; Yokoyama, K.; Ishikawa, K.; Ono, K.; Ejima, D.; Matsui, H.; Suzuki, E., Crystal structure of microbial transglutaminase from Streptoverticillium mobaraense. *J. Biol. Chem.* **2002,** *277*, 44252-44260.
2. Zotzel, J.; Pasternack, R.; Pelzer, C.; Ziegert, D.; Mainusch, M.; Fuchsbauer, H. L., Activated transglutaminase from Streptomyces mobaraensis is processed by a tripeptidyl aminopeptidase in the final step. *Eur. J. Biochem.* **2003,** *270*, 4149-4155.
3. Yang, M. T.; Chang, C. H.; Wang, J. M.; Wu, T. K.; Wang, Y. K.; Chang, C. Y.; Li, T. T., Crystal structure and inhibition studies of transglutaminase from *Streptomyces mobaraense*. *J. Biol. Chem.* **2011,** *286*, 7301-7307.
4. Fernandes, C. G.; Placido, D.; Lousa, D.; Brito, J. A.; Isidro, A.; Soares, C. M.; Pohl, J.; Carrondo, M. A.; Archer, M.; Henriques, A. O., Structural and Functional Characterization of an Ancient Bacterial Transglutaminase Sheds Light on the Minimal Requirements for Protein Cross-Linking. *Biochemistry* **2015,** *54*, 5723-5734.
5. Tagami, U.; Shimba, N.; Nakamura, M.; Yokoyama, K.; Suzuki, E.; Hirokawa, T., Substrate specificity of microbial transglutaminase as revealed by three-dimensional docking simulation and mutagenesis. *Protein Eng. Des. Sel.* **2009,** *22*, 747-752.
6. Lee, J. H.; Song, C.; Kim, D. H.; Park, I. H.; Lee, S. G.; Lee, Y. S.; Kim, B. G., Glutamine (Q)-peptide screening for transglutaminase reaction using mRNA display. *Biotechnol. Bioeng.* **2013,** *110*, 353-362.
7. Trott, O.; Olson, A. J., AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2010,** *31* (2), 455-461.
8. Yokoyama, K. I.; Nakamura, N.; Seguro, K.; Kubota, K., Overproduction of microbial transglutaminase in *Escherichia coli*, in vitro refolding, and characterization of the refolded form. *Biosci. Biotechnol. Biochem.* **2000,** *64*, 1263-1270.
9. Marx, C. K.; Hertel, T. C.; Pietzsch, M., Soluble expression of a pro-transglutaminase from Streptomyces mobaraensis in Escherichia coli. *Enzyme Microb. Technol.* **2007,** *40*, 1543-1550.

10. Bucher, M. H.; Evdokimov, A. G.; Waugh, D. S., Differential effects of short affinity tags on the crystallization of *Pyrococcus furiosus* maltodextrin-binding protein. *Acta Crystallogr. D Biol. Crystallogr.* **2002,** *58*, 392-397.
11. Oteng-Pabi, S. K.; Keillor, J. W., Continuous enzyme-coupled assay for microbial transglutaminase activity. *Anal. Biochem.* **2013,** *441*, 169-173.
12. Rickert, M.; Strop, P.; Lui, V.; Melton-Witt, J.; Farias, S. E.; Foletti, D.; Shelton, D.; Pons, J.; Rajpal, A., Production of soluble and active microbial transglutaminase in *Escherichia coli* for site-specific antibody drug conjugation. *Protein Sci.* **2016,** *25*, 442-455.
13. Ho, S. N.; Hunt, H. D.; Horton, R. M.; Pullen, J. K.; Pease, L. R., Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* **1989,** *77*, 51-9.
14. Poteete, A. R.; Rennell, D.; Bouvier, S. E., Functional significance of conserved amino acid residues. *Proteins* **1992,** *13*, 38-40.

Chapter 7 - Engineered, highly reactive substrates of microbial transglutaminase enable protein labeling within various secondary structure elements

7.1 Context

A topic that has come up repeatedly throughout this thesis is the elusive substrate specificity of microbial transglutaminase, particularly for its glutamine-containing substrate, as discussed in Chapter 2 and the following sections. In the previous chapter, we set out to reveal some of the determinants that constitute MTG's interactions with a glutamine substrate by attempting to create MTG constructs for crystallographic studies, which would be conducted in the presence of a binding partner. Ultimately, those attempts were not realized, and we sought out complimentary experimental means gain more knowledge about the manner in which MTG interacts with its glutamine substrates.

As we tailored our biocatalytic goals towards protein conjugation, our engineering goals converged in the same direction. As discussed in Chapter 1, the properties of an enzyme can be altered or, ideally, improved by engineering. We sought to apply an engineering approach to MTG to improve upon its protein conjugation ability, and established two possibilities: one could engineer MTG itself, or engineer its glutamine-containing protein substrate. The most difficult part about engineering MTG is selecting the residues to mutate, given that little is known of its substrate-binding mode. Engineering its glutamine-containing protein substrate, if successful, may reveal key information about the determinants of MTG's glutamine preference, in addition to yielding a reactive tag. A reactive protein tag should be small to maximize its applicability and to facilitate its extensive mutation, even in the absence of a high-throughput screen. During the course of my studies, I identified a small protein substrate exhibiting limited reactivity with MTG. In the following chapter, we describe our investigations into characterizing a variant library of this protein substrate to probe the effect secondary structure plays in MTG reactivity. In parallel, a small set of MTG variants was generated and

characterized. We then verified whether the best MTG variants reacted with the most effective variant substrate, seeking if the mutations on both elements of the system were additive.

This chapter is a reproduction of a manuscript currently under review after being submitted to the journal *Protein Science*, entitled: *Engineered, highly reactive substrates of microbial transglutaminase enable protein labeling within various secondary structure elements.* My contribution was the conceptualization and realization of laboratory experiments, performed in the laboratory of Prof. Joelle Pelletier. Dr. Daniela Quaglia contributed heavily to the conceptualization. Dr. Éric Lévesque synthesized the amine fluorophore used in all experiments, the synthesis of which was carried out in the laboratory of Prof. André Charette. The manuscript was drafted by myself with assistance from Prof. Pelletier. Supporting information associated with this manuscript is available for consultation in Annex 4 of this thesis.

Engineered, highly reactive substrates of microbial transglutaminase enable protein labeling within various secondary structure elements

Natalie M. Rachel[1,2,3], Daniela Quaglia[1,2,3], Éric Lévesque[1,2], André B. Charette[1,2], and Joelle N. Pelletier[1,2,3,4]

1 Département de Chimie, Université de Montréal, 2900 Boulevard Edouard-Montpetit, Montréal, Québec, H3T 1J4, Canada

2 CGCC, the Center in Green Chemistry and Catalysis, Montréal, H3A 0B8, Canada

3 PROTEO, the Québec Network for Protein Function, Structure and Engineering, Québec, G1V 0A6, Canada

4 Department de Biochimie, Université de Montréal, 2900 Boulevard Edouard-Montpetit, Montréal, Québec, H3T 1J4, Canada

Manuscript under review.

Corresponding Author: Joelle N. Pelletier <joelle.pelletier@umontreal.ca>

7.2 Abstract

Microbial transglutaminase (MTG) is a practical tool to enzymatically form isopeptide bonds between peptide or protein substrates. This natural approach to crosslinking the side-chains of reactive glutamine and lysine residues is solidly rooted in food and textile processing. More recently, MTG's tolerance for various primary amines in lieu of lysine have revealed its potential for site-specific protein labeling with aminated compounds, including fluorophores. Importantly, MTG can label glutamines at accessible positions in the body of a target protein, setting it apart from most labeling enzymes that react exclusively at protein termini. To expand its applicability as a labeling tool, we engineered the B1 domain of Protein G (GB1) to probe the the selectivity and enhance the reactivity of MTG towards its glutamine substrate. We built a GB1 library where each variant contained a single glutamine at positions covering all secondary structure elements. The most reactive and selective variants displayed a >100-fold increase in incorporation of a recently developed aminated benzo[*a*]imidazo[2,1,5-*cd*]indolizine-type fluorophore, relative to native GB1. None of the variants were destabilized. Our results demonstrate that MTG can react readily with glutamines in α-helical, β-sheet, and unstructured loop elements and does not favor one type of secondary structure. Introducing point mutations within MTG's active site further increased reactivity towards the most reactive substrate variant, I6Q-GB1, enhancing MTG's capacity to fluorescently label an engineered, highly reactive glutamine substrate. This work demonstrates that MTG-reactive glutamines can be readily introduced into a protein domain for fluorescent labeling.

7.3 Introduction

For decades, microbial transglutaminase (MTG) from Streptoverticillium mobaraense has found widespread use in industries ranging from food preparation to textile processing and regenerative medicine.[1] This breadth of applicability stems from two general characteristics: the first is its capacity to form amide bonds via the acyl-transfer reaction it catalyzes. In its native reaction, MTG catalyzes the reaction between the γ-carboxamide of a peptide- or protein-bound glutamine (referred to as the glutamine substrate) and the ε-amino group of a peptide- or protein-bound lysine residue (referred to as the lysine substrate). Their conjugation produces isopeptide bonds – or protein crosslinks – for peptide and protein modification purposes (Figure 7-1). The second characteristic is its robustness: MTG is relatively thermostable, co-factor independent, tolerant to organic co-solvents, and active over a range of pHs.[2] These attributes make it possible to incorporate MTG into a wide array of reaction media and conditions.

More recently, concerted efforts have been made to take advantage of MTG's inherent ability to covalently modify proteins to further develop it as a tool for site-specific peptide and protein conjugation.[3] Site-specific protein conjugation, which grants the researcher the ability to fine-tune the properties of a protein post-translationally, is an area of intense research interest. Such modifications can modulate enzymatic activities, molecular interactions and recognition, or introduce functionalities that extend beyond the naturally-encoded chemistry.[4] Among these, fluorescent labeling of biomolecules is of paramount interest.[5-8] One of the foundations of this approach is to optimize the incorporation efficiency of the label onto a protein of interest. MTG has been applied for fluorescent labeling[9-11] yet the determinants for its selective reactivity remains elusive. The deconvolution of these details holds great potential for improving MTG's labeling capacity.

Enzymes that are used to conjugate proteins are generally limited to using the *N*- or *C*-terminus as the site of modification.[12] The power of these enzymes stems from each enzymatic class having an amino acid recognition sequence that is targeted with high or exclusive selectivity, as long as this sequence is located at a protein terminus. Formylglycine generating enzyme, phosphopantetheinyl transferase, farnesyltransferase, biotin ligase, and lipoic acid ligase are examples of enzymes that catalyze such bioconjugations.[12] As a recent example,

formylglycine generating enzyme has been used to construct artificially glycosylated proteins,[13] and DNA-protein conjugates.[14]

MTG differs from these enzymes, as its targeted residue does not need to be terminally located. This is advantageous as it allows for a label or modification to be introduced at any accessible, reactive position on the protein. MTG can thus serve as a labeling device for protein substrates that are not amenable to modification at their termini, or where internal modification of a protein is desired. Notable examples of MTG-catalyzed conjugation yielding applied protein products include the synthesis of antibody-drug conjugates[15-16] and PEGylation of pharmaceutically relevant proteins.[17-19]

These successes result from MTG's high promiscuity toward its lysine substrates,[1, 3] with its ability to accept numerous primary amines being a key for the incorporation of diverse chemical functionality, such as bio-orthogonal functional groups to fluorophores. In contrast, MTG's glutamine reactivity is restricted to protein- and peptide-bound glutamine residues. Phage display screening of glutamine-containing peptides has yielded several 'glutamine tags'[20] that were successfully applied to channel MTG's reactivity during protein labeling;[10] we note that this example used a *C*-terminally expressed glutamine recognition tag rather than a reactive glutamine internal to the target protein. Nonetheless, those glutamine-containing sequences are diverse in composition, revealing no clear pattern in the primary structure surrounding the reactive glutamine.[20]

Further efforts made to elucidate MTG's mode of substrate recognition include the elucidation of two crystal structures[21-22] as well as an alanine scan of its broad active site cavity.[23] These have provided a greater understanding of the catalytic mechanism, kinetic parameters, and identifying key residues essential for activity yet did not clarify the characteristics of glutamine reactivity. An investigation of the impact of local secondary structure on glutamine reactivity comparing apomyoglobin, α-lactalbumin, and fragment 205-316 of thermolysin concluded that unstructured regions strongly favored reactivity.[24] Indeed, the majority of their multiple surface-exposed glutamines were not MTG-reactive. Consistent with this, we have observed no conjugation using the highly structured TEM-1 β-lactamase or *E. coli* asparaginase II as potential glutamine substrates, despite having 7 and 13 exposed glutamines, respectively (data not shown).

In the face of a clear need to map glutamine reactivity relative to its molecular environment and design highly reactive glutamine substrates, here, we designed a tightly controlled system to investigate relative glutamine reactivity. Glutamine residues were introduced within a single framework, at various positions within elements of secondary and tertiary structure of the B1 domain of Protein G (GB1). GB1 is a small self-folding domain of 6.2 kDa that has been extensively characterized as a model for protein folding and unfolding[25] and can also be used to aid in soluble expression of small proteins.[26] Native GB1 contains a single glutamine located on its unique α-helix (Figure 7-2). We recently determined that MTG can conjugate GB1 at this residue;[27] the efficiency of conjugation was poor, which we attributed to the glutamine belonging to a well-defined element of secondary structure. This presented us with the opportunity to use GB1 as a probe for investigating MTG's glutamine reactivity and identifying more highly reactive locations for a glutamine residue, towards making MTG a more effective tool for protein conjugation.

To this end, we employed a semi-rational approach[28] to engineer both GB1 and MTG. We produced a library of 24 GB1 variants in which a single glutamine residue was introduced at various locations within its α-helix, loop structures, and β-sheet. We identified four GB1 variants that are at least 100-fold more reactive than native GB1; to our surprise, all belonged to well-structured elements. In parallel, based on previous mutagenesis results,[23] we mutated three residues in the active-site area of MTG in the form of a small, focused library of six MTG variants. By those means, we identified one MTG variant that is significantly more reactive against native GB1 than native MTG. When tested against the most reactive GB1 substrate variant, two out of six MTG variants were observed to be 2.5-fold more reactive than native MTG. We thus demonstrate that highly MTG-reactive glutamines can be engineered into a well-folded protein scaffold without regard to secondary structure location, and that MTG can be engineered to be more reactive towards its glutamine substrates.

7.4 Results

7.4.1 Design of the Single-Glutamine-Containing GB1 Variants

Our objective was to compare the susceptibility of the different elements of secondary structure, namely α-helices, β-sheet, and unstructured loop elements to serve as backdrops for presenting a MTG-reactive glutamine. We targeted for mutagenesis a similar number of positions belonging to α-helical, β-sheet, and unstructured loop elements. A further criteria was that these positions were all solvent-exposed in the crystal structure (PDB ID: 3GB1). The crystal structure was visualized using PyMOL.[20] Glycine residues were omitted out of concern that substitutions would perturb the structure.

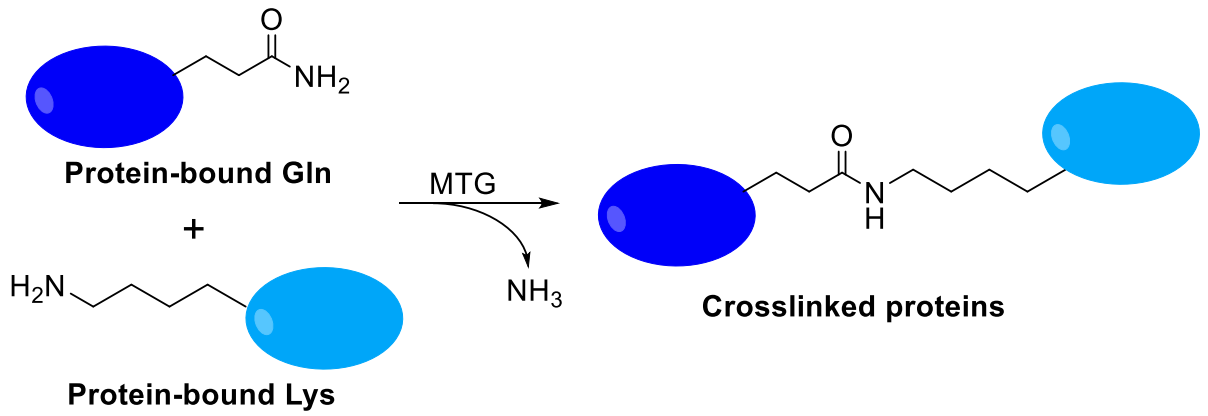


Figure 7-1 MTG-catalyzed protein crosslinking.

The first mutagenesis step consisted in replacing the sole native glutamine of GB1, Q32, with a structurally similar residue that MTG does not react with, asparagine. The Q32N knock-out served not only as the template for generating future mutants, but as a negative control to verify that no conjugation was occurring at other sites on GB1. This was confirmed by resolving on gel and by high-resolution MS (Table A 4-3). A single glutamine was then introduced at each of the 24 selected locations on the template. We confirmed that all the GB1 variants expressed solubly to similar levels as the native GB1 (Supporting Fig. A 4-1).

7.4.2 Fluorescent MTG Protein Assay

The establishment of a sensitive assay to monitor the efficiency of labeling of the GB1 variants was critical to the success of the study (Figure 7-2). We and others previously investigated MTG's ability to accept a variety of primary amines as substrates instead of lysine, and others have exploited this promiscuity as a tool to introduce diverse functionalities into proteins,[1, 3, 15, 29-30] providing us with considerable flexibility in the choice of our probe. Our standard methodology for monitoring the products of MTG-catalyzed conjugation has been based on the use of liquid chromatography-mass spectrometry (LC-MS).[31] However, visualizing and quantifying fluorescence output is more rapid and sensitive, and provides a direct means to screening for improved fluorescent protein labeling. To this effect, we recently reported a new class of highly tunable fluorescent compounds that can be readily functionalized to bear a primary amine.[27] These bright fluorescent dyes are characterized by an unusually high excitation-emission differential and are highly soluble in aqueous media, making them good candidates for bioconjugation. Although the primary amine of benzo[*a*]imidazo[2,1,5-*c,d*]indolizin-7-ylmethanaminium (**1**) is separated from the bulky, aromatic core by a single methylene, we have demonstrated that MTG can use it as a substrate to label two proteins, α-lactalbumin (α-LA) as well as GB1; while α-LA is well established to be highly reactive with MTG,[24, 32-33] GB1 had not yet been known to be a substrate prior to our investigation.[27] Here we extend this assay of fluorescent GB1 conjugation to the GB1 variants (Figure 7-2).

While MTG reacts with micromolar concentrations of protein, millimolar concentrations of small-molecule reagents are generally required for the reaction to proceed effectively.[22, 34] In the case of fluorescent labeling, use of fluorophore reagent 1 at a 100-fold excess relative to the GB1 protein substrates thus requires a means to remove excess unreacted **1**, to prevent it from masking visualization on tricine SDS-PAGE. Using a 20-fold excess of **1** resulted in suboptimal yields (data not shown). Microdialysis proved to be effective at removing excess 1 for visualization.[35]

The quantification of fluorescence is described herein according to two properties: selectivity and efficiency (Table 7-1). Selectivity refers to the degree to which GB1 is labeled in the presence of MTG relative to non-specific binding. When non-specific binding of **1** was observed, as in the case of native GB1 and some of the GB1 variants (Figure 7-3), the selectivity

ratio was calculated; the lower the background, the higher the selectivity. Efficiency, instead, compares the fluorescent output of a labeled GB1 variant to that of the labeled native GB1; it expresses the relative reactivity of the glutamine. We observed that selectivity tended to increase as efficiency increased.

7.4.3 Introduction of Glutamine into GB1 Loop Elements

Of the 24 glutamine-displaying GB1 variants prepared, eight of the targeted residues were located on flexible loops (Figure 7-4), with at least one mutation being made in each of

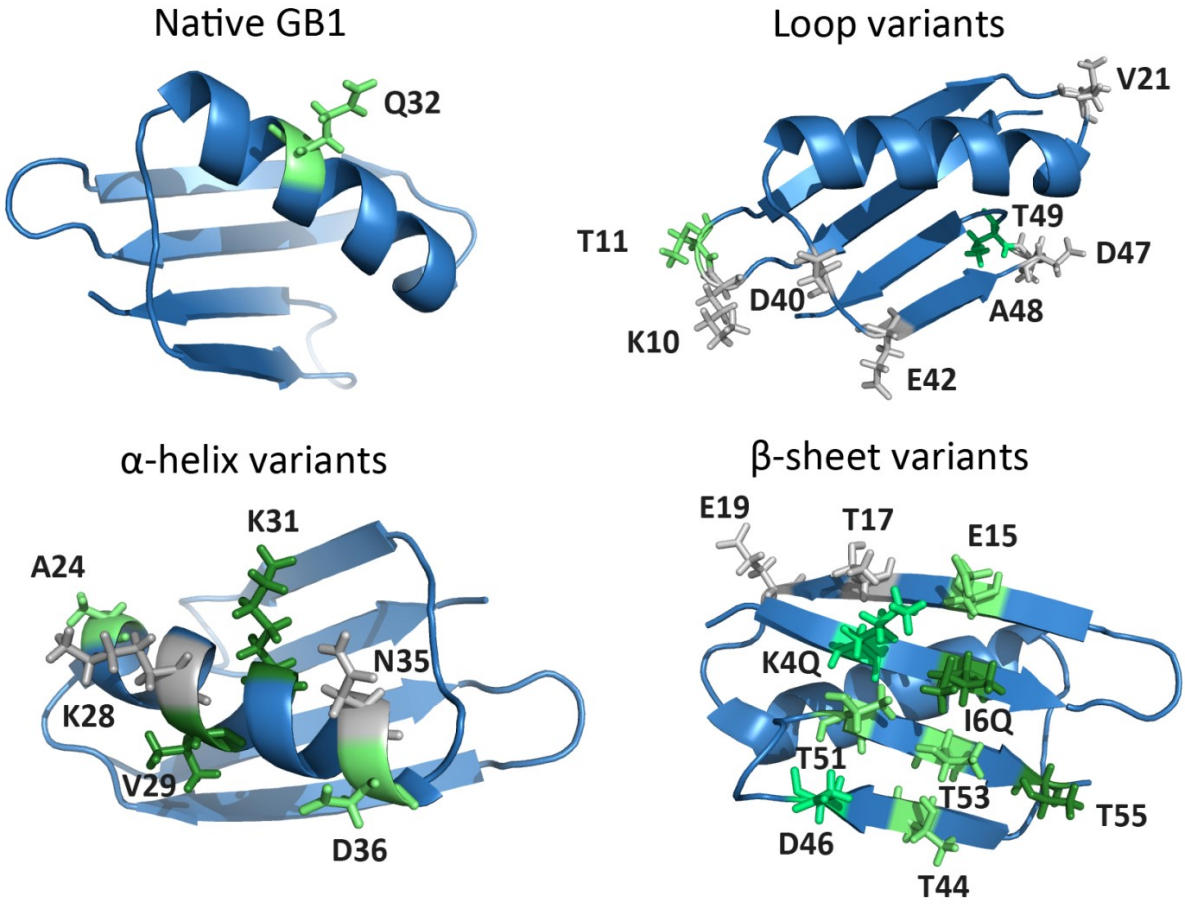


Figure 7-2 Structures of GB1.

Native GB1 (top left), as well as the residues which underwent glutamine substitution; loop variants (top right), helix variants (bottom left), and sheet variants (bottom right). The labeling efficiency of each residue is colored according to the results presented in Table 2; grey are inert.

the four loops present in GB1. Based on the report of higher glutamine reactivity in disordered regions,[24] we anticipated that this subgroup of variants should be the most reactive. While T49Q exhibited good fluorescence following conjugation with **1**, with both selectivity and efficiency well over one order of magnitude higher than native GB1, it was the only strongly improved loop variant. T11Q, located on a different loop, produced a modest increase compared to native GB1; all other loop variants were unreactive, or exhibited the same level of reactivity as the control lacking MTG (indicating non-specific binding of **1** to the GB1 variant). This is particularly surprising when closely observing the location of T49 within the crystal structure. Indeed, T49 is on the same loop as variants D47Q and A48Q; the former was inert to labeling, and A48Q was barely observable. These residues are all within a similar environment, making it difficult to rationalize the drastic difference in reactivity that MTG displays for its glutamine substrate. Similarly, the K10Q variant exhibited no reactivity despite being located beside T11 which, when substituted, was modestly more reactive than native GB1.

We hypothesized that stability of the GB1 variant could affect the likelihood of a glutamine residue being tagged by MTG: if the introduction of a glutamine into GB1 destabilizes the structure, the disorder may correlate with increased accessibility. To this end, we determined the thermal melting point (T_m) of each variant using differential scanning fluorimetry (DSF; Table 7-2).[36] DSF functions by the monitoring an increase in fluorescence upon binding of the dye, SYPRO Orange, to hydrophobic patches that become exposed as a protein unfolds. Variants that are more disordered should be less thermally stable, and display a lower T_m. The Tm calculated for all loop variants was essentially unchanged from the native GB1, allowing us to conclude that altered thermal stability of the variants is not a factor in the increased reactivity. These results indicate that there must be other determinants for glutamine reactivity beyond flexibility within its local environment.

Table 7-1 GB1 Q-library results after being treated with native MTG.

Selectivity represents the fold-increase of the ratio of fluorescence in a reaction to non-specific fluorescence in the control; the higher the selectivity, the lower the background. Efficiency is the fold-increase of the ratio between fluorescence of the GB1 variant to native GB1, labeled in increasingly saturated shades of green: 1 to 10-fold, pale green; 10 to 100-fold, bright green; greater than 100-fold, dark green. Signals that were lower than that of native GB1 are represented with a dash (-), and those that could not be accurately quantified due to saturation of the detector are indicated with an asterisk (*). N.D. = not detected, N/A = not available.

Secondary Structure	Mutation	Selectivity	Efficiency
α-helix	Native	1.3	1
	Q32N	N.D.	N.D.
Loop	K10Q	N.D.	N.D.
	T11Q	16	4
	V21Q	1.2	-
	D40Q	3.9	-
	E42Q	4.2	-
	D47Q	N.D.	N.D.
	A48Q	1.0	-
	T49Q	53	43
α-helix	A24Q	5.0	4.2
	K28Q	-	-
	V29Q	430	130
	K31Q	270	100
	N35Q	2.5	-
	D36Q	3.7	0.7
β-sheet	K4Q	98	79
	I6Q	190*	180
	E15Q	4.4	3.4
	T17Q	-	-
	E19Q	1.0	-
	T44Q	1.3	1.5
	D46Q	5.0	27
	T51Q	71	46

	T53Q	3.3	2.6
	T55Q	20	160

Table 7-2 Melting temperatures of GB1 variants, determined by differential scanning fluorimetry.

GB1 Variant	T_m, °C	Gln Location
WT	70.3 ± 0.3	α-helix
Q32N	70.0 ± 0.4	N/A
K10Q	70.0 ± 1.0	
T11Q	70.4 ± 1.3	
V21Q	69.5 ± 0.3	
D40Q	69.5 ± 1.1	Loop
E42Q	69.7 ± 1.4	
D47Q	69.8 ± 0.5	
A48Q	69.2 ± 0.7	
T49Q	70.2 ± 0.2	
A24Q	69.8 ± 0.3	
K28Q	70.0 ± 0.2	
V29Q	69.6 ± 0.4	α-helix
K31Q	69.6 ± 0.8	
N35Q	69.3 ± 0.2	
D36Q	69.9 ± 0.5	
K4Q	69.6 ± 1.5	
I6Q	70.1 ± 0.7	
E15Q	69.7 ± 0.1	
T17Q	70.0 ± 1.2	
E19Q	70.0 ± 1.0	β-sheet
T44Q	69.3 ± 1.1	
D46Q	70.1 ± 0.6	
T51Q	69.9 ± 0.5	
T53Q	70.0 ± 0.7	
T55Q	70.0 ± 0.1	

7.4.4 Relocating Glutamine in the α-Helix of GB1

Residue Q32 in native GB1 is located on the α-helix, with its side-chain exposed freely to the solvent (Figure 7-2). Based solely upon solvent accessibility (and, presumably, accessibility for MTG), K28Q, K31Q, the native Q32 and N35Q would be expected to be the most reactive among the α-helix variants. Upon screening, only K31Q was among the most reactive while K28Q, the native Q32, and N35Q were among the least effective positions assayed. The immediate neighbor of K28Q, V29Q, is less exposed, yet it and K31Q were two orders of magnitude more selective and efficient than native GB1. This demonstrates that solvent exposition is not a strong predictor of reactivity. A22Q and D34Q are at opposite ends of the helix, and both exhibited similar, modest increases in selectivity and efficiency. Similarly to the loop variants, the Tm calculated for all α-helix variants was essentially unchanged from the native GB1. We thus demonstrate that the well-structured and tightly packed α-helix of GB1 can harbour highly MTG-reactive glutamines.

7.4.5 Glutamine in the β-Sheet of GB1 Can Also Be Reactive

With the β-sheet being the largest single secondary structure element within GB1, over 40 % of the newly introduced glutamines were located within it. Upon examining the crystal

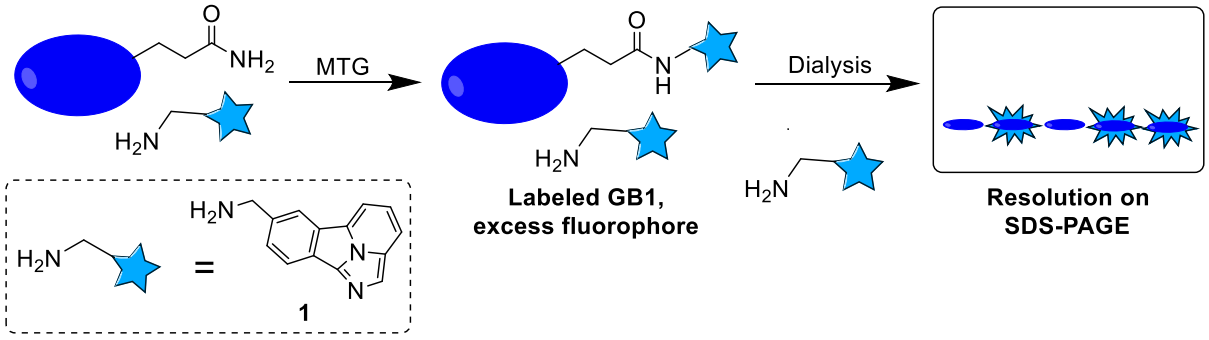


Figure 7-3 Diagram summarizing the assay used to conjugate GB1 variants with fluorescent probe 1.

GB1's single glutamine residue is targeted by MTG, forming an amide bond with the amine-bearing fluorophore. Excess fluorophore is removed by dialysis and reactivity is analyzed using SDS-PAGE.

structure, we speculated that many of these mutations would react poorly, particularly those located on the internal β-strands 1 and 4 because they belong to a flat protein surface that does not appear to be complementary to the crevice that forms MTG's active site (Figure 7-5).[23] This speculation was invalidated when the most selective and efficient variant was determined to be I6Q, located within β-strand 1. K4Q, the other variant introducing glutamine within β-strand 1, also exhibited high reactivity. T55Q also reacted strongly; that residue is located at the very edge of β-strand 4 and is not as tightly concealed within the structure as is I6Q. T17Q and E19Q lost reactivity, with the remaining four variants exhibiting activities on par with native GB1.

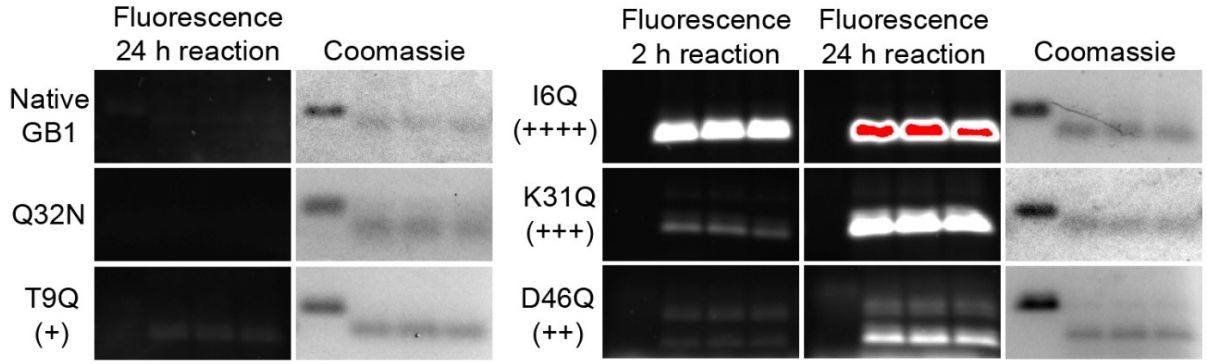


Figure 7-4 Representative SDS-PAGE analysis of fluorescently labeled GB1 variants.

Equal quantities of protein were loaded and excited for 5 s using a Cy2 excitation filter prior to Coomassie brilliant blue staining. GB1 variants exhibiting low fluorescent conjugation efficiencies were barely visible even after 24 h of reaction time (Native GB1, T9Q; Q32N served as a negative control); for this reason, the 2 h reactions were omitted. Red bands indicate saturation of the detector.

These results are surprising, as MTG has been reported to prefer glutamine-containing regions that are predominantly unstructured,[24] yet in the GB1 framework we observe the highest reactivity in α-helical and β-sheet regions. Therefore, secondary structure (or lack thereof) is not a strong predictor of glutamine reactivity. We attempted to identify patterns in the primary sequence flanking the reactive glutamines but failed to identify any potential markers to predict glutamine reactivity (Figure 7-6). We also considered tertiary structure, seeking patterns in surface charge and hydrophobicity (Figures A 4-2 to 4-4). No clear sequence or structural

pattern or trend was observed amongst the reactive glutamines, making it difficult to predict where MTG will bind.

7.4.6 Active-site Mutations in MTG Increase Reactivity Towards the Glutamine Substrate

Having obtained highly reactive glutamine variants of GB1 towards MTG-catalyzed conjugation, we sought to further improve the performance of MTG toward these GB1 substrates. There is a shortage of data indicating which residues play a role in binding MTG's glutamine-containing substrate. Among the most informative works is an alanine scan of 29 active-site residues, constituting 9 % of the apoenzyme's amino acid sequence.[23] A number of residues were found to be critical for activity, crippling MTG when substituted for alanine. Some alanine substitutions, however, resulted in an increase in activity, including the highly conserved W69 and the conserved Y75 and Y302. We selected these three aromatic residues for mutagenesis, introducing histidine as a semi-conservative modification (aromatic yet smaller and more hydrophilic), or glycine as a potentially more disruptive modification, ultimately yielding six MTG point mutants.

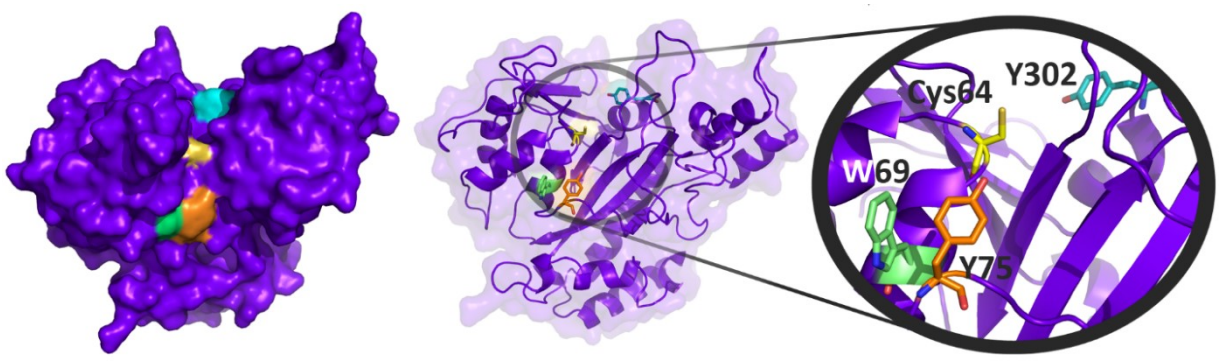


Figure 7-5 Location of residue substitutions in MTG.

Left : A top-down surface view into the active-site crevice, with W69, Y75, and Y302 colored green, orange, and cyan, respectively. Center: cartoon representation with the active site zoomed (right). The catalytic cysteine essential for enzymatic activity, shown in yellow, was not mutated. PDB coordinates 1IU4.

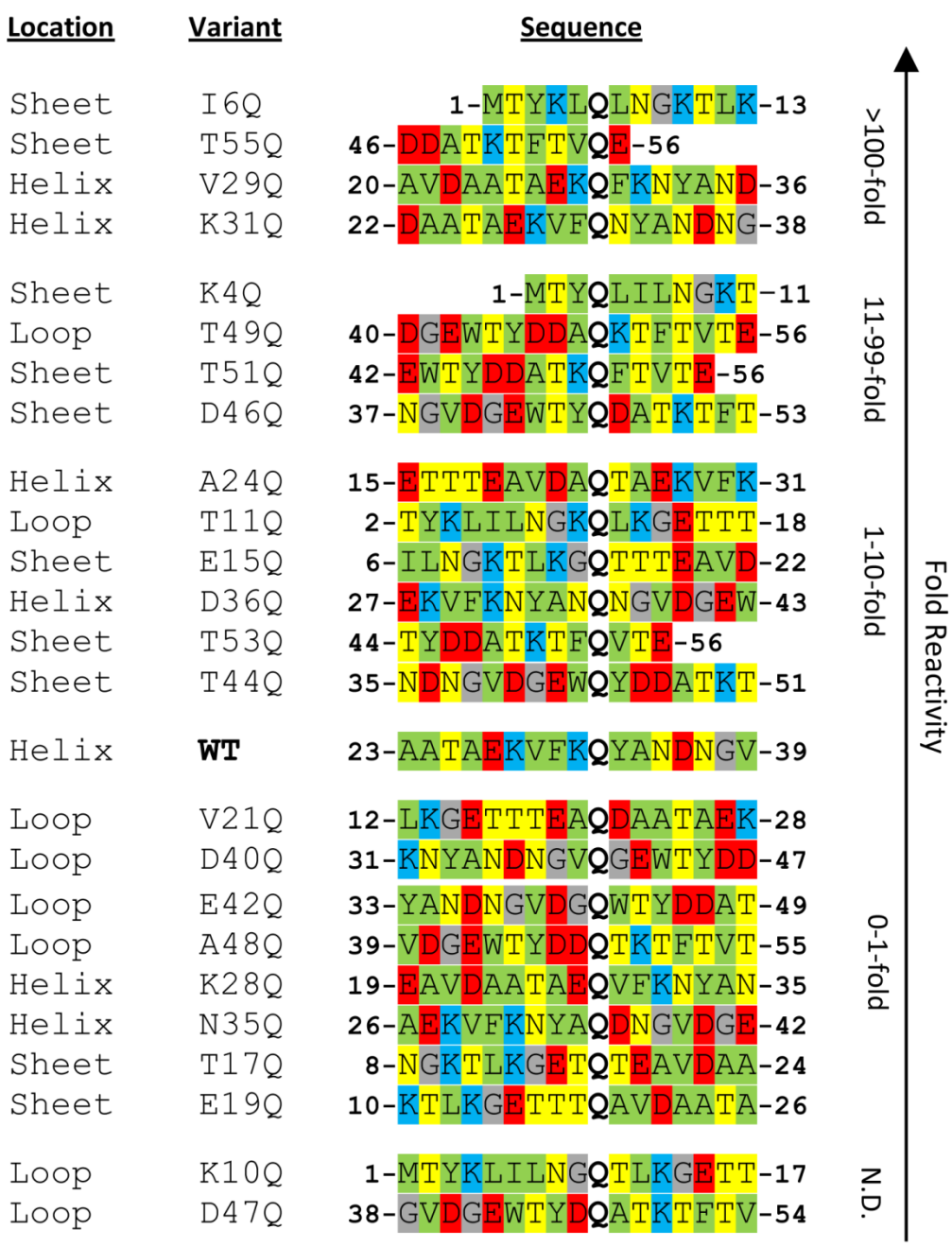


Figure 7-6 Primary amino acid sequence alignment of GB1 variants, centered on the glutamine residue present in the native or variant GB1s; residue numbering is indicated. *Variants are ranked according to their reactivity, with the most reactive variant presented first. Amino acids are colored according to the properties of their side chains: green = hydrophobic; yellow = polar; blue = basic; red = acidic; grey = glycine.*

Upon purification of the MTG variants, we verified activity using the standard hydroxamate assay with the Cbz-L-glutaminylglycine (ZQG) protected dipeptide substrate.[37] All six variants were not only active, but exhibited higher activity for ZQG than did native MTG (Table 7-3). This is consistent with observation of increased activity in the corresponding alanine variants.[23] Y302G displayed the greatest improvement, being nearly 3-fold more active. Despite this increase in activity toward ZQG, when the variants were assayed against native GB1, half were observed to have very modest increases in efficiency and selectivity, with both Y302 variants falling into this group (Table 7-4). When compared to the reactivity of native MTG, both Y75 and W69G substitutions decreased the conjugation efficiency.

Table 7-3 Specific activities of variant MTGs towards the model dipeptide, ZQG.

MTG Enzyme	Specific Activity (U/mg)	Activity Increase
Native	25.9	1
W69G	43.2	1.7-fold
W69H	43.1	1.7-fold
Y75G	52.0	2.0-fold
Y75H	46.6	1.8-fold
Y302G	70.4	2.7-fold
Y302H	68.4	2.6-fold

Building on these findings, we proceeded to assay the variants against the most reactive GB1 protein, I6Q. The W69 and Y75 MTG variants were all less active than native MTG, but both Y302G and Y302H MTG variants were moderately more active than native MTG, as had been the case when assayed against native GB1. When the efficiency of the six MTG variants on I6Q was compared to native GB1, they were all between one and two orders of magnitude more reactive towards I6Q GB1, maintaining the trend that I6Q GB1 is more reactive to labeling than native GB1.

The high reactivity of I6Q GB1 resulted in rapid saturation of the fluorescence signal after the exposure time we had determined to be optimal for quantification of most variants (5

s). To compare the reactivity of MTG and its variants more accurately, we recorded the fluorescent signal after 1 s of exposure, where saturation was not observed (Table 7-4). Three MTG variants, W69H, Y302G, and Y302H, reacted with I6Q GB1 as well as or better than native GB1. W69H maintained the same level of reactivity as native MTG, where Y302G and Y302H were twice as efficient. Taken together, these results demonstrate that engineered MTG and GB1 variants can be paired to create an effective protein labeling system: it is possible to engineer both the substrate and the catalyst towards higher efficiencies, and these effects of engineering the substrate and the enzyme are cumulative.

In summary, among a library of 24 different glutamine-containing point mutations, covering 43 % of the amino acid sequence, four GB1 variants were observed to be at least 100-fold more reactive towards MTG and four more were at least 10-fold more reactive. Thus, one-third of the glutamines tested in the well-folded, globular GB1 protein provided good substrates for MTG labeling, with the I6Q substitution being the most reactive among all. We were not able to identify any clear trend that MTG displays towards the environment in which the glutamine residue is located, whether considering the primary sequence flanking the glutamine (Figure 7-6) or tertiary structure properties (Figures A 4-2 to A 4-4). We initially expected that loop variants would be the most reactive, as the high flexibility of these elements would make them the most likely candidates to fit into MTG's active-site cleft. However, the loop variants underperformed relative to α-helix or β-sheet variants, leading us to hypothesize that if secondary structure plays a role in MTG's substrate recognition, there are other, more important factors that dominate MTG's glutamine selectivity. If MTG undergoes a significant structural rearrangement upon binding to its glutamine-bearing protein substrate, as does its mammalian TG2 counterpart,[38] then predicting their mode of interaction may require their co-crystallization.

To conclude, through a semi-rational approach, we constructed and improved a protein labeling system in which both the catalyst and substrate were optimized. Point mutations improved the reactivity of the substrate protein, GB1, by over two orders of magnitude, which was further enhanced when coupling with variants of the catalyst, MTG. Through this process, we probed the selectivity MTG displays for its glutamine-bearing protein substrate. Although no clear recognition pattern was observed, we have demonstrated the straightforward

engineering of MTG-reactive glutamines in a well-folded domain, suggesting that other proteins are amenable to similar modification to allow MTG-catalyzed protein labeling.

Table 7-4 MTG variant reactivity towards native and I6Q GB1.

For the top three tables, fluorescence was quantified after 5 s of exposure, while the last table quantified after 1 s of exposure to prevent saturation of the detector. Selectivity represents the fold-increase of the ratio of fluorescence in the reactions over non-specific fluorescence in the control; the higher the selectivity, the lower the background. Efficiency is the fold-increase of the ratio of fluorescence of the variant relative to native protein and is labeled in increasingly saturated shades of green: 1 to 10-fold, pale green; 10 to 100-fold, bright green; greater than 100-fold, dark green. Signals that were lower than that of native GB1 are represented with a dash (-), and those that could not be accurately quantified due to saturation of the detector are indicated with an asterisk (*). N/A = not available.

Selectivity

MTG variant		W69		Y75		Y302	
		G	H	G	H	G	H
GB1 substrate	Native	1.0	3.2	1.1	1.0	14	2.1
	I6Q	56	110*	5.1	39	140*	200*

Efficiency, compared to native MTG

MTG variant		W69		Y75		Y302	
		G	H	G	H	G	H
GB1 substrate	Native	-	2.1	-	-	2.5	1.1
	I6Q	-	1.1	-	-	1.6	1.6

Efficiency, compared to native GB1 substrate

MTG variant		W69		Y75		Y302	
		G	H	G	H	G	H
GB1 substrate	I6Q	230	31	42	80	50	200

I6Q GB1 substrate, 1 s exposure

MTG variant	Native	W69H	Y302G	Y302H
Selectivity	330	170	140	210
Efficiency	N/A	1.0	2.5	2.5

7.4.7 Acknowledgements

The authors thank Marie-Christine Tang and Alexandra Furtos of the Regional Mass Spectrometry Centre (Université de Montréal) for their technical aid. This work was supported by Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant RGPIN 227853.

7.5 Materials and Methods

7.5.1 Materials

The plasmid pDJ1-3 was kindly provided by Professor M. Pietzsch (Martin-Luther-Universität, Halle-Wittenberg, Germany). pDJ1-3 encodes the proenzyme of MTG from *S. mobaraensis* inserted between the *Nde*I and *Xho*I restriction sites of the vector pET20b.[39] The plasmid pQE80L-CysGB1Cys was kindly provided by Professor Hongbin Li (University of British Columbia, Vancouver, Canada). pQE80L-CysGB1Cys encodes GB1 with an *N*-terminal poly-histidine tag inserted into the *Bam*HI restriction site of the vector pQE80L. The plasmid also encodes an extra cysteine residue present just before and after the open reading frame of native GB1. The sequence served as a template for amplifying the native GB1 coding sequence. Deionized water (18Ω) was used for all experiments. Products used for the expression and purification of MTG and GB1 were of biological grade.

Other chemicals used were purchased from the suppliers listed below. Carboxybenzyl–L-glutaminyl–glycine (Z-Gln-Gly, or ZQG) was from Peptide Institute (Osaka, Japan). Glutathione (reduced) and thiamine were from Bioshop (Burlington, Canada). Dimethyl sulfoxide (99.7%) was purchased from Fisher Scientific (Ontatio, Canada). Formic acid (98 % purity) was from Fluka Analytical (St. Louis, USA). FastDigest *Nde*I, *Bam*HI, *Dpn*I, Phusion® High-Fidelity Polymerase and Fast AP Thermosensitive Alkaline Phosphatase were purchased from ThermoFisher Scientific (Waltham, MA, USA). Takara T4 DNA Ligase was purchased from Clontech (Mountain View, CA, USA). FastBreak™ Cell Lysis Reagent was purchased from Promega (Madison, WI, USA).

7.5.2 Expression and purification of MTG

MTG was expressed and purified as previously described.[39] Briefly, a 5-mL starter culture of *E. coli* BL21 (DE3) containing the plasmid pET20b-MTG, which expresses a *C*-terminally 6-His-tagged version of MTG, was propagated overnight at 37°C in ZYP-0.8G medium and shaking at 240 rpm. It was used to inoculate 500 mL of auto-inducing ZYP-5052 medium. After 2h of incubation at 37°C and 240 rpm, the temperature was reduced to 22°C overnight. Cells were collected by centrifugation and resuspended in 50 mM sodium phosphate buffer, 300 mM NaCl, pH 7.5. The cells were lysed using a Constant Systems cell disruptor set at 37 kPSI and cooled to 4°C. After further centrifugation to remove insoluble cellular matter, the inactive form of MTG was incubated with trypsin (1 mg/mL solution, 1:9 ratio of trypsin to MTG, v/v) for the purpose of cleaving its pro-sequence. Activated MTG was purified using a 5-mL His-trap nickel-nitrilotriacetic acid (Ni-NTA) column (GE Healthcare) equilibrated in 50 mM phosphate buffer, pH 7.5, with 300 mM NaCl, and eluted with an imidazole gradient (0 – 250 mM) using an Åtka FPLC (GE Healthcare). After purification, active MTG was dialyzed against 50 mM sodium phosphate buffer, 300 mM NaCl, pH 7.5. The average yield was 25 mg of activated MTG per litre of culture, with ~ 85% purity as estimated by SDS-PAGE and revelation with Coomassie blue stain. Aliquots were snap frozen and stored at -80°C in 15% glycerol.

7.5.3 MTG Mutagenesis

Plasmid pDJ1-3, encoding the open reading frame for MTG, was used as a template for mutagenesis. All mutants were obtained using the rolling circle approach.[40-41] Following mutagenesis with Phusion® High Fidelity polymerase, the amplified PCR product was treated with FastDigest *Dpn*I before being transformed into *E. coli* BL21 (DE3) for protein expression. Ampicillin (Amp) was used at 100 µg/mL for plasmid maintenance. Sequences were confirmed by DNA sequencing (ABI 3730 DNA sequencer, IRIC Genomic Platform at Université de Montréal). (Table S1).

7.5.4 Expression and purification of native GB1 and variants

A 2-mL starter culture of *E. coli* BL21 (DE3) containing the plasmid pET15b-GB1, or mutagenized plasmids expressing a variant within the same vector, which expresses an *N*-terminally 6x-His-tagged version of GB1, was propagated overnight at 37°C in LB medium containing 100 μg/mL ampicillin and shaking at 240 rpm. 500 μL of the starter culture was used to inoculate 50 mL of LB medium containing 100 μg/mL ampicillin. After 3h of incubation at 37°C and 240 rpm, IPTG was added to a final concentration of 0.2 mM, and expression was allowed to proceed for 3 hours. Cells were collected by centrifugation and resuspended in 2.7 mL of 50 mM sodium phosphate buffer, 300 mM NaCl, pH 7.5. FastBreak™ cell lysis reagent was added to the resuspended cells to a final volume of 3 mL, mixed by inversion, and incubated at room temperature for 10 min. After further centrifugation at 4°C to remove insoluble cellular matter, the clarified lysate was loaded onto 1 mL of Nickel-nitrilotriacetic acid (Ni-NTA) resin (GE Healthcare) equilibrated in 50 mM phosphate buffer, pH 7.5, with 300 mM NaCl. The resin was washed with 10 column volumes of the same buffer containing 15 mM imidazole, and eluted in 3 mL in the phosphate buffer containing 250 mM imidazole. After purification, GB1 was dialyzed against 50 mM sodium phosphate buffer, pH 7.5, 300 mM NaCl, 1 mM EDTA. The average yield was 3 mg of GB1 per 50 mL of culture, with ~ 90% purity as estimated by tricine SDS-PAGE[35] and revelation with Coomassie blue stain. Aliquots were snap frozen and stored at -80°C in 20% glycerol.

7.5.5 GB1 Mutagenesis

The pET15b-GB1 plasmid encoding the open reading frame for GB1 was used as a template for mutagenesis. The sequence for glutamine knock-out, Q32N, was generated first from native GB1, and was subsequently used as a template for amplification of all other GB1 mutants. Site overlap extension was used to generate mutant GB1 sequences.[42] The DNA fragments were digested with FastDigest *Nde*I and *Bam*HI restriction enzymes, and religated into pET15b which had been cut with the same enzymes and also dephosphoylated, and transformed in *E. coli* BL21 (DE3).

7.5.6 MTG Activity Assay

The activity of purified MTG was quantified using the hydroxamate assay.[37] Briefly, MTG was incubated with 30 mM Z-Gln-Gly and 100 mM hydroxamate at 37°C for 10 min. A concentrated acidic ferric chloride solution (2.0 M $FeCl_3 \cdot 6$ H2O, 0.3 M trichloroacetic acid, 0.8 M HCl) was used to quench the reaction in a 1:1 ratio (v/v) to the reaction, which was then vortexed and left to stand at room temperature for 10 min. The resulting iron complex was quantified by its absorbance at 525 nm, using the molar extinction coefficient 525 nm. One unit (U) of MTG produces 1 μmol of L-glutamic acid and γ-monohydroxamate per min at 37°C.

7.5.7 Fluorescent conjugation assays

Purified GB1 variants were quantified by measuring the A_{280}, using a molar extinction coefficient of 9970 $M^{-1}cm^{-1}$ as calculated using ExPASy's ProtParam module. Native GB1 or its variants (50 μM) were combined with 5 mM fluorophore **1** and 2.5 mM glutathione. The conjugation reaction was initiated by the addition of 2 U/mL of MTG, where control reactions had an equivalent volume of buffer (100 mM sodium phosphate, pH 7.5) added instead. The final volume of each reaction was 150 μL and all were incubated at 37°C for 24 h. Aliquots of 50 μL were taken after 2 h, 6 h, and 24 h of reaction time, and quenched with the addition of 2 μL formic acid. Excess, unreacted fluorophore was removed by dialysis using a Pierce™ 96-well microdialysis plate with a 3.5 kDa MWCO (ThermoFisher). To this effect, aliquots of 50 μL were dialyzed against three exchanges of 2 mL of buffer (100 mM sodium phosphate, 300 mM NaCl, 1 mM EDTA, pH 7.5 at 4°C.

7.5.8 High resolution mass spectrometry

Samples (5 μL) were injected onto an Aeris peptide XB-C18, 3.6-μm, 150 × 2.1 mm LC column (Phenomenex) and eluted with a 16-minute, 5-50% ACN/H2O gradient. Masses were detected under positive ionization (ESI) with a Synapt G2S (Q-TOF) triple quadrupole mass detector (Waters).

7.5.9 Differential Scanning Fluorimetry

Melting temperatures of GB1 proteins was determined using a LightCycler ® 480 real-time PCR platform (Roche) by thermally-induced incorporation of SYPRO Orange into the unfolding protein, as previously described.[36] Briefly, 6.66 × SYPRO Orange solution (Invitrogen) with 8 μM test protein was probed in a 96-well LightCycler plate (Sarstedt). SYPRO Orange and the protein were diluted with 50 mM sodium phosphate, pH 7.5, to a final volume of 20 μL per well. Controls contained SYPRO Orange in buffer. The plates were sealed using Optically Clear Sealing Tape (Sarstedt) and heated from 20°C to 95°C with a ramp speed of 0.04°C/sec and 10 acquisitions/°C. Fluorescence was monitored with a CCD camera, using λ_{exc} = 483 nm and λ_{em} = 568 nm and a 1 s exposure time. Any curve showing a maximum fluorescence plateau during denaturation was excluded from the T_m calculation.

For the T_m calculations, both temperature and fluorescence data were smoothed.[43] The first derivatives dFluo or dTemp were calculated using the cubic spline interpolation. The preliminary maximum was determined to obtain the half-values to the left and right of it. The linear fit for the curve outside the half-values was calculated, followed by the calculation of the average deviation from the fit. If the maximum was below the detection limit (fit value + 3 × deviation), the T_m determination was considered uncertain. The quadratic fit around the maximum was then calculated as follows to obtain T_m. The first derivative of the quadratic fit function (y-value) was set to 0 and the x-axis value (temperature) was resolved. Then, the average deviation of the curve points around the maximum from the quadratic fit was calculated. If the relative deviation was greater than 5%, the T_m values were rejected if the corresponding maximum was below the detection limit. However, T_m values with a maximum above the detection but a relative deviation greater than 5% were defined as uncertain.

7.6 Conflicts of Interest

The authors declare no competing financial interests.

7.7 References

1. Rachel, N. M.; Pelletier, J. N., Biotechnological applications of transglutaminases. *Biomolecules* **2013,** *3*, 870-888.

2. Yokoyama, K.; Nio, N.; Kikuchi, Y., Properties and applications of microbial transglutaminase. *Appl. Microbiol. Biotechnol.* **2004,** *64*, 447-454.
3. Strop, P., Versatility of microbial transglutaminase. *Bioconjug. Chem.* **2014,** *25*, 855-62.
4. Krall, N.; da Cruz, F. P.; Boutureira, O.; Bernardes, G. J., Site-selective protein-modification chemistry for basic biology and drug development. *Nat. Chem.* **2016,** *8* (2), 103-113.
5. Hoffman, R. M., The multiple uses of fluorescent proteins to visualize cancer *in vivo*. *Nat. Rev. Cancer* **2005,** (10), 796-806.
6. Kerppola, T. K., Visualization of molecular interactions by fluorescence complementation. *Nat. Rev. Mol. Cell Biol.* **2006,** *7*, 449-456.
7. Chan, J.; Dodani, S. C.; Chang, C. J., Reaction-based small-molecule fluorescent probes for chemoselective bioimaging. *Nat. Chem.* **2012,** *4*, 973-984.
8. Zhang, G.; Zheng, S.; Liu, H.; Chen, P. R., Illuminating biological processes through site-specific protein labeling. *Chem. Soc. Rev.* **2015,** *44*, 3405-3417.
9. Kamiya, N.; Abe, H., New fluorescent substrates of microbial transglutaminase and its application to peptide tag-directed covalent protein labeling. *Methods Mol. Biol.* **2011,** *751*, 81-94.
10. Oteng-Pabi, S. K.; Pardin, C.; Stoica, M.; Keillor, J. W., Site-specific protein labelling and immobilization mediated by microbial transglutaminase. *Chem. Commun.* **2014,** *50*, 6604-6606.
11. Siegmund, V.; Schmelz, S.; Dickgiesser, S.; Beck, J.; Ebenig, A.; Fittler, H.; Frauendorf, H.; Piater, B.; Betz, U. A.; Avrutina, O.; Scrima, A.; Fuchsbauer, H. L.; Kolmar, H., Locked by Design: A Conformationally Constrained Transglutaminase Tag Enables Efficient Site-Specific Conjugation. *Angew. Chem., Int. Ed.* **2015,** *54*, 13420-13424.
12. Rashidian, M.; Dozier, J. K.; Distefano, M. D., Enzymatic labeling of proteins: techniques and approaches. *Bioconjug. Chem.* **2013,** *24*, 1277-1294.
13. Liang, S. I.; McFarland, J. M.; Rabuka, D.; Gartner, Z. J., A modular approach for assembling aldehyde-tagged proteins on DNA scaffolds. *J. Am. Chem. Soc.* **2014,** *136*, 10850-10853.
14. Smith, E. L.; Giddens, J. P.; Iavarone, A. T.; Godula, K.; Wang, L. X.; Bertozzi, C. R., Chemoenzymatic Fc glycosylation via engineered aldehyde tags. *Bioconjug. Chem.* **2014,** *25*, 788-795.
15. Strop, P.; Liu, S. H.; Dorywalska, M.; Delaria, K.; Dushin, R. G.; Tran, T. T.; Ho, W. H.; Farias, S.; Casas, M. G.; Abdiche, Y.; Zhou, D.; Chandrasekaran, R.; Samain, C.; Loo, C.; Rossi, A.; Rickert, M.; Krimm, S.; Wong, T.; Chin, S. M.; Yu, J.; Dilley, J.; Chaparro-Riggers, J.; Filzen, G. F.; O'Donnell, C. J.; Wang, F.; Myers, J. S.; Pons, J.; Shelton, D. L.; Rajpal, A., Location matters: site of conjugation modulates stability and pharmacokinetics of antibody drug conjugates. *Chem. Biol.* **2013,** *20*, 161-167.
16. Dennler, P.; Chiotellis, A.; Fischer, E.; Bregeon, D.; Belmant, C.; Gauthier, L.; Lhospice, F.; Romagne, F.; Schibli, R., Transglutaminase-based chemo-enzymatic conjugation approach yields homogeneous antibody-drug conjugates. *Bioconjug. Chem.* **2014,** *25*, 569-578.
17. Fontana, A.; Spolaore, B.; Mero, A.; Veronese, F. M., Site-specific modification and PEGylation of pharmaceutical proteins mediated by transglutaminase. *Adv. Drug. Deliv. Rev.* **2008,** *60*, 13-28.

18. Maullu, C.; Raimondo, D.; Caboi, F.; Giorgetti, A.; Sergi, M.; Valentini, M.; Tonon, G.; Tramontano, A., Site-directed enzymatic PEGylation of the human granulocyte colony-stimulating factor. *FEBS J.* **2009,** *276*, 6741-6750.
19. Mero, A.; Spolaore, B.; Veronese, F. M.; Fontana, A., Transglutaminase-mediated PEGylation of proteins: direct identification of the sites of protein modification by mass spectrometry using a novel monodisperse PEG. *Bioconjug. Chem.* **2009,** *20*, 384-389.
20. Sugimura, Y.; Yokoyama, K.; Nio, N.; Maki, M.; Hitomi, K., Identification of preferred substrate sequences of microbial transglutaminase from *Streptomyces mobaraensis* using a phage-displayed peptide library. *Arch. Biochem. Biophys.* **2008,** *477*, 379-383.
21. Kashiwagi, T.; Yokoyama, K.; Ishikawa, K.; Ono, K.; Ejima, D.; Matsui, H.; Suzuki, E., Crystal structure of microbial transglutaminase from Streptoverticillium mobaraense. *J. Biol. Chem.* **2002,** *277*, 44252-44260.
22. Yang, M. T.; Chang, C. H.; Wang, J. M.; Wu, T. K.; Wang, Y. K.; Chang, C. Y.; Li, T. T., Crystal structure and inhibition studies of transglutaminase from *Streptomyces mobaraense*. *J. Biol. Chem.* **2011,** *286*, 7301-7307.
23. Tagami, U.; Shimba, N.; Nakamura, M.; Yokoyama, K.; Suzuki, E.; Hirokawa, T., Substrate specificity of microbial transglutaminase as revealed by three-dimensional docking simulation and mutagenesis. *Protein Eng. Des. Sel.* **2009,** *22*, 747-752.
24. Spolaore, B.; Raboni, S.; Ramos Molina, A.; Satwekar, A.; Damiano, N.; Fontana, A., Local unfolding is required for the site-specific protein modification by transglutaminase. *Biochemistry* **2012,** *51*, 8679-8689.
25. Shea, J. E.; Brooks, C. L., 3rd, From folding theories to folding proteins: a review and assessment of simulation studies of protein folding and unfolding. *Annu. Rev. Phys. Chem.* **2001,** *52*, 499-535.
26. Cheng, Y.; Patel, D. J., An efficient system for small protein expression and refolding. *Biochem. Biophys. Res. Commun.* **2004,** *317*, 401-405.
27. Lévesque, É.; Bechara, W. S.; Constantineau-Forget, L.; Pelletier, G.; Rachel, N. M.; Pelletier, J. N.; Charette, A. B., General C-H Arylation Strategy for the Synthesis of Tunable Visible Light Emitting High Stokes Shift Benzo[*a*]imidazo[2,1,5-*c,d*]indolizine Fluorophores. *J. Org. Chem.* **2017,** *82*, 5046–5067.
28. Chica, R. A.; Doucet, N.; Pelletier, J. N., Semi-rational approaches to engineering enzyme activity: combining the benefits of directed evolution and rational design. *Curr. Opin. Biotechnol.* **2005,** *16*, 378-384.
29. Ohtsuka, T.; Sawa, A.; Kawabata, R.; Nio, N.; Motoki, M., Substrate specificities of microbial transglutaminase for primary amines. *J. Agric. Food. Chem.* **2000,** *48*, 6230-6233.
30. Gundersen, M. T.; Keillor, J. W.; Pelletier, J. N., Microbial transglutaminase displays broad acyl-acceptor substrate specificity. *Appl. Microbiol. Biotechnol.* **2014,** *98*, 219-230.
31. Rachel, N. M.; Pelletier, J. N., One-pot peptide and protein conjugation: a combination of enzymatic transamidation and click chemistry. *Chem. Commun.* **2016,** *52*, 2541-2544.
32. Matsumura, Y.; Chanyongvorakul, Y.; Kumazawa, Y.; Ohtsuka, T.; Mori, T., Enhanced susceptibility to transglutaminase reaction of alpha-lactalbumin in the molten globule state. *Biochim Biophys Acta* **1996,** *1292* (1), 69-76.
33. Nieuwenhuizen, W. F.; Dekker, H. L.; de Koning, L. J.; Groneveld, T.; de Koster, C. G.; de Jong, G. A., Modification of glutamine and lysine residues in holo and apo alpha-lactalbumin with microbial transglutaminase. *J. Agric. Food Chem.* **2003,** *51*, 7132-7139.

34. Oteng-Pabi, S. K.; Keillor, J. W., Continuous enzyme-coupled assay for microbial transglutaminase activity. *Anal. Biochem.* **2013,** *441*, 169-173.
35. Schagger, H.; von Jagow, G., Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Anal Biochem* **1987,** *166* (2), 368-79.
36. Niesen, F. H.; Berglund, H.; Vedadi, M., The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat. Protoc.* **2007,** *2*, 2212-2221.
37. Folk, J. E.; Cole, P. W., Mechanism of action of guinea pig liver transglutaminase. I. Purification and properties of the enzyme: identification of a functional cysteine essential for activity. *J. Biol. Chem.* **1966,** *241*, 5518-5525.
38. Pinkas, D. M.; Strop, P.; Brunger, A. T.; Khosla, C., Transglutaminase 2 undergoes a large conformational change upon activation. *PLoS Biol.* **2007,** *5*, e327.
39. Marx, C. K.; Hertel, T. C.; Pietzsch, M., Soluble expression of a pro-transglutaminase from Streptomyces mobaraensis in Escherichia coli. *Enzyme Microb. Technol.* **2007,** *40*, 1543-1550.
40. Fire, A.; Xu, S. Q., Rolling replication of short DNA circles. *Proc. Natl. Acad. Sci. U.S.A.* **1995,** *92*, 4641-4615.
41. Lizardi, P. M.; Huang, X.; Zhu, Z.; Bray-Ward, P.; Thomas, D. C.; Ward, D. C., Mutation detection and single-molecule counting using isothermal rolling-circle amplification. *Nat. Genet.* **1998,** *19*, 225-232.
42. Ho, S. N.; Hunt, H. D.; Horton, R. M.; Pullen, J. K.; Pease, L. R., Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* **1989,** *77*, 51-9.
43. Savitzky, A.; Golay, M. J. E., Smoothing + Differentiation of Data by Simplified Least Squares Procedures. *Anal Chem.* **1964,** *36*, 1627-1639.

Chapter 8 - Discussion and future work

Our efforts throughout this thesis, both chemical and biochemical, worked towards the common goal illustrating the value and versatility of transglutaminase-catalyzed bioconjugation, and to build upon these characteristics by further improving its utility. MTG has already found high commercial value in the food industry, and we hypothesize that its bioconjugation ability has immense value within applications including biological imaging, modulating pharmacokinetic properties of biological therapeutics, or the construction of antibody-drug conjugates. Specifically, we focused on how transglutaminase has the ability to introduce a covalent linkage at various internal points of a protein's structure to fluorescently label proteins, making it distinct from other enzymatic labeling systems which are spatially restricted to a terminal labeling site. It can be combined with bio-orthogonal chemistries, even in a one-pot format, to enhance its labeling applicability. The topics investigated included biocatalysis, fundamental characterization and enzyme engineering; here, the key findings of these topics will be discussed with respect to our overreaching goal.

8.1 Biocatalysis

Our biocatalytic investigations started with TG2 before progressing to MTG. We sought to open up a biocatalytic avenue for TG2, which was the formation of peptide bonds and is the topic of Chapter 3. Gratifyingly, we were able to confirm the peptide synthase capacity of TG2. However, its narrow substrate range, sensitivity to reaction and storage conditions, susceptibility to accepting water as a nucleophile, and co-factor dependence were drawbacks, without having a distinctive advantage. Conversion rates were low, such that an LC-MS assay was required to acquire substrate conversion data with greater accuracy. It is possible that TG2 could benefit from engineering to enhance its peptide synthase activity, but its other attributes makes me question the feasibility of the task. As work with MTG had been progressing concurrently, it became evident that TG2 was more cumbersome than its microbial counterpart. We wanted to devote our efforts towards an enzymatic system that was simple and versatile, making it more likely to be applicable and embraced. For this to occur, we decided that the enzyme with which to continue was MTG.

Initially, we had hopes to develop MTG as a general biocatalyst for amide bond synthesis, expanding its substrate scope to other amides, or even esters, other than glutamine. Esters cannot be utilized by MTG, with the exception of basal activity detected with some ester analogs of glutamine, as determined previously in our group.[1] Experimentation with MTG by myself and others determined MTG exhibits stringent selectivity towards its amide substrate; it will only accept glutamine. Not only this, but MTG has the added restriction of this glutamine needing to be part of a larger peptide or protein structure; it will not accept free glutamine. From a biocatalytic perspective, this is a damaging drawback as it precludes small-molecule chemistry.

Nonetheless, we hypothesized that this rigid glutamine reactivity was the foundation of an appealing protein labeling tool. Indeed, with MTG displaying a broad scope of reactivity towards its amine substrate, this meant that a peptide or protein could be labeled with a diverse array of reactive amines.[2] In the end, we pushed to expand the scope of substrate diversity by coupling MTG with downstream chemical reactions, resulting in additional functionalization of a protein conjugated with a reactive functional group. To continue this work, rather than investigating additional complementary reactions, I'd propose selecting the most promising one out of the four already characterized, and work to improve its effectiveness. Propargylamine was one of the most highly reactive amines, reaching full conversion, making it an appealing substrate, although it has the disadvantage of requiring copper for the downstream click step. Even so, the high reactivity of propargylamine paired with the CuAAC makes it a work-around labeling strategy to direct conjugation with a fluorophore, because as shown in Chapters 5 and 7, direct labeling with bulky amines displays poor reactivity.

8.2 Structural determinants

If a substrate-bound crystal structure of MTG existed, our engineering efforts would have been different, as it would provide valuable information. For example, it's possible that MTG may undergo a conformational change to accommodate its substrates, especially when reacting with larger proteins. If such were the case, the current apoenzyme crystal structure is of little use as a guide for engineering.

The inconsistency we observed between different batches of proteases for processing the signal sequence of MTG was frustrating and difficult to explain. As discussed in Chapter 6, the recently reported work focusing on weakening the interactions between active MTG and its cleaved pro-sequence[3] could make purification and recovery using our, or a similar strategy, achievable. Since MTG can only be solubility expressed as a zymogen, a protease activation step will remain a requirement. To address poor digestion efficiency, it's possible that screening a greater number of proteases would reveal one that is most suitable. For example, Rickert and colleagues observed success with 3C protease.[3] Using a *C*-terminally tagged construct of MTG, possibly coupled with an additional polishing chromatography step (e.g., gel filtration, ion exchange), is another alternative if attempting to crystallize a tagged protein is acceptable.

8.3 Evolvability

Engineering MTG to make it a more effective catalyst is key to increasing its versatility and applicability. The engineering efforts we made with MTG were conservative. Without a high-throughput or automated screen, our experimental output was limited. Even so, we were astonished to find that even within a small collection of point mutants, modest improvements in activity were found. This suggests that MTG is susceptible to improvement by engineering and evolution, although not enough data exists at the moment to determine the extent of such an improvement. A protein that is susceptible and tolerant to mutations is described as 'evolvable'. Additionally, thermostability is a characteristic that greatly aids a protein's propensity to being engineered or evolved.[4] MTG's inherent thermostability makes it a good candidate for tolerating mutations. The development of a high-throughput screening assay for MTG would be an enormous advancement in this respect. It's possible that a fluorescent assay, be it the one used in Chapter 7 or another, could be adapted to be carried out fully in a 96-well format. Robotic automation would be essential for such a strategy. To begin, MTG would have to be expressed, lysed (using lysis buffer, as was done in Chapter 7), and cleaved within crude lysate. If cleavage occurs, then a volume of the crude lysate would be transferred to another plate containing the buffer, reactive amine probe, and protein substrate. It's at this point, the most troublesome step is revealed: removal of excess, unreacted fluorophore to quantify any conjugated protein. Even

low concentrations of remaining fluorophore will likely mask a signal. Considering this, before trying anything else, I would concentrate efforts to addressing this logistical obstacle.

Improving activity is one thing, but a far more significant challenge would be to improve the site-specific selectivity of MTG. Indeed, the largest stumbling block plaguing the field of protein labeling is non-specific or background labeling. MTG's ability to react with side chains at numerous locations within a protein, as well as not having a defined recognition sequence structure makes it, in my view, an incredibly interesting starting point for site-specific labeling. I've always wondered if MTG could be evolved to specifically recognize, hypothetically, any accessible glutamine residue. If so, MTG would serve as a starting point for evolving variants for customizable labeling: targeting any specific glutamine-containing sequence, and particularly, where this would be a native protein sequence, eliminating the need for an additional encoded recognition tag. This would require screening MTG reactivity towards a glutamine-containing sequence of interest in the presence of other reactive proteins, and being capable of detecting increased reactivity towards your desired sequence and that sequence only. This exact concept was applied to another similar enzyme, sortase, indicating that it's theoretically possible.[5-6] Of course, such ambitions require a tremendous investment, and the uncertainty of payoff makes it intimidating. Indeed, as with evaluating reactivity, without having an effective screen, such efforts are challenging, or even unfeasible.

There's a common saying in the field of enzyme engineering: "you get what you screen for". To expand this within the context of research, I would state: you get what you try for. Now that I've reached the end of this thesis, I can state concretely that I still believe what I sought out at the beginning: that biocatalysis is powerful technology that can, and has, provided solutions to complex challenges. I hope to see biocatalysis embraced more widely in the future.

8.4 References

1. Gundersen, M. T. New insights into the substrate specificities of microbial transglutaminase: a biocatalytic perspective. Université de Montréal, Montréal, 2014.
2. Gundersen, M. T.; Keillor, J. W.; Pelletier, J. N., Microbial transglutaminase displays broad acyl-acceptor substrate specificity. *Appl. Microbiol. Biotechnol.* **2014,** *98*, 219-230.
3. Rickert, M.; Strop, P.; Lui, V.; Melton-Witt, J.; Farias, S. E.; Foletti, D.; Shelton, D.; Pons, J.; Rajpal, A., Production of soluble and active microbial transglutaminase in *Escherichia coli* for site-specific antibody drug conjugation. *Protein Sci.* **2016,** *25*, 442-455.

4. Bloom, J. D.; Labthavikul, S. T.; Otey, C. R.; Arnold, F. H., Protein stability promotes evolvability. *Proc. Natl. Acad. Sci. U.S.A.* **2006,** *103*, 5869-5874.
5. Chen, I.; Dorr, B. M.; Liu, D. R., A general strategy for the evolution of bond-forming enzymes using yeast display. *Proc. Natl. Acad. Sci. U.S.A.* **2011,** *108*, 11399-11404.
6. Dorr, B. M.; Ham, H. O.; An, C.; Chaikof, E. L.; Liu, D. R., Reprogramming the specificity of sortase enzymes. *Proc. Natl. Acad. Sci. U.S.A.* **2014,** *111*, 13343-13348.

Annex 1 – Chapter 3 Supplemental information

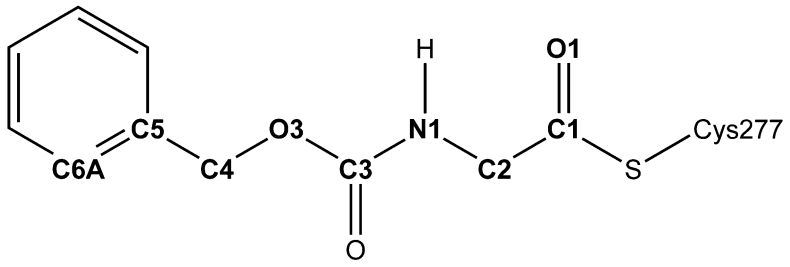


Figure A 1-1 Atom names for Cys277 acylated with Cbz-glycyl moiety. Atoms used to define torsions are highlighted in bold.

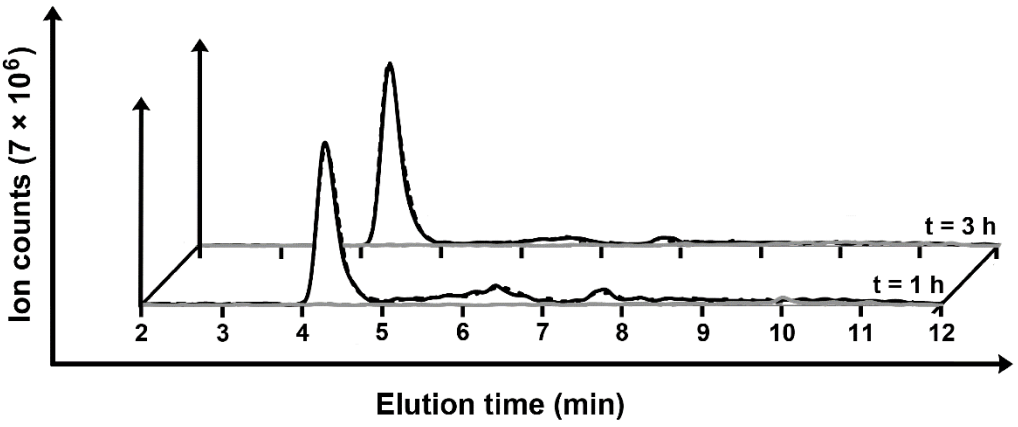


Figure A 1-2 LC-MS traces of the reaction mixture of Cbz-GlyNH$_2$ with GlyNH$_2$ in the presence and absence of gTG2.

Cbz-GlyNH$_2$ remaining at different time points after incubation in the presence (solid lines) or absence (dashed lines) of gTG2. No substrate consumption or dipeptide product formation (grey line) was observed, demonstrating that Cbz-GlyNH$_2$ is not a substrate of gTG2.

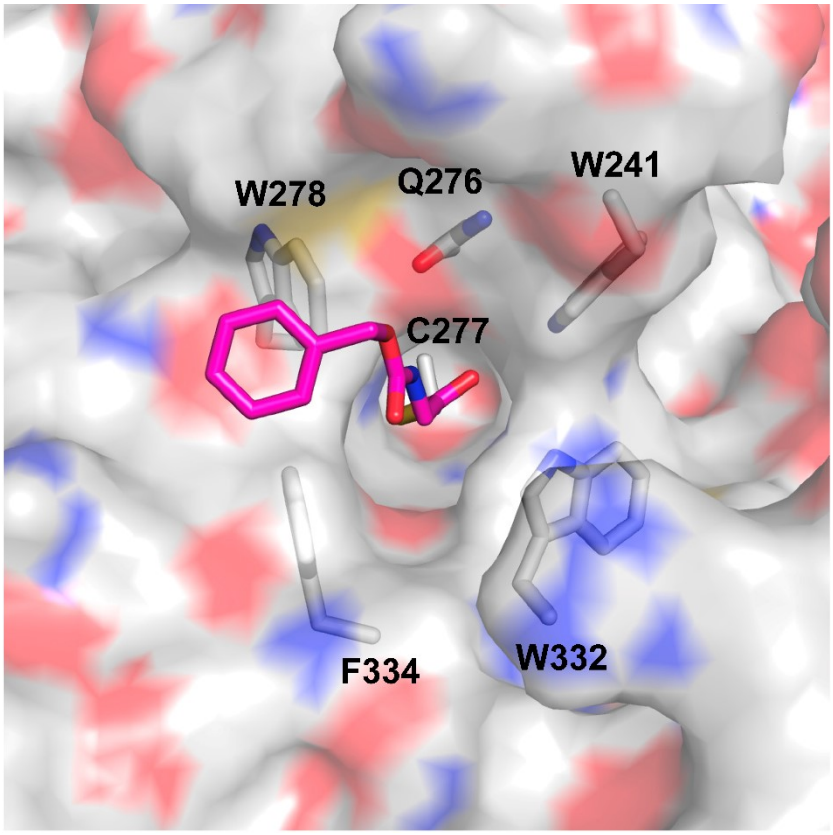


Figure A 1-3 Active site tunnel of gTG2. The Cbz-Gly moiety (magenta) covalently attached to the catalytic Cys277 residue (white) fits inside a tunnel formed by residues Trp241, Gln276, Trp278, Trp332 and Phe334.

The gTG2 surface is colored according to atom types, with carbon, oxygen, nitrogen, and sulfur atoms colored white, red, blue, and yellow, respectively.

Annex 2 – Chapter 4 Supplemental information

Materials

The plasmid pDJ1-3 was kindly provided by Professor M. Pietzsch (Martin-Luther-Universität, Halle-Wittenberg, Germany). pDJ1-3 encodes the proenzyme of MTG from *S. mobaraensis* inserted between the *Nde*I and *Xho*I restriction sites of the vector pET20b. Deionized water (18Ω) was used for all experiments. HPLC solvents were of analytical grade, and products used for the expression and purification of MTG were of biological grade.

Other chemicals used were purchased from the suppliers listed below. Carboxybenzyl–L-glutaminyl–glycine (Z-Gln-Gly, or ZQG) was from Bachem (Bubendorf, Switzerland). Glutathione (reduced) and thiamine were from Bioshop (Burlington, Canada). Azidopropylamine was synthesized previously in the laboratory of Jeffrey Keillor, according to published protocols. Propargylamine hydrochloride (95%) and 4-pentynoic acid (98%) were purchased from Acros Organics (New Jersey, USA). Formic acid (98 % purity) was from Fluka Analytical (St. Louis, USA). 6-Azidohexanoic acid was purchased from Merck Millipore (Darmstadt, Germany). Copper (II) sulfate pentahydrate (≥ 98%), (+)-sodium L-ascorbate (≥ 98%) and α-lactalbumin from bovine milk (calcium saturated) were purchased from Sigma Aldrich (St. Louis, MO, USA). Sulfo-cyanine5 azide (Cy5-azide) and sulfo-cynanine5 alkyne (Cy5-alkyne) were purchased from Lumiprobe (Hallandale Beach, FL, USA).

MTG Expression and Purification

MTG was expressed and purified as previously described. Briefly, a 5-mL starter culture of *E. coli* BL21 (DE3) containing the plasmid pET20b-MTG, which expresses a *C*-terminally 6-His-tagged version of MTG, was propagated overnight at 37°C in ZYP-0.8G medium and shaking at 240 rpm. It was used to inoculate 500 mL of autoinducing ZYP-5052 medium. After 2h of incubation at 37°C and 240 rpm, the temperature was reduced to 22°C overnight. Cells were collected by centrifugation and resuspended in 0.2 M Tris-HCl, pH 6.0. The cells were lysed using a Constant Systems cell disruptor set at 37 kPSI and cooled to 4°C. After further centrifugation to remove insoluble cellular matter, the inactive form of MTG was incubated with trypsin (1 mg/mL

solution, 1:9 ratio of trypsin to MTG, v/v) for the purpose of cleaving its pro-sequence. Activated MTG was purified using a 5-mL His-trap nickel-nitrilotriacetic acid (Ni-NTA) column (GE Healthcare) equilibrated in 50 mM phosphate buffer, pH 8.0, with 300 mM NaCl, and eluted with an imidazole gradient (0 – 250 mM) using an Åtka FPLC (GE Healthcare). After purification, active MTG was dialyzed against 0.2 M Tris-HCl buffer, pH 6.0. The average yield was 25 mg of activated MTG per litre of culture, with ~ 85% purity as estimated by SDS-PAGE and revelation with Coomassie blue stain. Aliquots were snap frozen and stored at -80°C in 15% glycerol.

MTG Activity Assay

The activity of purified MTG was quantified using the hydroxamate assay. Briefly, MTG was incubated with 30 mM Z-Gln-Gly and 100 mM hydroxamate at 37°C for 10 min. A concentrated acidic ferric chloride solution (2.0 M $FeCl_3 \cdot 6\ H_2O$, 0.3 M trichloroacetic acid, 0.8 M HCl) was used to quench the reaction, which was then vortexed and left to stand at room temperature for 10 min. The resulting iron complex was quantified by its absorbance at 525 nm. One unit (U) of MTG produces 1 μmol of L-glutamic acid and γ-monohydroxamate per min at 37°C.

Conjugation Assays

Amide, amine and complementary azide or alkyne substrates (30 or 60 mM, as indicated) were combined with 2.5 mM $CuSO_4 \cdot 5\ H_2O$, 25 mM sodium ascorbate, and 5 mM glutathione. The conjugation reaction catalyzed by MTG was initiated by the addition of 1 U/mL of MTG, where control reactions had an equivalent volume of buffer (200 mM Tris-acetate, pH 7.5) added. The final volume of each reaction was 350 μL and all were incubated at 37°C for the time indicated.

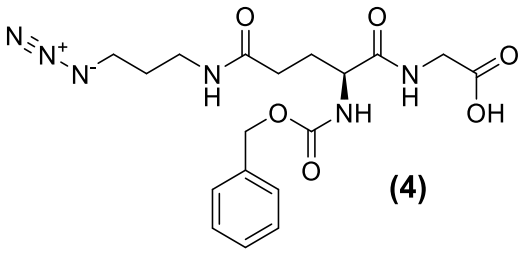
For labelling of α-lactalbumin, azidopropylamine or propargylamine (2 mM), 5 mM glutathione, 1 mM $CuSO_4$, 10 mM sodium ascorbate, and 2 mM Cy5-alkyne or Cy5-azide were mixed with α-lactalbumin such that its final concentration was 4 mg/mL, in 200 mM Tris-acetate buffer, pH 7.5. The final volume of each reaction was 200 μL. Reactions were incubated at 37°C for 24 – 48 h. The reactions were washed 6 times over a Spin-X® UF microfuge concentrator, 10k

MWCO (Corning), using 200 mM Tris-acetate buffer, pH 7.5, containing 2 mM EDTA. Washed and concentrated sample (10 µL out of 75-100 µL) was resolved using tricine SDS-PAGE. The fluorescent bands were visualized and recorded using a Bio Rad ChemiDoc™ MP Imaging System using an excitation filter of 625 nm with a 30 nm bandpass. The gels were then stained with Coomassie brilliant blue to reveal the protein.

Characterization of Product Formation

Samples were prepared for HPLC-MS analysis by taking an aliquot at specific time points indicated by transferring 10 µL of each reaction to a vial containing 10 µL of formic acid. 960 µL of 18.2 mΩ deionized water and 20 µL of internal standard solution (5 g/L of 4-methoxybenzamide in neat DMSO) were added and mixed by vortex. Samples (10 µL) were injected onto a Synergi 4-µm polar-RP 80 Å, 150 × 4.60 mm LC column (Phenomenex), using an Agilent 1200 series HPLC apparatus and eluted with a 5-70% MeOH/H_2O gradient. Masses were detected under positive ionization with a single quadrupole mass detector. Concentrations of amide substrate and reaction products were quantified by comparison to a standard curve constructed with the corresponding compound and the internal standard.

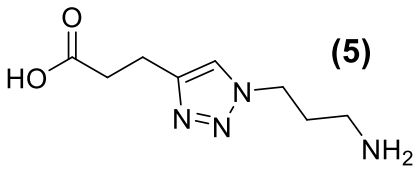
Synthesis of Products



(4)

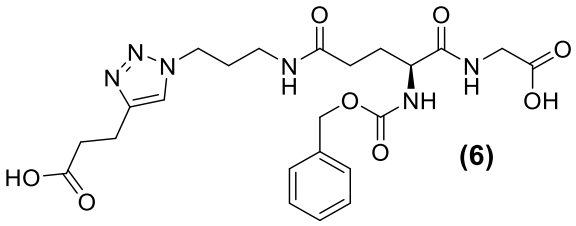
ZQG-APA (4). An aqueous solution (1 mL) buffered by 200 mM potassium phosphate, pH 7.5, containing 50 mM ZQG (1) and 90 mM azidopropylamine (3) was prepared. The conjugation reaction was initiated by adding 1 U/mL of MTG, and incubated at 37°C for a minimum of 2 hours. An additional 1 U/mL of MTG was then added to the reaction, vortexed, and incubated at the same temperature for 16 hours. The reaction volume was centrifuged in a microfuge (10 min, 13,000 rpm, 4°C) to remove insoluble debris, and the supernatant was divided

into two 500 µL volumes. They acidified by adding approximately 5 µL 6M HCl to each mixture. A white precipitate formed shortly after acidification, and it was centrifuged again the supernatant transferred to new tubes. Dichloromethane (500 µL) was added to each tube and shaken vigorously. The mixtures were centrifuged, and more white precipitate was observed sitting at the interface between the two phases. The aqueous phase was removed and the precipitate transferred into a new tube. This precipitate was dried at 37°C, and its structure analyzed by NMR and its mass determined by MS. ^{1}H NMR (700 MHz, DMSO-d_6): δ 8.09 (t, J = 5.4 Hz, 1H), 7.88 (t, J = 5.3 Hz, 1H), 7.46 (d, J = 8.1 Hz, 1H), 7.39 – 7.31 (m, 5H), 5.04 (d, J = 12.6 Hz, 1H). 5.02 (d, J = 10.1 Hz, 1H), 4.01 (m, 1H), 3.76 (dd, J = 9.1, 5.6 Hz, 1H), 3.67 (dd, J = 9.1, 5.6 Hz, 1H), 3.34 (t, J = 6.8 Hz, 2H), 3.12 – 3.05 (m, 2H), 2.19 – 2.11 (m, 2H), 1.95 – 1.90 (m, 1H), 1.76 – 1.70 (m, 1H), 1.63 (quint, 6.8 Hz, 2H) ppm. ^{13}C NMR (176 MHz, DMSO-d_6): δ 172.2, 172.0, 171.6, 156.3, 137.4, 128.8, 128.25, 128.17, 65.9, 54.7, 48.9, 41.5, 36.3, 32.2, 28.9, 28.4 ppm.



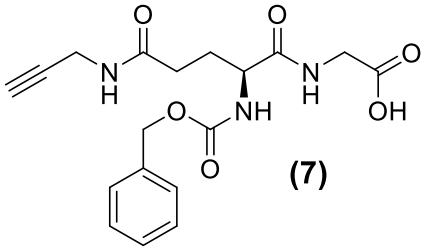
(5)

Triazole-APA (5). An aqueous solution (3 mL) containing 300 mM azidopropylamine (3), 300 mM 4-pentynoic acid (2), 5 mM $CuSO_4$ and 50 mM sodium ascorbate was prepared. The reaction was incubated at 37°C for 24 hours, and then lyophilized. ^{1}H NMR (700 MHz, D_2O): δ 7.68 (s, 1H), 4.42 (t, J = 6.7, 2H), 2.86 (q, J = 7.0 Hz, 4H), 2.442 (t, J = 7.2 Hz, 2H), 2.17 (quint, J = 7.0 Hz, 2H) ppm. ^{13}C NMR (176 MHz, D_2O): δ 181.3, 147.9, 123.1, 47.1, 36.56, 36.53, 27.3, 21.6 ppm.



(6)

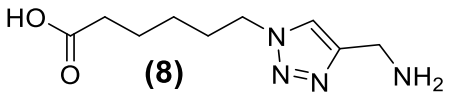
ZQG-Triazole-APA (6). ZQG-APA was resuspended in 200 mM potassium phosphate, pH 7.5, to a final concentration of 100 mM, to which 110 mM 4-pentynoic acid, 2.5 mM $CuSO_4$ and 25 mM sodium ascorbate was also added and mixed by vortexing, and incubated for 24 hours at 37°C. A volume of methanol containing 0.5% formic acid was added in 2.5-fold excess, vortexed, and incubated at -20°C for two hours. A precipitate was observed, and the solution was centrifuged at 4°C for 10 minutes at 13,000 rpm. A small, white and blue grainy pellet was observed, and the supernatant was kept and transferred to a new microfuge tube. The volume was evaporated down to approximately 100 µL, more methanol was added in a 5-fold excess, and the procedure repeated. Acetonitrile was added in 10-fold excess and a thick brown precipitate immediately formed. The mixture was incubated on ice for 1 hour, centrifuged, and the supernatant removed. The pellet was left to dry and was subsequently analyzed by NMR and HPLC-MS. ^{1}H NMR (700 MHz, DMSO-d_6): δ 8.29 (t, J = 5.4 Hz, 1H), 7.84 (s, 1H), 7.69 (t, J = 8.5 Hz, 1H), 7.56 (br, s, 1H), 7.37 – 7.30 (m, 5H), 5.04 (d, J = 11.7 Hz, 1H), 5.10 (d, J = 12.9 Hz, 1H), 4.29 (m, 2H), 3.91 (q, J = 7, 7.6 Hz, 1H), 3.23 (s, 1H), 3.20 (dd, J = 16.5, 4.2 Hz, 2H), 2.99 (m, 1H), 2.73 (t, J = 7.4 Hz, 2H), 2.18 (br, 2H), 2.13 (t, J = 7.6 Hz, 2H), 1.89 (m, 2H), 1.76 (m, 1H) ppm. ^{13}C NMR (176 MHz, DMSO-d_6): δ 174.7, 172.4, 170.8, 159.0, 156.3, 137.5, 128.8, 128.20, 128.15, 122.6, 65.8, 63.7, 60.8, 55.0, 47.2, 44.3, 36.1, 32.3, 30.3, 28.7, 23.2 ppm.



(7)

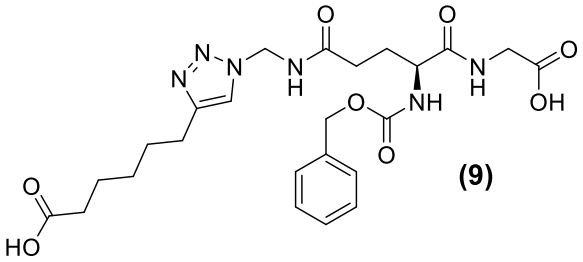
ZQG-PRO (7). An aqueous solution (3 mL) buffered by 200 mM Tris acetate, pH 7.5, containing 40 mM ZQG and 80 mM propargylamine was prepared. The reaction was initiated by adding 1

U/mL of MTG, and incubated at 37°C for 16 hours. Formic acid was added to the reaction (1%), vortexed, and placed on ice for 3 hours. A white precipitate formed, and was separated by centrifugation for 10 minutes at 4°C and 13,000 rpm. The supernatant was discarded and the pellet was washed with ice-cold water containing 1 % formic acid four times. The pellet was left to dry and was subsequently analyzed by NMR and HPLC-MS. ^{1}H NMR (700 MHz, DMSO-d_6): δ 8.23 (t, J = 5.4 Hz, 1H), 8.18 (t, J = 5.8 Hz, 1H), 7.44 (d, J = 8.2 Hz, 1H), 7.39-7.31 (m, 5H), 5.04 (d, J = 12.6 Hz, 1H), 5.02 (d, J = 12.9 Hz, 1H), 4.02 (m, 1H), 3.83 (dd, J = 5.4, 2.5 Hz, 2H), 3.80 (dd, J = 17.8, 6.3 Hz, 1H), 3.72 (dd, J = 17.5, 5.6 Hz, 1H), 3.08 (t, J = 2.4 Hz, 1H), 2.51 (dt, J = 3.5, 1.7 Hz, 1H), 2.21-2.13 (m, 2H), 1.95-1.90 (m, 1H), 1.76-1.70 (m, 1H). ^{13}C NMR (176 MHz, DMSO-d_6): δ 172.4, 171.8, 171.6, 156.4, 137.4, 128.8, 128.26, 128.18, 81.7, 73.3, 65.9, 54.6, 41.1, 31.9, 28.3 ppm.



(8)

Triazole-PRO (8). An aqueous solution (3 mL) containing 300 mM propargylamine, 300 mM 6-azidohexanoic acid, 5 mM $CuSO_4$ and 50 mM sodium ascorbate was prepared. The reaction was incubated at 37°C for 24 hours and then lyophilized. ^{1}H NMR (700 MHz, D_2O): δ 8.00 (s, 1H), 4.36 (t, *J* = 6.9 Hz, 2H), 4.22 (s, 2H), 2.05 (t, *J* = 7.3 Hz, 2H), 1.82 (quint, *J* = 7.3 Hz, 2H), 1.45 (quint, *J* = 7.5 Hz, 2H), 1.13 (quint, *J* = 7.6 Hz, 2H) ppm. ^{13}C NMR (176 MHz, D2O): δ 183.4, 139.5, 125.2, 50.4, 37.1, 33.9, 28.9, 25.2, 24.9 ppm.



(9)

ZQG-Triazole-PRO (9). ZQG-PRO was resuspending in water to a final concentration of 150 mM, to which 150 mM 6-azidohexanoic acid, 2.5 mM $CuSO_4$ and 25 mM sodium ascorbate was also added and mixed by vortexing, and incubated for 24 hours at 37°C. A small, dark brown pellet

was observed, and the supernatant was kept and transferred to a new microfuge tube. Ice-cold 300 µL water containing 1 % formic acid was added to the mixture. A white precipitate began to form after 15 minutes, and the solution was incubated on ice for 3 hours. The mixture was centrifuged once more, and the supernatant discarded. The pellet was left to dry and was subsequently analyzed by NMR and HPLC-MS. ^{1}H NMR (700 MHz, DMSO-d_6): δ 8.27 (t, J = 5.8 Hz, 1H), 8.19 (t, J = 5.7 Hz, 1H), 7.89 (s, 1H), 7.45 (d, J = 8.2 Hz, 1H), 7.37-7.30 (m, 5H), 5.02 (br, 2H), 4.28 (m, 4H), 4.02 (m, 1H), 3.80 (dd, J = 17.5, 7 Hz, 1H), 3.71 (dd, J = 17.5, 7 Hz, 1H), 2.23-2.15 (m, 4H), 1.97-1.92 (m, 1H), 1.80-1.71 (m, 3H), 1.51 (quint, J = 7.5 Hz, 2H), 1.26-1.21 (br, m, 2H). ^{13}C NMR (176 MHz, DMSO-d_6): δ 174.8, 172.4, 171.9, 171.6, 156.4, 145.3, 137.4, 128.8, 128.24, 128.16, 123.1, 65.9, 54.6, 49.6, 41.1, 34.7, 33.9, 32.1, 29.9, 28.4, 25.9, 24.3 ppm.

Nuclear Magnetic Resonance

Compounds **5** and **8** were dissolved in D_2O, and **4**, **6**, **7** and **9** in DMSO-d_6. ^{1}H and ^{13}C NMR spectra were obtained with Bruker Avance II 700 MHz NMR spectrometer. Chemical shifts are reported in parts per million (ppm) downfield from tetramethylsilane. Coupling constants are reported in Hertz (Hz).

Table A 2-0-1 Relative activity of MTG in the presence of CuAAC reagents.

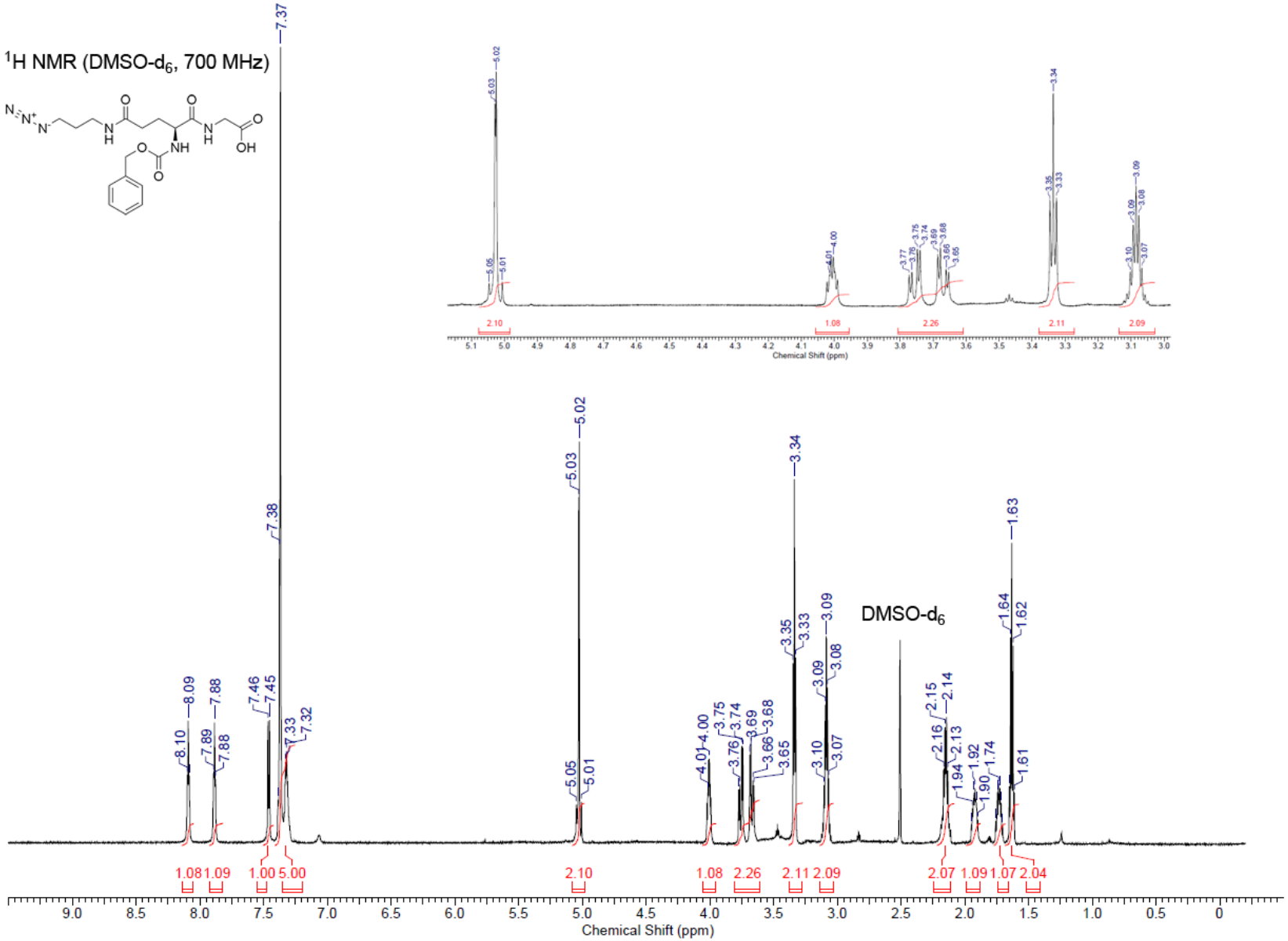
Components	Relative Activity (%)
200 mM Tris-Acetate pH 7 30 mM ZQG 100 mM hydroxylamine	100
200 mM Tris-Acetate pH 7 30 mM ZQG 100 mM hydroxylamine **+ 25 mM sodium ascorbate**	55
200 mM Tris-Acetate pH 7 30 mM ZQG 100 mM hydroxylamine **+ 2.5 mM $CuSO_4$**	22
200 mM Tris-Acetate pH 7 30 mM ZQG 100 mM hydroxylamine **+ 25 mM sodium ascorbate, 2.5 mM $CuSO_4$**	14
200 mM Tris-Acetate pH 7 30 mM ZQG 100 mM hydroxylamine **+ 25 mM sodium ascorbate, 2.5 mM $CuSO_4$, 10 mM glutathione**	276

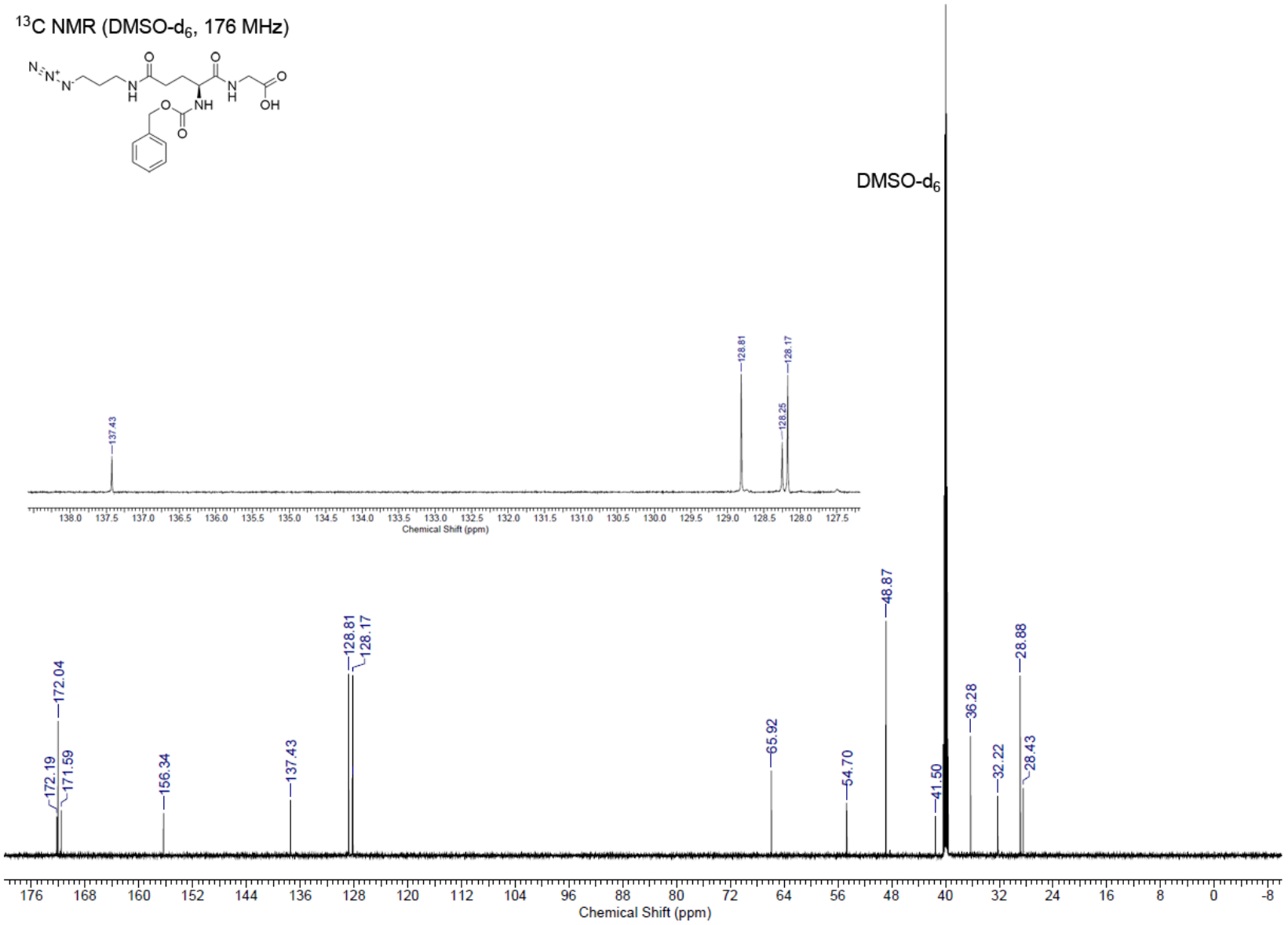
Activities were measured using the hydroxamate assay and were done in triplicate. A control reaction for each set of conditions was done containing all reaction components, excluding MTG, and subtracted from the reactions containing MTG.

Table A 2-0-2 Triazole product formation in the absence or presence of glutathione.

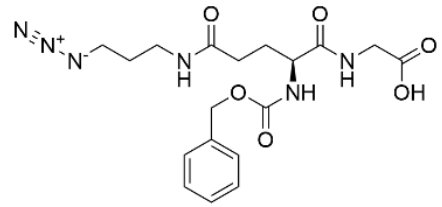
Sample	Product formation (%)
APA Control	82.7 %
APA Reactions	94.7 % ± 1.3
PRO Control	75.7 %
PRO Reactions	86.9 ± 6.3

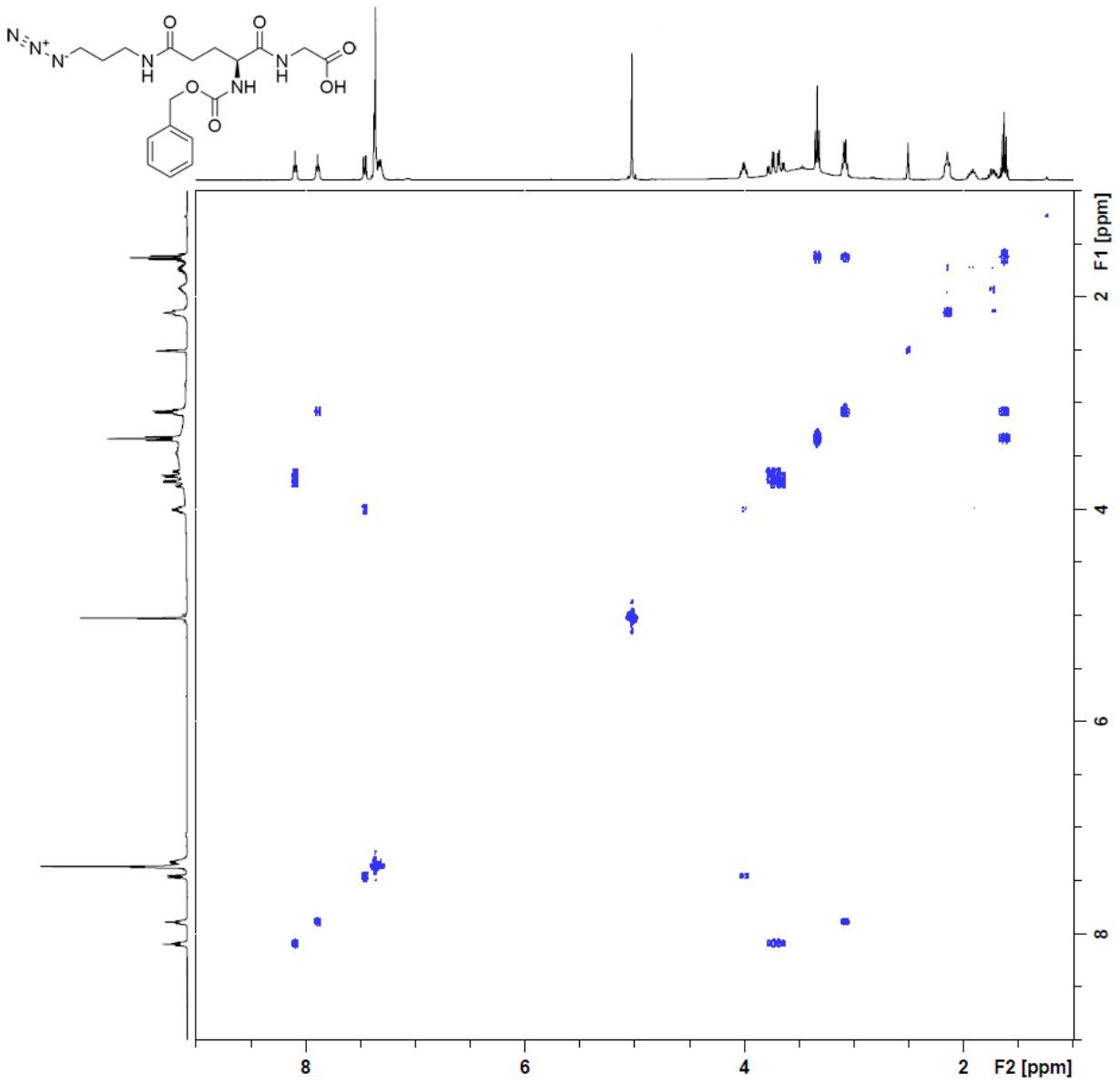
Reactions were prepared in 300 µL volumes. Each reaction contained 30 mM azidopropylamine (APA reactions) or 30 mM propargylamine (PRO reactions), 30 mM 4-pentynoic acid (APA reactions) or 30 mM 6-azidohexanoic acid (PRO reactions), 2.5 mM $CuSO_4$, 25 mM Na^+ ascorbate in 200 mM Tris acetate buffer, pH 7.5. The reactions contained 5 mM glutathione, and the control contained no glutathione; it was substituted by an equivalent volume of buffer. The reactions were done in triplicate, and the control was done as a single reaction. After 24hr incubation time at 37°C, aliquots were taken for LC-MS analysis, and the product formation quantified according to the appearance of the mass of the respective triazole product.

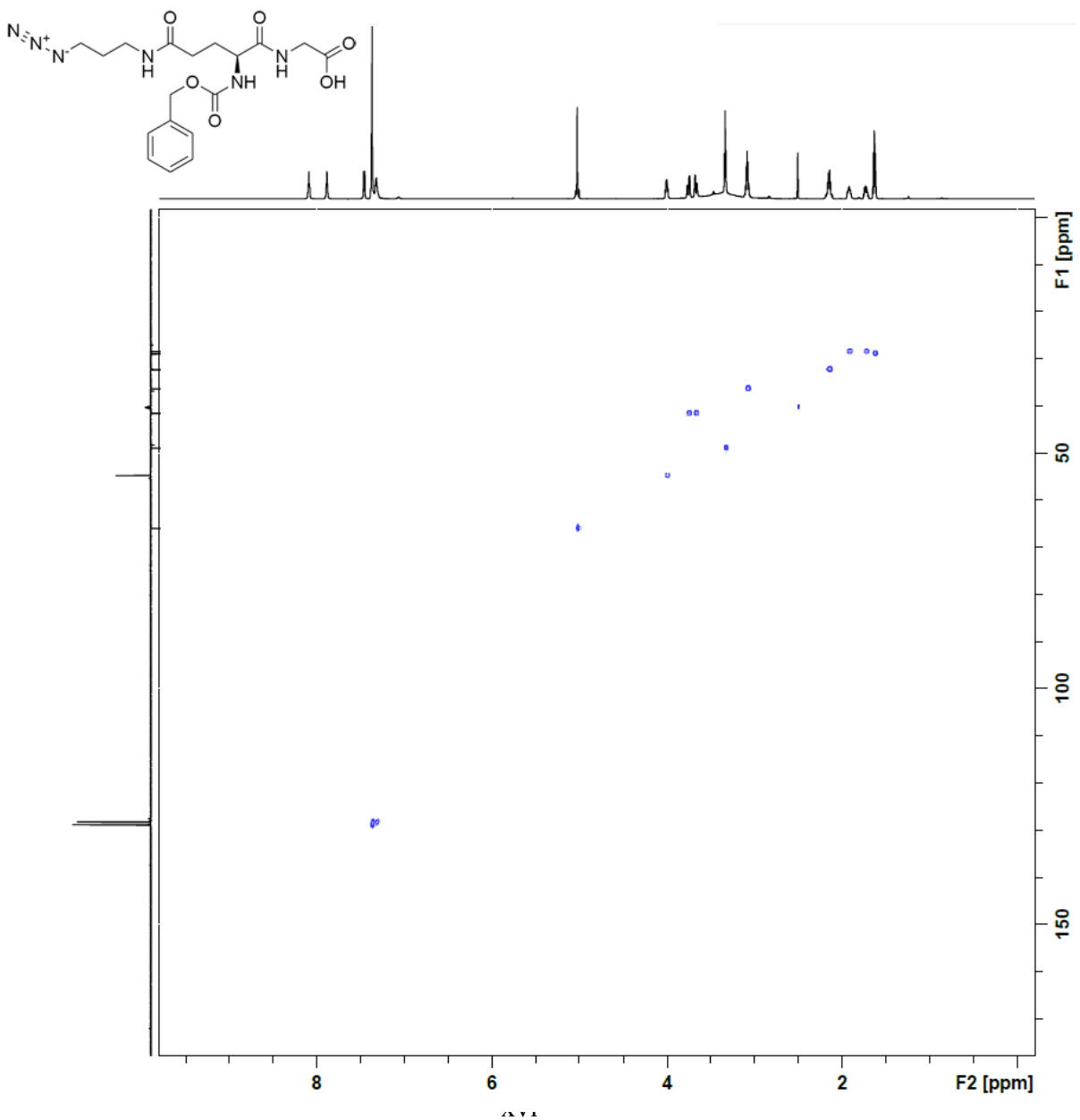

^{1}H NMR (DMSO-d$_6$, 700 MHz)
DMSO-d$_6$
Chemical Shift (ppm)

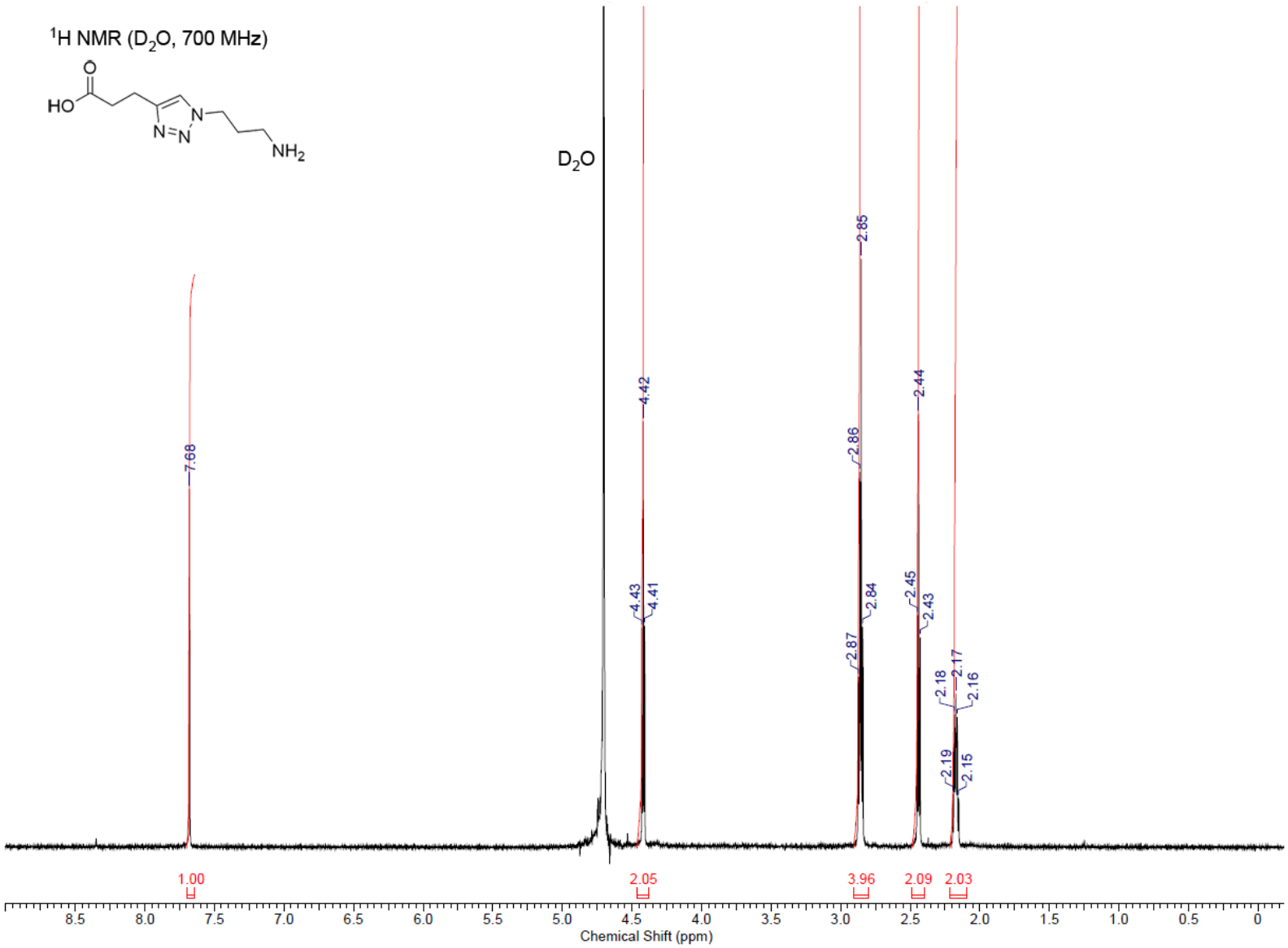

^{13}C NMR (DMSO-d$_6$, 176 MHz)
DMSO-d$_6$
172.19
172.04
171.59
156.34
137.43
128.81
128.17
65.92
54.70
48.87
41.50
36.28
32.22
28.88
28.43
Chemical Shift (ppm)

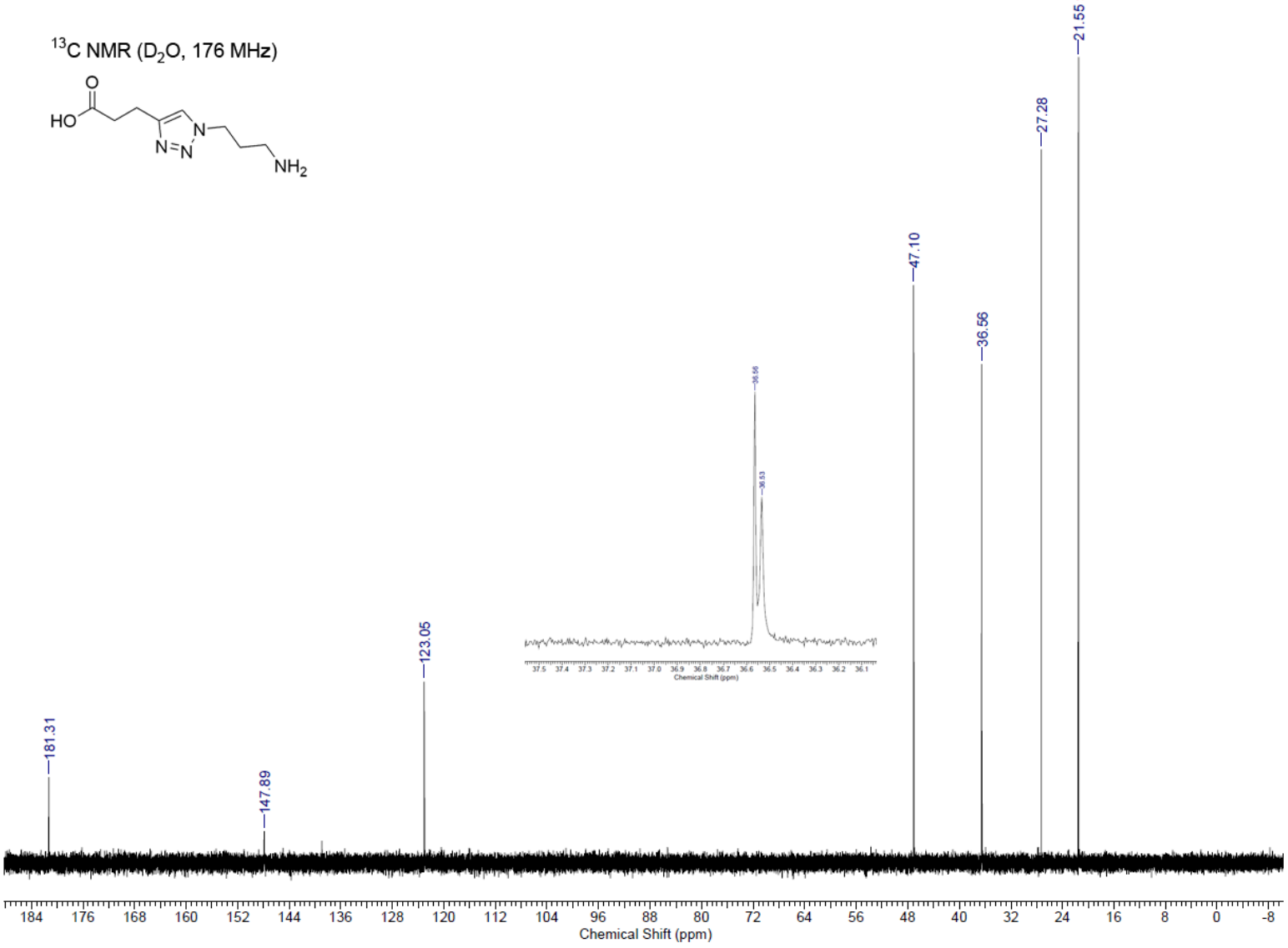
DEPT NMR (DMSO-d_6, 176 MHz)

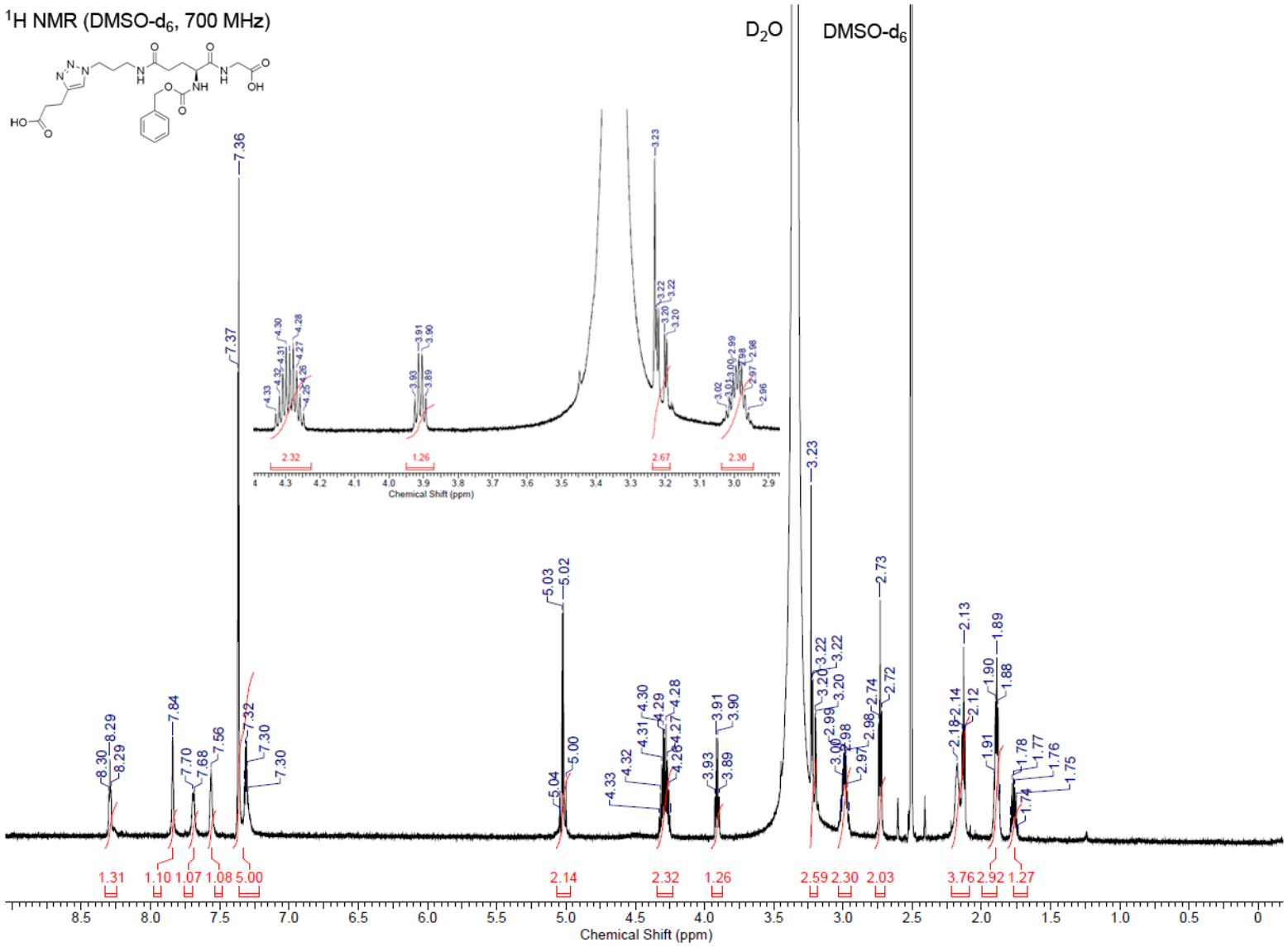


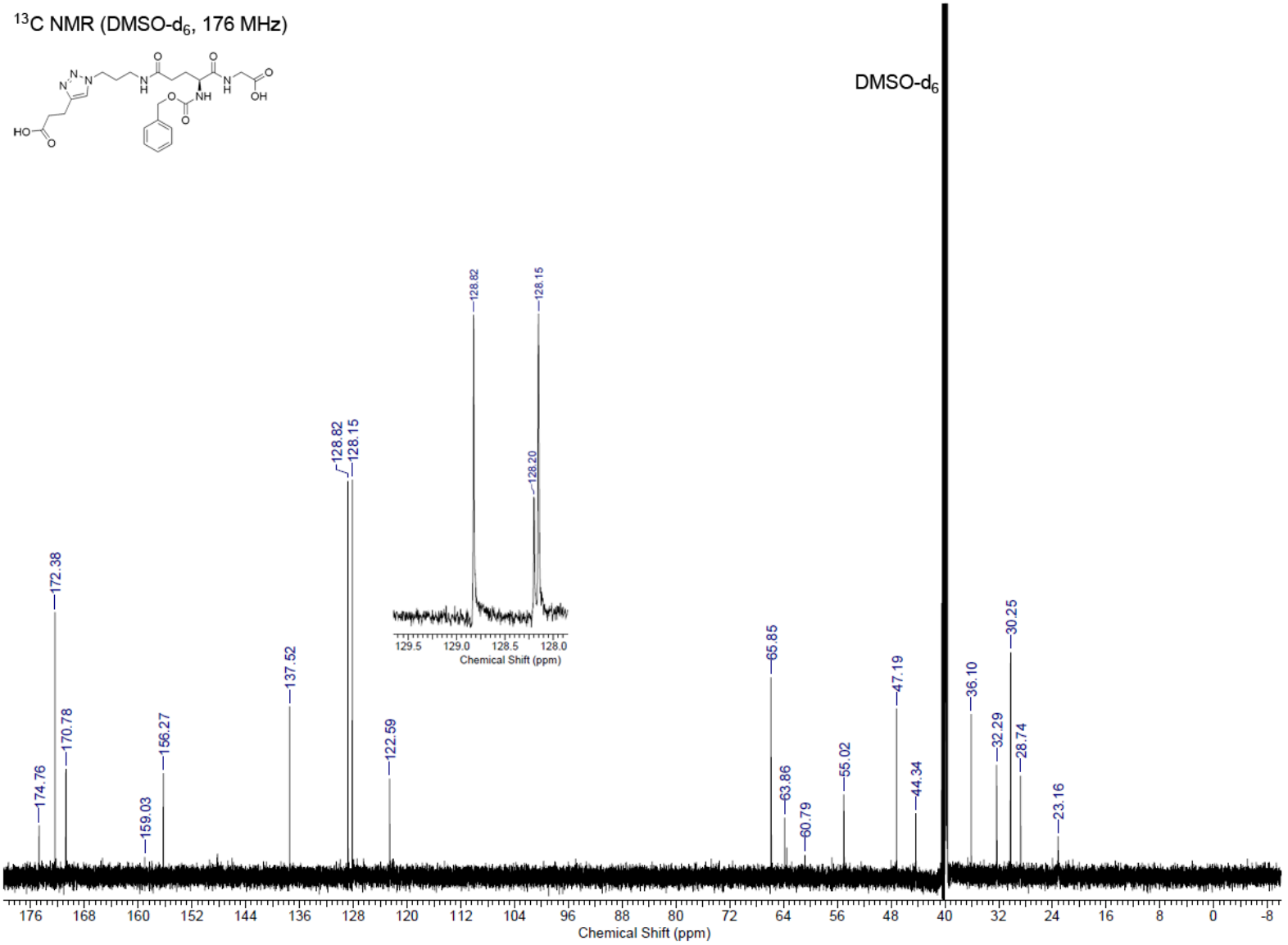

F1 [ppm]
2
4
6
8
F2 [ppm]
8
6
4
2


F1 [ppm]
50
100
150
8
6
4
2
F2 [ppm]

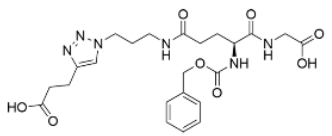

^{1}H NMR (D_2O, 700 MHz)
HO
O
N
N=N
NH$_2$
D_2O
7.68
4.43
4.42
4.41
2.87
2.86
2.85
2.84
2.45
2.44
2.43
2.19
2.18
2.17
2.16
2.15
1.00
2.05
3.96
2.09
2.03
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0
Chemical Shift (ppm)

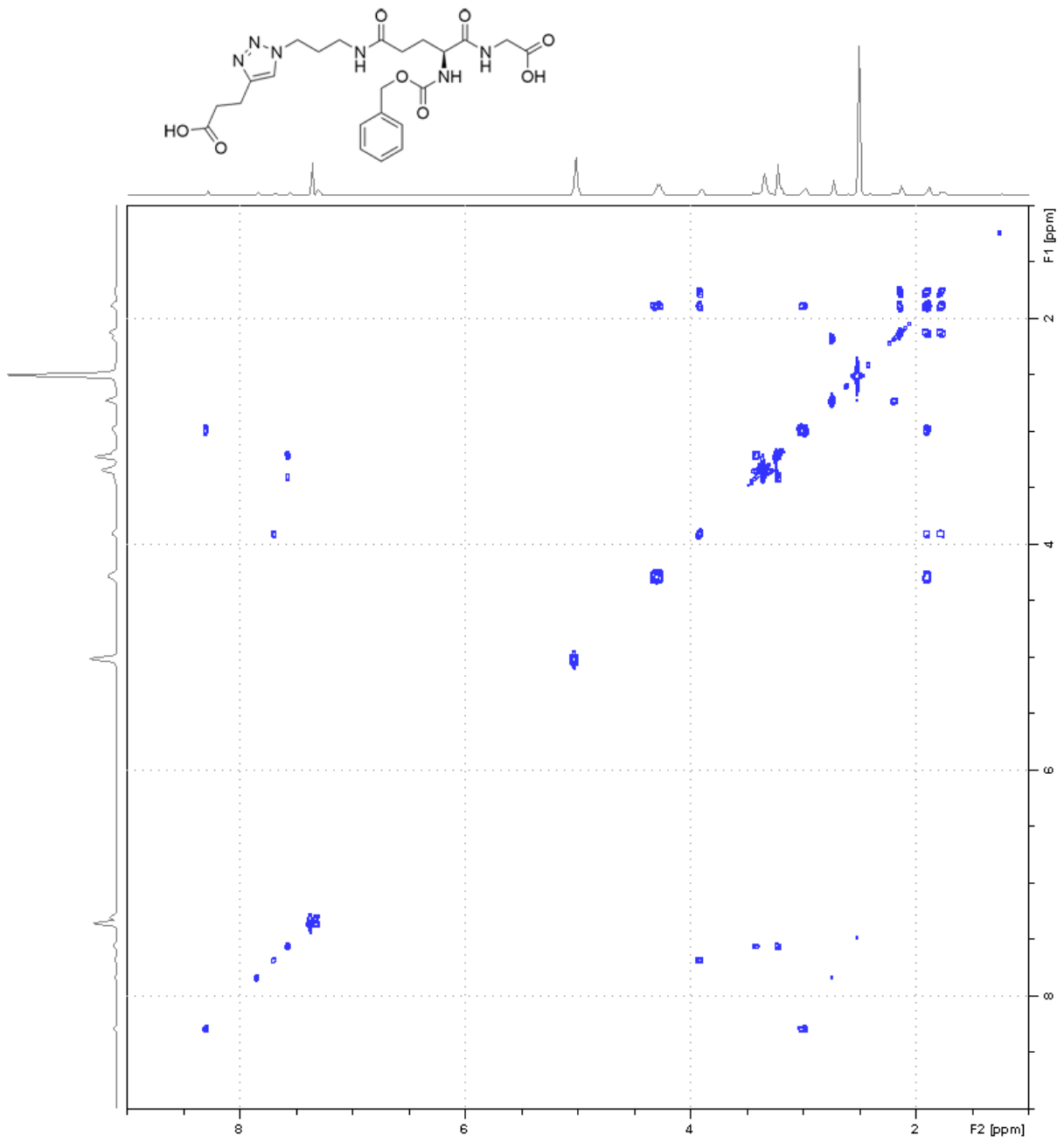

^{13}C NMR (D_2O, 176 MHz)
HO
O
N
N=N
NH2
181.31
147.89
123.05
47.10
36.56
27.28
21.55
Chemical Shift (ppm)

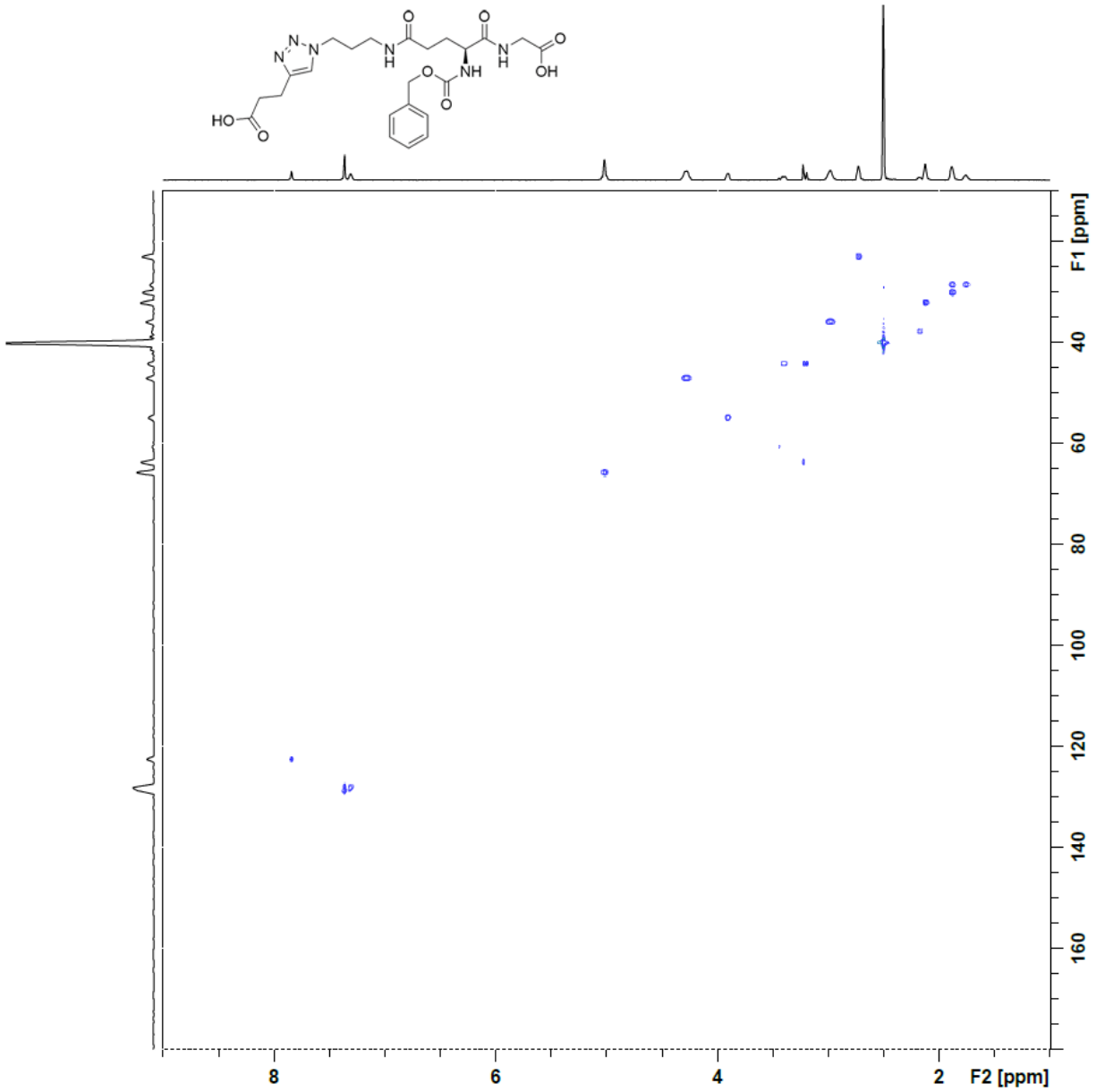

^{1}H NMR (DMSO-d$_6$, 700 MHz)
D$_2$O
DMSO-d$_6$
Chemical Shift (ppm)

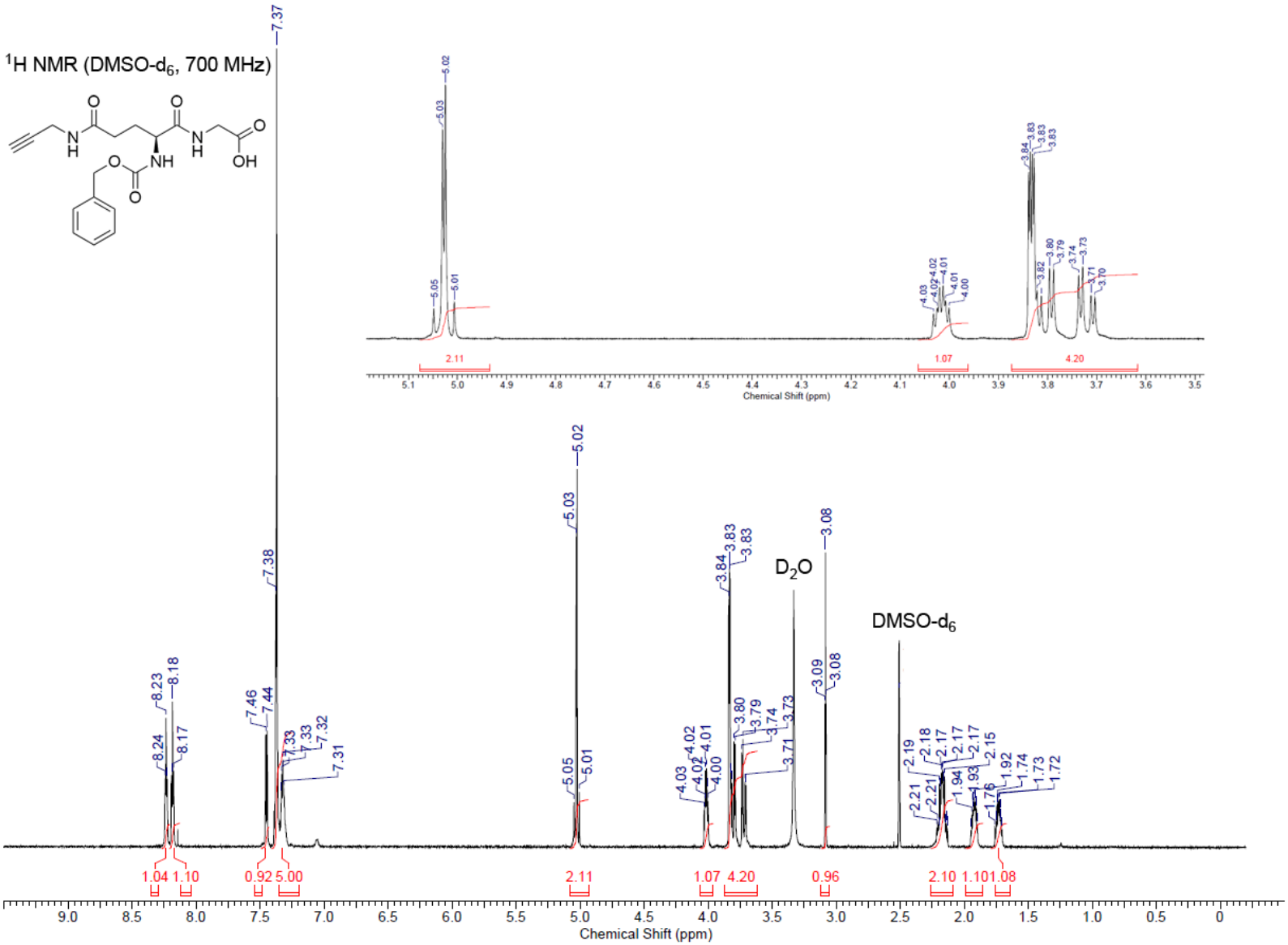

^{13}C NMR (DMSO-d$_6$, 176 MHz)
DMSO-d$_6$
128.82
128.15
128.20
129.5
129.0
128.5
128.0
Chemical Shift (ppm)
172.38
170.78
174.76
159.03
156.27
137.52
128.82
128.15
122.59
65.85
63.86
60.79
55.02
47.19
44.34
36.10
32.29
30.25
28.74
23.16
176
168
160
152
144
136
128
120
112
104
96
88
80
72
64
56
48
40
32
24
16
8
0
-8
Chemical Shift (ppm)

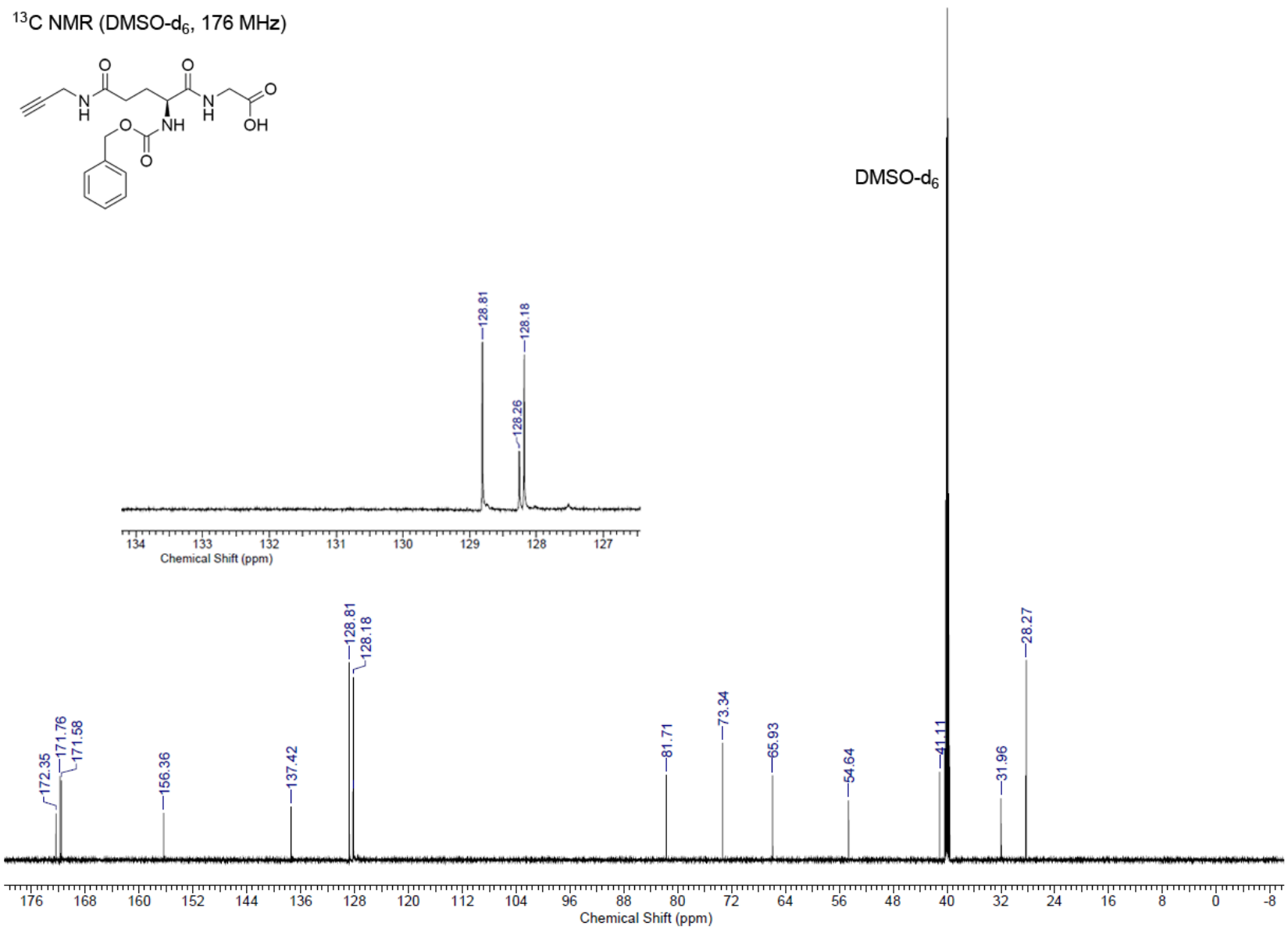
DEPT NMR (DMSO-d_6, 176 MHz)

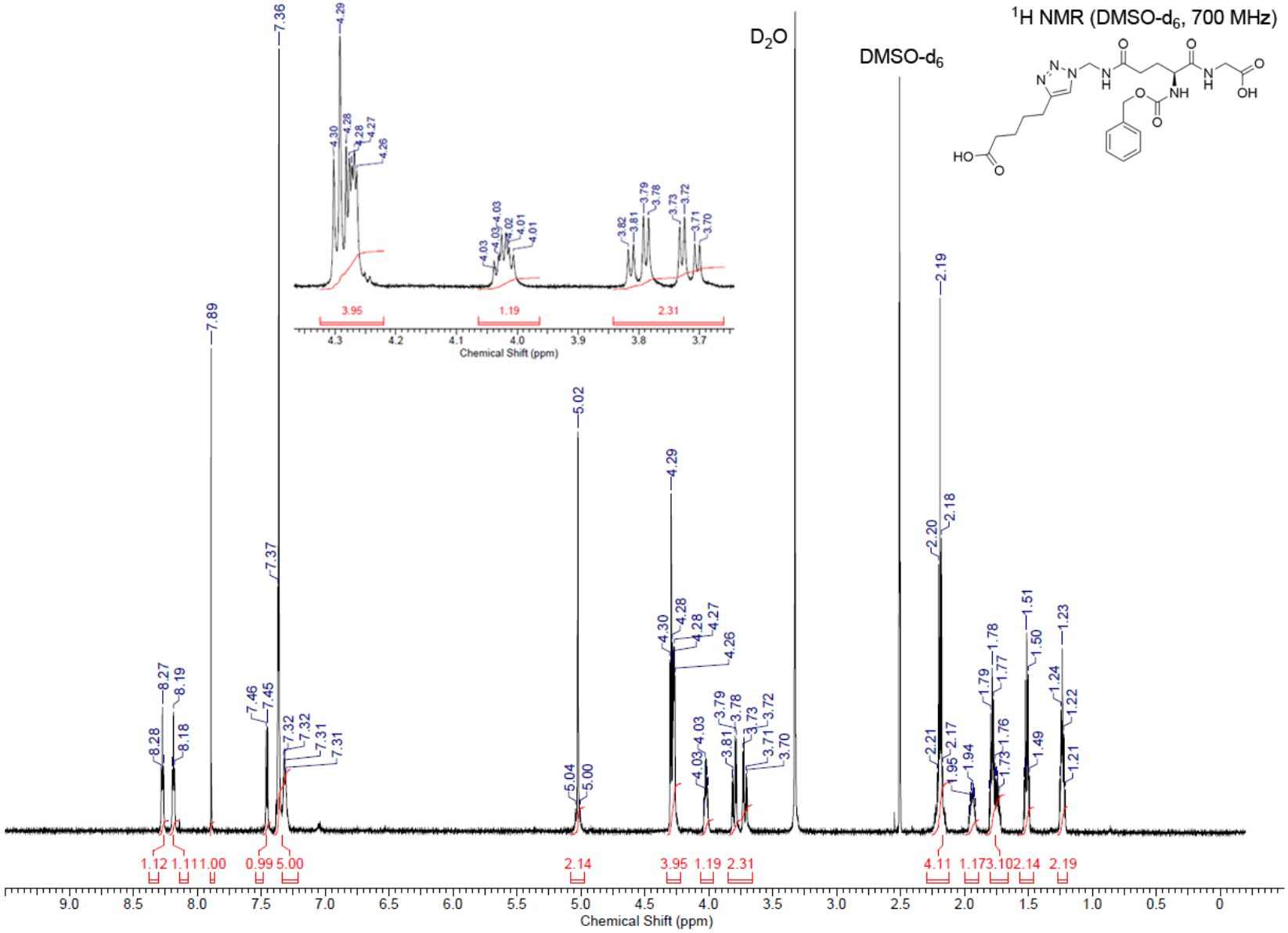


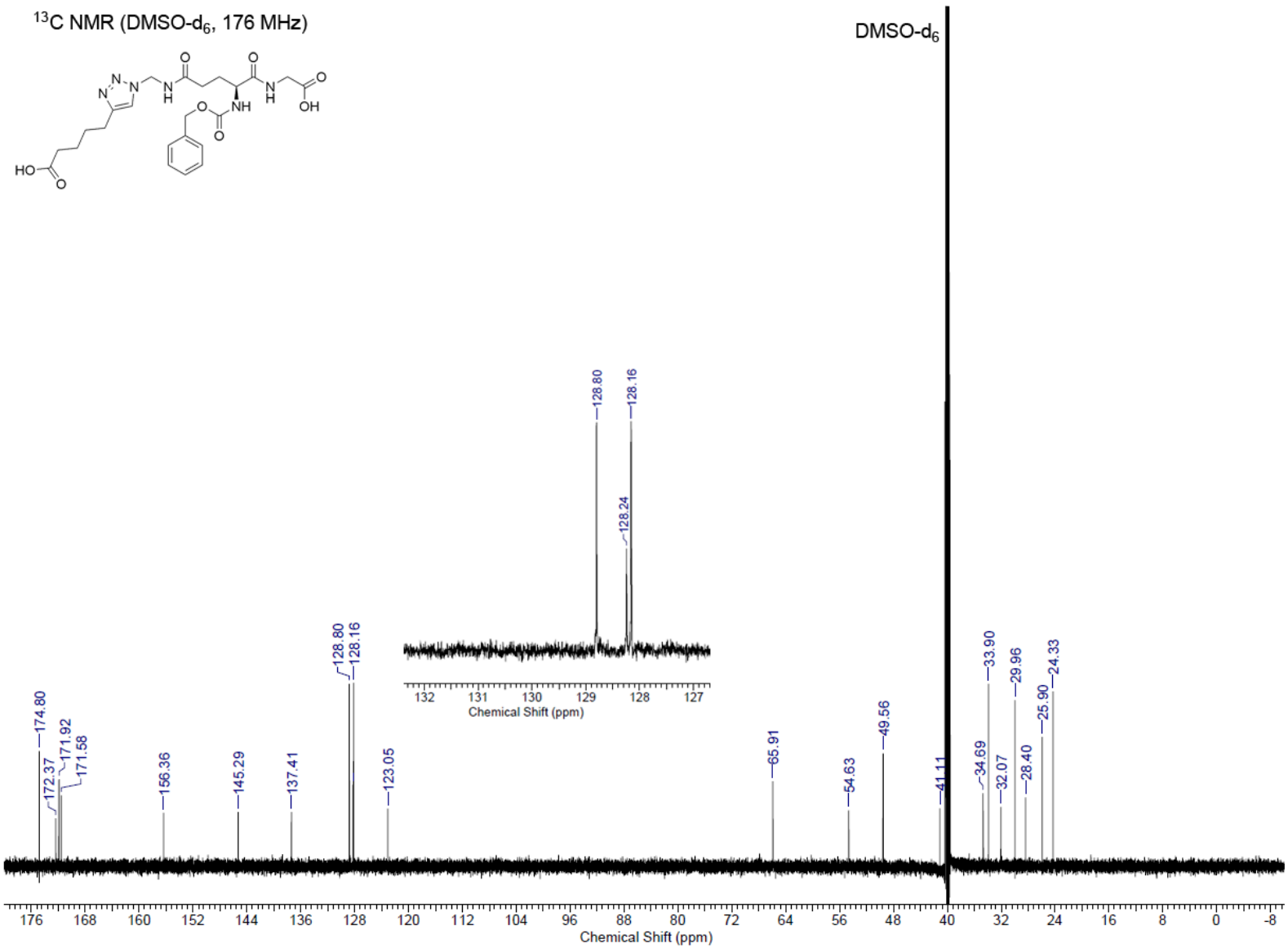

F1 [ppm]
F2 [ppm]
2
4
6
8

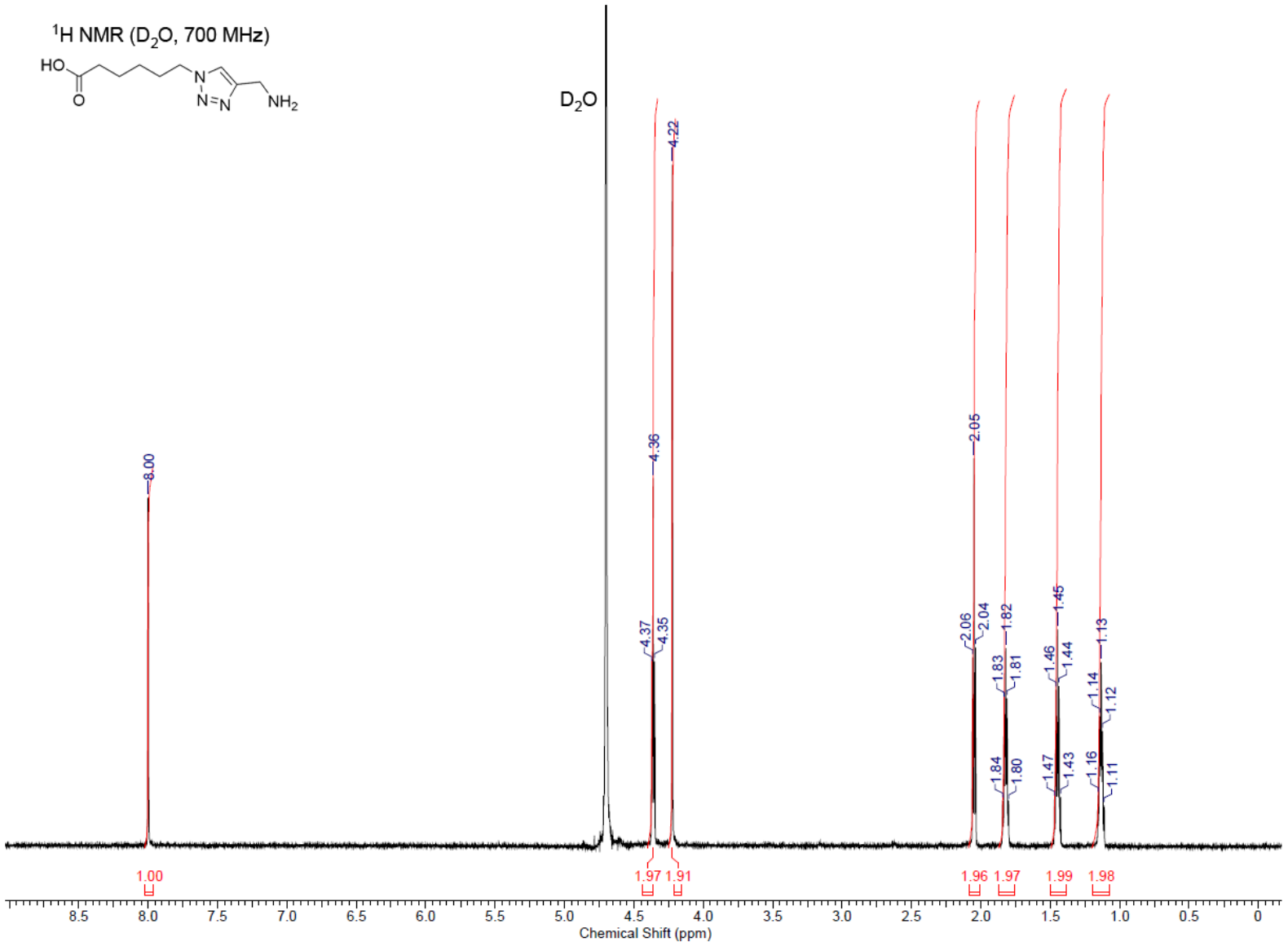

F1 [ppm]
40
60
80
100
120
140
160
8
6
4
2
F2 [ppm]

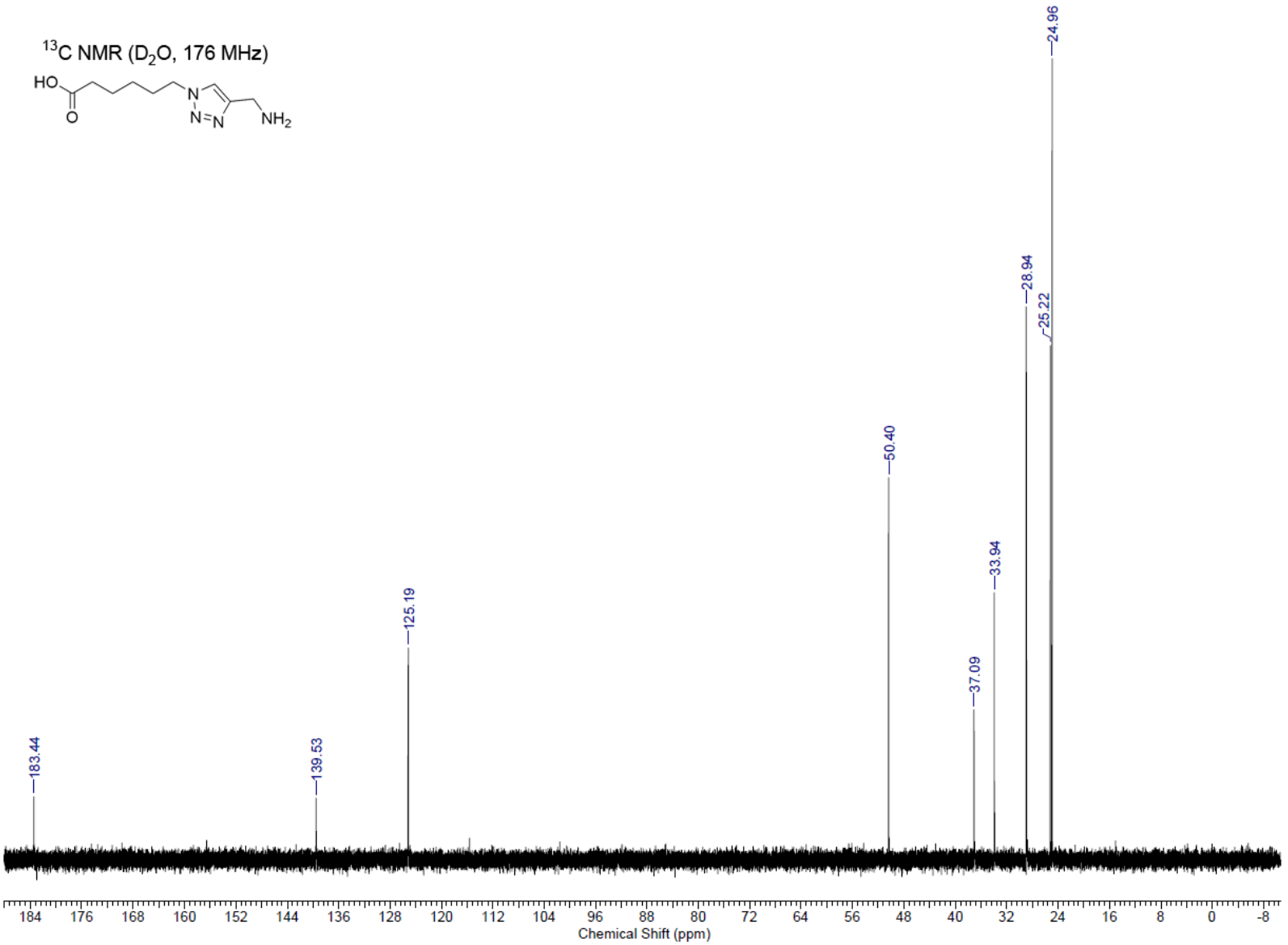

^{1}H NMR (DMSO-d$_6$, 700 MHz)
D$_2$O
DMSO-d$_6$
Chemical Shift (ppm)

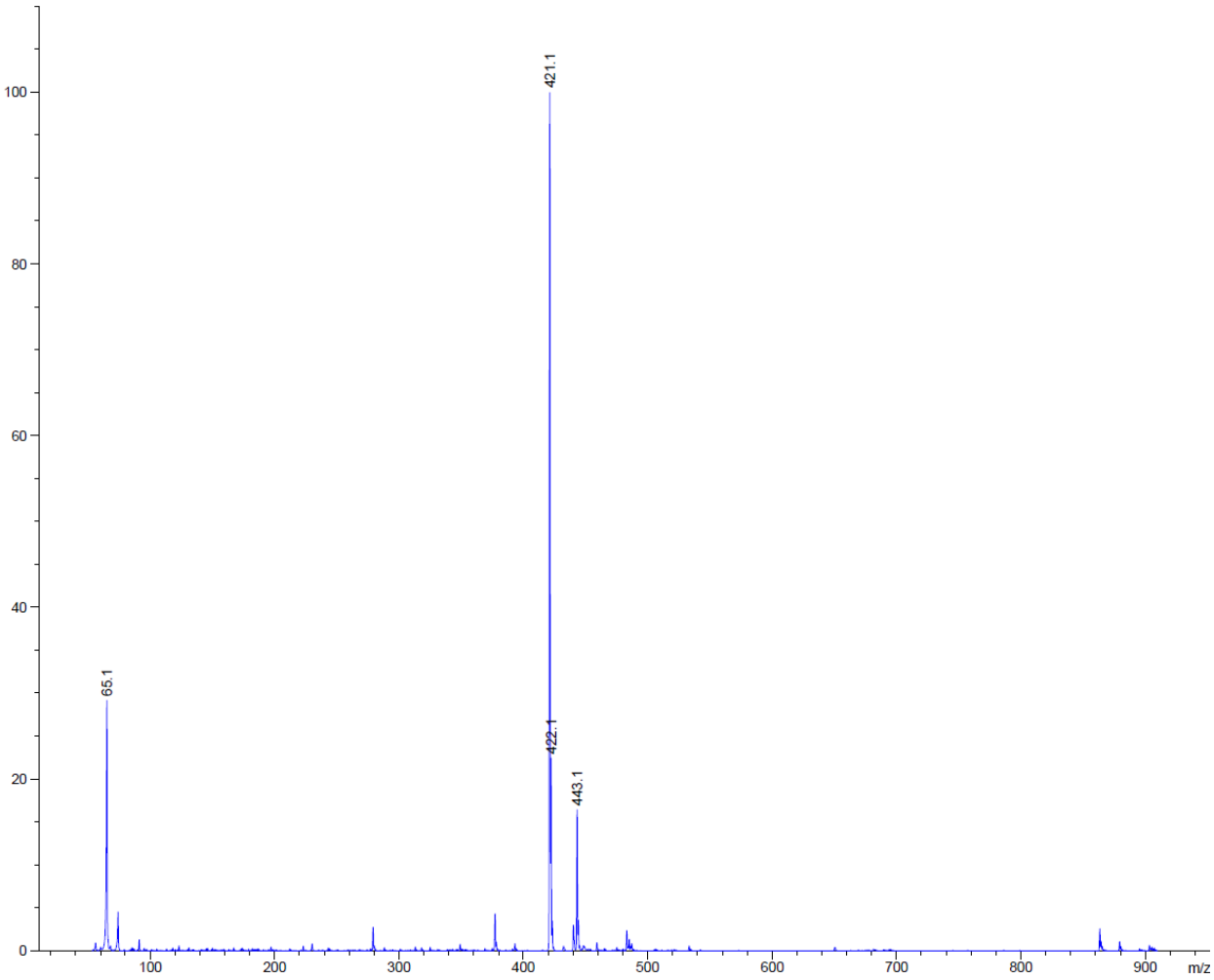

13C NMR (DMSO-d6, 176 MHz)
DMSO-d6
128.81
128.18
128.26
134
133
132
131
130
129
128
127
Chemical Shift (ppm)
172.35
171.76
171.58
156.36
137.42
128.81
128.18
81.71
73.34
65.93
54.64
41.11
31.96
28.27
176
168
160
152
144
136
128
120
112
104
96
88
80
72
64
56
48
40
32
24
16
8
0
-8
Chemical Shift (ppm)

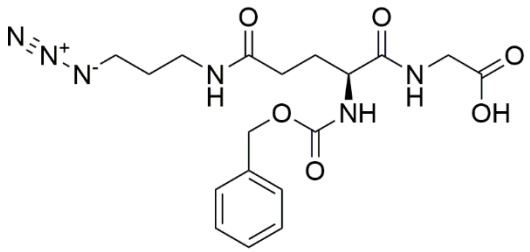

^{1}H NMR (DMSO-d$_6$, 700 MHz)
D$_2$O
DMSO-d$_6$
Chemical Shift (ppm)


^{13}C NMR (DMSO-d$_6$, 176 MHz)
DMSO-d$_6$
174.80
172.37
171.92
171.58
156.36
145.29
137.41
128.80
128.16
123.05
65.91
54.63
49.56
41.11
34.69
33.90
32.07
29.96
28.40
25.90
24.33
128.80
128.16
128.24
Chemical Shift (ppm)
Chemical Shift (ppm)

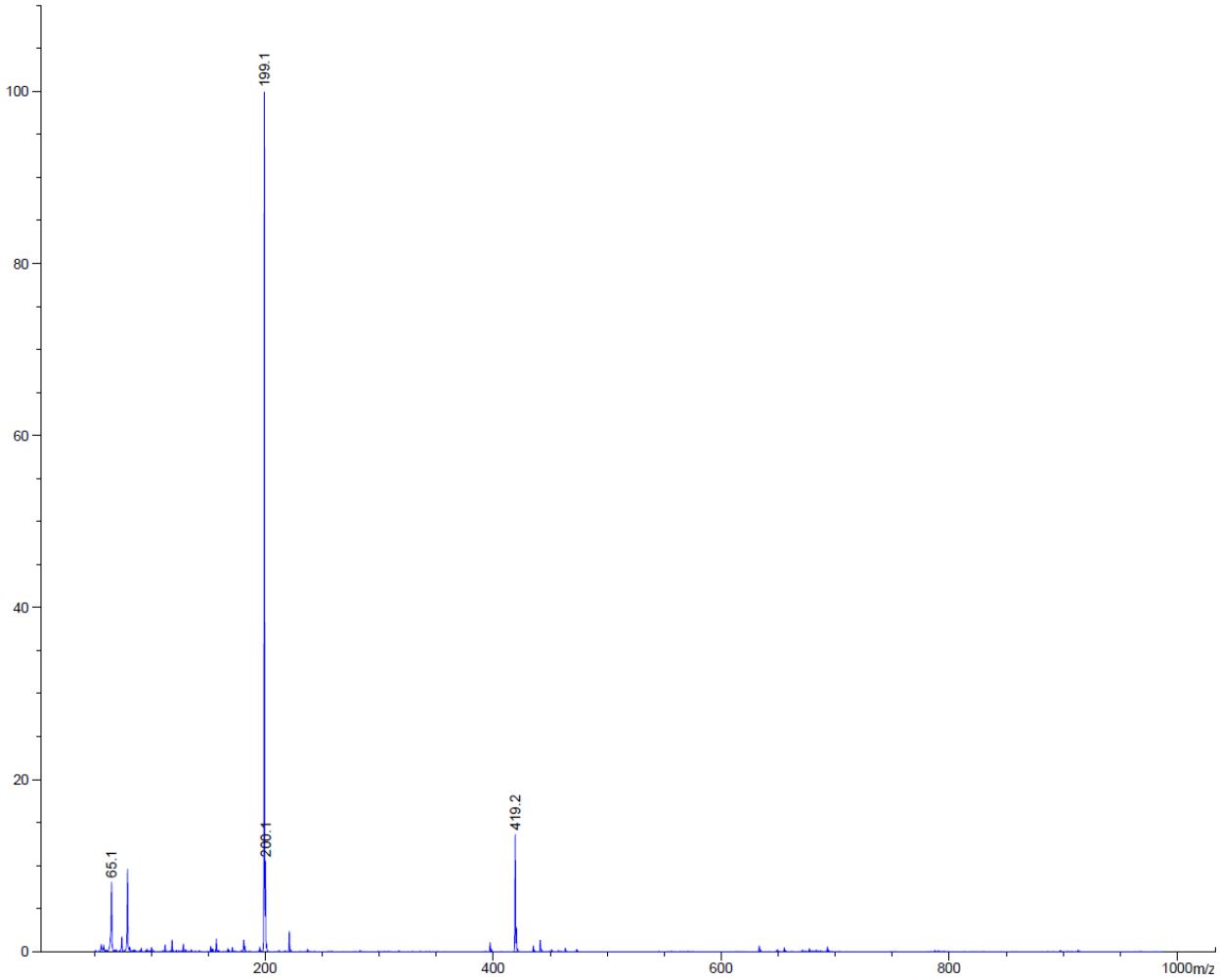

^{1}H NMR (D_2O, 700 MHz)
HO
O
N
N=N
NH$_2$
D_2O
8.00
4.36
4.22
2.05
4.37
4.35
2.06
2.04
1.82
1.45
1.83
1.81
1.46
1.44
1.13
1.14
1.12
1.84
1.80
1.47
1.43
1.16
1.11
1.00
1.97
1.91
1.96
1.97
1.99
1.98
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0
Chemical Shift (ppm)

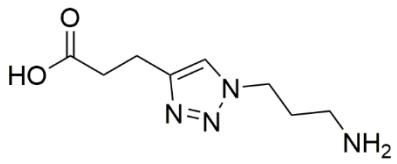

^{13}C NMR (D_2O, 176 MHz)
HO
O
N
N=N
NH2
24.96
28.94
25.22
50.40
33.94
37.09
125.19
139.53
183.44
184 176 168 160 152 144 136 128 120 112 104 96 88 80 72 64 56 48 40 32 24 16 8 0 -8
Chemical Shift (ppm)



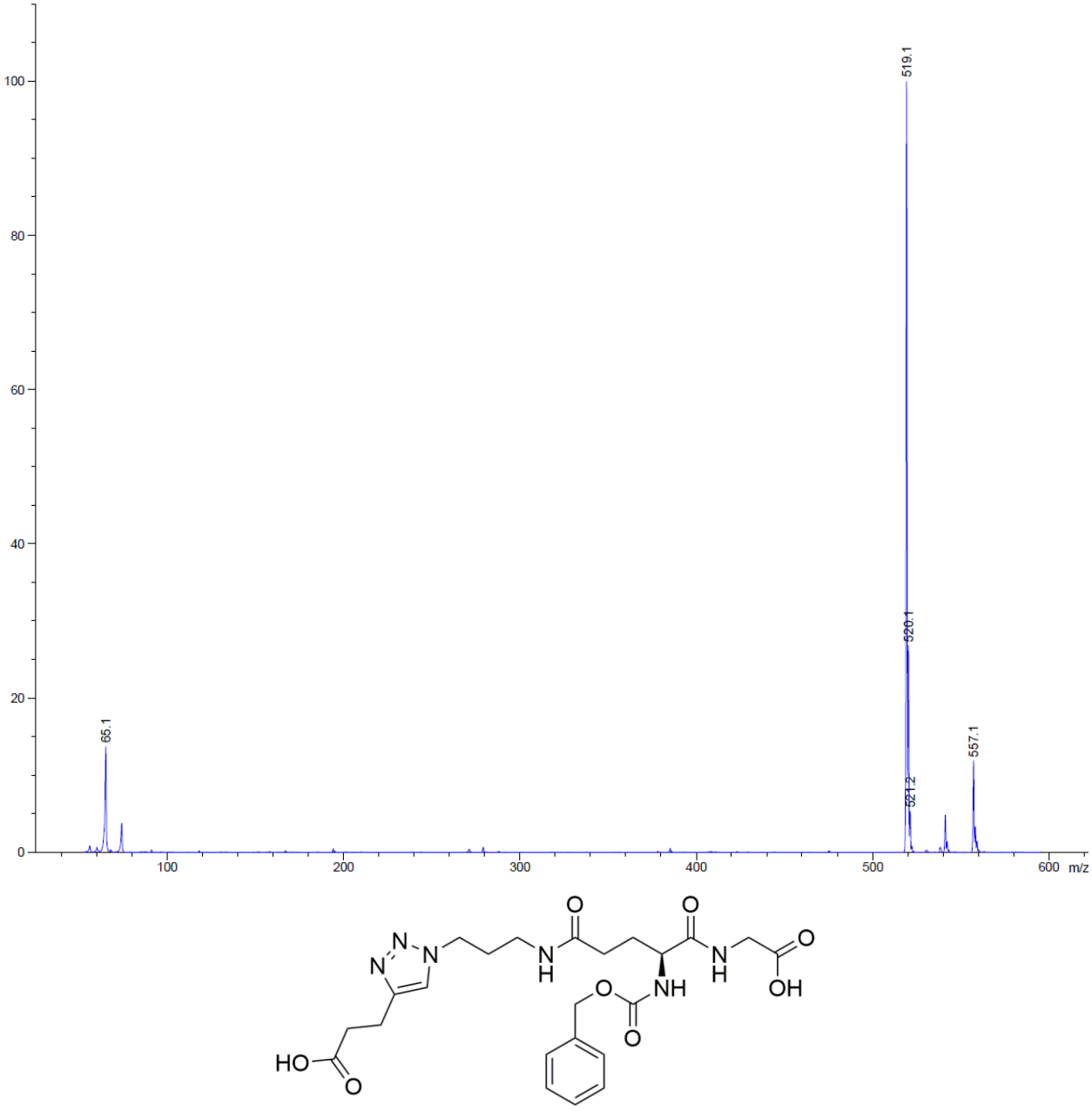


LRMS (ESI) m/z calculated for $C_{18}H_{24}N_6O_6$ $[M+H]^+$, 421.18; found: 421.1.

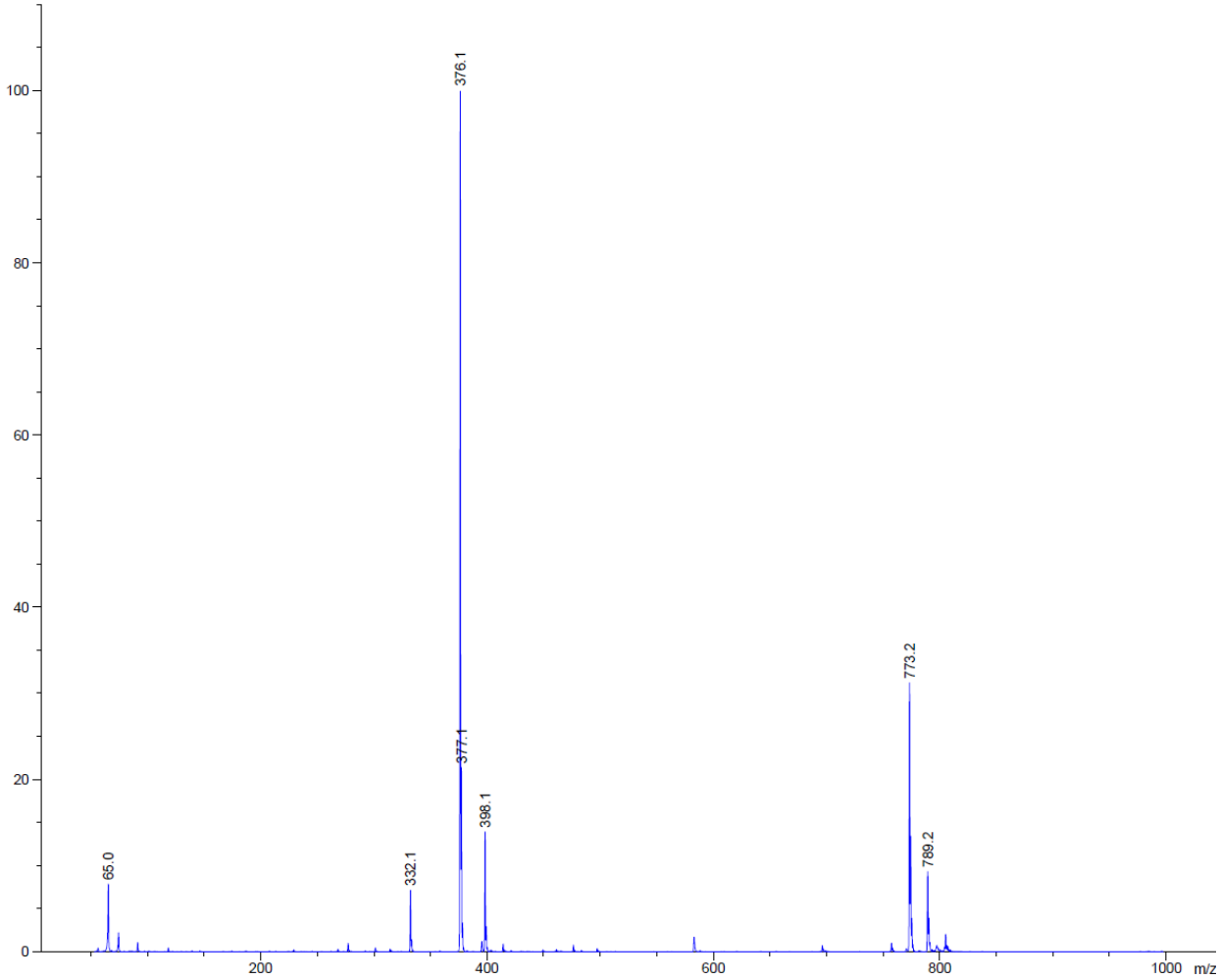


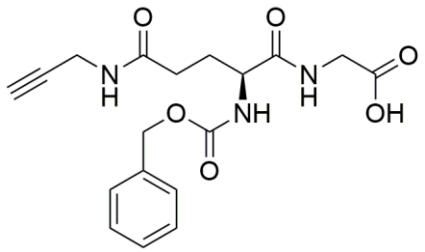


LRMS (ESI) m/z calculated for $C_8H_{14}N_4O_2$ $[M+H]^+$, 199.12; found: 199.1.

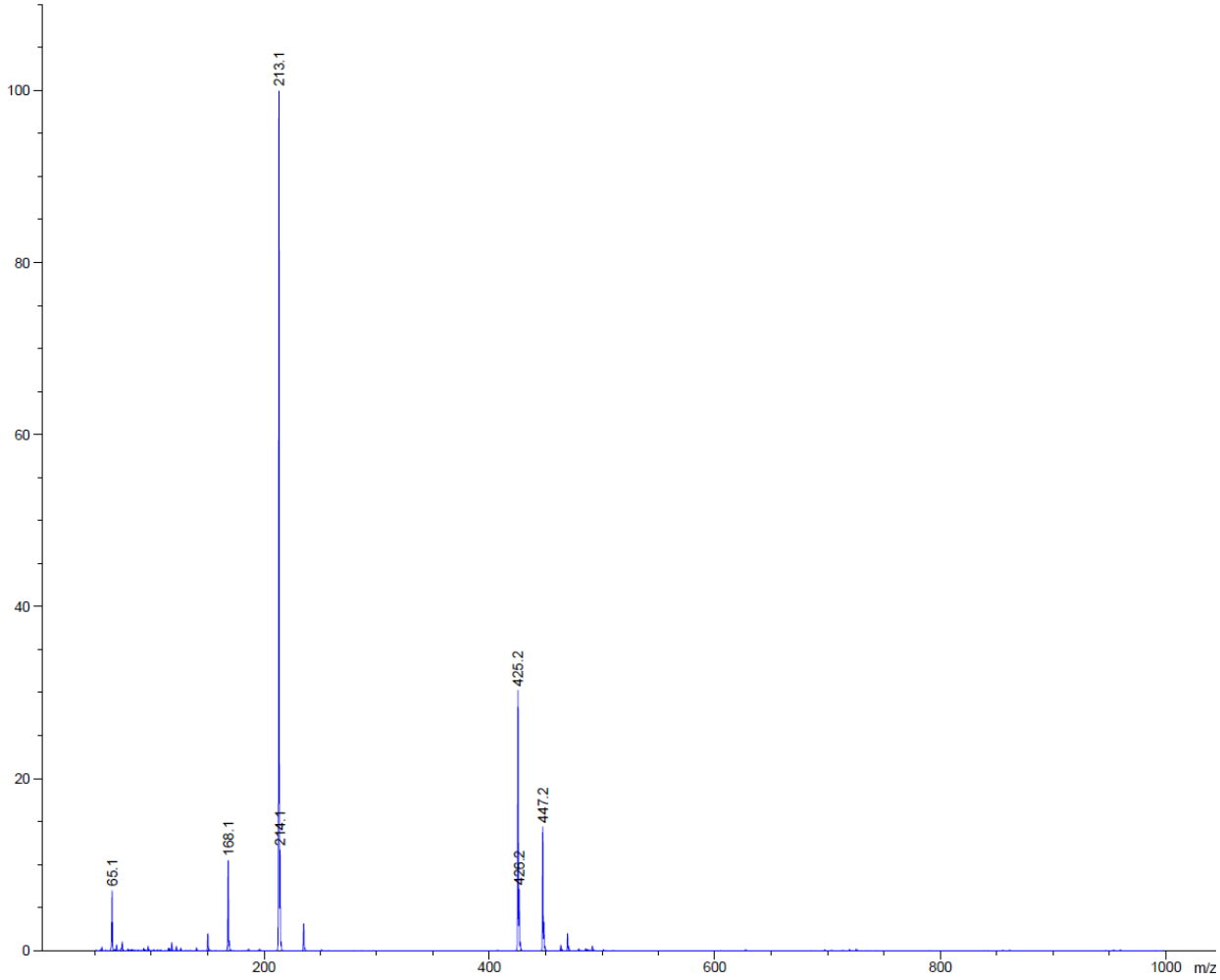


LRMS (ESI) m/z calculated for $C_{23}H_{30}N_6O_8$ $[M+H]^+$, 519.22; found: 519.1.

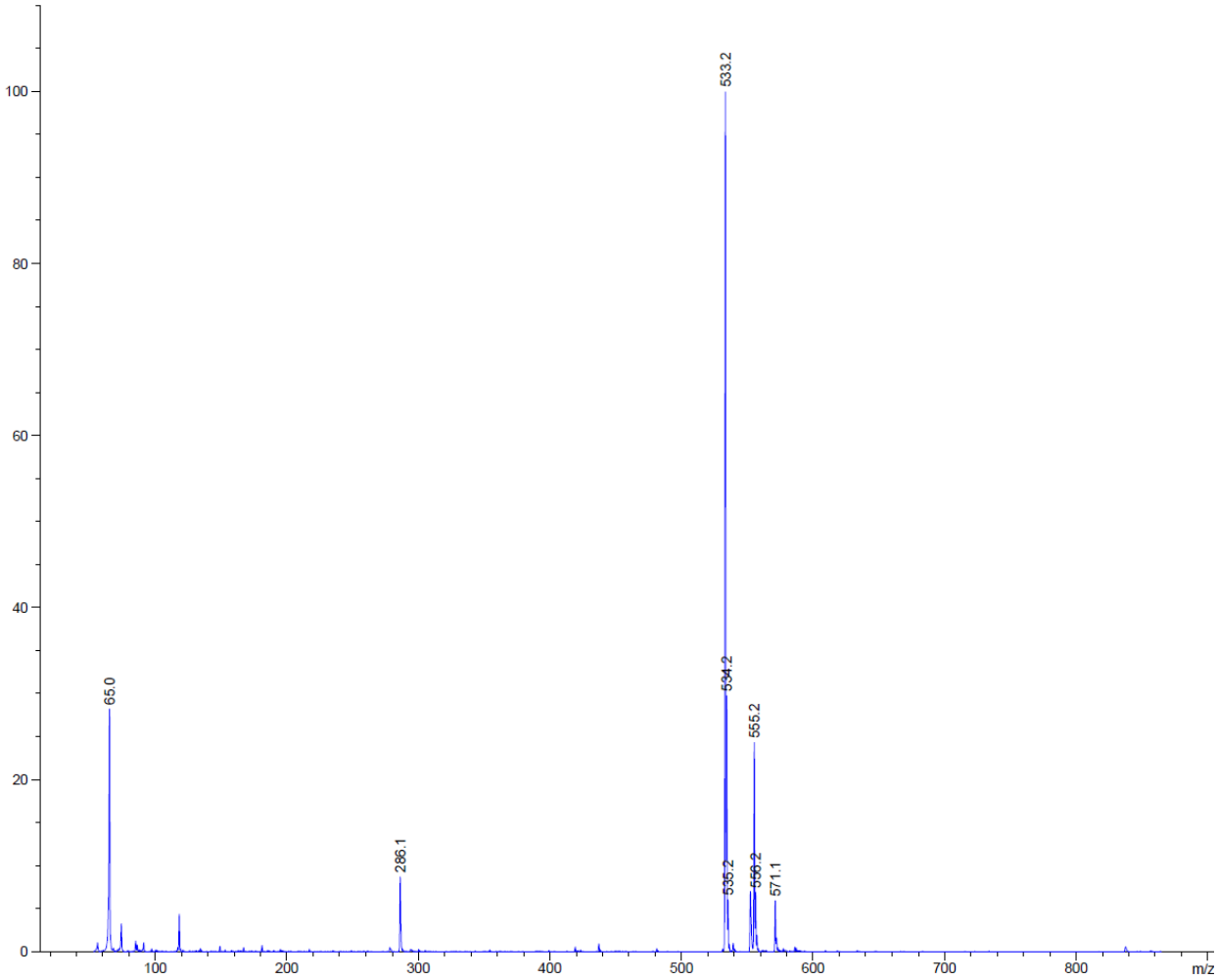


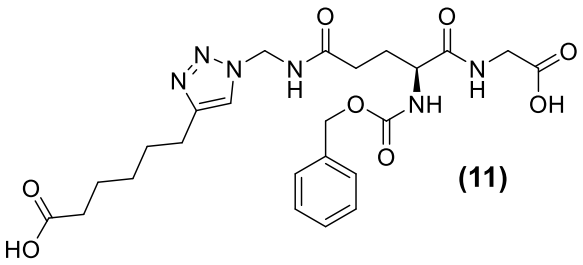


LRMS (ESI) m/z calculated for $C_{18}H_{21}N_4O_2$ $[M+H]^+$, 376.15; found: 376.1.



LRMS (ESI) m/z calculated for $C_4H_{26}N_4O_2$ $[M+H]^+$, 213.13; found: 213.1.





(11)

LRMS (ESI) m/z calculated for $C_{24}H_{32}N_6O_8$ $[M+H]^+$, 533.23; found: 533.2.

Annex 3 – Chapter 5 Supplemental information

Materials

The plasmid pDJ1-3 was kindly provided by Professor M. Pietzsch (Martin-Luther-Universität, Halle-Wittenberg, Germany). pDJ1-3 encodes the proenzyme of MTG from *S. mobaraensis* inserted between the *Nde*I and *Xho*I restriction sites of the vector pET20b. Deionized water (18Ω) was used for all experiments. HPLC solvents were of analytical grade, and products used for the expression and purification of MTG were of biological grade.

Other chemicals used were purchased from the suppliers listed below. Carboxybenzyl–L-glutaminyl–glycine (Z-Gln-Gly, or ZQG) was from Peptide Institute (Osaka, Japan). Glutathione (reduced) and thiamine were from Bioshop (Burlington, Canada). Dimethyl sulfoxide (99.7%) was purchased from Fisher Scientific (Ontatio, Canada). Dibenzylcyclooctyne-PEG4-Amine, methyltetrazine-PEG4-Amine, *trans*-cyclooctene-Cy5 were purchased from Click Chemistry Tools (Arizona, USA). Formic acid (98 % purity) was from Fluka Analytical (St. Louis, USA). α-lactalbumin from bovine milk (calcium saturated), cadaverine dihydrochloride (98%), and 2-(Diphenylphosphino)terephthalic acid 1-methyl 4-pentafluorophenyl diester were purchased from Sigma Aldrich (St. Louis, MO, USA). Sulfo-cyanine5 azide (Cy5-azide) was purchased from Lumiprobe (Hallandale Beach, FL, USA).

MTG Expression and Purification

MTG was expressed and purified as previously described. Briefly, a 5-mL starter culture of *E. coli* BL21 (DE3) containing the plasmid pET20b-MTG, which expresses a *C*-terminally 6-His-tagged version of MTG, was propagated overnight at 37°C in ZYP-0.8G medium and shaking at 240 rpm. It was used to inoculate 500 mL of auto-inducing ZYP-5052 medium. After 2h of incubation at 37°C and 240 rpm, the temperature was reduced to 22°C overnight. Cells were collected by centrifugation and resuspended in 50 mM sodium phosphate buffer, 300 mM NaCl, pH 7.5. The cells were lysed using a Constant Systems cell disruptor set at 37 kPSI and cooled to 4°C. After further centrifugation to remove insoluble cellular matter, the inactive form of MTG was incubated with trypsin (1 mg/mL solution, 1:9 ratio of trypsin to MTG, v/v) for

the purpose of cleaving its pro-sequence. Activated MTG was purified using a 5-mL His-trap nickel-nitrilotriacetic acid (Ni-NTA) column (GE Healthcare) equilibrated in 50 mM phosphate buffer, pH 7.5, with 300 mM NaCl, and eluted with an imidazole gradient (0 – 250 mM) using an Åtka FPLC (GE Healthcare). After purification, active MTG was dialyzed against 50 mM sodium phosphate buffer, 300 mM NaCl, pH 7.5. The average yield was 25 mg of activated MTG per litre of culture, with ~ 85% purity as estimated by SDS-PAGE and revelation with Coomassie blue stain. Aliquots were snap frozen and stored at -80°C in 15% glycerol.

MTG Activity Assay

The activity of purified MTG was quantified using the hydroxamate assay. Briefly, MTG was incubated with 30 mM Z-Gln-Gly and 100 mM hydroxamate at 37°C for 10 min. A concentrated acidic ferric chloride solution (2.0 M $FeCl_3 \cdot 6\ H_2O$, 0.3 M trichloroacetic acid, 0.8 M HCl) was used to quench the reaction, which was then vortexed and left to stand at room temperature for 10 min. The resulting iron complex was quantified by its absorbance at 525 nm. One unit (U) of MTG produces 1 μmol of L-glutamic acid and γ-monohydroxamate per min at 37°C.

hDHFR Expression and Purification

Recombinant human chromosomal DHFR (hDHFR) was overexpressed in Escherichia coli BL21 (DE3) and purified as previously described, with the following minor modifications. The expression was done in Terrific Broth. The purification buffer was 10 mM potassium phosphate pH 8.0 for DEAE, and 50 mM potassium phosphate pH 7.5.

Conjugation Assays

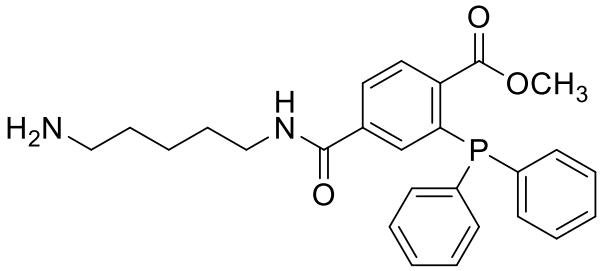
Protein substrate (50 μM; α-Lactalbumin, hDHFR), amine, and complementary azide or *trans*-cyclooctene Cy5 substrates (100 μM) were combined with 5 mM glutathione in 100 mM sodium phosphate buffer, pH 7.5. The conjugation reaction catalyzed by MTG was initiated by the addition of 1 U/mL of MTG, where control reactions had an equivalent volume of buffer

added. A second control reaction was performed in addition, in which MTG was added, but the amine substrate was omitted. The final volume of each reaction was 300 µL and all were incubated at 37°C for 24h. 50 µL aliquots were taken after 10 min, 1h, 4h, 8h, to which 3 µL of formic acid was added to quench the reaction. The remaining volume for the 24h aliquot was quenched by adding 6 µL of formic acid. Samples were stored at 4°C for short-term storage, or at -20°C for long-term storage if necessary.

For reactions prepared in a subsequent fashion, the same protocol as above applies, with the exception that the Cy5 substrates were added only after aliquots had been quenched, and then incubated at 4°C overnight prior to analysis.

Aliquots were resolved using tricine SDS-PAGE. The fluorescent bands were visualized and recorded using a Bio Rad ChemiDoc™ MP Imaging System using an excitation filter of 625 nm with a 30 nm bandpass. The gels were then stained with Coomassie brilliant blue to reveal the protein.

Synthesis and Purification of 1



A solution (1 mL) containing DMSO, equimolar (12.5 mM) cadaverine dihydrochloride and 2-(Diphenylphosphino)terephthalic acid 1-methyl 4-pentafluorophenyl diester was prepared. After mixing, the solution was left to sit for 48h in the dark at 4°C. The reaction progress was checked using HPLC-MS. 5 µL was transferred to 495 µL of 18.2 mΩ deionized water containing 0.1% formic acid, and mixed by pipetting. Diluted sample was injected (5 µL) onto a Synergi 4-µm polar-RP 80 Å, 150 × 4.60 mm LC column (Phenomenex), using an Agilent 1200 series HPLC apparatus and eluted with a 12 minute 5-70% MeOH/H_2O gradient. Masses were detected under positive ionization with a single quadrupole mass detector. The reaction

had not gone to completion, and so additional cadaverine (to a total of 25 mM) was added to push the consumption of the phosphine substrate. The reaction was incubated once more for 48h at 4°C, and injected again on the LC-MS. When reaction progress was observed to be sufficient, the reaction mixture was injected onto preparative LC-MS. A 200 µL volume of the DFFT-CAD solution was injected onto a Synergi polar-RP 80 Å, 100 × 21.20 mm AXIA-packed column (Phenomenex), using a Waters 1525 HPLC and Waters 3100 single quadrupole mass detector. Elution was performed with a 12-minute 10-70% MeOH/H_2O gradient. Compounds with the mass corresponding to the expected product were collected and pooled. Methanol was evaporated, the remaining solution was lyophilized to yield the purified product.

Supplementary Tricine SDS-PAGE

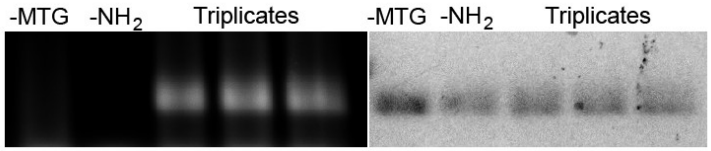


Figure A 3-1 One-pot chemoenzymatic labelling of α-LA using the SPAAC in the absence of glutathione.

After 24h, the reactions were quenched as usual with formic acid, and resolved using denaturing tricine SDS-PAGE. The gels were excited with a Cy5 imaging filter, photographed (left panel), and then stained with Coomassie blue to reveal the presence of α-LA (right panel).

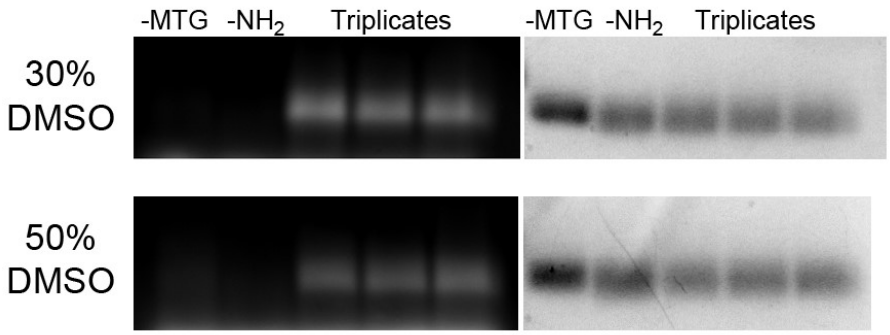


Figure A 3-2 One-pot chemoenzymatic labelling of α-LA using the SPAAC.

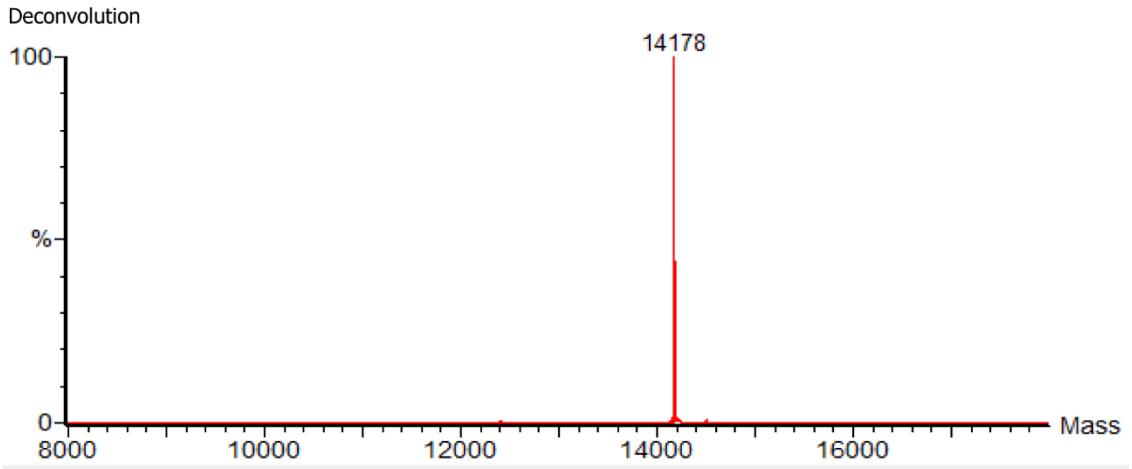
The reactions were performed in presence of either 30% or 50% DMSO. After 24h, the reactions were quenched as usual with formic acid, and resolved using denaturing tricine SDS-PAGE. The gels were excited with a Cy5 imaging filter, photographed (left panel), and then stained with Coomassie blue to reveal the presence of α-LA (right panel).

High Resolution MS Spectra of Conjugated Protein Products

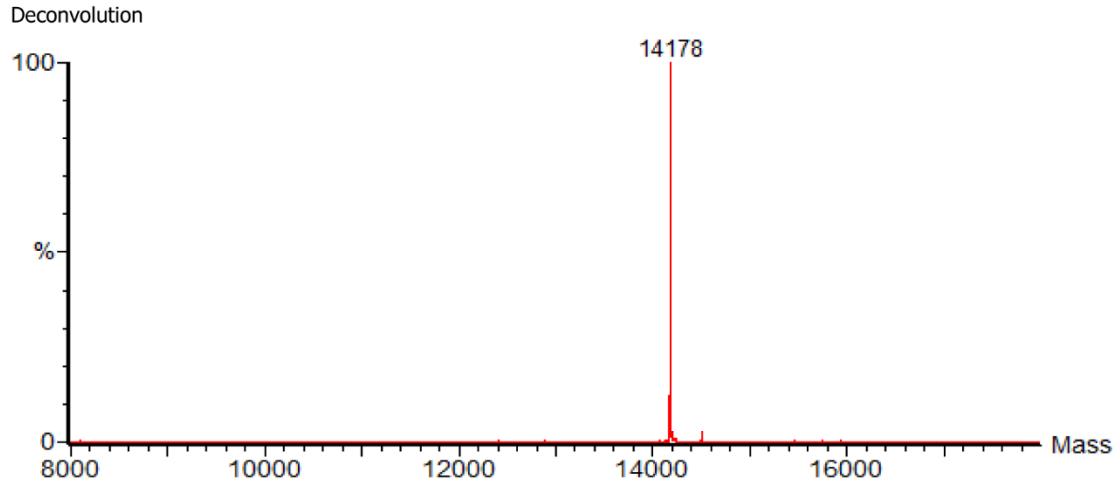
Each page displays a set of reactions, for which there are two spectra: the first is the control in which MTG is absent, preventing conjugation. The second is the reaction. For the latter, two potential products can be observed: the protein conjugated with its amine (compounds **1-3** in Figure 1 in the main text), and/or the protein conjugated with its amine after it has been clicked with its corresponding probe (compounds **4** or **5** in Figure 1). The masses for these products are calculated in the captions.

In some spectra, many weaker peaks are observed. They are observed in our control reaction as well as our reaction sample, and for this reason, were not listed; we listed only the most prominent peaks.

➔ ***SPAAC on α-Lactalbumin***

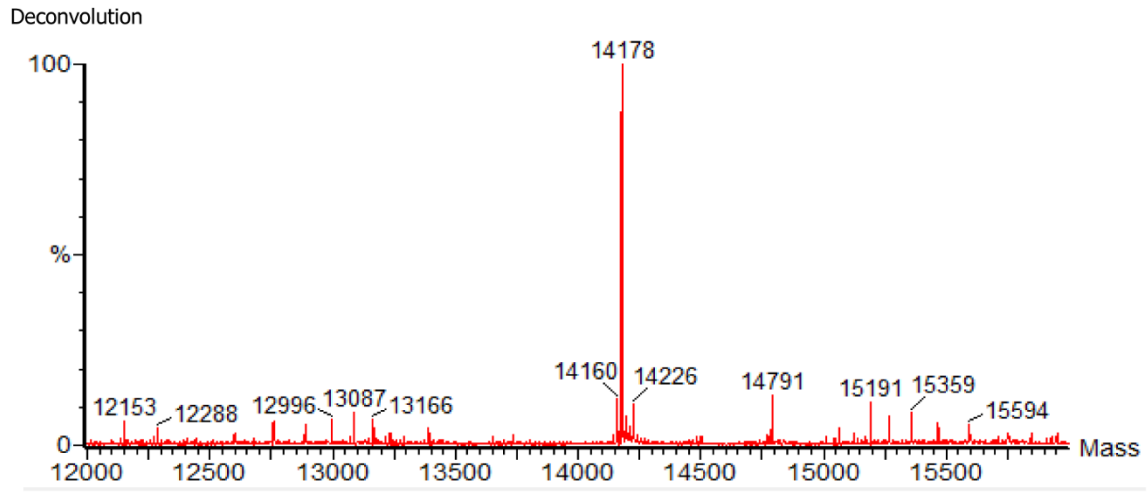


HRMS (ESI) m/z calculated for control containing α-Lactalbumin + 2, no MTG: 14178; found 14178.

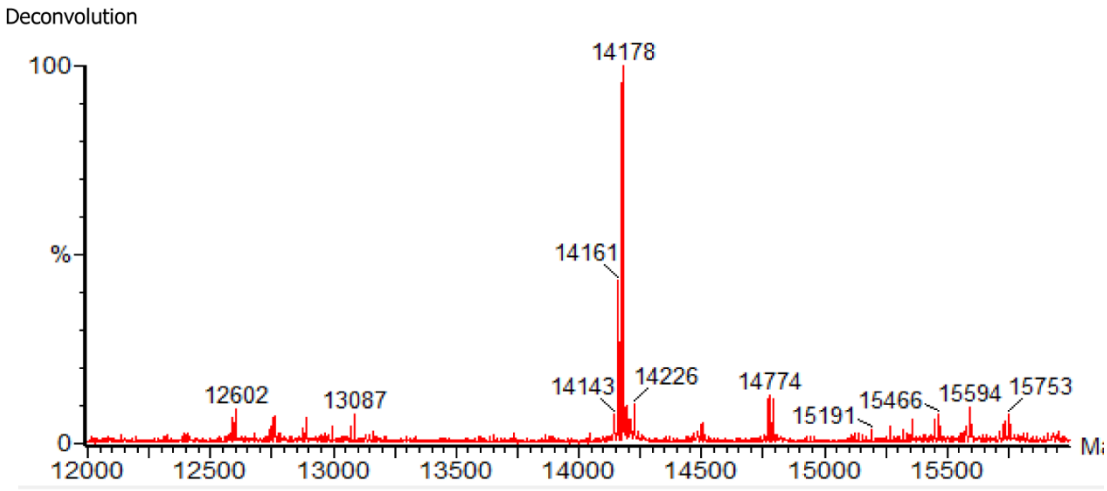


HRMS (ESI) m/z calculated for reaction containing α-Lactalbumin + 2: 14685; m/z calculated for reaction containing α-Lactalbumin + 2 + 4: 15433. Found: 14178.

➔ ***SPAAC on α-Lactalbumin in 30% DMSO***

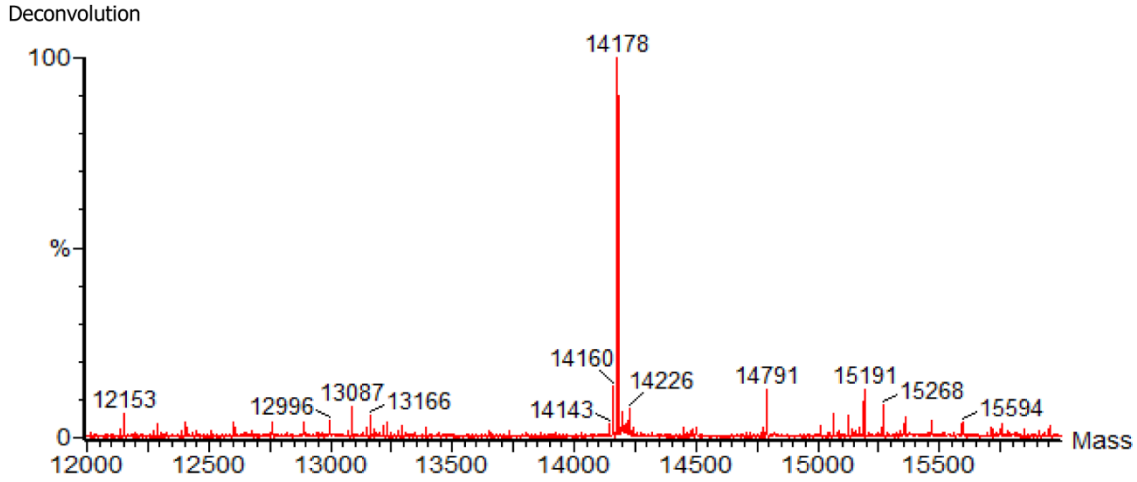


HRMS (ESI) m/z calculated for control containing 30% DMSO, α-Lactalbumin + 2, no MTG: 14178; found 14178.

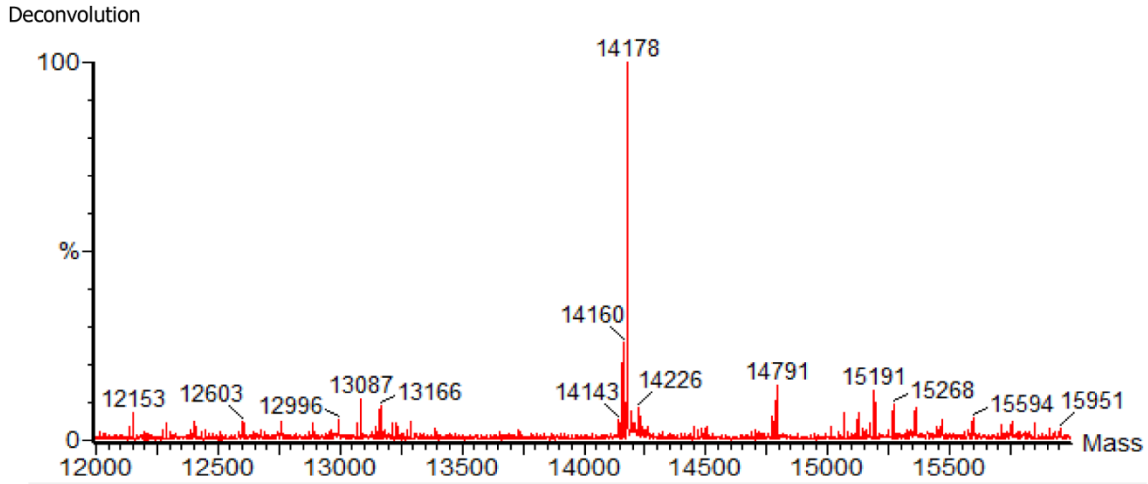


HRMS (ESI) m/z calculated for reaction containing 30% DMSO, α -Lactalbumin + 2: 14684; m/z calculated for reaction containing α-Lactalbumin + 2 + 4: 15433. Found 14178.

➔ ***SPAAC on α-Lactalbumin in 50% DMSO***

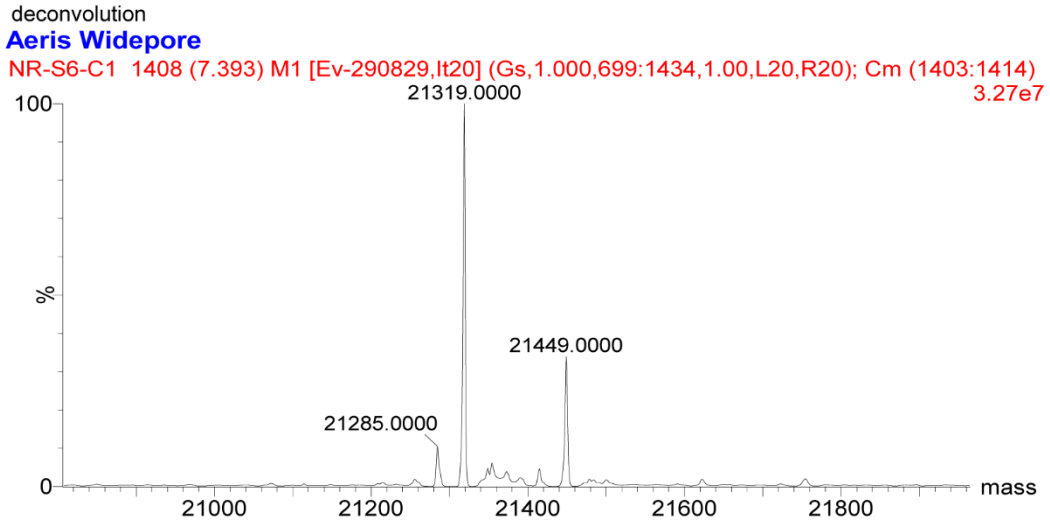


HRMS (ESI) m/z calculated for control containing 50% DMSO, α-Lactalbumin + 2, no MTG: 14178; found 14178.

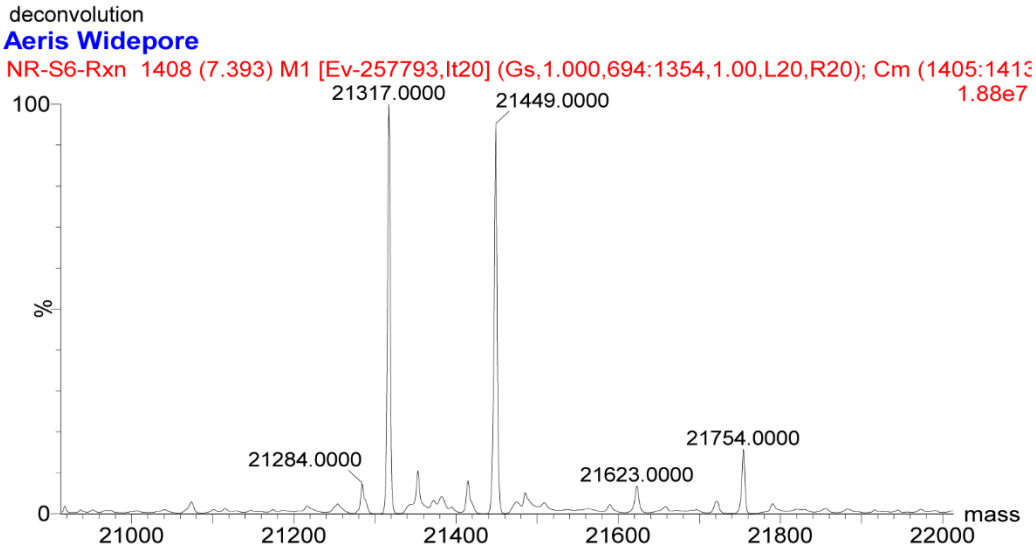


HRMS (ESI) m/z calculated for reaction containing 50% DMSO, α -Lactalbumin + 2: 14684; m/z calculated for reaction containing α-Lactalbumin + 2 + 4: 15433. found 14178.

➔ ***SPAAC on hDHFR***

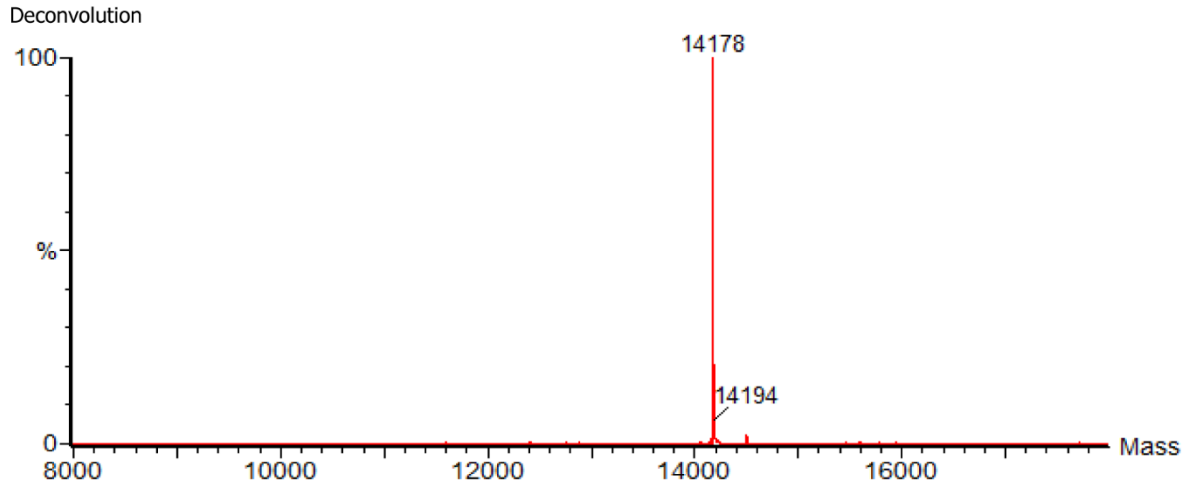


HRMS (ESI) m/z calculated for control containing hDHFR + 2, no MTG: 21324; found 21319.0, 21449.0.

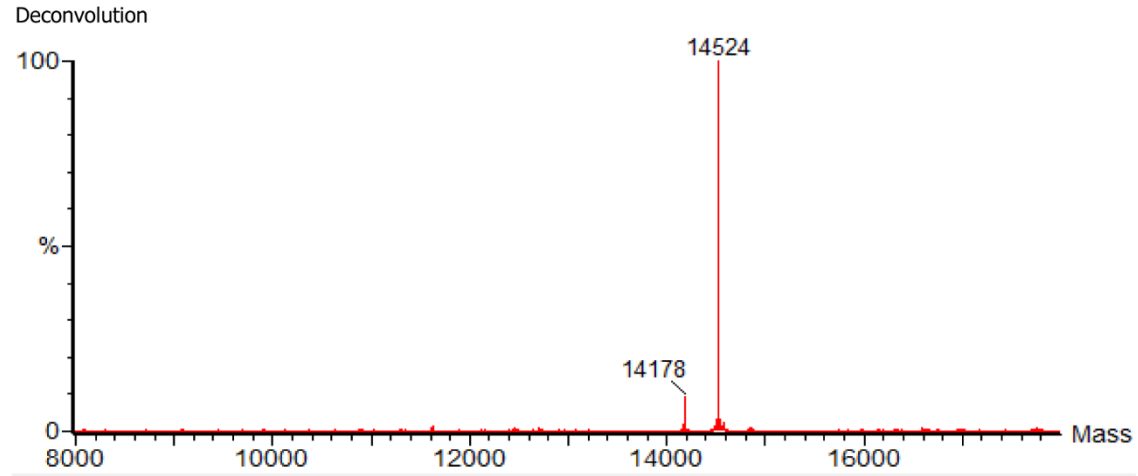


HRMS (ESI) m/z calculated for reaction containing hDHFR + 2: 21829; m/z calculated for reaction containing hDHFR + 2 + 4; 22574. Found 21317.0, 21449.0.

➔ ***TL on α-Lactalbumin***

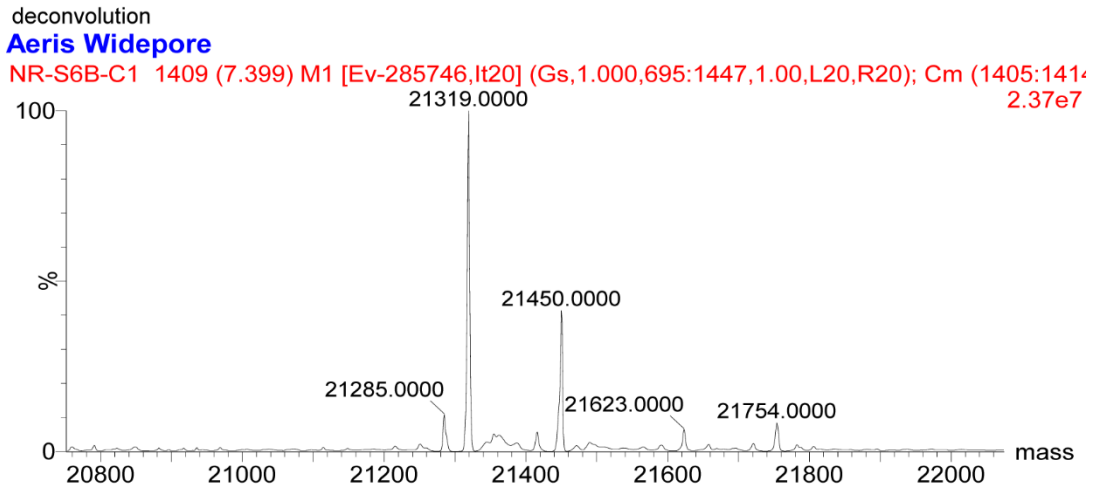


HRMS (ESI) m/z calculated for control containing α-Lactalbumin + 3, no MTG: 14178; found 14178.

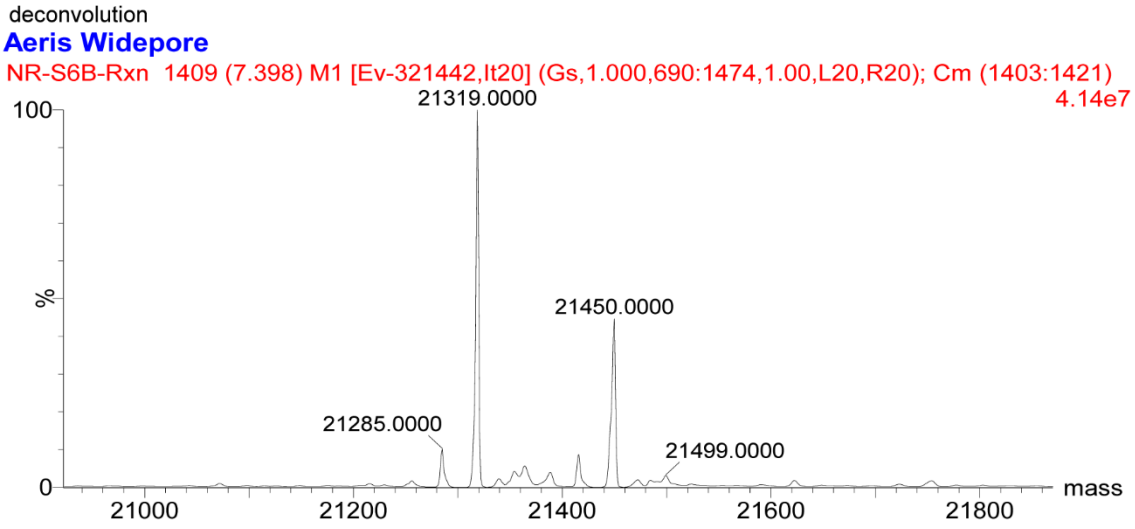


HRMS (ESI) m/z calculated for reaction containing α-Lactalbumin + 3: 14524; found 14524, 14178.

➔ ***TL on hDHFR***

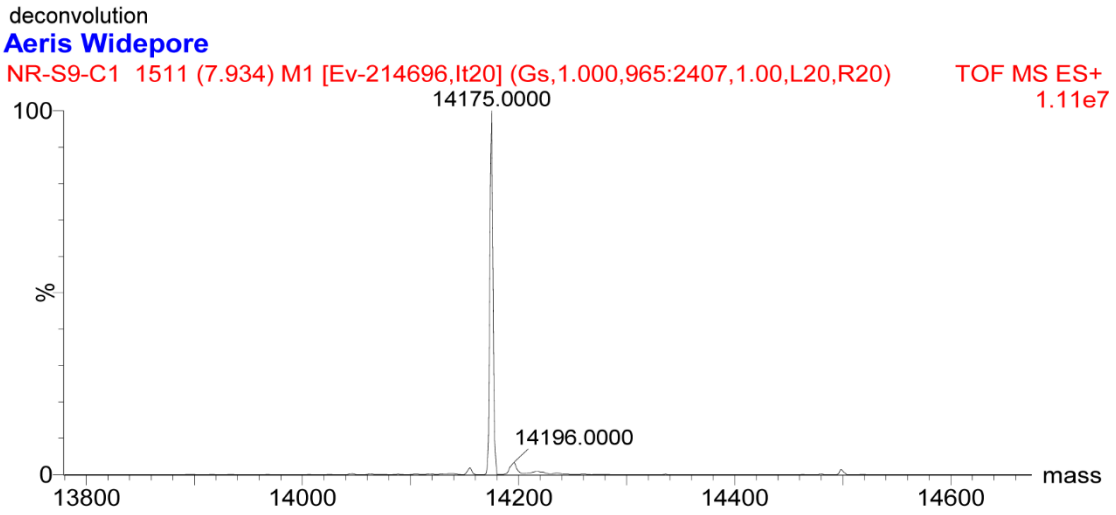


HRMS (ESI) m/z calculated for control containing hDHFR + 3, no MTG: 21324; found 21319.0, 21450.0.

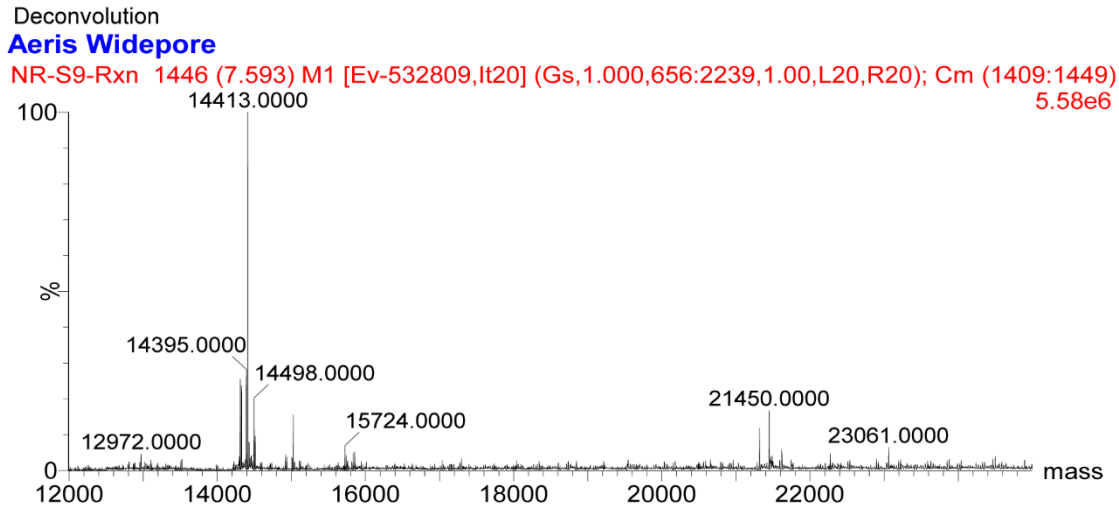


HRMS (ESI) m/z calculated for reaction containing hDHFR + 3: 21668; found 21319.0, 21450.0.

➔ ***Staudinger ligation on α-Lactalbumin***

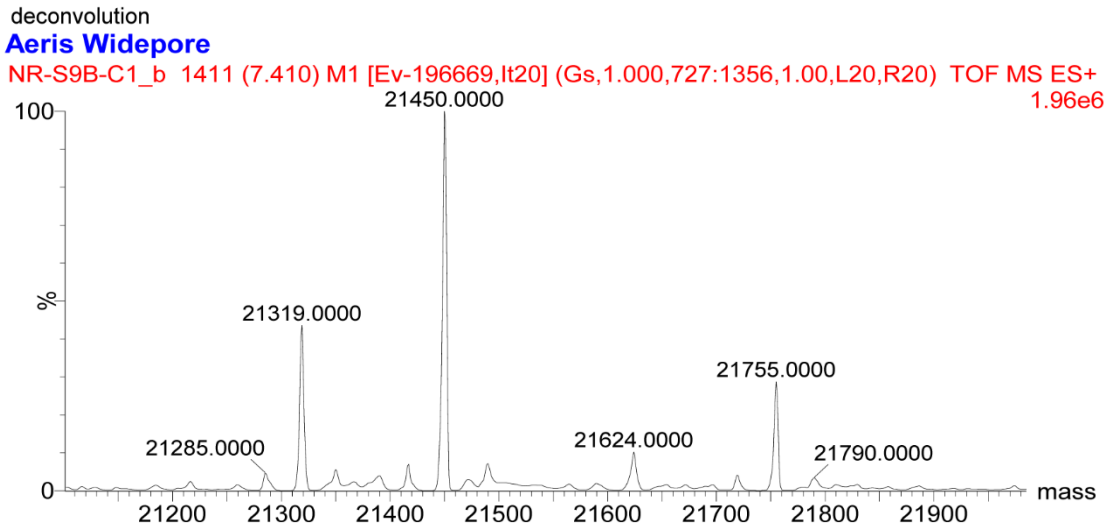


HRMS (ESI) m/z calculated for control containing α-Lactalbumin + 1, no MTG: 14178; found 14175.0.

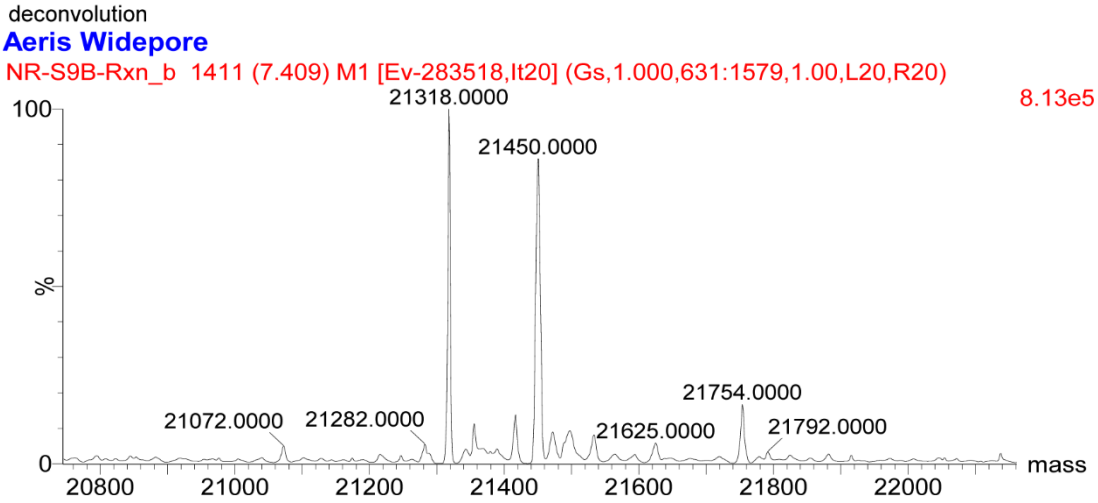


HRMS (ESI) m/z calculated for reaction containing α-Lactalbumin + 1: 14610; m/z calculated for reaction containing α-Lactalbumin + 1 + 4: 15291; m/z calculated for reaction containing α-Lactalbumin + 1 (two molecules) + 4 (one molecule): 15724. Found 14413.0, 14498, 15724.

➔ ***Staudinger ligation on hDHFR***



HRMS (ESI) m/z calculated for control containing hDHFR + 1, no MTG: <u>21324</u>; found <u>21319.0</u>, <u>21450</u>.



HRMS (ESI) m/z calculated for reaction containing hDHFR + 1: <u>22437</u>; m/z calculated for reaction containing hDHFR + 1 + 4: <u>23184</u>. Found <u>21318.0</u>, <u>21450</u>.

Annex 4 - Chapter 7 Supplemental information

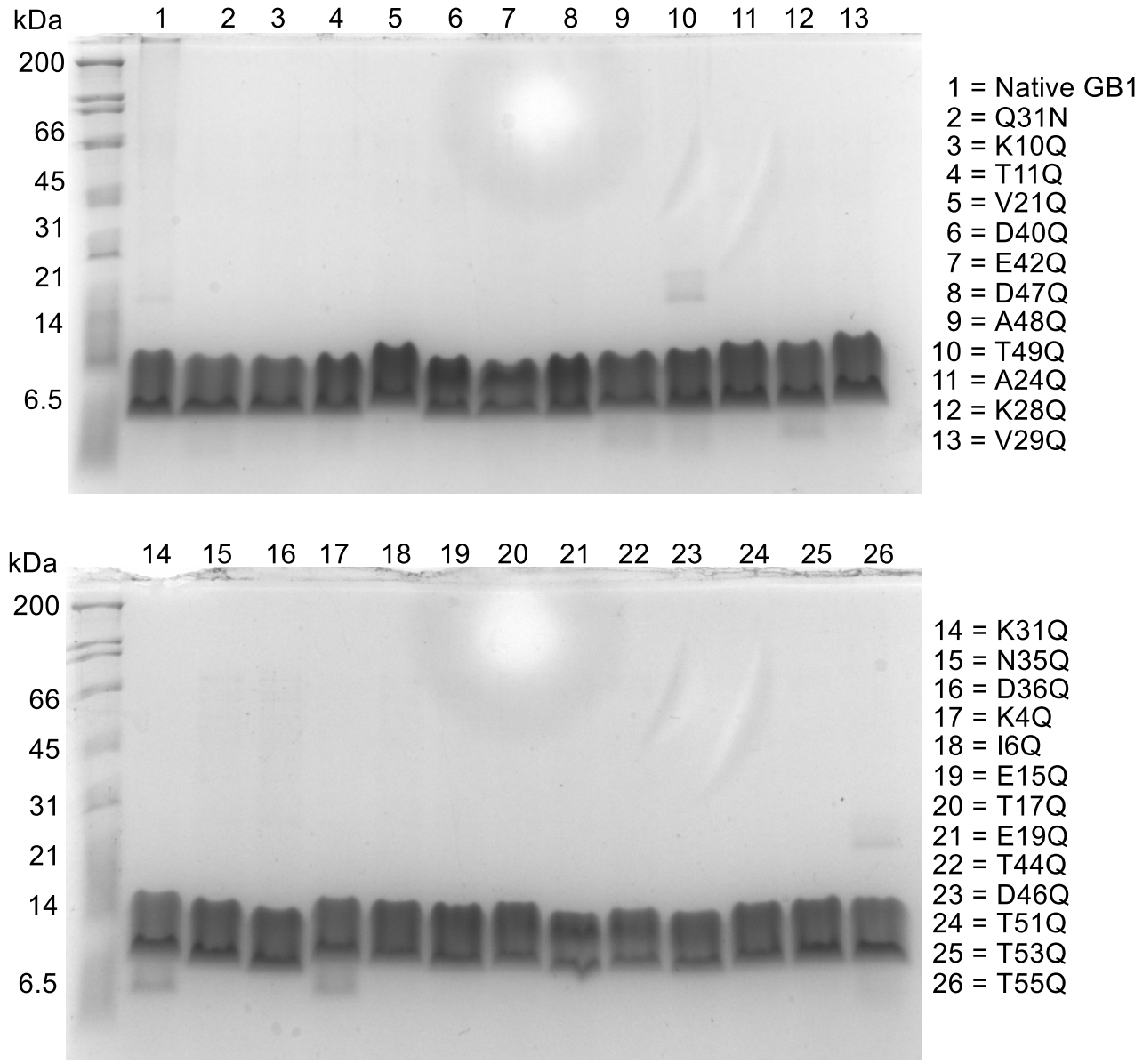


Figure A 4-1 Purified and dialyzed GB1 proteins resolved using tricine SDS-PAGE.

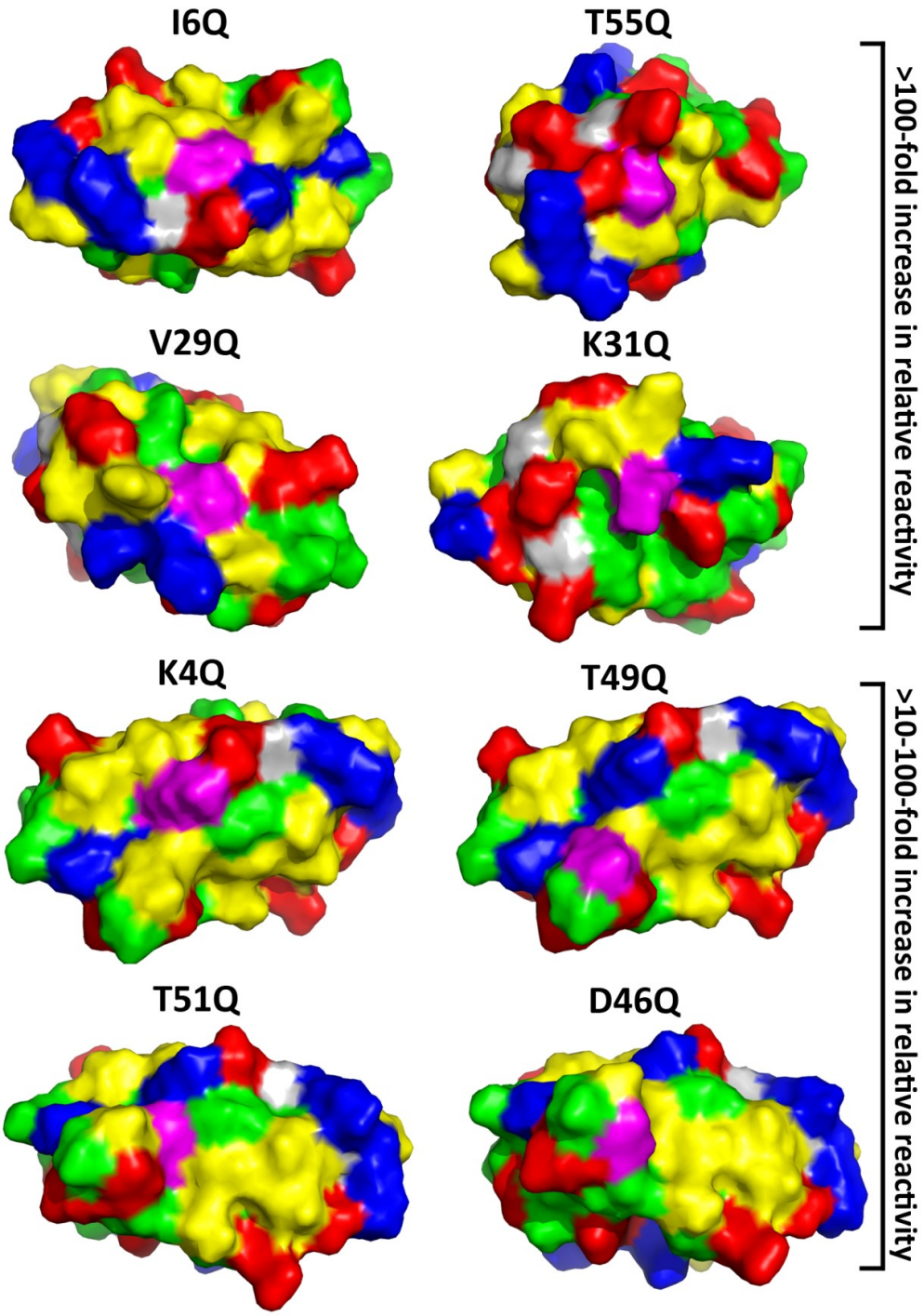


Figure A 4-2 Surface representations of >10-fold more reactive GB1 variants illustrating the molecular environment surrounding the glutamine residue.

Magenta corresponds to glutamine; green, hydrophobic residues; yellow, polar residues; blue, basic residues; red, acidic residues; grey, glycine.

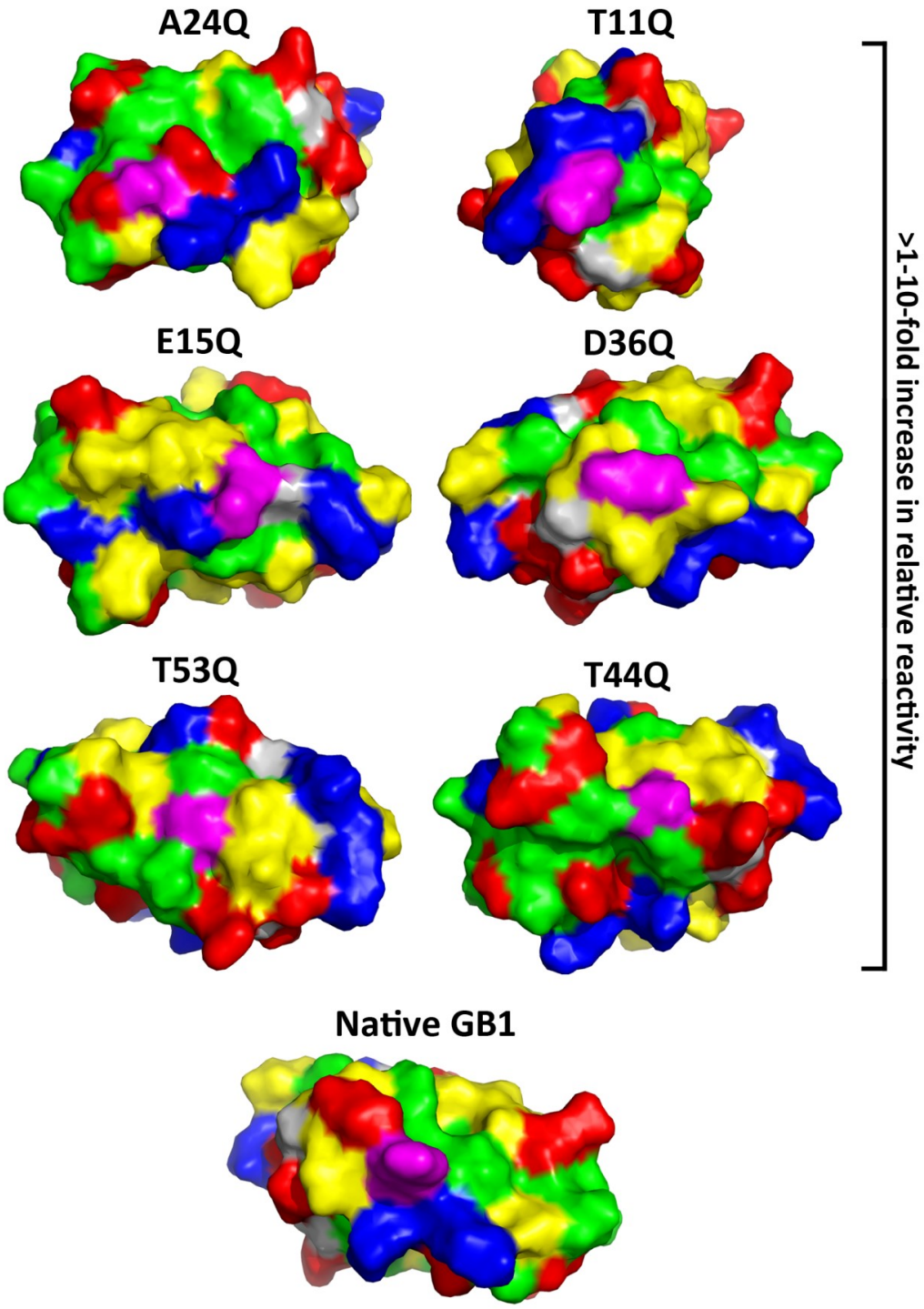


Figure A 4-3 Surface representations of 1-10 fold more reactive GB1 variants illustrating the molecular environment surrounding the glutamine residue.

Magenta corresponds to glutamine; green, hydrophobic residues; yellow, polar residues; blue, basic residues; red, acidic residues; grey, glycine.

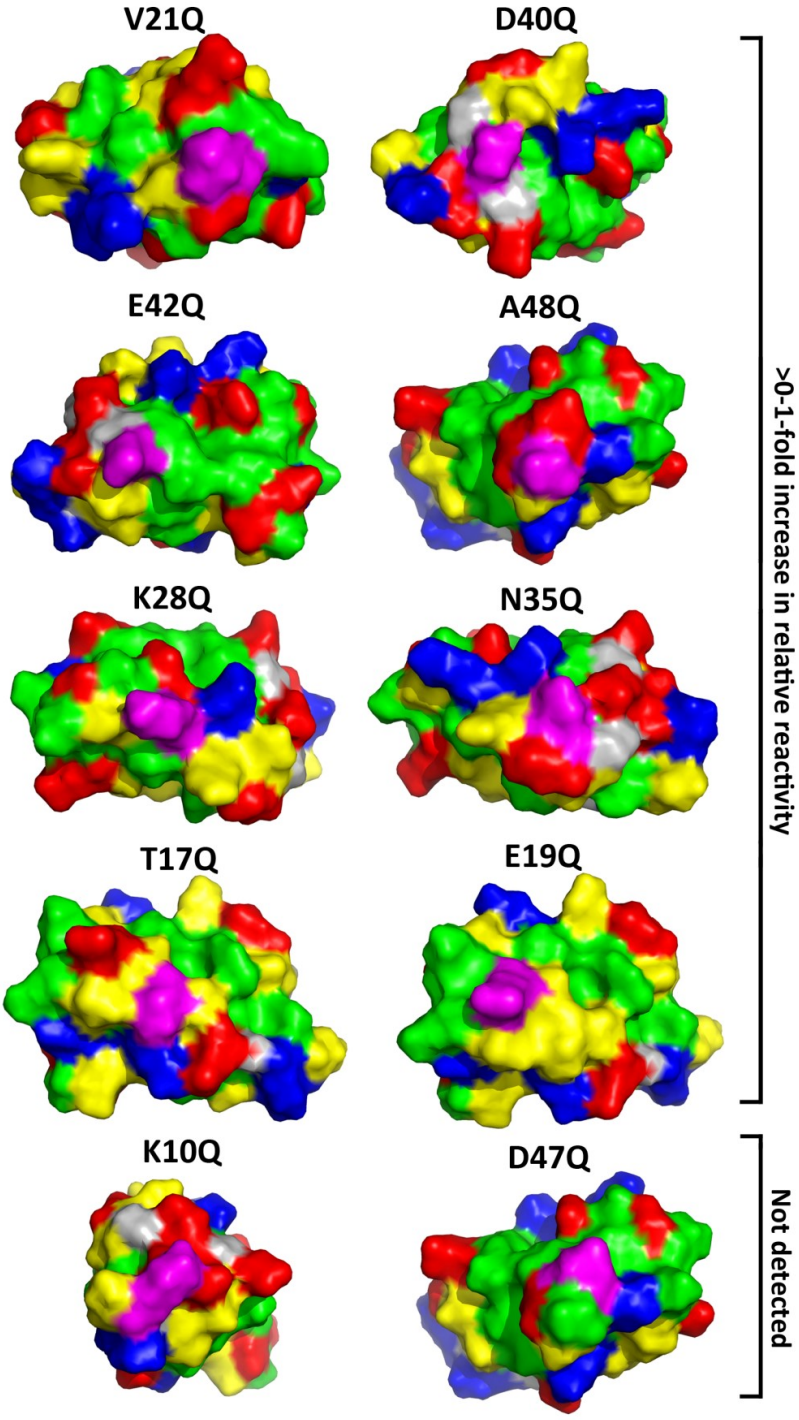


Figure A 4-4 Surface representations of poorly reactive GB1 variants illustrating the molecular environment surrounding the glutamine residue.

Magenta corresponds to glutamine; green, hydrophobic residues; yellow, polar residues; blue, basic residues; red, acidic residues; grey, glycine.

MTG mutagenesis primers

Table A 4-0-1 Mutagenic primers for MTG.

MTG Variant	Direction	Oligonucleotide sequence (5' → 3')
W69G	Forward	TCGGTGTCACCGGGGTCAATTCGGGT
	Reverse	ACCCGAATTGACCCCGGTGACACCGA
W69H	Forward	GCGTCGGTGTCACCCACGTCAATTCGGGTCAG
	Reverse	CTGACCCGAATTGACGTGGGTGACACCGACGC
Y75G	Forward	CAATTCGGGTCAGGGCCCGACGAACAGAC
	Reverse	GTCTGTTCGTCGGGCCCTGACCCGAATTG
Y75H	Forward	CAATTCGGGTCAGCACCCGACGAACAGAC
	Reverse	GTCTGTTCGTCGGGTGCTGACCCGAATTG
Y302G	Forward	CAACTGGTCCGAGGGTGGCTCGGACTTCGACC
	Reverse	GGTCGAAGTCCGAGCCACCCTCGGACCAGTTG
Y302H	Forward	CAACTGGTCCGAGGGTCACTCGGACTTCGACC
	Reverse	GGTCGAAGTCCGAGTGACCCTCGGACCAGTTG

Table A 4-0-2 Mutagenic primers for GB1.

GB1 Variant	Direction	Overlap oligonucleotide sequence (5' → 3')
Native	Forward	GGAATTCCATATGGACACCTACAAACTGATCC
	Reverse	CGGGATCCTTATTCGGTAACCGTGAAGGT
Q32N	Forward	GAAAAAGTTTTCAAAAACTACGCTAACGACAAC
	Reverse	GTTAGCGTAGTTTTTGAAAACTTTTTCTGCAGTAG
K4Q	Forward	CGGCAGCCATATGGACACCTACCAACTGATCC
	Reverse	CGGGATCCTTATTCGGTAACCGTGAAGGT
I6Q	Forward	CGGCAGCCATATGGACACCTACAAACTGCAGCTGAACGGTA
	Reverse	CGGGATCCTTATTCGGTAACCGTGAAGGT
K10Q	Forward	ACTGATCCTGAACGGTCAAACCCTGAAA
	Reverse	GTTTCACCTTTCAGGGTTTGACCGTT
T11Q	Forward	ACTGATCCTGAACGGTAAACAGCTGAAAGGTGAA
	Reverse	GTGGTGGTTTCACCTTTCAGCTGTTTACCGTTC
E15Q	Forward	CTGAAAGGTCAAACCACCACCGAAG
	Reverse	ACAGCTTGGGTGGTGGTTTCACCTT
T17Q	Forward	AAGGTGAAACCCAGACCGAAGCTGTAGACGC
	Reverse	AGCTTCGGTCTGGGTTTCACCTTTCAGGGTT
E19Q	Forward	AACCACCACCCAAGCTGTAGACGCT
	Reverse	ACAGCTTGGGTGGTGGTTTCACCTT
V21Q	Forward	ACCGAAGCTCAAGACGCTGCTACTG
	Reverse	AGCGTCTTGAGCTTCGGTGGTGGTT
A24Q	Forward	GCTGTAGACGCTCAGACTGCAGAAAAAGTTT
	Reverse	TCTGCAGTCTGAGCGTCTACAGCTTCGGTG
K28Q	Forward	TGCTACTGCAGAACAAGTTTTCAAAAACTACG
	Reverse	TGAAAACTTGTTCTGCAGTAGCAGCGTCTA
V29Q	Forward	TACTGCAGAAAAACAGTTCAAAAACTACGCTAACG
	Reverse	GTAGTTTTTGAACTGTTTTTCTGCAGTAGCAGCGT
K31Q	Forward	TTAGCGTAGTTTTGGAAAACTTTTTCTGCAGTA
	Reverse	GAAAAAGTTTTCCAAAACTACGCTAACGACAAC
N35Q	Forward	AACTACGCTCAGGACAACGGTGTCGACG
	Reverse	ACCGTTGTCCTGAGCGTAGTTTTTGAAAACT
D36Q	Forward	TACGCTAACCAGAACGGTGTCGACGGTGA
	Reverse	CGACACCGTTCTGGTTAGCGTAGTTTTTGA
D40Q	Forward	CAACGGTGTCCAAGGTGAATGGACCTACG
	Reverse	CATTCACCTTGGACACCGTTGTCGTTAGC
E42Q	Forward	GTCGACGGTCAATGGACCTACGACG
	Reverse	GTCCATTGACCGTCGACACCGTTGT

T44Q	Forward	CGGTGAATGGCAGTACGACGACGCTACCA
	Reverse	TCGTCGTACTGCCATTCACCGTCGACACC
D46Q	Forward	ATGGACCTACCAGGACGCTACCAAAACCTTCA
	Reverse	TTGGTAGCGTCCTGGTAGGTCCATTCACCGT
D47Q	Forward	CGTGAAGGTTTTGGTAGCCTGGTCGTAGGTC
	Reverse	GGTGAATGGACCTACGACCAGGCTACCAAAA
A48Q	Forward	GAATGGACCTACGACGACCAGACCAAAACCTT
	Reverse	AACCGTGAAGGTTTTGGTCTGGTCGTCGTAG
T49Q	Forward	TACGACGACGCTCAGAAAACCTTCACGGTT
	Reverse	AAGGTTTTCTGAGCGTCGTCGTAGGTCCAT
T51Q	Forward	GGAATTCCATATGGACACCTACAAACTGATCC
	Reverse	CGGGATCCTTATTCGGTAACCGTGAACTGTTTGGTAGCGTC
T53Q	Forward	GGAATTCCATATGGACACCTACAAACTGATCC
	Reverse	CGGGATCCTTATTCGGTAACCTGGAAGGTTTTGGTAGCG
T55Q	Forward	GGAATTCCATATGGACACCTACAAACTGATCC
	Reverse	CGGGATCCTTATTCCTGAACCGTGAAGGTTTTGGTAGC

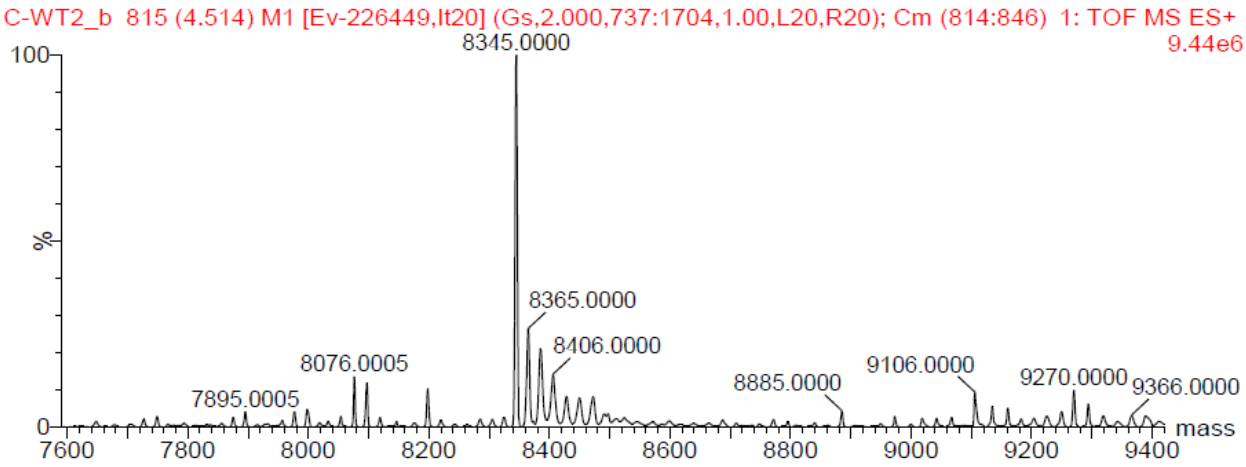
High resolution mass spectrometry

Table A 4-0-3 Masses of GB1 variants as determined by high resolution LC-MS.

Masses of GB1 variants as determined by high resolution LC-MS. Only native GB1, the glutamine knock-out, Q32N, and variants observed to have visible fluorescence were sent for analysis. Variants for which no conjugated species is observed are highlighted in orange; these are all variants exhibiting less than an order of magnitude increase in fluorescence intensity, or the knock-out. The control samples (3rd column from left), retain their poly-histidine tag as there is no MTG present to cleave it. Calculated masses were obtained by inputing the amino acid sequence for each variant into ExPASy's ProtParam module. Observed conjugated masses are presented in bold.

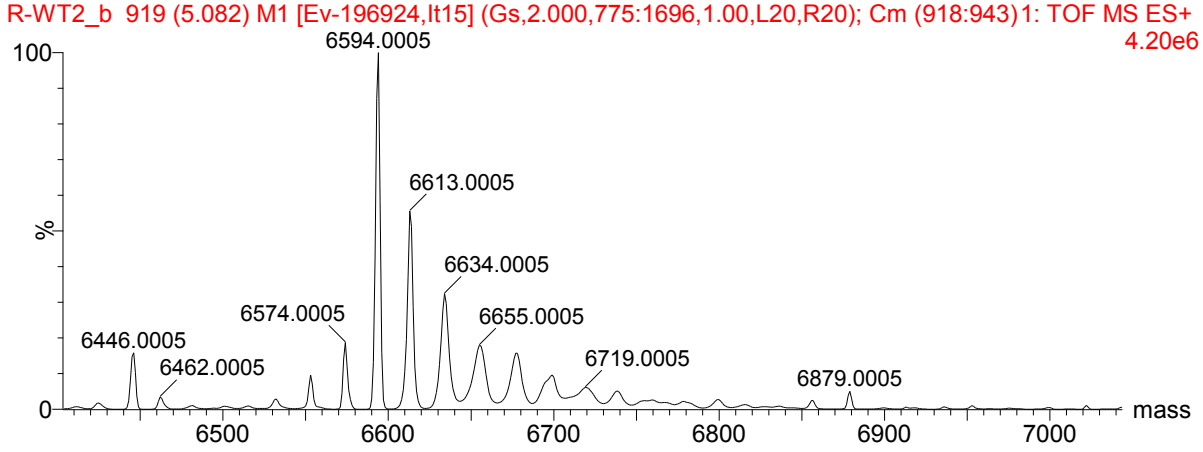
Variant	Calculated mass, with His-Tag (Da)	Observed mass, control (No MTG; Da)	Calculated mass, His-Tag cleaved (Da)	Calculated mass, cleaved variant, conjugated (Da)	Observed mass, reaction (Da)
Native	8343.0	8345.0	6592.2	6800.2	6594.0, **6800.0**
Q32N	8329.0	8329.0	6578.2	6786.2	6579.0
K4Q	8329.0	8330.0	6578.1	6786.1	6579.0, **6788.0**
I6Q	8344.0	8345.0	6593.1	6801.1	6597.0, **6900.0**
T11Q	8356.0	8357.0	6605.2	6813.2	6606.0
E15Q	8328.0	8329.0	6577.2	6785.2	6578.0
A24Q	8386.0	8386.0	6635.2	6843.2	6636.0
V29Q	8358.0	8358.0	6607.1	6815.1	6608.0, **6816.0**
K31Q	8329.0	8330.0	6578.1	6786.1	6579.0, **6787.0**
T44Q	8356.0	8356.0	6605.2	6813.2	6606.0
D46Q	8342.0	8343.0	6591.2	6799.2	6593.0, **6800.0**
T49Q	8356.0	8356.0	6605.2	6813.2	6606.0, **6814.0**
T51Q	8356.0	8356.0	6605.2	6813.2	6606.0, **6815.0**
T53Q	8356.0	8357.0	6605.2	6813.2	6606.0, **6814.0**

Mass spectra of GB1 variants

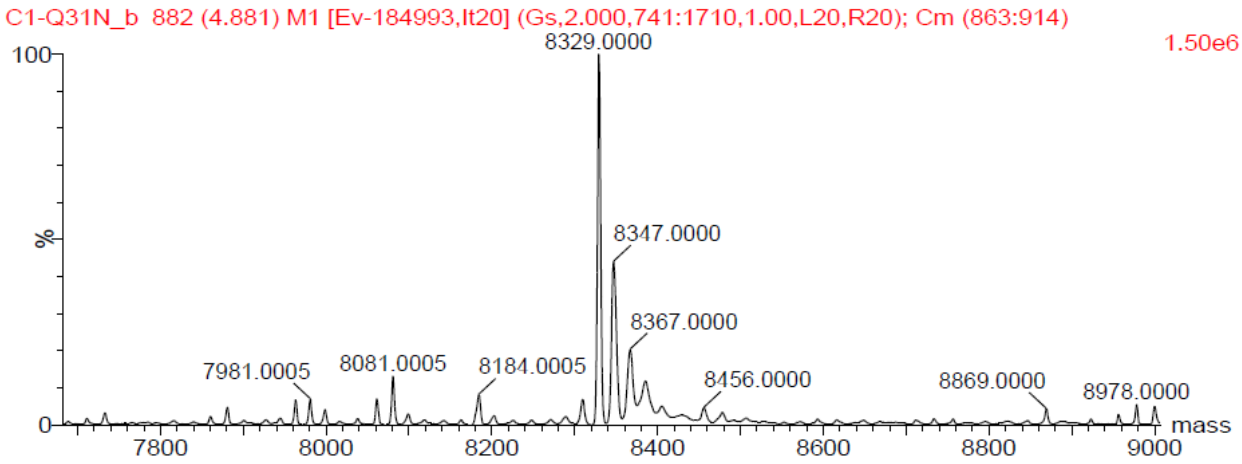


Native GB1, No MTG

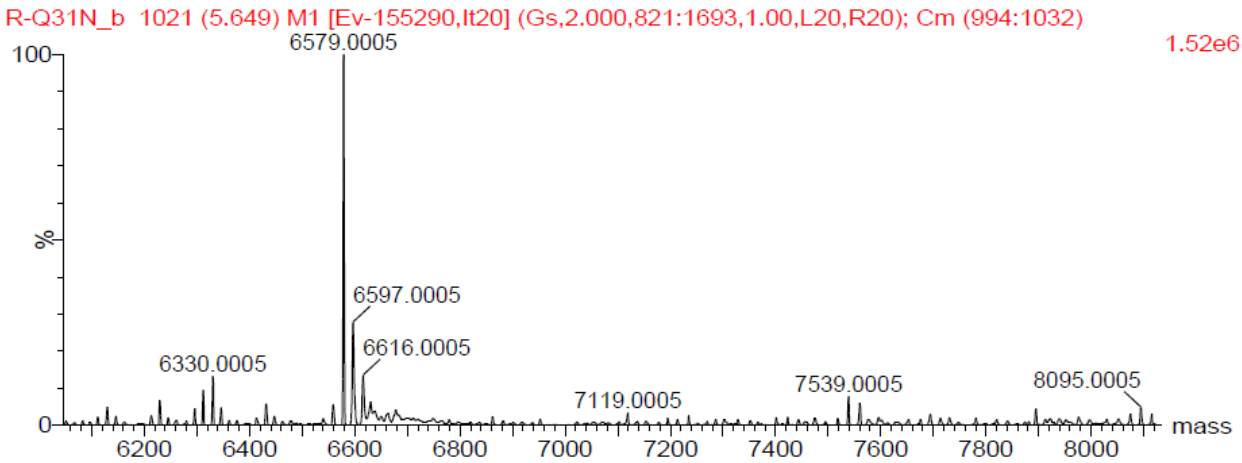
Aeris Peptide, 150x2.1mm



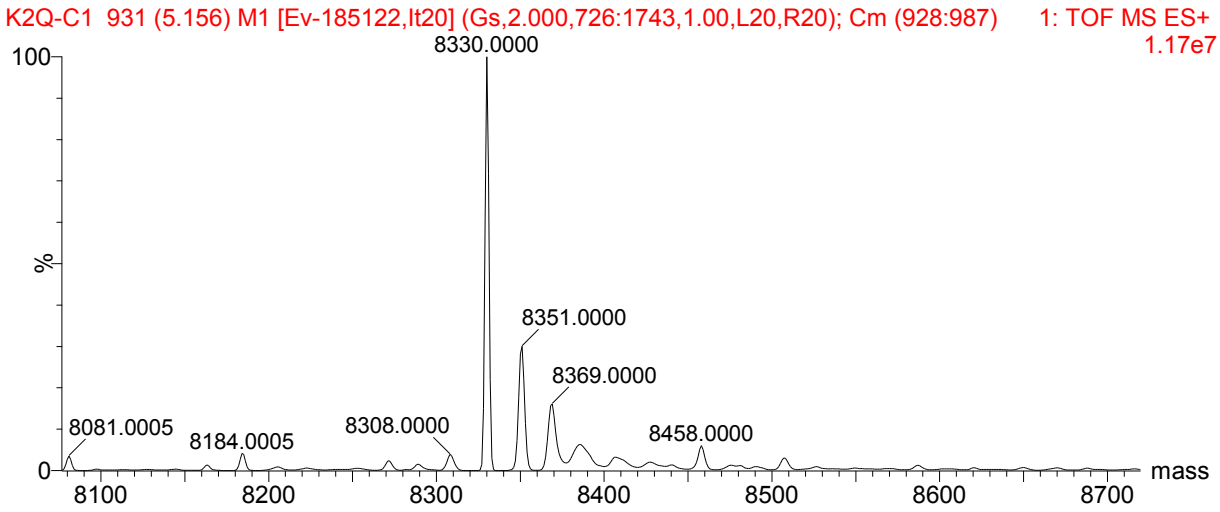
Native GB1, MTG present



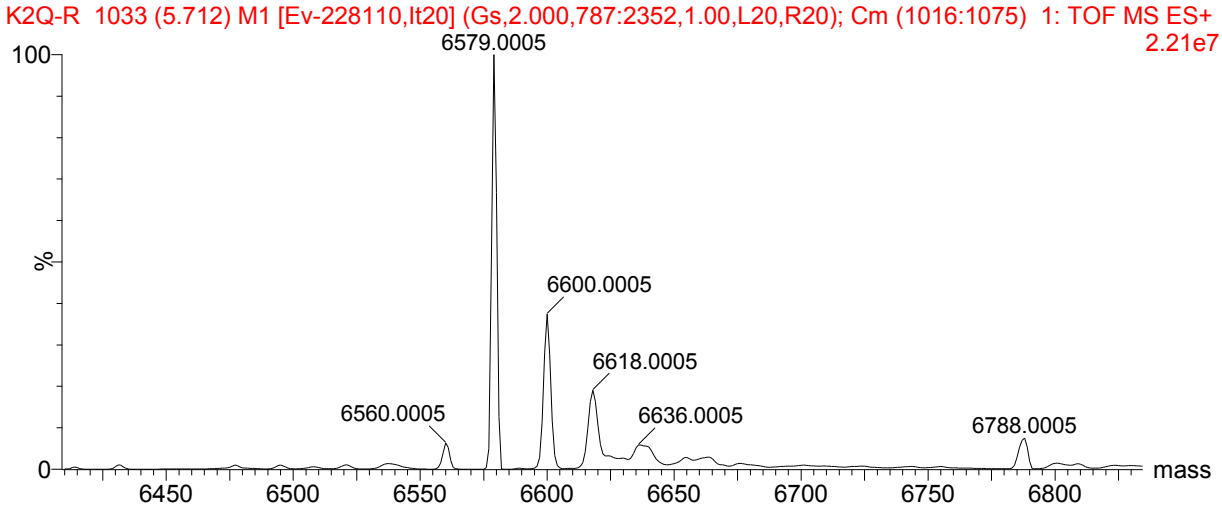
Q32N, No MTG



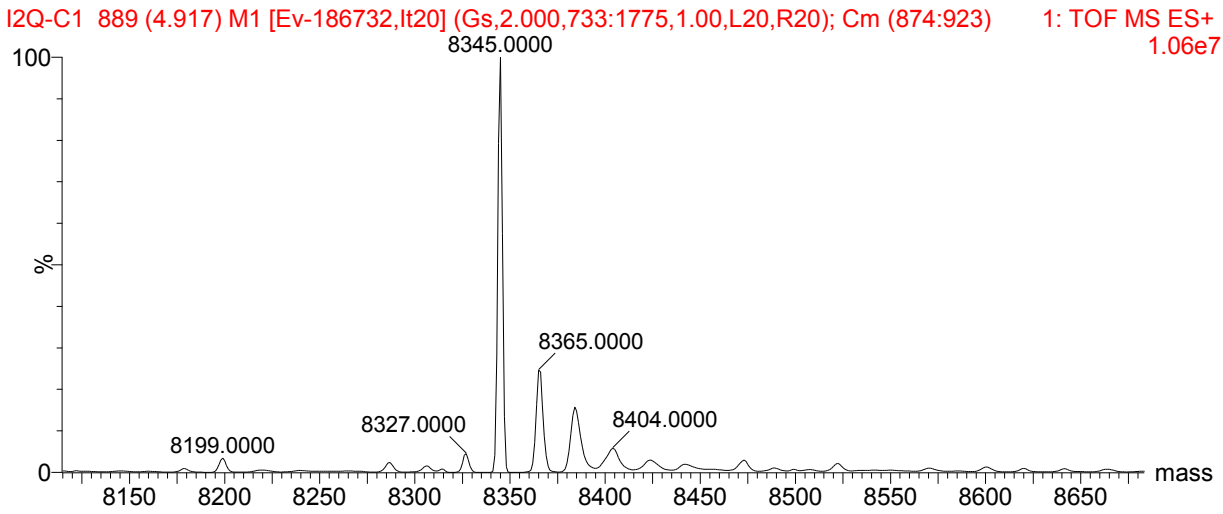
Q32N, MTG present



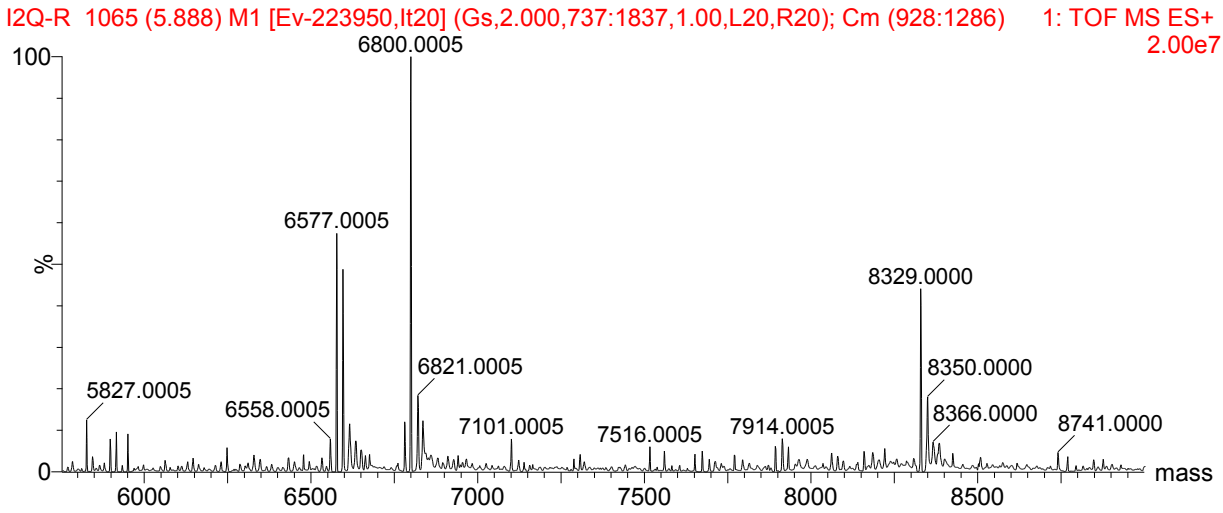
K4Q, No MTG



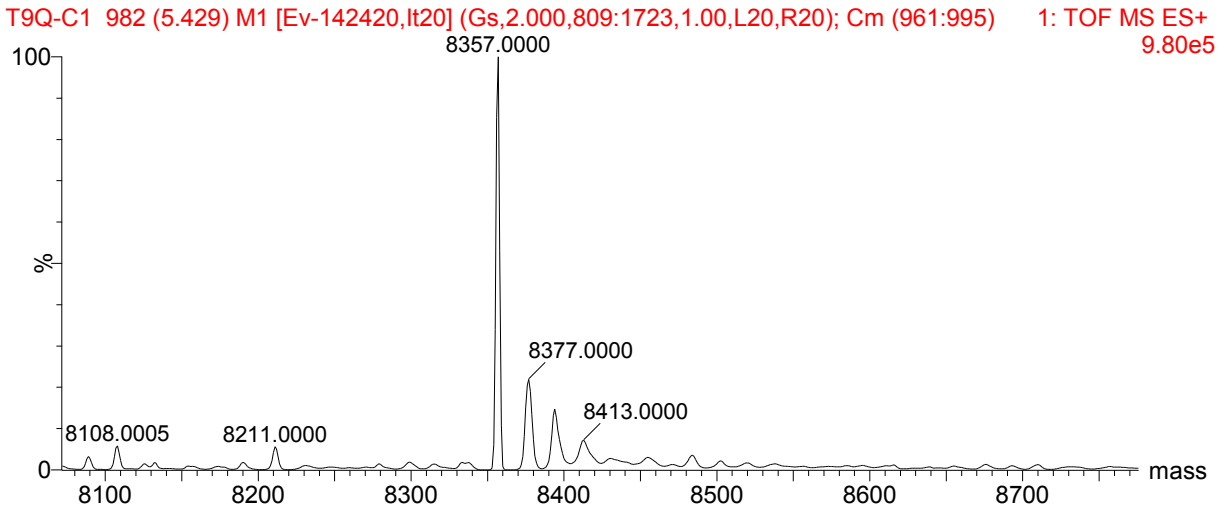
K4Q, MTG present



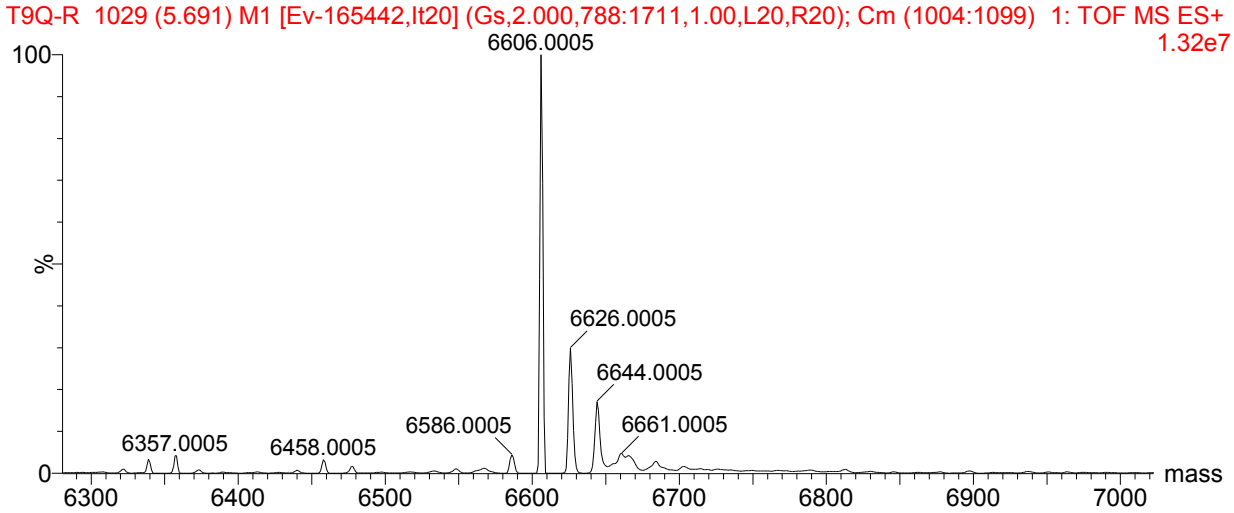
I6Q, No MTG



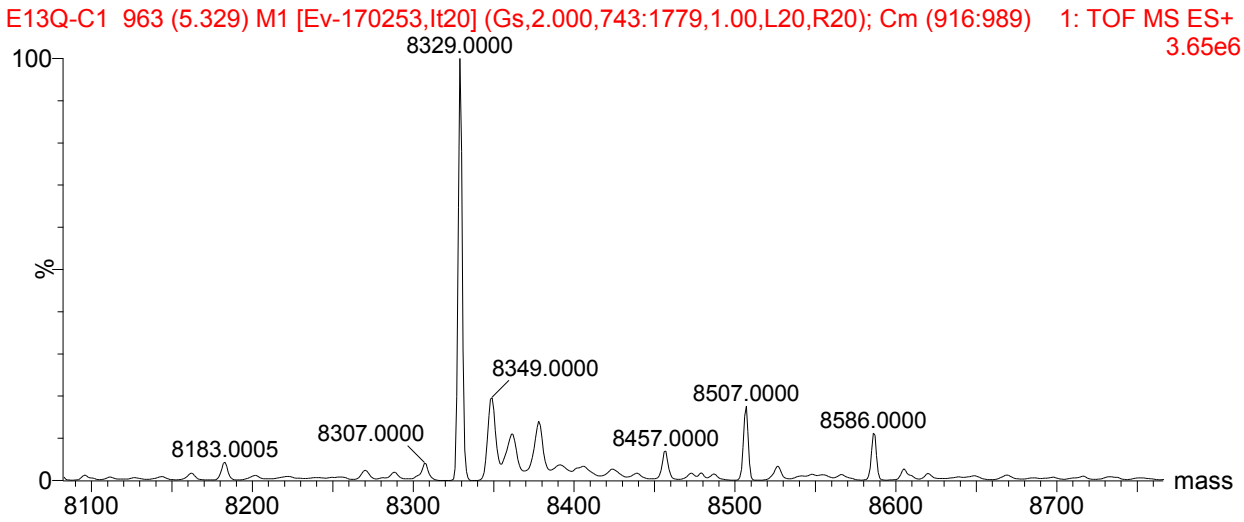
I6Q, MTG present



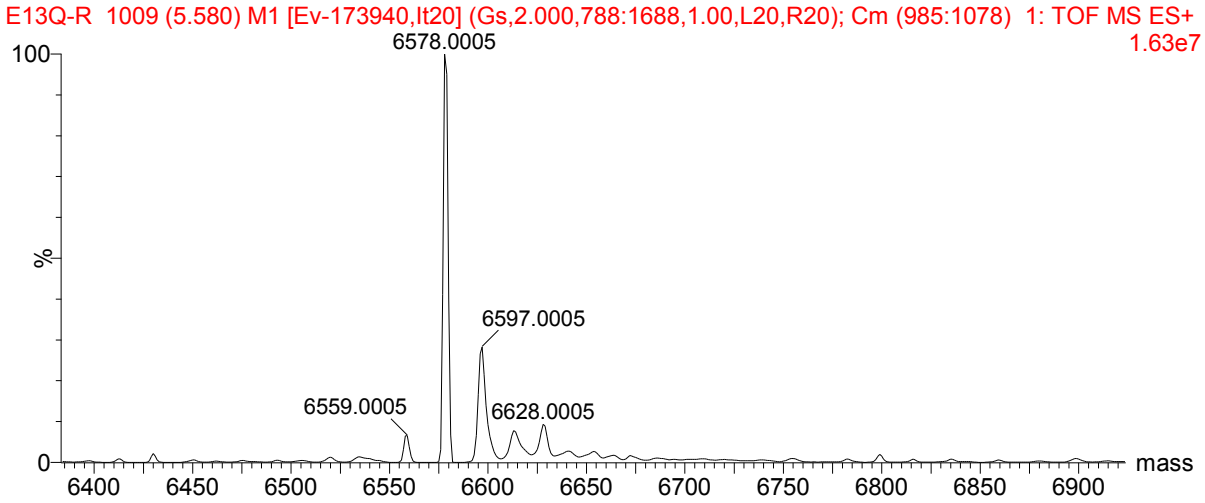
T11Q, No MTG



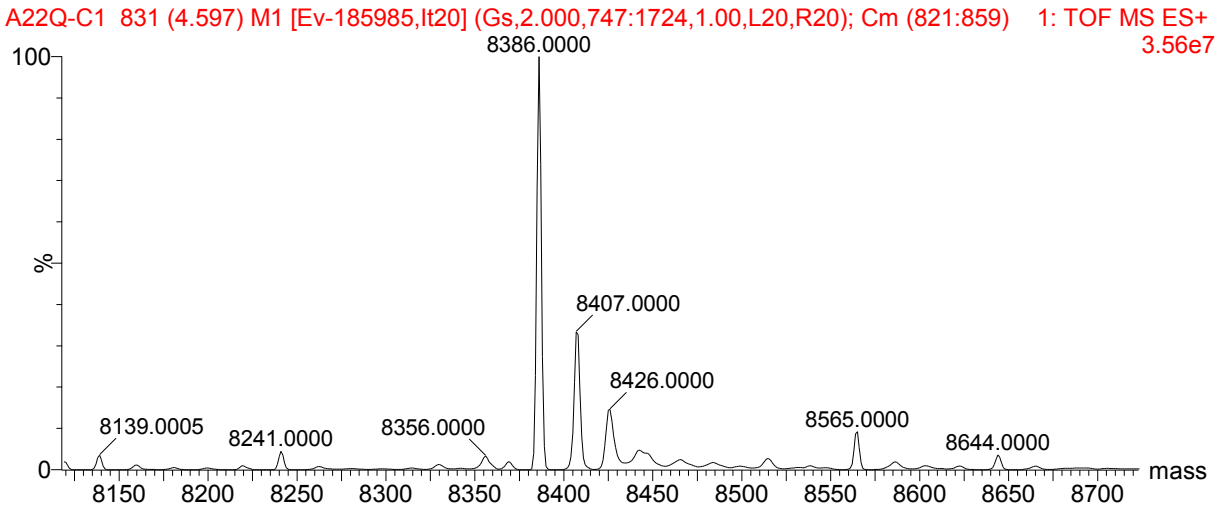
T11Q, MTG present



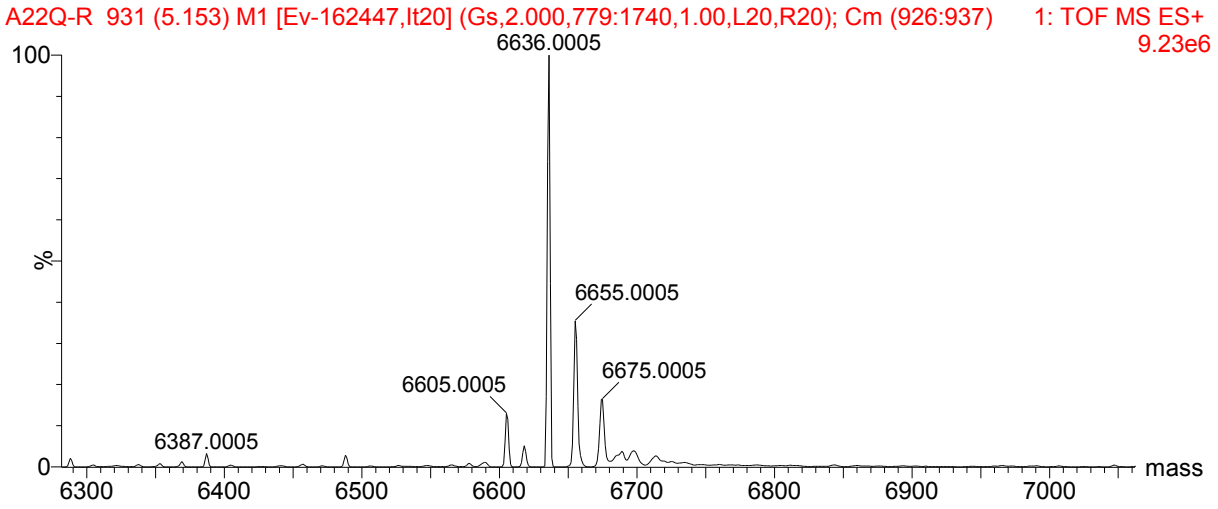
E13Q, No MTG



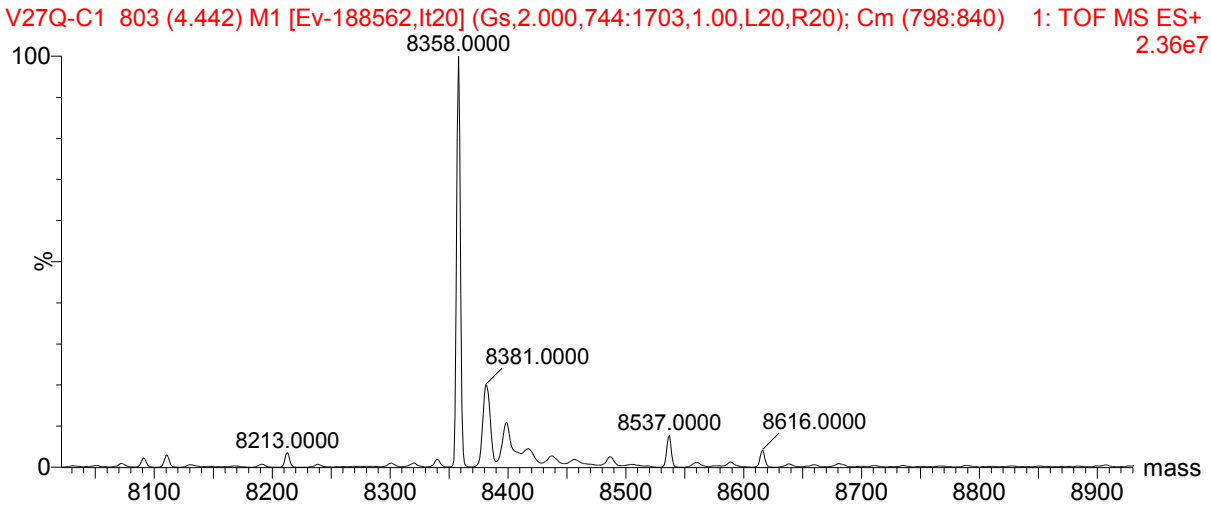
E13Q, MTG present



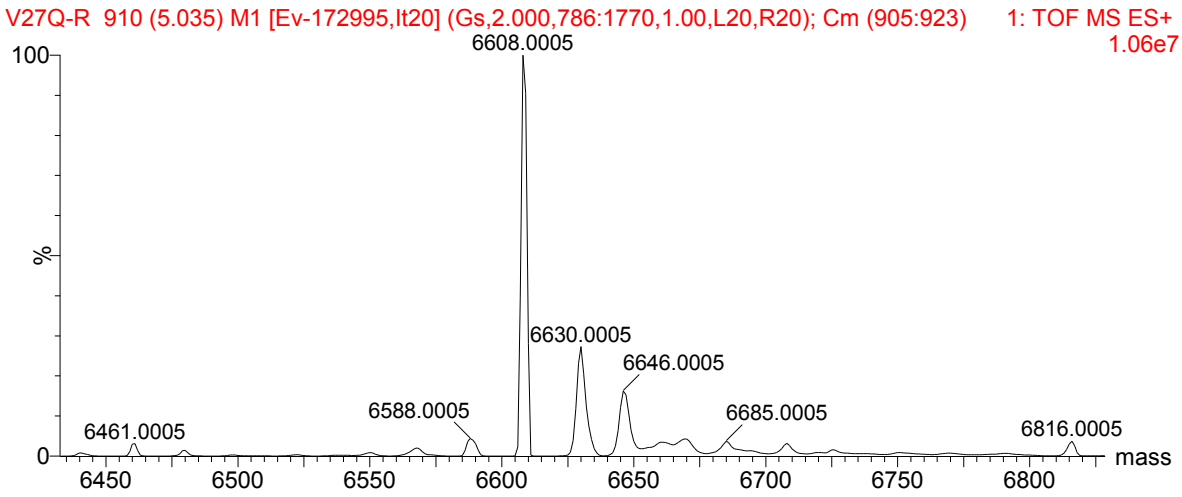
A24Q, No MTG



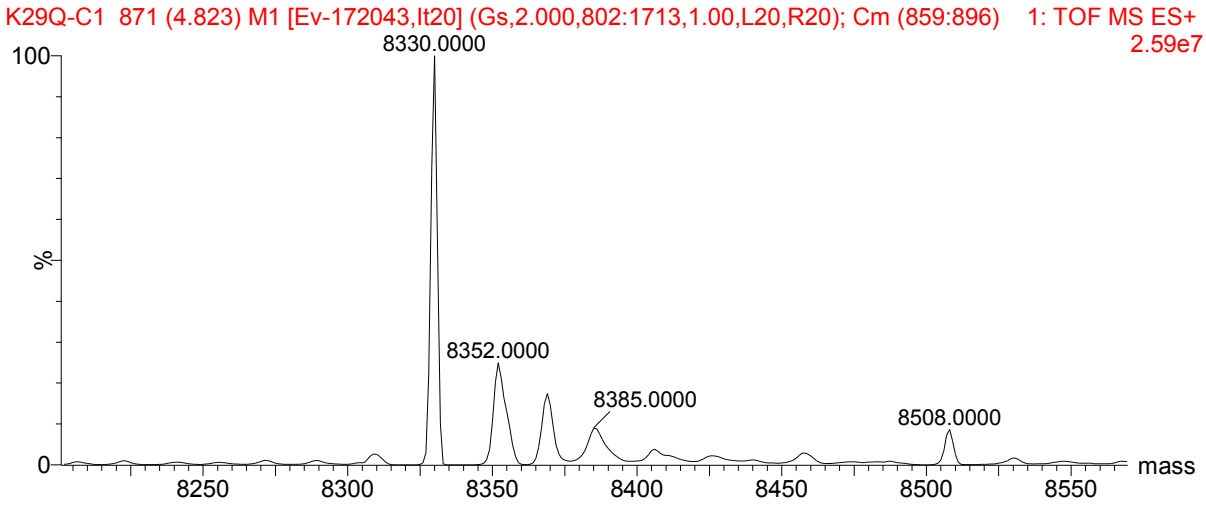
A24Q, MTG present



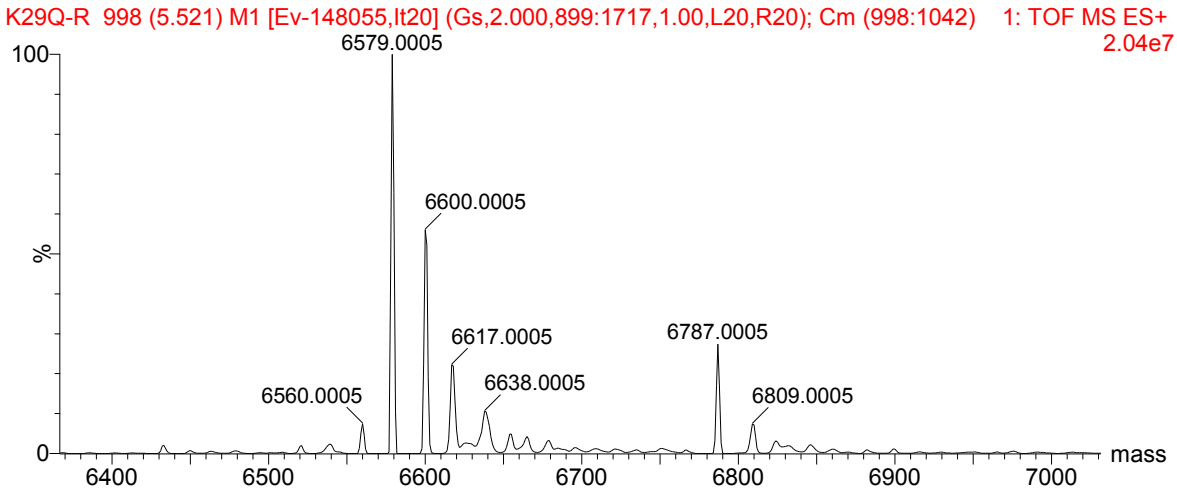
V29Q, No MTG



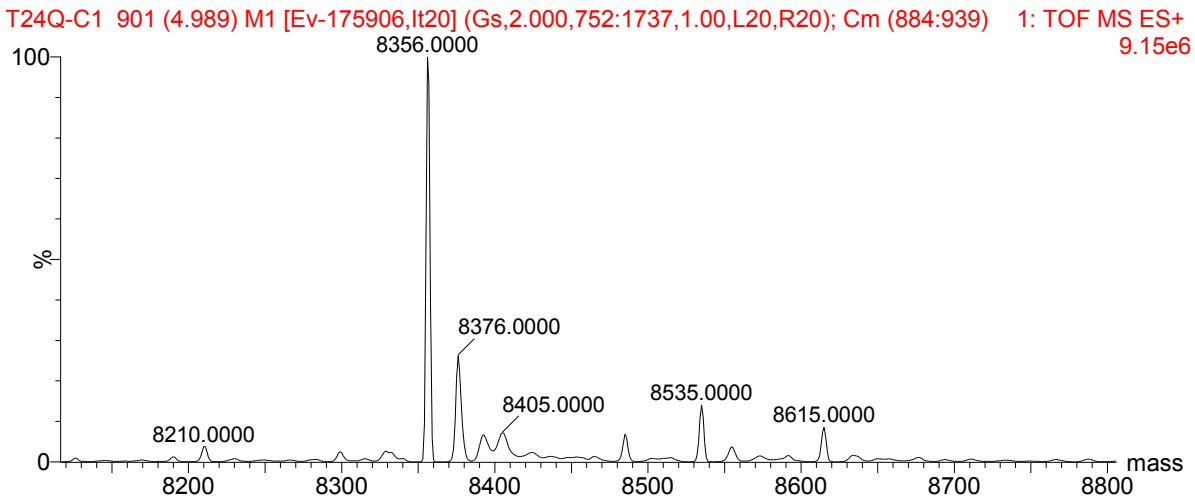
V29Q, MTG present



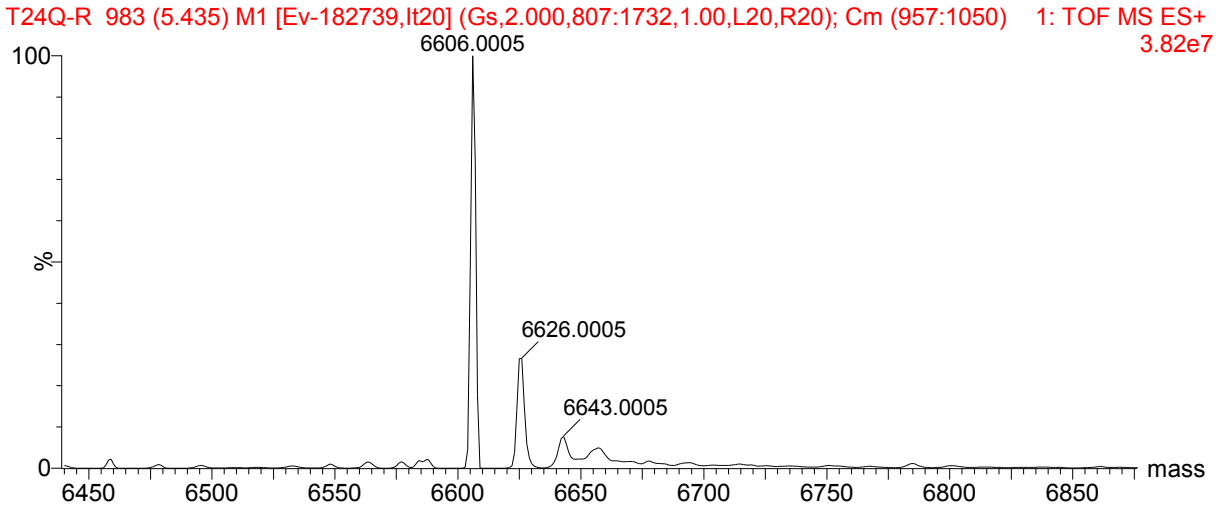
K31Q, No MTG



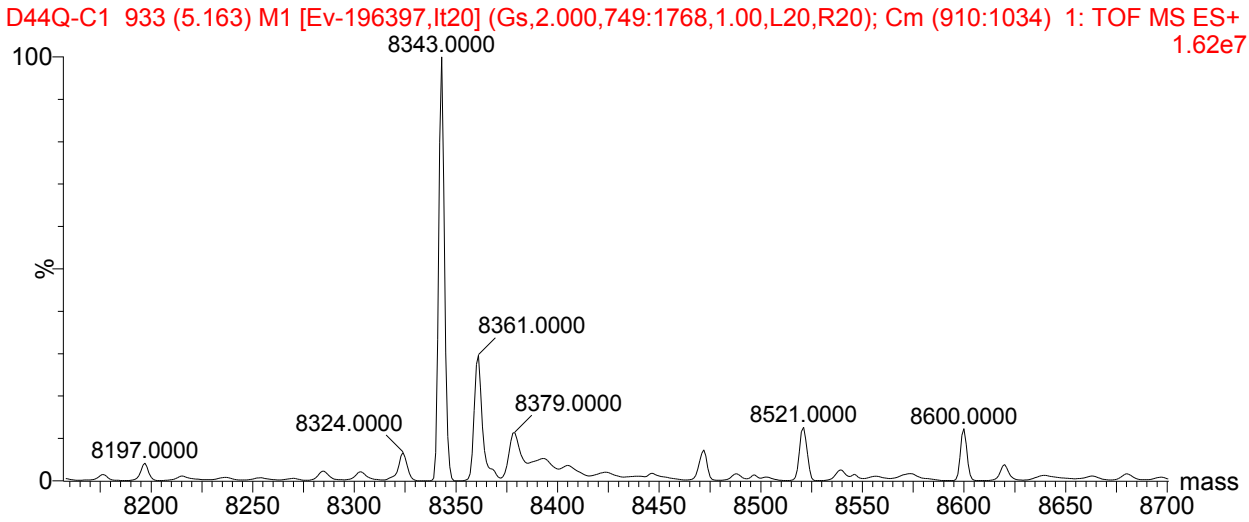
K31Q, MTG present



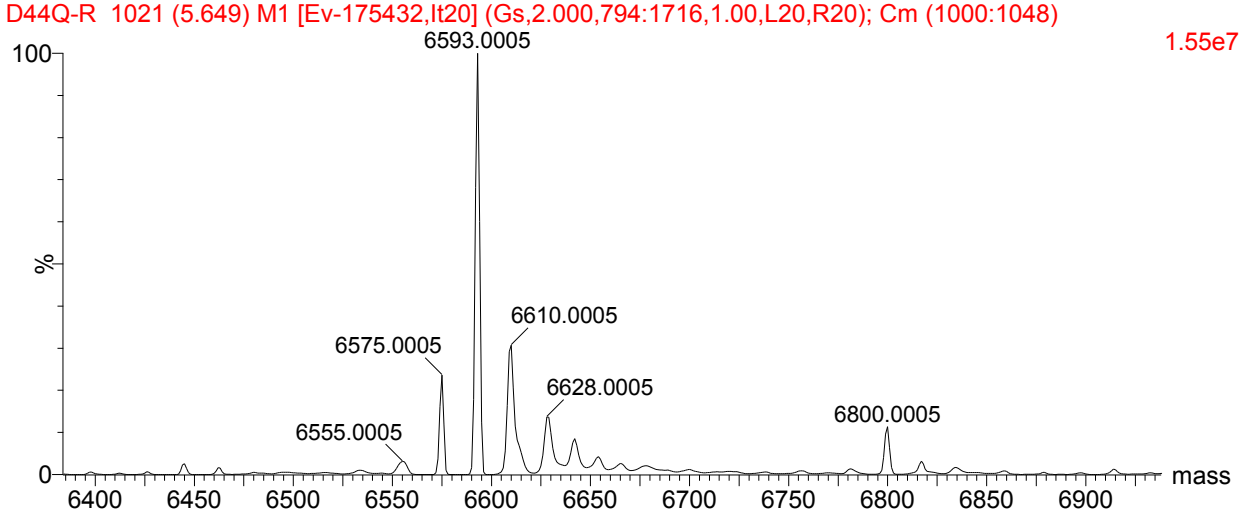
T44Q, No MTG



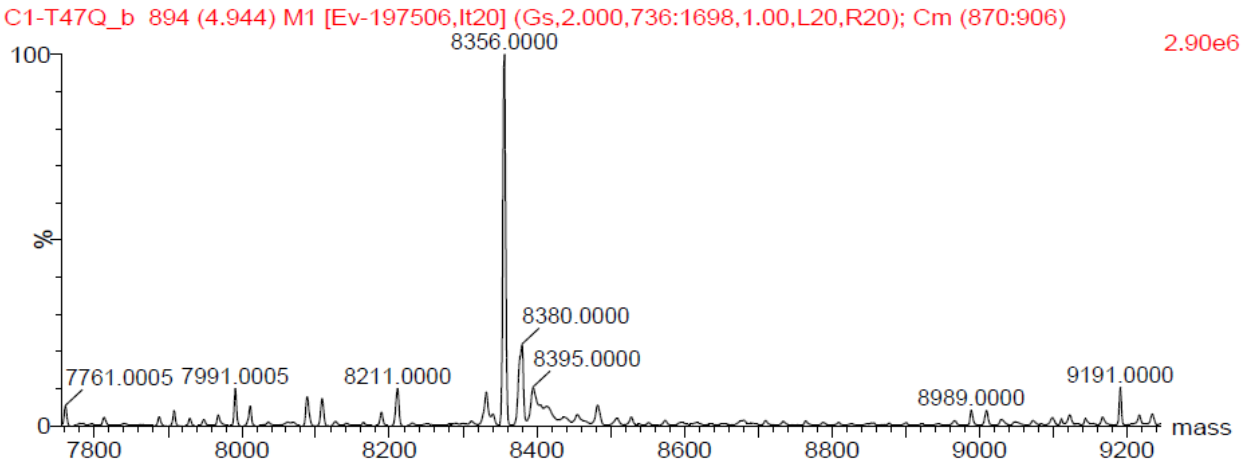
T44Q, MTG present



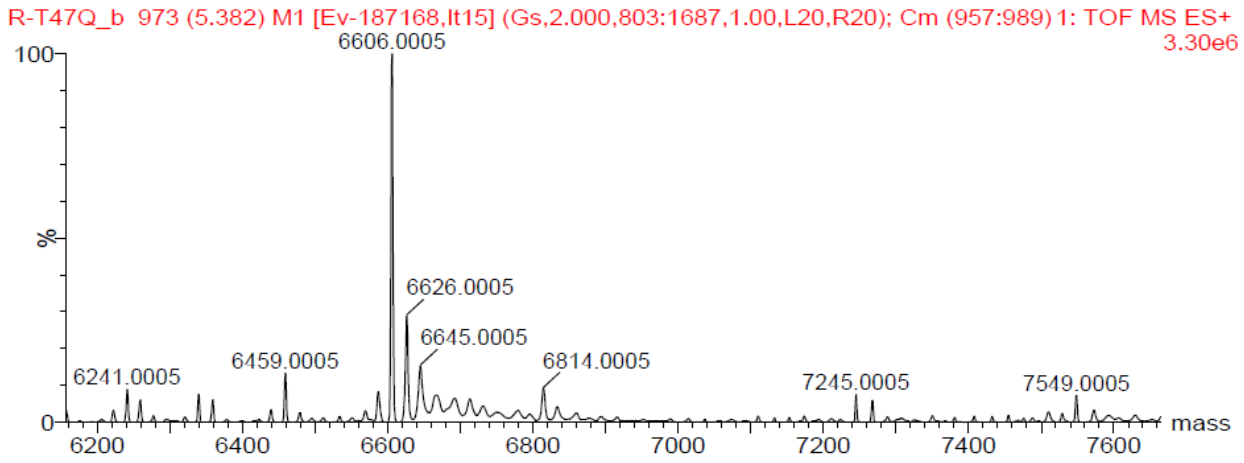
D46Q, No MTG



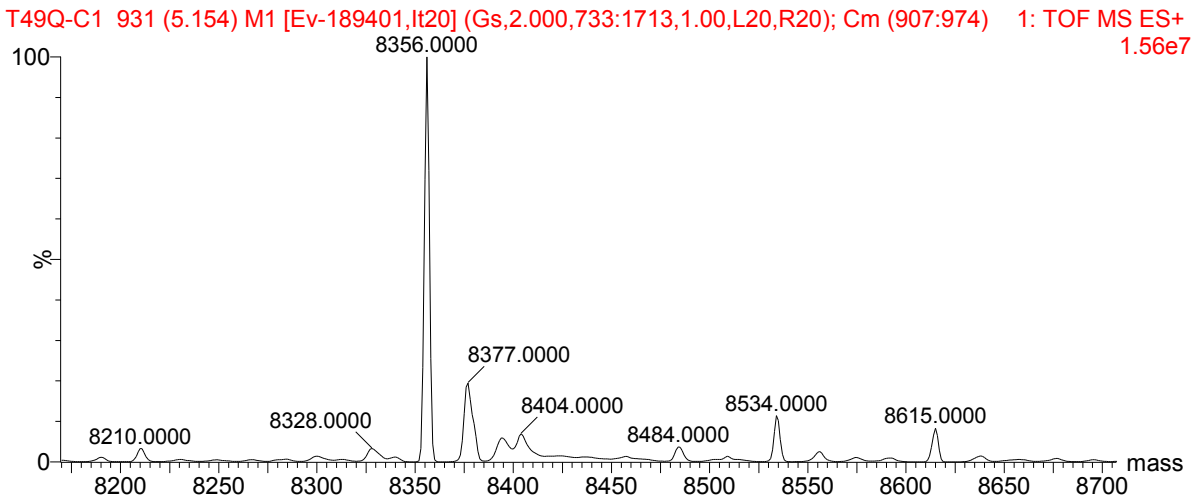
D46Q, MTG present



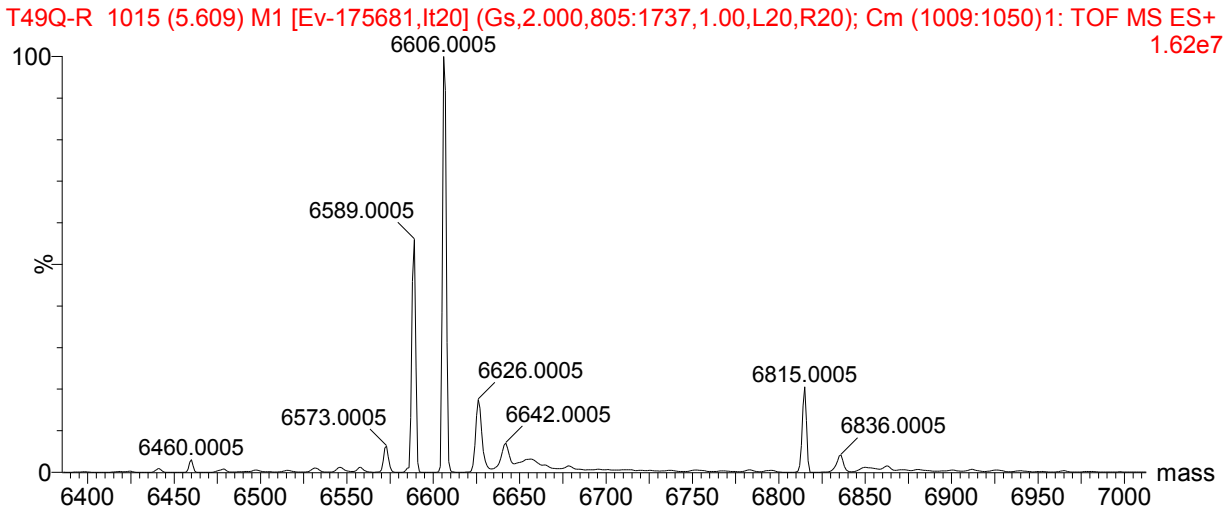
T49Q, No MTG



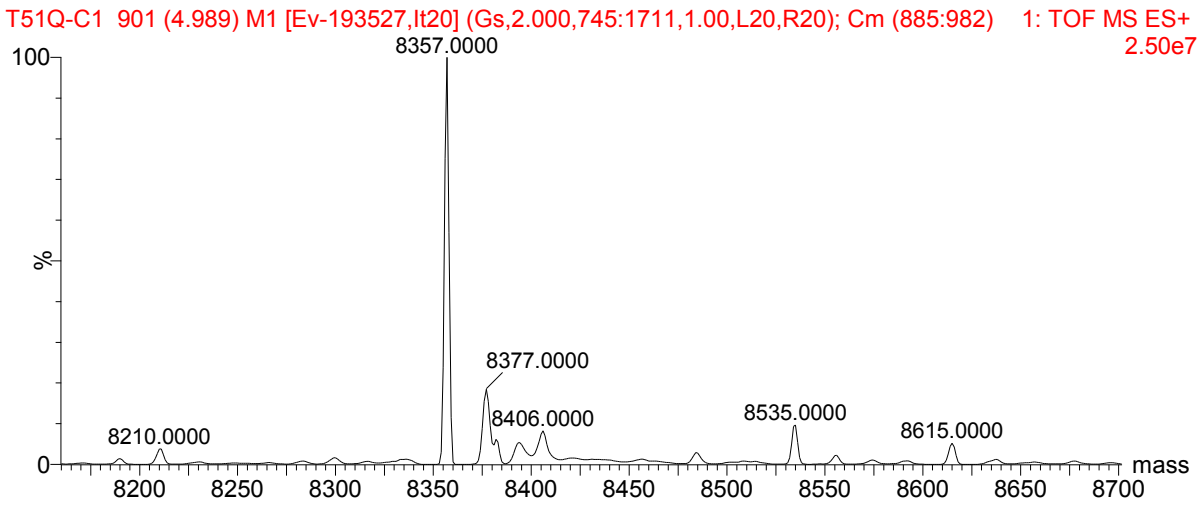
T49Q, MTG present



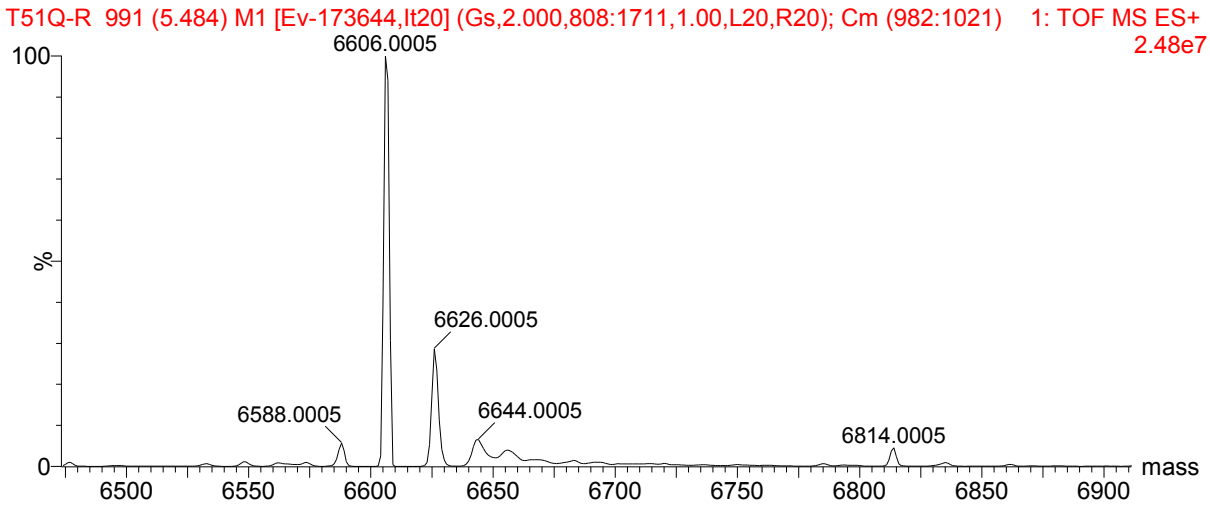
T51Q, No MTG



T51Q, MTG present



T53Q, No MTG



T53Q, MTG present