Identification and characterization of peptide substrates of bacterial transglutaminases for use in bio-conjugation and bio-catalytic applications

SAMUEL OTENG-PABI

Thesis submitted to the

Faculty of Graduate & Postdoctoral Studies

in partial fulfillment of the requirements

for the Doctorate in Philosophy degree in Chemistry

Department of Chemistry and Biomolecular Sciences

Ottawa-Carleton Chemistry Institute

Faculty of Science

University of Ottawa

Abstract

Transglutaminases (protein-glutamine:amine y-glutamyl- transferase, EC 2.3.2.13) are a family of calcium-dependent enzymes which catalyze an acyl transfer between glutamine residues and a wide variety of primary amines. When lysine acts as the acyl-acceptor substrate, α-glutamyl-lysine isopeptide bond is formed. Isopeptide catalyzation results in protein cross-linkage which is prevalent throughout biological processes. Microbial transglutaminase (mTG) is a bacterial variant of the transglutaminase family, distinct by virtue of its calcium-independent catalysis of the isopeptidic bond. Furthermore, mTGs promiscuity in donor substrate preference highlights its biocatalytic potential. To realize the potential of the enzyme, a high-reactivity tag was necessary for protein labelling. To address this, an enzyme-coupled assay was developed to characterize peptides in the hopes of developing orthogonal substrates to facilitate mTG-mediated labelling and biocatalysis. The discovery of high-reactivity peptide tags allowed the realization of *in vitro* protein labelling- facilitated by mTG. The 7M48 peptide was fused to a test protein, where it was subsequently propargylated with propargyl amine to fluorescently label or immobilize a test protein. Although there are endless possibilities for *in vitro* bio-conjugation through mTG, proteolytic activation limits any in-cell labelling strategies with this enzyme. To circumvent this issue, development of an alternative bacterial enzyme, *Bacillus subtilis* transglutaminase (bTG), was chosen to replace mTG. bTG maintains the advantages associated with mTG but is expressed in its active form. Unlike mTG, there is limited preliminary research associated with the enzyme or its substrate scope. To better understanding substrate reactivity, a FRET-based assay was developed allows for the discovery of new high-reactivity peptides for bTG. These peptides were then used in labelling strategies to demonstrate the potential bTG-mediated bioconjugation. This strategy includes the added advantage of potential for *in cellulo* labelling.

Acknowledgements

I don't think there are enough words to describe how thankful I am for my PhD supervisor, Dr. Jeffrey Keillor. I am the first student that was accepted into his research group when he transitioned to the University of Ottawa and I will be forever thankful that he took a shot on me. Throughout my research, I've wanted to be pessimistic on many occasions, but knowing that Jeff would be in my corner constantly gave me the confidence to keep pushing through. His versatility and knowledge ensured that any project we tackled would always have a shot at success. What he lacks in height, he makes up for in compassion. His concern for his student's well-being and mental health was an unexpected bonus that I will not soon forget. He also happens to be one of the few people who are more fun to be around than myself, and that is saying a lot, because I am great time.

I would like to thank the Keillor research group, both past and present members. When I joined the group, it was filled with senior students who helped shape the researcher I have become today. Mirka was an invaluable source of knowledge that pulled me through my initial graduate school nerves and an intriguing conversationalist that constantly kept me entertained. Amina was always available when I needed help with chemistry or figuring out what to eat for dinner; our Brampton connection will always keep us in contact. Not enough can be said about Chris, he deserves his own acknowledgment section. I cannot imagine making it through my graduate studies without Chris. I will just say that his guidance and friendship are more valued then he'll ever know.

As my time in the Keillor group progressed, I met a variety of high-quality researchers that also happened to be amazing people. Kim, Abdullah and Kelvin, Dan, Nicole and Mark all brought an energy to the lab that always made showing up for work enjoyable.

A special thanks to the entire University of Ottawa research community and the department of Chemistry and Biomolecular Sciences in particular. The quality of professors and graduate students fosters an encouraging research environment that helped keep me motivated throughout my research.

Finally, I'd like to thank Lindsay for supporting me throughout my entire PhD. Grad school was more intensive and time-consuming then I ever could have imagined, and she stood by me throughout all the complaining and long hours in the lab with a patient smile and words of encouragement. For that, I will forever be grateful.

Dedication

This is dedicated to my parents. I've always wanted to make you proud…

And all you've asked of me is to be happy and follow my passion....

Thank you.

Table of Contents

List of Tables

List of Figures and Illustrations

List of Symbols, Abbreviations and Nomenclature

Symbol	Definition
PTM	Post-translation modification
POI	Protein of interest
FP	Fluorescent protein
GFP	Green fluorescent protein
RFP	Red fluorescent protein
CuAAc	Copper-catalyzed alkyne-azide cycloadditions
E. coli	*Escherichia coli*
FGE	Formylglycine Generating Enzyme
FGly	Formylglycine
PPtase	Phosphopantetheinyltransferase
Sfp	Phosphopantetheinyltransferase (from *Bacillus subtilis*)
ACP	Acyl carrier protein
PCP	Peptide carrier protein
SrtA	Sortase A
FPP	Farnesyl diphosphate
PFTase	Protein farnesyltransferase
PGGTase	Protein geranylgeranyl transferase
BirA	Biotin ligase (derived from *E. coli*)
LplA	Lipoic acid ligase
TGase	Transglutaminase

mTG	Microbial transglutaminase
CBz-Gln-Gly	N-α-benzyloxycarbonyl-L-glutaminylglycine
bTG	*Bacillus subtilis* transglutaminase
NADH	Nicotinamide adenine dinucleotide
GDH	Glutamate dehydrogenase
Gly-OMe	Glycine methyl ester
α-KG	α-ketoglutarate
SPPS	Solid phase peptide synthesis
MCS	Multiple cloning site
FRET	Förster resonance energy transfer
7M48	Heptamer peptide (WALQRPH)
6K	Hexa-lysine peptide
FACS	Fluorescence-activated cell sorting
SDS-PAGE	Sodium dodecyl sulfate- polyacrylamide gel electrophoresis
clickable	Functional group that can undergo cycloaddition
n.d.	Not detected

Epigraph

"Never say can't, it's better to try"

- Jadakiss, *Keep Ya Head Up*

Chapter One: **Introduction to Protein Labelling**

1.1 Importance of protein labelling

An understanding of gene function cannot be identified strictly through the sequencing of DNA. To definitively ascertain the functional role of proteins, reading their corresponding primary sequences provides limited information.

For most genes, assigning functional roles will require methods for monitoring the activity and interactions of the gene products. Throughout nature, the post-translation modification (PTM) of proteins creates a surge in biological functionality that would otherwise be impossible. PTM's encompass a wide range of biochemical mechanisms that result in a covalently modified amino acid residue, typically on the surface of a protein. Although there are various functions of PTM, they usually involve some form of cellular information processing. The complexity of cellular signalling dictates that nature develop numerous pathways to PTM, depending on the circumstances. Nature has provided "tools" to facilitate the formation of specific covalent bond through enzymes. The highly specific and selective nature of enzymes allow them to mediate otherwise difficult reactions with precision and speed. Enzymatic modifications to a protein display definitive advantages that cannot be replicated by any other means[1]. A wide range of methods for the chemical modification of proteins has also been developed[2]; however, these methods usually result in the selective labelling of a given functional group, rather than the site-specific labelling of one residue of a specific protein. Conversely, labelling proteins in living cells, can garner critical information regarding their expression, localization and trafficking. With PTMs providing the blueprint, enzymatic modifications of a protein of interest (POI) to a reporter is an advantageous method to elucidate information.

1.2 Fluorescent Labelling

In order to effectively monitor a POI, fluorescence tags have been used as a tool for detection[2,3]. Fluorescence is the phenomenon of an electron absorbing enough energy to leave its ground state and enter an excited state orbital, subsequently followed by a return to its original ground state and a release of energy, via emission of a photon[4]. The emitted photon results in the photometric observation that can be monitored visually[4]. The multitude of fluorophores that have been discovered or synthesized affords the opportunity for versatility in fluorescent labelling[5]. The methods in which fluorescent tags are crosslinked is always evolving; however, the fluorescent tags generally fall into two categories: Small fluorescent molecules and fluorescent proteins. Small fluorescent molecules and fluorescent proteins will be discussed here as they are the main analytical tools used for biomolecule detection.

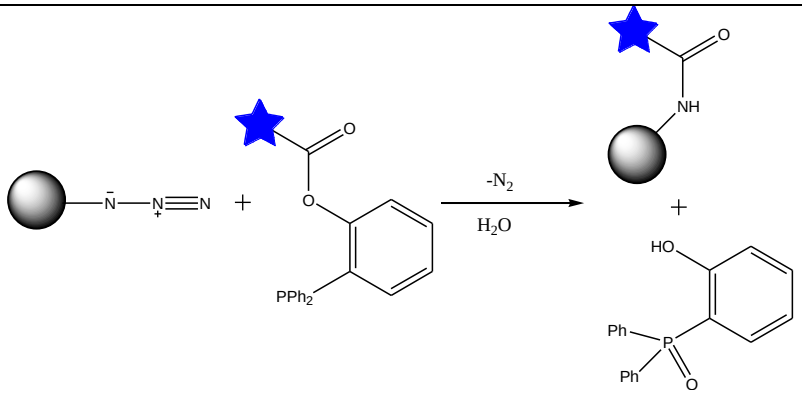
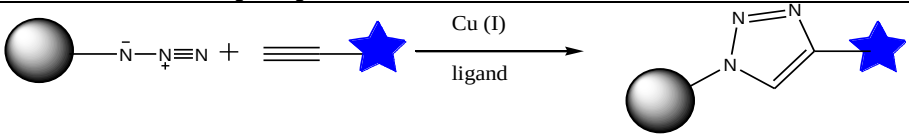
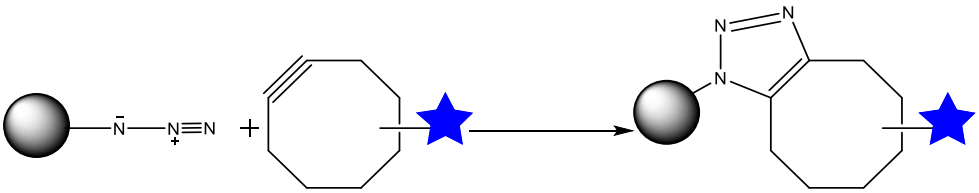
The most common method for protein labelling involves the incorporation of an intrinsically fluorescent protein (FP) through genetic fusion onto a POI[6]. The natural fluorescence of the FP allows localization of a POI through visualization of the FP. Many studies have characterized a wide range of FPs that emit at various wavelengths along the visual spectrum. The predecessor to all fluorescent protein biochemistry is the green fluorescent protein (GFP) originally isolated and cloned from *Aequorea victoria*[7]. Currently, there are a multitude of FP's that have exhibited enhanced properties such as elevated quantum yield and stability. Unfortunately, slow maturation, tendency to aggregate, and concerns over intracellular trafficking interference limit the functionality of FPs[8]. Small fluorescent molecules eliminate many of these issues, although toxicity is a concern in some cases[9,10].

Covalently tagging small fluorescent molecules or fluorescent proteins onto a POI has been thoroughly studied and refined in recent history. Regardless of the fluorescent probe, a multitude of chemical and enzymatic tagging techniques have been characterized[2,5,11,12]. With either strategy, there are inherent limitations. Non-enzymatic modifications may not be the most efficient or selective; furthermore, the conditions required to facilitate the reaction may limit the scope of the labelling strategy[12]. Enzymatic modifications are generally highly efficient and selective, but are limited in the scope of substrate specificity to fluorescent proteins. Depending on concerns over brightness, background, photo-stability and spectral overlap, fluorescent proteins may not be sufficient[8]. To widen the scope of labelling solutions, chemo-enzymatic labelling strategies have been developed. Ideally, these hybrid strategies reduce the limitations of either method while maintaining their advantages.

1.2.1 Bio-orthogonal reactions for protein labelling

A critical factor to the success of chemo-enzymatic solutions is that any reaction used to covalently attach a biological probe to a POI must be bio-orthogonal. Bio-orthogonal reactions are selected for protein modification to ensure that any labelling events are in response to a POI that has been enzymatically modified to introduce a chemical "handle" on the protein[13]. Once modified, the newly-introduced functional group is now susceptible to chemical manipulation. **Table 1.1** identifies a few commonly used bio-orthogonal reactions,

Table 1.1 Bio-orthogonal Reactions for Protein Labelling

Reaction	Scheme
Staudinger ligation[2,14] (traceless)	 - *in cellulo*, phosphines are oxidation-sensitive
CuAAC (Cu^I-catalyzed alkyne-azide cycloadditions)[15]	 - *In vitro*
SPAAC (strain-promoted alkyne-azide cycloadditions)[16]	 - *In cellulo*

= Protein of Interest

= Fluorophore/Fluorescent Protein/Solid Support

1.3 Site-Specific Protein Labelling

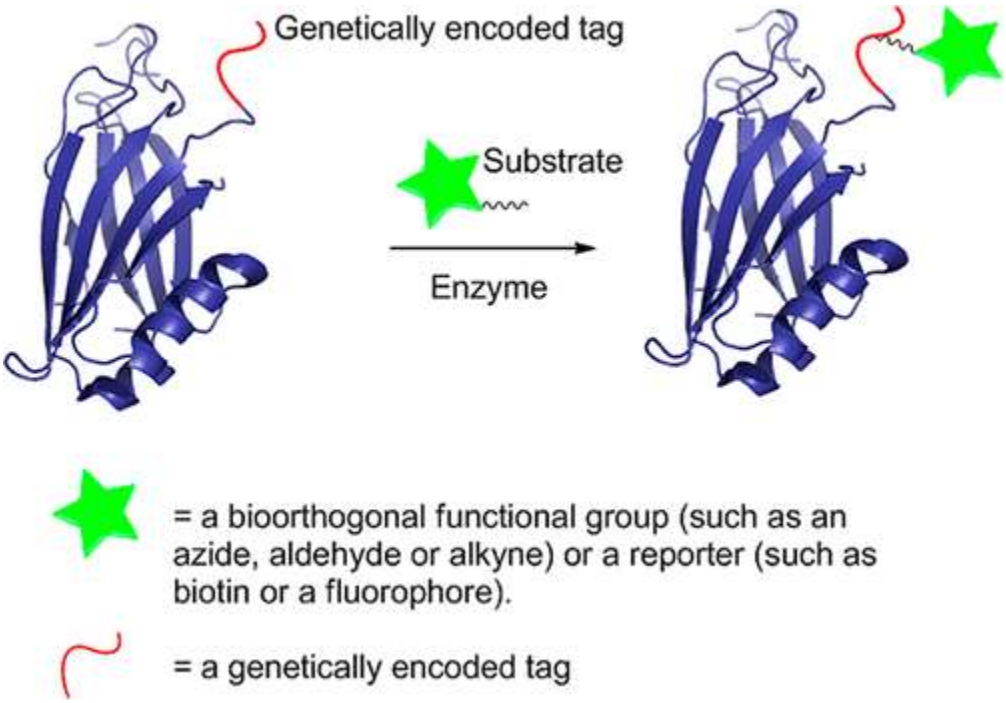


Figure 1.1 Illustration of site-specific covalent modification of POI through a genetically encoded tag. Enzyme incorporation of substrate introduces bio-orthogonal functional group to protein. (Adapted with permission from Rashidian, et.al., Bioconjugate Chemistry, 2013. Copyright 2017 American Chemical Society)

Several methods for site-specific labelling of specific proteins have emerged over time[12]. One of the most important of these techniques involves the genetic fusion of a short peptide sequence onto a POI[17], followed by the site-specific modification of this peptide tag, using a co-expressed enzyme.

Ideally, bio-conjugation catalysts achieve a high level of efficiency while maintaining a broad substrate specificity. This ensures that the methodology is flexible to the array of biochemical probes available to monitor or immobilize a POI. Several enzymes have been used to this end,

including farnesyl transferase, biotin ligase, myristyl transferase, formylglycine- generating enzyme, sortase, lipoic acid ligase and transglutaminase[12]. Herein, the covalent modification of POI's is described as it relates to each of these enzymes. Furthermore, the advantages and disadvantages of each chemoenzymatic method is discussed as well as biological applications of the technique.

1.3.1 Formylglycine Generating Enzyme

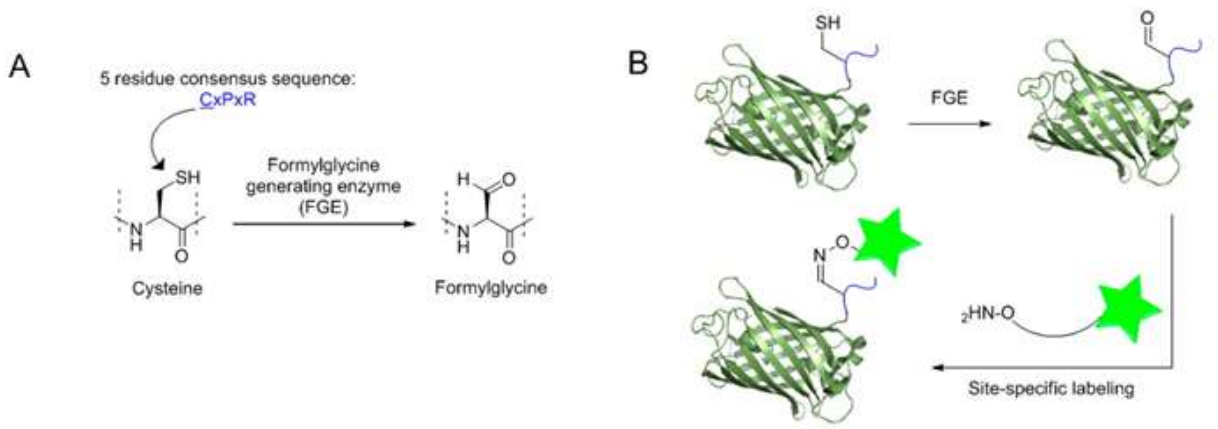


Figure 1.2 Formylglycine Generating Enzyme. Illustration of formylglycine generating enzyme reaction. Covalent modification of genetically encoded aldehyde tag. A) Oxidation of cysteine to formylglycine by FGE. B) FGE-facilitated protein modification and subsequent oxime ligation reaction to label POI. (Adapted with permission from Rashidian, et.al., Bioconjugate Chemistry, 2013. Copyright 2017 American Chemical Society)

Formylglycine Generating Enzyme (FGE) natively modifies type I sulfatases through the conversion of a cysteine, within a **C**xPxR amino acid sequence, to an aldehyde-containing formylglycine residue (FGly). After the post-translational modification has taken place, the protein is now functionalized for site specific addition. The second step to this chemo-enzymatic

modification takes advantage of aminooxy- or hydrazide- functionalized reagents, leading to a new covalent bond[18].

Many groups have worked on the refining the FGE strategy, including the Bertozzi group, whose screening of FGEs has identified new peptides sequences that diverge from the known FGE-motif[18,19].

FGE modification can be utilized as a versatile tool for site-specific modification, specifically for cytosolic and membrane-associated proteins[19]. The relatively short peptide tag presents a non-intrusive method to generate modified proteins of interest. The varying length of the sequence (6-13 residues) is based on the necessity for Cys to FGly conversion efficiency[20]. Although shorter tags are recognized, a lengthier recognition motif improves the accessibility of the tag to the FGE active site. Although the modification is simple, it does require the exogenous addition of a tagging reagent once the FGE-based alteration is complete. This limitation can be critical depending on the locale of the desired protein[18,20].

1.3.2 Phosphopantetheinyltransferase

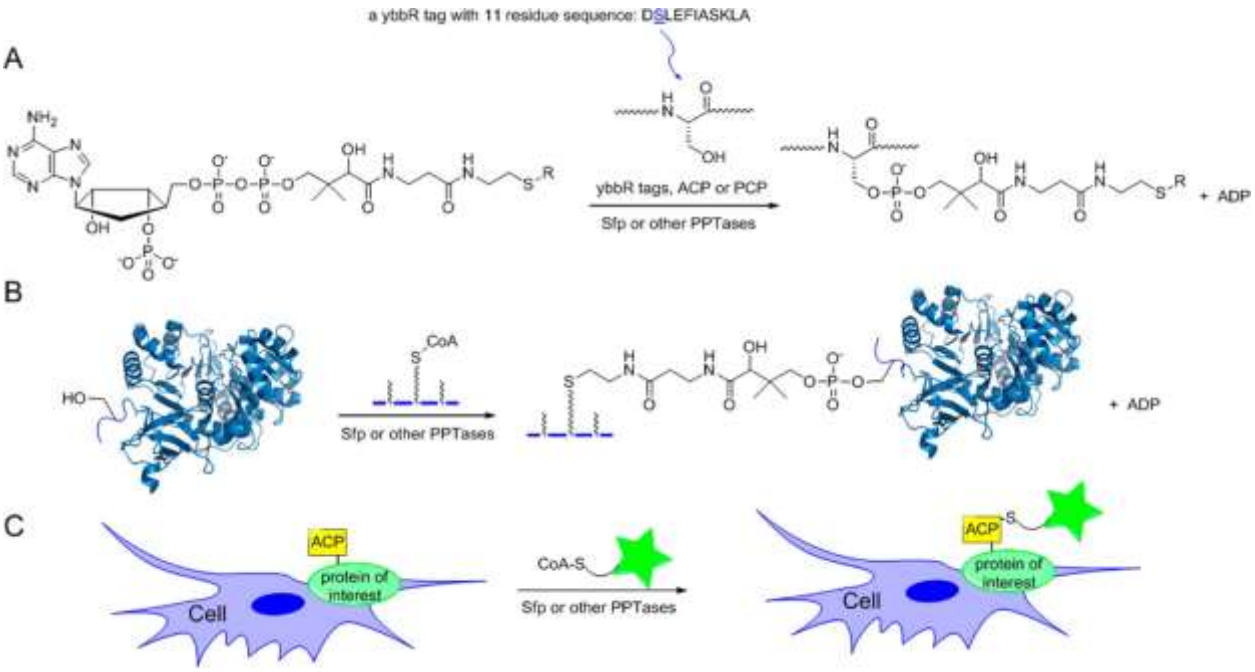


Figure 1.3 Phosphopantetheinyltransferase. A) Modification of PCP or ACP at serine residue by Sfp. B) Protein immobilization through an immobilized CoA and a POI containing recognition sequence for Sfp/PPTases. C) Cell surface labelling facilitated by PPtase, through ACP/PCP/ybbR tag and CoA-fluorophore conjugate. (Adapted with permission from Rashidian, et.al., Bioconjugate Chemistry, 2013. Copyright 2017 American Chemical Society)

Phosphopantetheinyltransferases (PPTases) are a family of enzymes that are responsible for the post-translation modification of peptide carrier protein (PCP) and acyl carrier protein (ACP)[21]. PCP and ACP are crucial in the biosynthetic pathways of nonribosomal peptide synthetases (NRPSs), fatty acid synthases (FASs) and polyketide synthase (PKSs)[21]. PPTase enzymatically modifies all PCPs and ACPs, leading to the activation of these domains[22]. PPTase introduces a phosphopantetheinyl (Ppant) prosthetic group onto the surface of the protein. The Ppant group is covalently attached through a phosphodiester bond to the hydroxyl group of a conserved serine within the N- or C- terminus. Coenzyme A (CoASH) is the native substrate for PPTases and is

precursor from which the Ppant group is derived[23]. Once linked to the ACP/PCP domain, Ppant acts as an anchor for the elongation of peptide, fatty acid chain or polyketide as the chains grow and travel through their biosynthetic machinery[23].

Several research groups have exploited PPTases for their potential as a tool for protein labelling. Sfp, a member of the PPTase family and derived from *B. subtilis*, maintains broad substrate specificity when a range of molecules are conjugated via the thiol linker of Coenzyme A (CoA). Fusing PCP/ACP to a protein of interest permits the targeting of proteins with a variety of labelling or immobilizing agents. This technique is highly efficient while maintaining the promiscuity for small molecule probes. Furthermore, the recognition domain is significantly smaller than that of GFP (80 AA to 238 AA). Unfortunately, the PCP/ACP domain is still significantly larger than common peptide recognition sites. To remedy thisissues, an 11-residue tag, ybbR (DSLEFIASKLA), was identified as a substrate for Sfp via a phage display of genomic *B. subtilis* library. ybbR maintains the high levels of efficiency of the full-length domains, while introducing a level of flexibility in the locale of the recognition tag on the protein of interest[24].

The Sfp strategy for labelling allows for broad specificity and high efficiency, two important factors in labelling. However, limiting the substrates to CoA- thiol moiety may not be ideal; a more flexible handle to introduce probes to a POI is desired. Furthermore the additional synthesis required to introduce a probe onto the CoA- thiol moiety is added complication[25].

1.3.3 Sortase

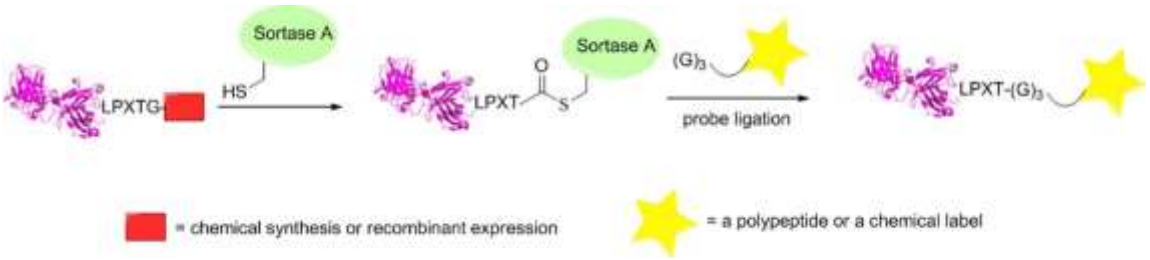


Figure 1.4 Sortase. Illustration of sortase-facilitated covalent modification. Sortase A forms covalent acyl-enzyme intermediate with LPXTG recognition sequence. Intermediate is collapsed by attack from G_3 probe, leading to new peptide bond between probe and LPXT sequence. (Adapted with permission from Rashidian, et.al., Bioconjugate Chemistry, 2013. Copyright 2017 American Chemical Society)

Natively, sortases are essential for the biosynthesis of cell walls as well as the covalent linkage of proteins to peptidoglycan cell wall[26]. As a tool for protein labelling, sortases have generated a lot of interest. The ability to tag proteins without perturbing the structure is essential to any effective tagging solution. Sortases catalyze crosslinkage via a transpeptidase reaction between a small peptide recognition sequence and the alpha-amino group of the N terminal glycine residue[27]. Sortase A (SrtA) is derived from *S. aureus* and is the most commonly used sortase for protein labelling. SrtA recognizes the pentapeptide sequence LPXTG and facilitates the cleavage of amide bond between threonine and glycine, leading to a thioester intermediate[28]. The thiol moiety is then displaced by an N-terminal oligo-glycine substrate, creating a new covalent bond, LPXT-GGG[28]. This method is referred to as "sortagging" and has been demonstrated on cell surfaces as well as *in cellulo*[28].

This tagging method is very effective; however, the by-product of the first reaction step, glycine, is highly reactive. The introduction of a reactive glycine can lead to a nucleophilic attack, resulting in the re-formation of the original species. This limitation results in an inability to label

in a quantitative manner. To address this issue, excess G_3-probe can be used for labelling. This however is not an efficient use of material. If the substrate is costly or time consuming to prepare, this is not a viable solution. As a response to this problem, a pentapeptide was synthesized where the amide bond between threonine and glycine was replaced with an ester bond[28]. Theoretically, this would have little effect on the forward reaction, while significantly reducing the nucleophilicity of the released product. This alteration to the recognition sequence was successful, permitting labelling at a quantitative efficiency. Unfortunately, utilizing a synthetic tag, rather than a genetically-encodable sequence limits the scope of Sortase A as tool for bio-conjugation.

1.3.4 Farnesyltransferase

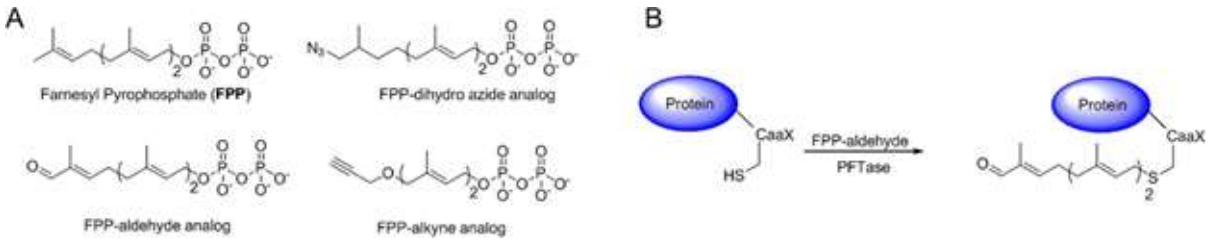


Figure 1.5 Farnesyltransferase. A) Structures of farnesyl diphosphate (FPP) and FPP azide, aldehyde and alkyne derivatives. B) Prenylation of protein containing CaaX recognition tag on C-terminus with FPP-aldehyde. (Adapted with permission from Rashidian, et.al., Bioconjugate Chemistry, 2013. Copyright 2017 American Chemical Society)

Protein farnesyltransferase (PFTase) catalyzes the transfer of a farnesyl isoprenoid group to form farnesyl pyrophosphate (FPP) on the C terminus of a protein. This covalent attachment is forged through a thioester bond with the sulfur atom of a cysteine residue[12]. C- terminal cysteine

residues are specifically recognized when found within a "CaaX" tetrapetide recognition sequence (C= cysteine, a= aliphatic amino acid, X= one of a variety of amino acids)[29]. PGGTase catalyzes the same reaction as PFTase, but with the farnesyl substrate replaced by the larger geranylgeranyl isoprenoid substrate. Both substrates can be functionalized with a variety of groups that enable subsequent biorthogonal chemistry[30].

PFTase and PGGTase have been utilized *in vitro* and show high levels of promiscuity to modifications in the FPP GGPP structures[31]. Furthermore, the short peptide recognition sequence is ideal for protein labelling as it will have little to no effect on the structure/function of the POI. A drawback to this technique has been identified in the azido-FPP substrate[32]. Azide containing analogues undergo a thermal rearrangement that yields a combination of isomers, because of the allylic azide[32]. Azides are the most commonly used functional groups in bio-orthogonal chemistry, so the necessity of removing this site selective modification is a detriment to the technique[33]. However, researchers have addressed this drawback by replacing allylic azides with aliphatic azides (dihydroazide analogue)[33]. Another deterrent to isoprenoid transferases are the increased hydrophobicity of the modified proteins, especially with GGPP[34]. More polar substrates were utilized as a substitute that contain a single isoprenoid unit linked to an alkyne[34,35]. However, this substrate appears to be far less efficient. Counteracting this by increasing the concentration of enzyme used has proven to be an effective solution, but more costly[34].

1.3.5 Biotin Ligase

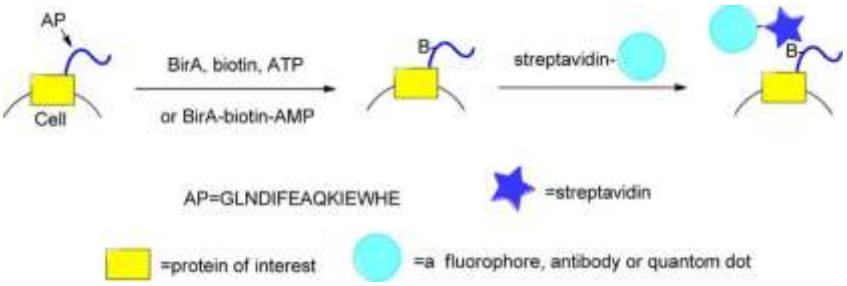


Figure 1.6 Biotin Ligase. Schematic representation of biotin ligase recognizes acceptor peptide that has been incorporated onto POI, Biotinylation of protein allows for the incorporation of fluorophore through streptavidin-biotin binding. (Adapted with permission from Rashidian, et.al., Bioconjugate Chemistry, 2013. Copyright 2017 American Chemical Society)

Biotin ligases facilitate the attachment of biotin onto proteins through a recognition peptide sequence[12]. The site-specific modification of proteins is only made possible through a genetically fused tag onto a POI. A covalent bond between a biotin and a lysine residue within the recognition sequence sets up a functionalized handle of the POI to which avidin can bind. Although the biotin-avidin interaction is noncovalent, the strength of the newly formed conjugate (K_d= 10^{-15} M) is sufficient to label a protein[36].

Biotin ligase derived from *E. coli*, BirA, is the most common ligase used to attach biotin to a POI. The acceptor peptide, AP= GLNDIFEAQKIEWHE, is a highly efficient substrate[37]. Further studies identified biotin analogues containing azide/alkyne that are also act as substrates for BirA. This increase in substrate scope affords this technique a multitude of bio-orthogonal handles to modify the POI. However, the use of BirA as an enzymatic labelling tool has only been reported on *in cellulo*[38].

1.3.6 Lipoic Acid Ligase

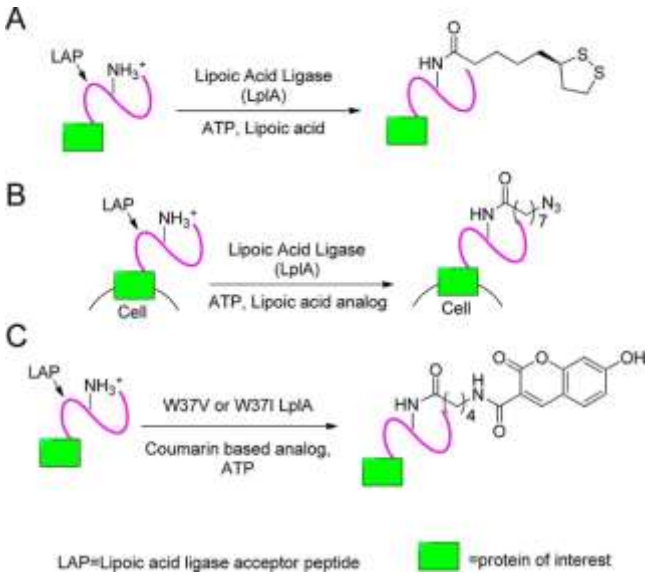


Figure 1.7 Lipoic Acid Ligase. Illustration of lipoic acid ligase (LplA) protein labelling. A) Native ligation of lipoic acid, catalyzed by LplA. B) LplA ligation of lipoic acid fused with azide to LAP domain on the surface of a cell. C) Mutated coumarin ligase, site-specifically incorporating lipoic acid coumarin probe to protein. (Adapted with permission from Rashidian, et.al., Bioconjugate Chemistry, 2013. Copyright 2017 American Chemical Society)

Similar to biotin ligase, lipoic acid ligase, LplA, takes advantage of a lipoic acid and a naturally occurring lysine to site-specifically modify a POI. The amine on the sidechain of lysine is covalently linked to the carboxylic lipoic acid[39]. Functionalized lipoic acid was also tested as a potential substrate for LplA; although most analogues result in varying levels of activity, the azide-containing analogue yielded the highest activity. Naturally, LplA recognizes a large domain to identify the lysine, which may interfere structurally or with the functionality of the POI. Therefore, studies were done to identify a smaller peptide tag, resulting in the discovery of a 22-amino acid sequence, referred to as LAP[39]. LAP maintained high levels of efficient reactivity without any significant drawbacks to the POI. Although the LAP tag is smaller than fusion domains, there is still potential for interference from the 22 residue tag[39].

1.3.7 N-Myristoyltransferase

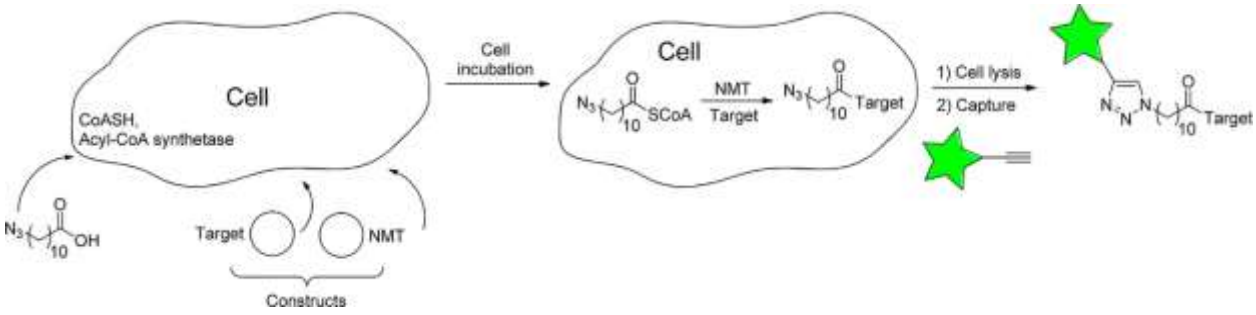


Figure 1.8 N-Myristoyltransferase. Schematic representation of N-Myristoyltransferase-mediated protein labelling. Cells expressing N-Myristoyltransferase (NMT) are incubated with myristic acid analogues so that the analogues are converted to CoA derivatives and fused to the target protein. Modified proteins are labelled or captured using chemoenzymatic reactions. (Adapted with permission from Rashidian, et.al., Bioconjugate Chemistry, 2013. Copyright 2017 American Chemical Society)

The process of N-myristoylation involves the exchange of myristate from myristoyl-CoA via an acyl transfer to the N-terminal of a glycine, resulting in an amide bond[40]. N-myristoyltransferase (NMT) catalyzes this reaction through the recognition site of GXXXS/T, where X can be any number of amino acids[41]. Initially thought to occur during protein translation, recent evidence suggest that N-myristoylation may also occur post-translationally. Given that the known targets for N-myristoylation include tyrosine kinases and HIV-1 matrix protein to name a few, utilizing NMT to label POI may prove to be useful for therapeutics[42]. NMT is also known to be somewhat promiscuous in the myristoylate analogues that are recognized, specifically azide and alkyne derivatives[42]. Through metabolic incorporation, *E. coli* cells could express NMT and the POI, followed by the addition of a functionalized lipid moiety site-specifically labelling a POI. Overall, this technique provides an efficient and convenient method for protein labelling at the

N-terminus of POIs. However, this may prove to be limiting in systems where the N-terminus is not readily available for modification[42].

1.3.8 Transglutaminase

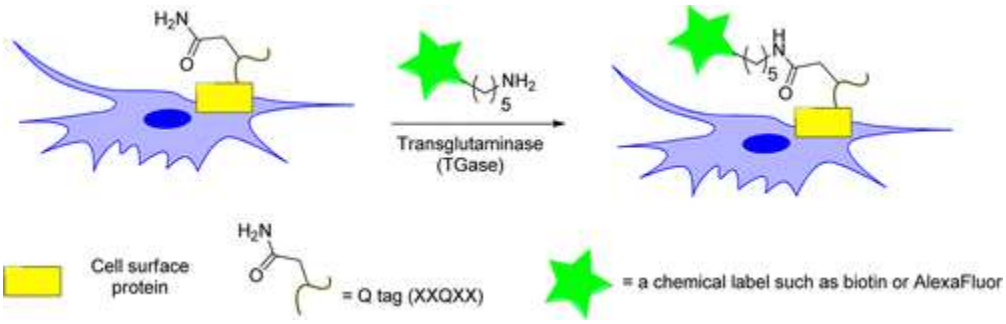


Figure 1.9 Transglutaminase. Site-specific modification of cell-surface protein mediated by transglutaminase activity. Q-tag recognition sequence permits selective incorporation of iso-peptide bond with amine-probe to fluorescently label protein. (Adapted with permission from Rashidian, et.al., Bioconjugate Chemistry, 2013. Copyright 2017 American Chemical Society).

Transglutaminases (TGases) are a family of calcium-dependant enzymes that have a multitude of biological roles. The TGase class of enzymes catalyzes an acyl transfer between glutamine's amide side chain and a primary amine, typically the ε-amino group of lysine[43]. The resulting carboxamide cross-linkage opens the door to a wide variety of protein-conjugates[44]. Natively, glutamines substrates are found within flexible regions within proteins and are flanked by a TGase recognition site. A wide variety of diseases have been linked to TGase abnormal activity, making TGase inhibition very relevant to therapeutic studies[45–47]. TGase is known to have a

broad specificity in regard to the primary amine substrate. This has been taken advantage of in many studies that demonstrate the potential for TGase as a catalyst for bio-conjugation[43].

TGase has been utilized to develop numerous conjugates, including protein-small molecule, and protein-biomolecules[48]. It has also been identified as a tool for immobilization and *in cellulo* labelling[49]. As mentioned previously, a glutamine peptide recognition sequence is required for TGase activity; this sequence is TGase-dependant but is generally limited to a XXQXX pentapeptide motif[12]. It is also worth noting that not all TGase sequences are specific to their TGases and may have some level of activity with other homologs[12].

1.3.9 Summary of common enzymes used for site-specific modifications

After reviewing a variety of enzymes capable of covalently modifying POI's, multiple differences can be observed. Length and composition of substrate recognition sequence, substrate specificity, and position of covalent modification all vary depending on the enzyme.

Table 1.2 Summary of enzymes used for protein labelling. Summary of site-specific protein labelling (Adapted with permission from Rashidian, et.al., Bioconjugate Chemistry, 2013. Copyright 2017 American Chemical Society)

Enzyme	**Recognition tag**	**k_{cat} min^{-1}**	**K_M µM**	**k_{cat}/K_M min^{-1} μM^{-1}**	**Site of peptide tag**
Formylglycine generating enzyme	CXPXR	-	-	-	C or N terminus
Phosphopantetheinyl transferase	ACP, PCP or ybbR	-	500 uM (ybbR)	-	C or N terminus or flexible loop
Myristyl transferase	GXXXS/T	-	-	-	N-terminus
Sortase	LPXTG	14.7	60.8	0.242	Any section
Farnesyltransferase	CaaX	31.2	1.71	18.2	C-terminus
Biotin ligase	GLNDIFEAQKIEWHE	7.8	1.87	4.17	C or N terminus
Lipoic acid ligase	GFEIDKVWYDLDA	9.6	4.2	2.3	C or N terminus or flexible loops
Transglutaminase	XXQXX	45	7	6	Any section

Table 1.2 highlights these differences in a clear concise manner. For example: Phosphopantetheinyl transferase, biotin ligase and lipoic acid ligase recognize large 13-mer or more tags, whereas farnesyl transferase, myristyl transferase, sortase, formylglycine-generating enzyme and TGase recognize minimal 4-6 residue tags. Larger tags can be potentially limiting protein modification; *in cellulo* studies require the least amount of perturbation to the function of the POI. The larger the tag, the less likely that a modified POI will mimic the function of its native precursor[12] (in general, longer tags do improve selectivity). Additionally, it is important to consider the functional groups associated with specific residues as they may lead to unintended consequences. The polarity or charge of amino acid side chains can have a detrimental impact on the POI or the reaction environment.

Apart from the size of the genetically encodable tag, the location is the peptide sequence can be equally as important. Most tags are placed at either terminus of the POI due to ease of accessibility and reduced likelihood to affect protein folding. However, many proteins have termini that are buried after tertiary structures are formed[50]. In this case, flexibility in tag location is advantageous. Only TGase and Sfp can modify the side-chain of its peptide substrate sequence regardless of where this tag is located on the POI, setting them apart from the rest as having the most versatility in regards to protein targets[51,52].

Not only is the size of the tag important but so is the size of the addition to the POI. Similar to the concerns about a tag's ability to perturb function, a large enzymatic addition may also cause unwanted drawbacks to the POI[53]. On the other hand, addition of a peptide, polymer or protein may be the desired consequence; this also leads to considering limitations of certain enzymatic modification strategies. Sortase and phosphopantetheinyltransferase are the recommended enzymes when flexibility of addition is desired. Sortase has no limits in addition size, and

PPTase is flexible in regards to large modifications[54]. The other methods of labelling must catalyze the incorporation of analogues of similar size to their native substrates for catalysis to occur. TGase can bypass its active site limitation; lysine analogues have been used for bio-conjugation, regardless of the size of the rest of the substrate[55].

Finally, the rate at which the reaction occurs is important when deciding on which enzyme is the most viable for protein labeling. Depending on the half-life of the POI, it may be critical to expeditiously modify the protein. Farnesyl transferase and TGase maintain relatively rapid enzymatic reaction rates, while other enzymes may be sluggish and/or unstable over the duration of a labelling experiment[55,56].

After reviewing the different variables of the highlighted protein labeling solutions, transglutaminase appears to have the most potential as tool for bio-conjugation and merits further exploration.

1.4 Transglutaminase Family

1.4.1 Tissue Transglutaminase

As mentioned previously, transglutaminases (EC 2.3.2.13,amine-γ-glutamyltransferases) are a family of Ca^{2+}-dependent enzymes. Natively, TGase catalyzes the acyl transfer of the carboxamide moiety of a glutamine residue (acting as an acyl donor) to a free amine, typically the side chain of lysine (acting as an acyl acceptor) [45]. This activity results in the formation of an intermolecular amide crosslinking. The most commonly studied member of this family of enzymes is human tissue transglutaminase (TG2). Due to its ubiquitous nature, its activity has been characterized in many different contexts of unregulated activity that may signal a multitude of diseases[45]. TG2 has been identified as a contributor to cataracts and celiac disease; increasing evidence suggest a link to autoimmune disease[57] and cancer metastasis[58]. Furthermore, TG2 is believe to play a role in neurodegenerative diseases such as Huntington's disease[59], Alzheimer's disease[60] and Parkinson's disease[61]. Due to its role in a wide variety of abnormal behaviour, TG2 is considered a therapeutic target for many researchers.

As mentioned previously, TGase, specifically TG2, has been identified as a tool for protein labelling and has been utilized to site-specifically label a POI, *in vitro*. The multiple advantages to TG2 have been outline; however, there are many drawbacks that limit the potential of this enzyme's versatility as a tool for bio-conjugation. Allosteric regulation by calcium and guanosine-5'-triphosphate (GTP) limit the conditions in which labeling with TG2 can occur[47]. Also, the ubiquitous nature of TG2 insists that endogenous expression levels of the enzyme must be considered when developing an enzymatic labeling strategy. TG2 is also a relatively large

protein, at 76 kDa[62], there are concerns with functional interference during in-cell labelling. Overall, these factors limit the *in cellulo* application of the enzyme.

1.4.2 Microbial Transglutaminase

Another member of the TGase family includes bacterial transglutaminases. Discovered through the screening of micro-organisms, bacterial transglutaminases have been extensively adopted by the food industry (see below). Although bacterial TGases catalyze the same reaction as their mammalian homologs, they bear no structural homology to their counterparts. More importantly, bacterial transglutaminases are not restricted by the same allosteric regulation as TG2 or other mammalian TGases[63]. Along with the broader substrate specificity and lower deamidation activity, bacterial transglutaminases have great potential as a biocatalyst for protein labelling.

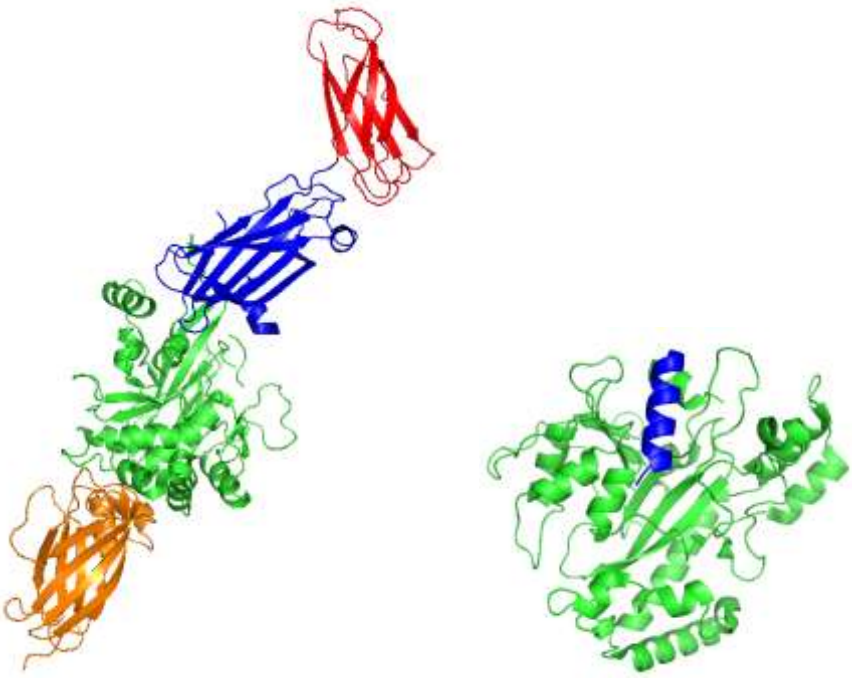


Figure 1.10 TG2 vs mTG. Crystal structures of TG2**.** PDB: 2QEZ (left) vs. mTG, PDB: 3IU0 (right).

1.4.3 Structure and Function

The most prevalent bacterial transglutaminase is derived from *Streptomyces mobaraensis*, and named microbial transglutaminase (mTG). mTG has long been used in the food industry to act as a protein binding agent for meat and fish[64]. This is beneficial to the processing of different foods (ex. Sausages) as well improving the texture and utilizing lower quality meats. Many high quality reviews have been published on the application of mTG in the food industry[64].

Although its biological function is not well understood, it has been hypothesized that the cross-linking activity acts as defense mechanism against host proteases. mTG is postulated to target

inhibitory proteins and deactivate them through crosslinking. This allows the formation of spores to persist, sustaining the life cycle of *Streptomyces mobaraensis*[65].

Structurally, mTG has limited consensus with TG2; at 38 kDa, minimal homology is expected. **Figure 1.10** demonstrates the lack of similarity between the two proteins. TG2 is a well characterized enzyme and is known to contain four domains flanking a Cys-His-Asp catalytic triad. Once activated allosterically by calcium, TG2 undergoes a conformation change that opens the enzyme, exposing its active site[66]. mTG is much less complex, containing a central eight-stranded β-sheet that is surrounded by 11 α-helices as shown in **Figure 1.10**. The active site is found at the bottom of a 16-Å cleft; at its base, the cysteine catalytic triad can be found. The mechanism of mTG is not known but given the similarities in the active site, it is proposed to undergo a similar mode of reaction as TG2, which is reminiscent of the acyl-transfer reactions catalysed by cysteine proteases[67]. **Figure 1.11** highlights the proposed mechanism of action for mTG.

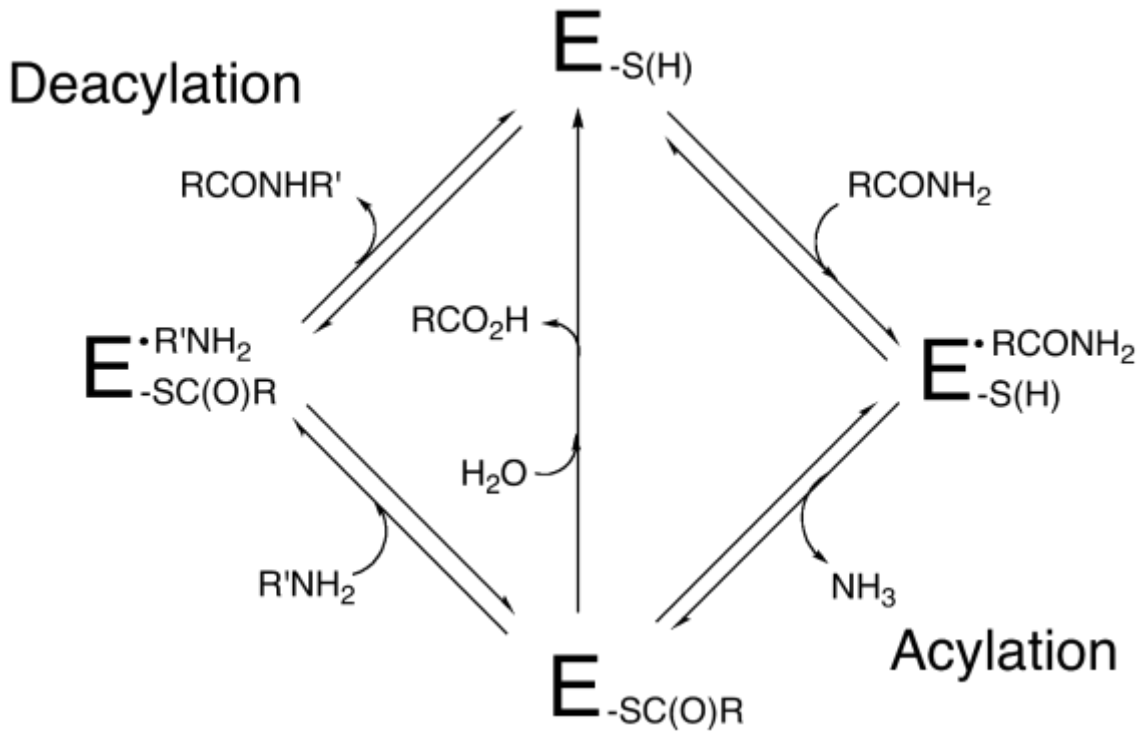


Figure 1.11 Proposed Mechanism of action of microbial transglutaminase.

The transamidation mechanism begins with the nucleophilic attack by the thiol group of an active-site cysteine residue on the donor substrate carboxamide group, resulting in the loss of an equivalent of ammonia and formation of a covalent thioester intermediate. The acyl group of the transient thioester is transferred to the acceptor amine substrate in the second step to complete transamidation. At a much slower rate, water is used as a nucleophile to eliminate the thioester intermediate, leading to hydrolysis of the carboxamide bond[62].

Due to expression and solubility issues, mTG is primarily expressed as a zymogen. The inactive enzyme is activated through cleavage of an N-terminal propeptide that can be found shielding the hydrophobic cleft of mTG[68]. Natively, cleavage is delayed until the enzyme has properly folded and has been secreted. Once secreted, mTG is cleaved by a protease, typically tripeptidyl

aminopeptidase or a metalloprotease[68]. Mature mTG is compared with its zymogen precursor in **Figure 1.12**, demonstrating the only difference as the removal of the propeptide.

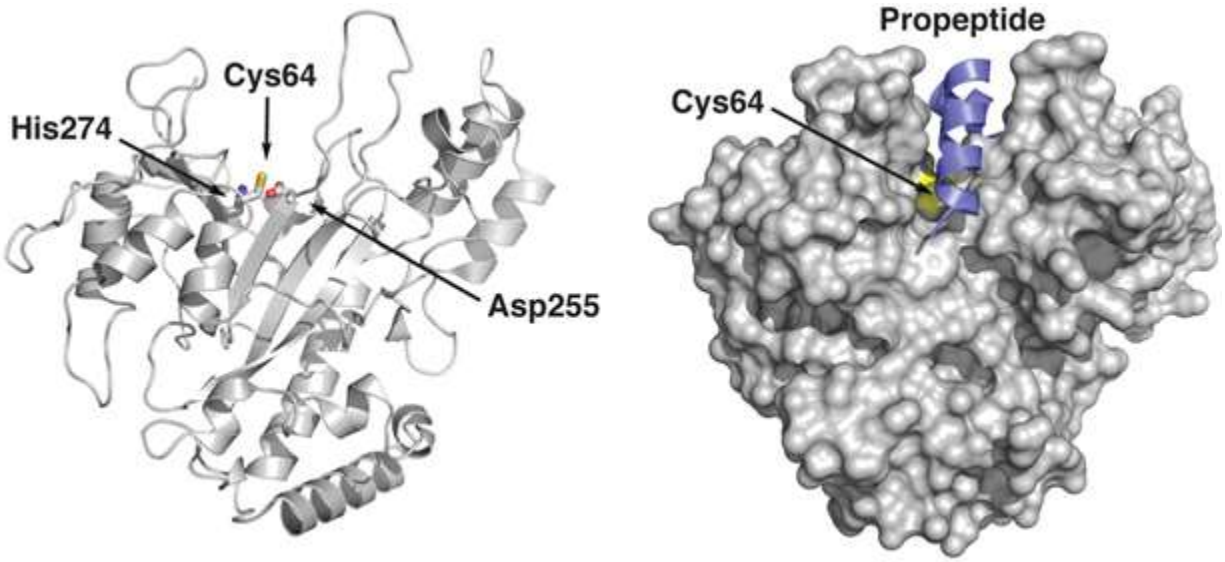


Figure 1.12 Crystal structure of microbial transglutaminase. Left: mature mTG with catalytic triad highlighted (Cys64, His274 and Asp255). Shown in cartoon representation. Right: Zymogen mTG, highlighting catalytic cysteine within active site covered by α-helical pro-peptide. (Adapted with permission from Strop, P., Bioconjugate Chemistry, 2014. Copyright 2017 American Chemical Society)

1.4.4 Substrate Specificity

As previously discussed, an effective tool for bio-conjugation must display a relatively quick reaction time, and efficient reactivity. The key to mTG substrate recognition relies on the peptide sequence flanking the reactive glutamine of any peptide substrate. Studies to elucidate the substrate scope of the acyl acceptor have also been carried out and it is evident that the closer a substrate resembles lysine, the greater the reactivity for mTG[69]. However, similarly conclusive studies have not been carried out on the acyl donor substrate. Initially, heptapeptides, GGGQGGG, were investigated for substrate recognition[70]. It was discovered that an increase in hydrophobicity at the N-terminus of the Gln peptide yielded elevated reactivity[70]. This study did not result in a peptide with greater reactivity then N-α-benzyloxycarbonyl-$_L$-glutaminylglycine (CBz-Gln-Gly). CBz-Gln-Gly is considered the industry standard for acyl donor substrate reactivity for mTG with an apparent K_M of 52.6 mM and k_{cat}/K_M of 40.4 μM^{-1} s^{-1} [71]. To properly develop a versatile tool for bio-conjugation, improving upon this relative poor substrate reactivity is crucial.

A more recent study by Hitomi and co-workers attempted to understand the acyl donor substrate preference of mTG by screening a phage-displayed $1.5x10^{11}$ peptide library, comprised of 12-mer peptides[72]. Once screened, certain trends in substrate specificity were identified.

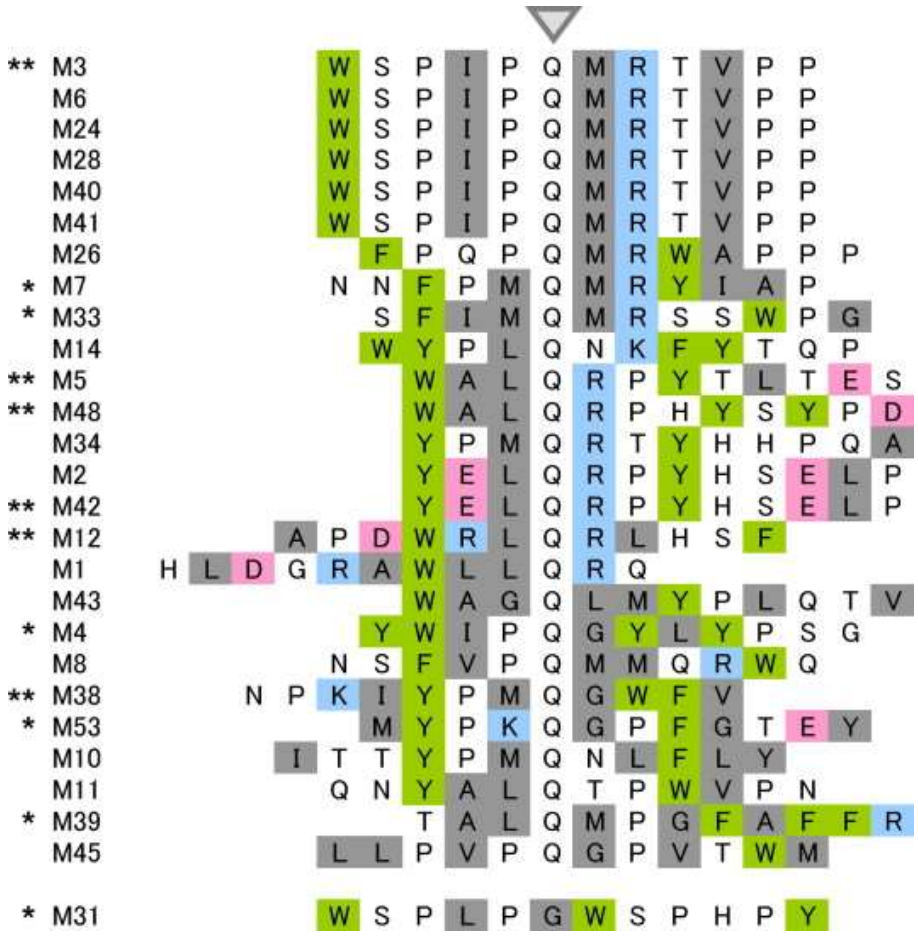


Figure 1.13 Candidate peptide sequences of the preferred substrate for mTG. Amino acid sequences of the peptides displayed on the phage clones that were selected by the screening are aligned per the potential reactive glutamine residues (bold arrowhead). Basic and acidic charged amino acid residues are highlighted as blue and red, respectively, and aromatic acids are highlighted as green. The hydrophobic amino acids, except for proline, are heavily shaded. (Adapted with permission from Sugimura, Y., et.al., Archives of Biochemistry and Biophysics, 2008. Copyright 2017 Elsevier)

After reviewing **figure 1.13**, it is evident that the presence of aromatic residues at positions -5 to -3, an arginine or hydrophobic residue at positions +1 or +2 and hydrophobic residues at the -2 and -1 positions were observed throughout the library[72].

Another important test in this study was the determination of orthogonality for any peptides of interest. Given the ubiquitous nature of mammalian TGases, identifying preferred mTG

sequences that are *not* substrates for TG2 and Factor XIIIa may prevent any unwanted cross-linking for future *in vivo* applications.

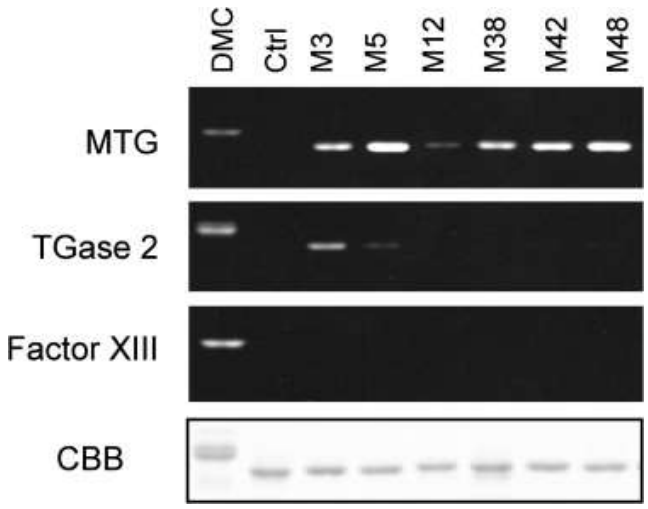


Figure 1.14 Reactivities of the selected preferred sequences in the catalytic reaction by mammalian TGases. Peptide-GST(QN)-fusion proteins were reacted with Dansyl-Cd for 10 min in the presence of each TGase. In each reaction, MTG, TGase 2 and Factor XIIIa were included at similar enzymatic activities in the reaction mixtures[73,74]. (Adapted with permission from Sugimura, Y., *et al*., Archives of Biochemistry and Biophysics, 2008. Copyright 2017 Elsevier)

Figure 1.14 highlights M12/38/42/48 as mTG substrate sequences that are orthogonal to other common TGases. This work is fundamental to the understanding of mTG specificity; however, it does not address how reactive the enzyme is to any of these peptides. Given the promiscuity of mTG, it is not enough to rely on specificity alone, and the question of reactivity must be addressed. This distinction suggest that further characterization is necessary for the peptide candidates.

1.5 Objectives of Thesis

Protein labelling is a diverse field that touches on a variety of disciplines in order to achieve the end goal of visually monitoring a POI. In Chapter 1, many of the widely used techniques for protein labelling were discussed including their versatility, advantages and drawbacks. Regardless of the many strategies to label a protein, there are still inherent disadvantages to each method. Chemo-enzymatic labelling strategies attempt to combine the selectivity and efficiency of enzymatic modifications with the broad substrate specificity a chemical toolbox will afford. Within these enzymatic solutions, transglutaminase balances the efficient turnover rate, versatile substrate scope and minimal recognition tag that an effective biocatalyst demands. However, there are still concerns with using TGase for protein labelling that must be addressed. Microbial transglutaminase effectively answers many of the limitations found with mammalian TG2. The bacterial transglutaminase's broad substrate scope and robust constitution give it the most potential for protein labeling applications. However, unlike the well-characterized TG2, mTG has more unknowns that require further research before this tool can be properly utilized. With these challenges in mind, we intended to achieve the following goals: 1) Develop an assay in order to measure the activity of mTG with any Gln-containing peptide sequence and determine whether peptide candidates are high-reactivity substrates for mTG; 2) Demonstrate proof-of-principle (POP) applications of mTG-mediated protein labelling, in a site-specific modification of a Q-tagged protein; 3) Express, purify and characterize the transglutaminase from *Bacillus subtilis* (bTG) and identify high-reactivity substrates using a FRET-based assay; 4) Demonstrate proof-of-principle (POP) applications of bTG-mediated protein labelling, in the site-specific

modification of a Q-tagged protein. The challenges these goals address, as well as the related results will be detailed in the following chapters:

Chapter 2: Development of continuous enzyme-coupled assay for mTG activity

In this chapter, we present the development and optimization of a modified glutamate dehydrogenase assay with the intention of analyzing potential high-reactivity peptides. This direct continuous assay presents significant advantages over the commonly used hydroxamate assay, including generality, sensitivity, and ease of manipulation. Furthermore, we identified 7M48 (Ac-WALQRPH-NH_2), a high-reactivity peptide that shows greater reactivity with mTG (K_M = 3 mM) than the commonly used Cbz-Gln-Gly (K_M = 58 mM), attesting to its potential for application in biocatalysis and bio-conjugation.

Chapter 3: Site-specific labelling and immobilization, mediated by mTG

We have identified that microbial transglutaminase (mTG) shows broad substrate specificity that is amenable to *in vitro* bio-conjugation applications. Herein, test proteins were genetically fused with peptide tags, followed by mTG-mediated propargylation of their reactive Gln residues. The propargylated proteins were then subjected to copper-assisted azide-alkyne cycloaddition to demonstrate either fluorescent labelling or immobilization.

Chapter 4: Characterization of transglutaminase from *Bacillus subtilis* and identification of substrate sequences

In vitro applications of mTG have yielded success and mTG has been proven as an effective tool for bio-conjugation. However, the expression of mTG as a zymogen introduces a limitation that cannot be easily overcome. To address this limitation, a new bacterial transglutaminase, bTG, was identified that can be expressed in its active form. In this chapter, the expression and purification of bTG is detailed. Furthermore, we present the development of an in-cell FRET-based assay to screen a library of bTG acyl-donor substrate candidates.

Chapter 5: Site-specific labeling and immobilization, mediated by bTG

In chapter 3 we demonstrate the bio-catalytic potential of mTG-mediated labelling, using the peptide 7M48 as a tag. In chapter 5, bTG-specific Q-tags were fused to a POI, followed by bTG-facilitated propargylation. As was performed following mTG-mediated propargylation, site-specifically modified proteins were then subjected to acceptor substrates to demonstrate the enzyme's labelling potential. bTG-mediated labelling strategies have the added advantage of being easily amenable to *in cellulo* applications. We will discuss the observations of the work and our belief in enzyme as a bio-catalytic tool.

Chapter 6: Conclusions and Perspectives

In the final chapter, a reflection on the work will be presented to determine the success of the research. Following the summation of each chapter, future perspectives based on the conclusion of the research will be suggested. This is an opportunity to discuss the direction of the study and identify questions that the research has not addressed.

Chapter Two: **Development of continuous enzyme-coupled assay for mTG activity**

The majority of the work presented in this chapter was published in:

Oteng-Pabi, S.K. & Keillor, J. Continuous enzyme-coupled assay for microbial transglutaminase activity. *Analytical Biochemistry* **441,** 169-173 (2013).

2.1 Introduction

In this chapter, the tools needed to identify high reactivity substrates for mTG are examined. Given the availability of screening techniques previously used to study the activity of the enzyme, it was evident that a new method must be developed, not only to study activity but also to screen potential peptide substrates in a high-throughput manner.

As mentioned in the introduction, transglutaminases (TGs, protein-glutamine:amine gamma-glutamyl- transferase, EC 2.3.2.13) are a family of calcium-dependent enzymes that catalyze an acyl transfer between glutamine residues and a wide variety of primary amines[75]. When a protein-bound lysine residue acts as the acyl-acceptor substrate, a gamma-glutamyl-epsilon-lysine isopeptide bond is formed between two proteins (**Figure 2.1**). TG-mediated protein cross-linking is critical to several biological processes and diseases, including extracellular matrix stabilization[76], blood coagulation[77],formation of cataracts[78], type 2 diabetes[79],neurodegenerative diseases[80] and certain cancers[81].

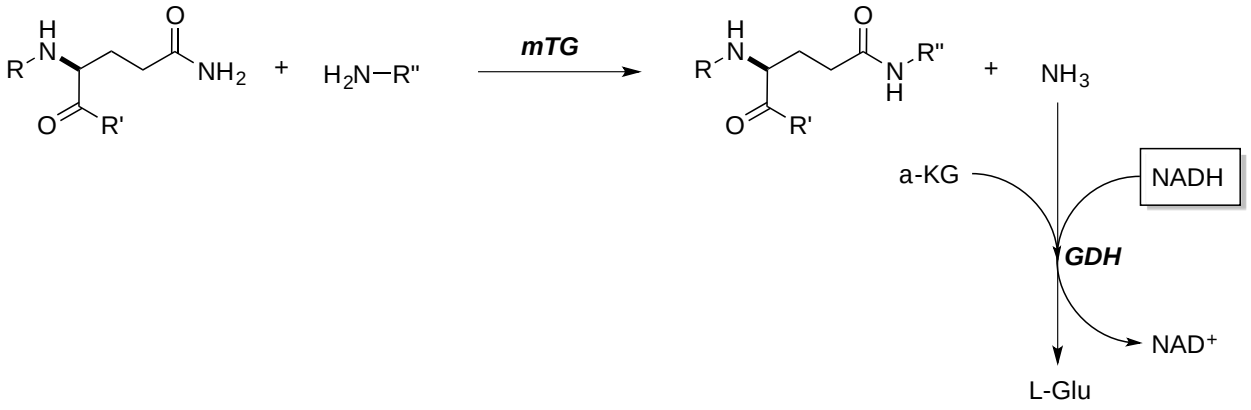


Figure 2.1 GDH-coupled mTG activity assay. mTG catalyses the acyl-transfer reaction from glutamine to a primary amine. (With Cbz-Gln-Gly, R = Cbz and R' = Gly. With 7M48, R = WAL and R' = RPH. R''-NH_2 = Gly-OMe.) Released ammonia acts as a substrate in the GDH-catalyzed reductive amination of alpha-ketoglutarate (α-KG). This reaction results in the disappearance of NADH, monitored by its absorbance at 340 nm.

Several TGs have been discovered in humans; while they play many diverse physiological roles, the requirement of calcium as a co-factor is pervasive with all mammalian members[82]. Microbial transglutaminases (mTG) are a distinct variant of the transglutaminase family, by virtue of their calcium-independent catalysis of the isopeptide bond formation[83]. Moreover, the 38-kDa mTG[84] from *Streptomyces mobaraensis* has been isolated and shown to have very little sequence similarity to any mammalian TG, whose molecular weights are >70 kDa[85]. mTG also exhibits more robust catalytic properties than its mammalian counterparts, including high stability throughout a broad range of pH and temperatures in comparison to other transglutaminases[86].

The broad substrate specificity of mTG makes it amenable as a tool for conjugation and opens its potential as a biocatalyst. For instance, mTG cross-linking has been exploited in the conjugation of biodegradable polymers. Hydroxyethyl starch (HES) is a polymer that has been used as an effect agent for therapeutic drug delivery[87]. The conjugation of the polymer to drugs and proteins is normally executed chemically; however, mTG-mediated cross-linking presents a more selective, less expensive alternative[87].

mTG's promiscuity with respect to its acyl-acceptor substrate highlights its bio-catalytic potential. The acyl-donor substrate, however, is more limited in its scope; the amino acid sequences flanking glutamine residues dramatically affect their activity. For example, it has been shown that the most likely sequences for mTG specificity have similarities at residues -3, -1, +1, and +2 relative to the position of the glutamine residue (namely aromatic, leucine, arginine, and proline residues, respectively)[72]. However, very few specific peptide sequences encompassed by these general preferences have been studied in enough detail to provide *quantitative* information regarding their affinity for mTG. For example, while Cbz-Gln-Gly is used as a standard acyl-donor substrate, its affinity for mTG is relatively low[71], requiring that it be used at relatively high

concentrations. This illustrates the importance of identifying a high-reactivity acyl-donor substrate for mTG, if its full potential as a bio-conjugation catalyst is to be realised.

2.2 *In vitro* Hydroxamate Activity Assay

The identification of a high-reactivity acyl-donor substrate requires a rapid and convenient activity assay that can be applied to the screening of multiple peptides. However, the assay most commonly used for measuring mTG activity is the hydroxamate activity assay (**Figure 2.2**)[88], a discontinuous method. The hydroxamate activity assay introduces added costs to the kinetic characterization of an enzyme and its substrates due to the relatively high quantity of reagents that are consumed. However, to obtain reliable results, a high volume of consumables is necessary. Furthermore, discontinuous assays are not ideal when attempting to monitor enzymatic rates of a reaction. A sensitive, continuous and rapid assay for mTG activity is therefore urgently required. The development of such an assay would not only contribute to expanding the scope of knowledge within this field, it would permit such possibilities in a fraction of the time the hydroxamate assay would allow.

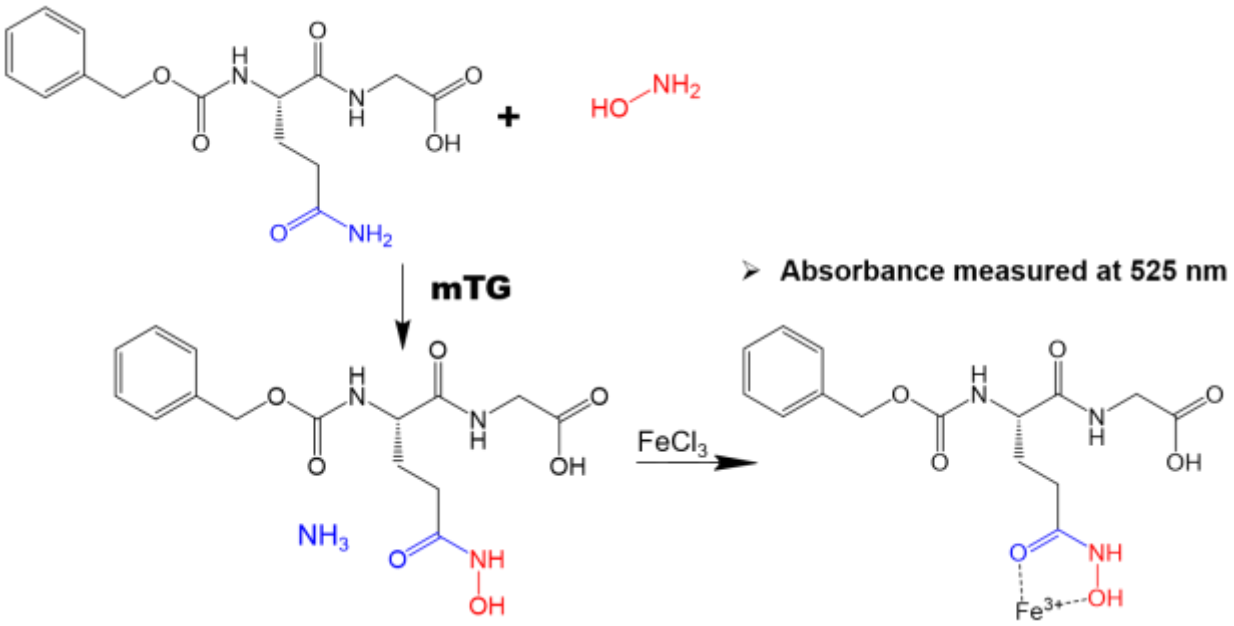


Figure 2.2 Hydroxamate activity assay. mTG catalyses the acyl-transfer reaction from glutamine within Cbz-Gln-Gly to hydroxylamine over the course of 10 min, after which the reaction is quenched by a $FeCl_3$ solution (15% TCA, 5% $FeCl_3$, 2.5 M HCl). Transamidation product can be quantified by measuring the absorbance at 525 nm, corresponding to the Fe^{3+} hydroxamate complex[89]. Activity of mTG is calculated using the equation:

$$\text{Activity} = \frac{\left((A - A_0)/340\ M^{-1}\right) \times 10^{-3}\,L \times 10^{6}\ \mu mol/mol}{10\ min} = \boxed{\left((A - A_0)/340\right) \times 10^{2}}\ \mu\text{mol per min}$$

Where A= absorbance after 10 min of the reaction, A_o= absorbance without enzyme after 10 min, 340 M^{-1}= extinction coefficient of Cbz-Gln-Gly in a 1-cm cuvette.

To that end, we have developed a coupled-enzyme assay to measure the activity of mTG with any Gln-containing peptide sequence. Based on a microtiter plate scale, this assay uses little material and is amenable high-throughput screening applications. We have applied it herein to quantitatively measure the kinetic parameters for several potential high-reactivity peptide sequences.

2.3 Results and Discussion

2.3.1 Relative Rates of Coupled Enzymes

Glutamate dehydrogenase (GDH) catalyses the reductive amination of alpha-ketoglutarate with concomitant consumption of the co-enzyme NADH. In the past we have coupled this reaction to the catalytic activity of tissue transglutaminase in the development of an assay[90] that has since been applied to kinetic and inhibition studies[91,92]. The adaptation of this assay for application to mTG requires that several conditions must be met.

Most importantly, for this assay to report on mTG activity, it is crucial that the GDH-mediated consumption of NADH and ammonia is faster than the mTG-mediated release of ammonia during acyl transfer (**Figure 2.1**). To establish this, the effect of GDH concentration on the observed mTG reaction rate was studied at a saturating concentration of Cbz-Gln-Gly as acyl-donor substrate (45 mM). As shown in **Figure 2.3**, the observed reaction rate (using 0.1 units of mTG) was found to increase linearly with added GDH, up to the addition of 2 units. When the concentration of GDH exceeds 2 units, no further increase in reaction rate was observed, indicating that the reaction rate is limited by mTG-mediated release of ammonia under these conditions.

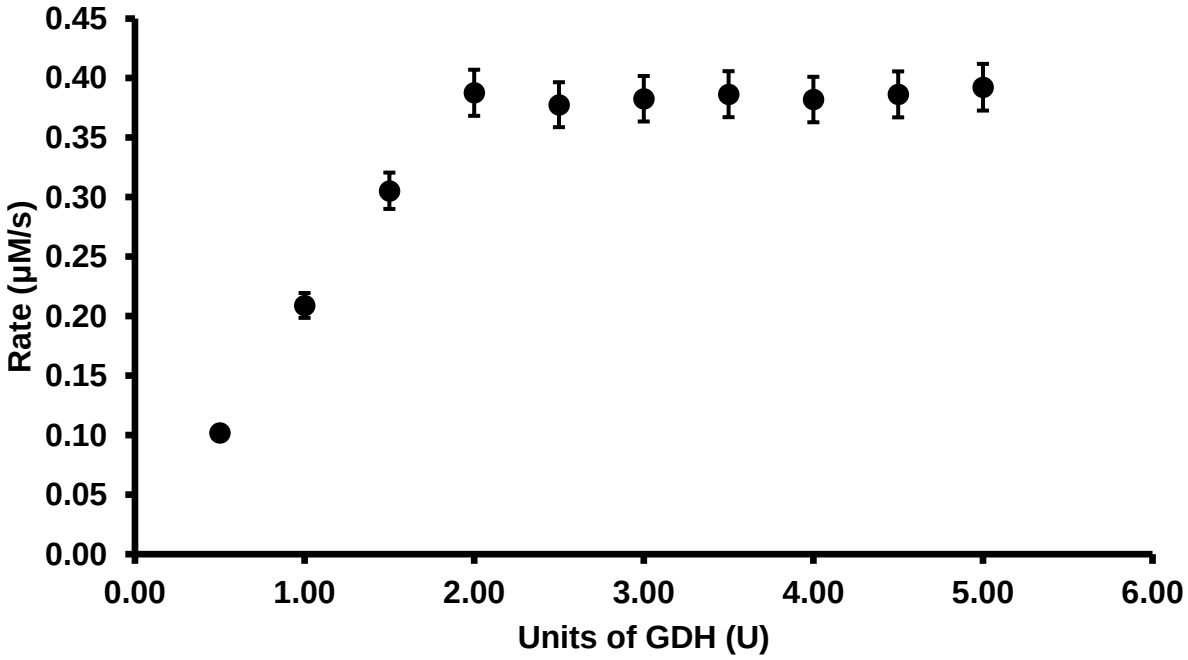


Figure 2.3 Rate of reaction of 0.1 units MTG as a function of added GDH. Rate values were measured in triplicate, by monitoring the decrease of NADH over time. Error bars represent the standard error of the mean values.

2.3.2 Substrate Concentration Optimization

Another condition that must be established is that the reactants and products of both enzyme have no effect on the activity of the other. First, the reaction of GDH was monitored in the absence and presence of the reagents required for the mTG reaction. **Figure 2.4** shows the negligible effect of the presence of mTG on the GDH reaction. The rate of the positive control reaction of GDH, upon addition of 0.1 mM ammonium chloride (**Figure 2.4**), is shown in **Table 2.1**. When this reaction was repeated in the presence of 40 mM added Cbz-Gln-Gly and 10 mM Gly-OMe, the reaction rate was found be the same, within experimental error. When the reaction was repeated in the presence of added mTG, the reaction rate was similarly unaffected.

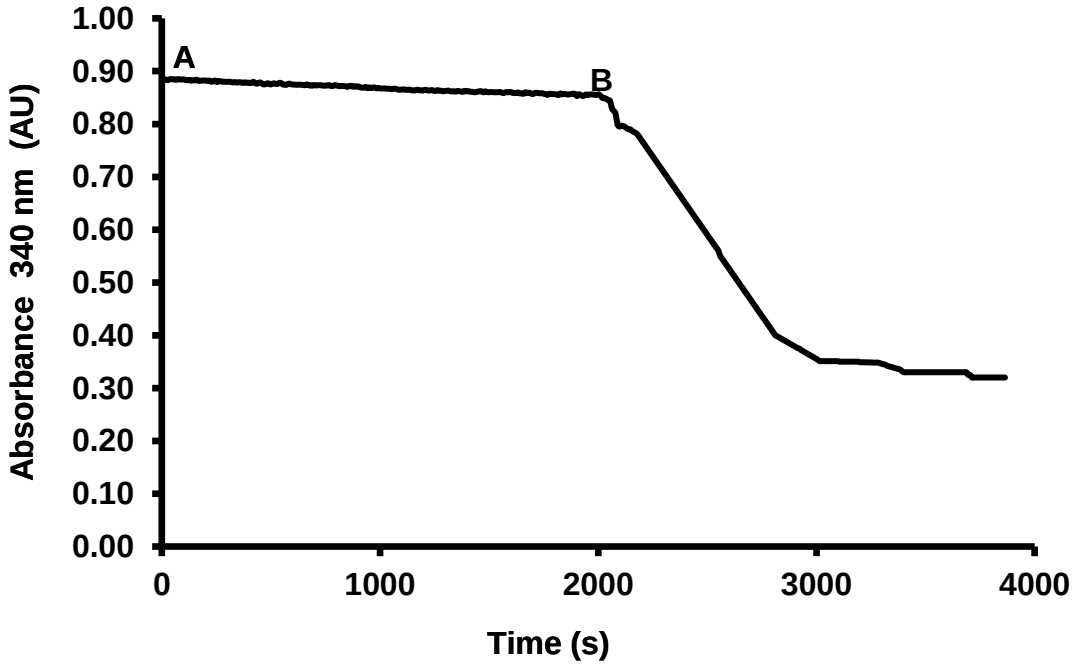


Figure 2.4 Effect of mTG and mTG substrates of the activity the GDH assay. The consumption of NADH was monitored over time, A) in the presence of mTG and its substrates, before and B) after the addition of 0.1 mM ammonium chloride, final concentration.

Secondly, the activity of mTG was determined using the well-known hydroxamate assay, in the absence and presence of added reagents required for the GDH-coupled assay. Specifically, the rate of mTG-mediated hydroxamate formation was determined to be 0.823 ± 0.042 μM/min over a 10-minute period in the presence of 0.8 units of mTG. When the reaction was repeated in the presence of the additives required for the GDH reaction (namely 500 μM NADH, 3.34 mM α-KG, and 2 units of GDH), the rate was determined to be 0.793 ± 0.038 μM/min. The similarity of these results show that mTG and GDH, as well as their respective substrates, have a negligible effect on each other's activity.

Table 2.1 Reaction rates of the GDH assay. Rates of reaction according to GDH-coupled assay in the absence and presence of added reagents required for mTG reaction. Rates were measured in triplicate and reported errors are standard deviation of the mean values.

Added Reagents	**Rate[a] (μM/s)**
ammonia	0.85 ± 0.04
Cbz-Gln-Gly, GlyOMe +ammonia	0.83 ± 0.03
mTG + ammonia	0.82 ± 0.03

[a] initial rate measured after the addition of ammonia

2.3.3 Limits of Detection

Under the conditions of the optimized activity assay, 0.1 U of mTG were used. However, in order to explore the full scope of our microtiter plate-based assay, its sensitivity and lower limits of detection were also determined. As shown in **Figure 2.5**, the observed rate of reaction decreases linearly in the presence of decreasing amounts of added mTG. Remarkably, the assay was found to have a linear dynamic range of over 100-fold in mTG concentration. However, in the presence of less than 0.02 U of mTG, it is difficult to discern the rate of mTG-mediated ammonia release.

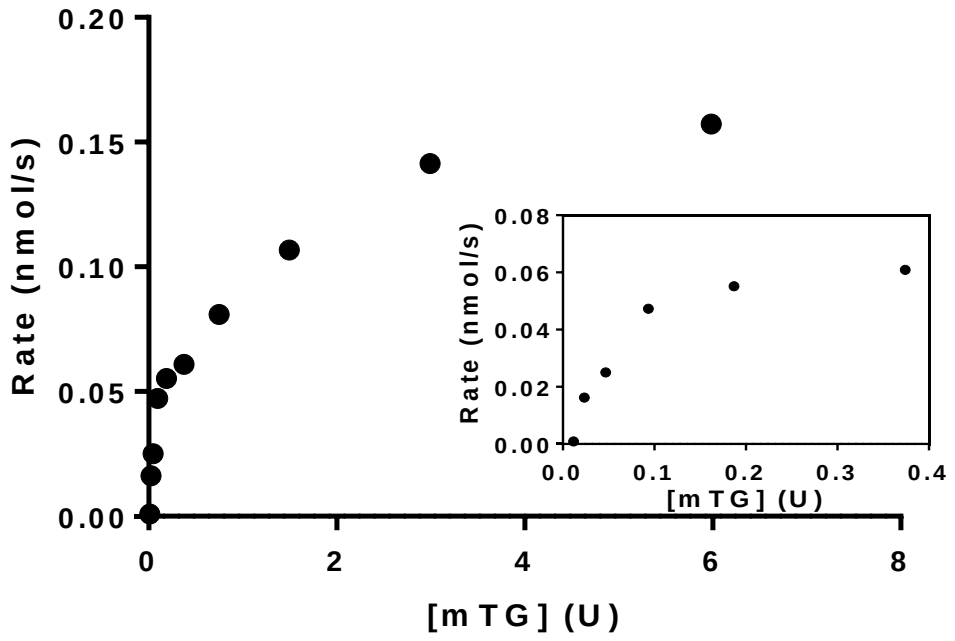


Figure 2.5 Monitoring the lower limit of detection of GDH-coupled assay. Initial rate was measured by the GDH assay as a function of units of mTG (from ~0.01 to ~6 U) in the presence of 25mM CBz-Gln-Gly, over 20 min, performed in triplicate on a microplate reader.

2.3.4 pH Dependence

The effect of pH was then studied to determine the sensitivity of the coupled enzyme assay. Using the same concentrations of substrates and enzymes as in the optimized assay (see Experimental), the pH of the reaction buffer was varied between 2.5 and 9. As shown in **Figure 2.6,** negligible differences in reaction rates were observed between pH 6 and 9, suggesting that neither of the coupled enzymes are greatly affected by variations in pH over this broad range, centred around physiological pH. This is consistent with a previous report that mTG has an optimal activity between pH 6-9[86,93].

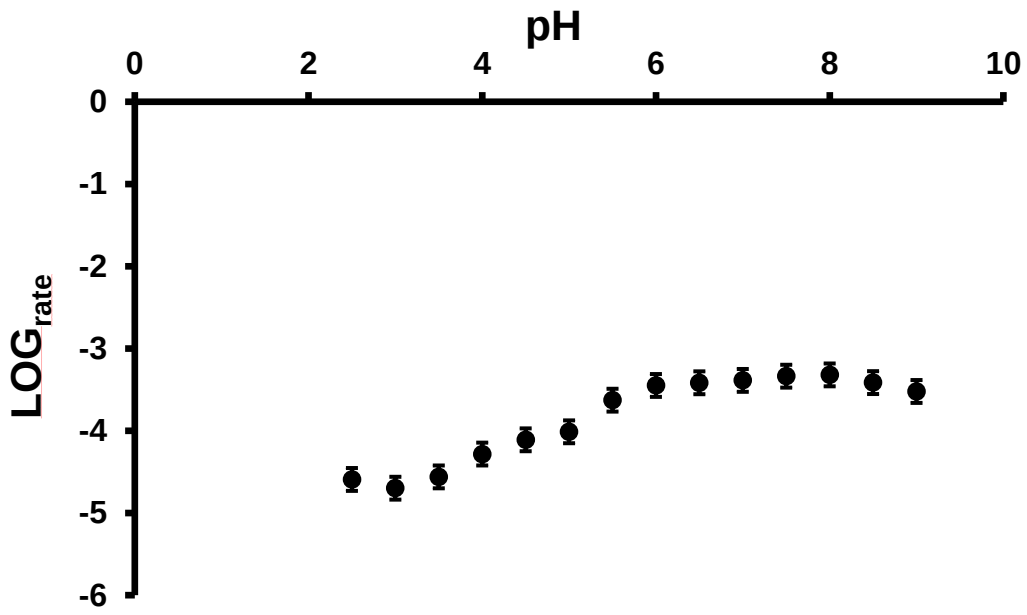


Figure 2.6 pH-rate profile of the GDH-coupled assay. pH vs LOG(rate). Rate is reported in units of mM/s. pH-profile of mTG-mediated transamidation, measured by coupled GDH assay, over pH range 2.5-9, with 0.10 UmTG, 50 mM Cbz-Gln-Gly and 10 mM Gly-OMe.

2.3.5 Application of K_M Measurement

As shown above, the optimized GDH-coupled assay is a convenient, rapid, sensitive and robust method for measuring mTG activity. As such, it offers several significant advantages over the widely applied hydroxamate assay. Namely, it is a continuous rather than discontinuous assay that requires less material, can be carried out quickly and conveniently in microtiter plate format, and is readily amenable for high-throughput screening. Furthermore, since it is based on the release of ammonia from the glutamine-donor substrate, it is independent of the precise peptide sequence of the glutamine sequence and the identity of the acyl-acceptor substrate. As a simple demonstration of this advantage, the GDH-coupled assay was applied to the facile kinetic

characterization of alternative Gln-donor peptide substrates allowing their comparison with Cbz-Gln-Gly, the standard acyl-donor substrate used in many mTG studies.

Table 2.2 Peptides screened by hydroxamate activity assay. Activity and specific activity of high reactivity peptides and CBz-Gln-Gly, obtained from transamidation, mediated by mTG and monitored via the hydroxamate activity assay. Results obtained were performed in triplicate, using a spectrophotometer. Reported errors are stand errors of the mean values.

Substrate	**Activity (µmol/min)**	**Specific Activity (µmol/min/mg)**
Cbz-Gln-Gly	0.412 ± 0.02	10.2 ± 0.920
5M38*	0.493 ± 0.09	11.8 ± 1.03
5M42*	0.512 ± 0.03	16.7 ± 0.521
5M48*	.720 ± 0.04	18.1 ± 0.755
7M38**	0.913 ± 0.09	23.4 ± 1.82
7M42**	1.62 ± 0.02	42.9 ± 0.781
7M48**	1.75 ± 0.06	55.1 ± 0.922

* Truncated peptides (pentamers), ** Truncated peptides (heptamers)

The sequences of these alternative peptide substrates were based on the high reactivity peptides discovered by Hitomi and co-workers through their phage-display studies[72]. Notably, peptides named M42 and M48 were found to be among the most reactive of all of peptides screened in their phage-displayed library, according to the hydroxamate assay (**Table 2.2**). Initially, pentapeptide variants were truncated from the -2 to +2 positions of the parent sequences, relative

to the central reactive Gln residue. These peptides, named 5M38 (Ac-PMQGW-NH_2), 5M42 (Ac-ELQRP-NH_2) and 5M48 (Ac-ALQRP-NH_2), were synthesized according to standard Fmoc-based solid-phase peptide synthesis (SPPS) protocols, as described in the Experimental section. Enzyme activity was then determined using the hydroxamate assay. Surprisingly, substrate reactivity for mTG was found to be as low as that of Cbz-Gln-Gly (**Table 2.2**). Given the suspected relative importance of the +3 residue[72], this lack of reactivity for the pentapeptides tested was reasonable. To address this issue, heptapeptide variants were designed from the -3 to +3 positions. Specifically, peptides 7M42 (Ac-YELQRPY-NH_2) and 7M48 (Ac-WALQRPH-NH_2) were prepared by SPPS and characterized kinetically using both the hydroxamate assay (**Table 2.2**) as well as the GDH-coupled assay (**Table 2.3**). In **Table 2.3** the K_M value of 7M42 was determined to be several-fold lower than that of Cbz-Gln-Gly, while the K_M value of 7M48 was found to be nearly 20-fold lower (**Figure 2.7** and **Figure 2.8**). While this experiment demonstrates the convenience of the application of the coupled-enzyme assay to detailed kinetic studies, it also illustrates the potential for the development of simple peptidic substrates that react much more efficiently with mTG than Cbz-Gln-Gly.

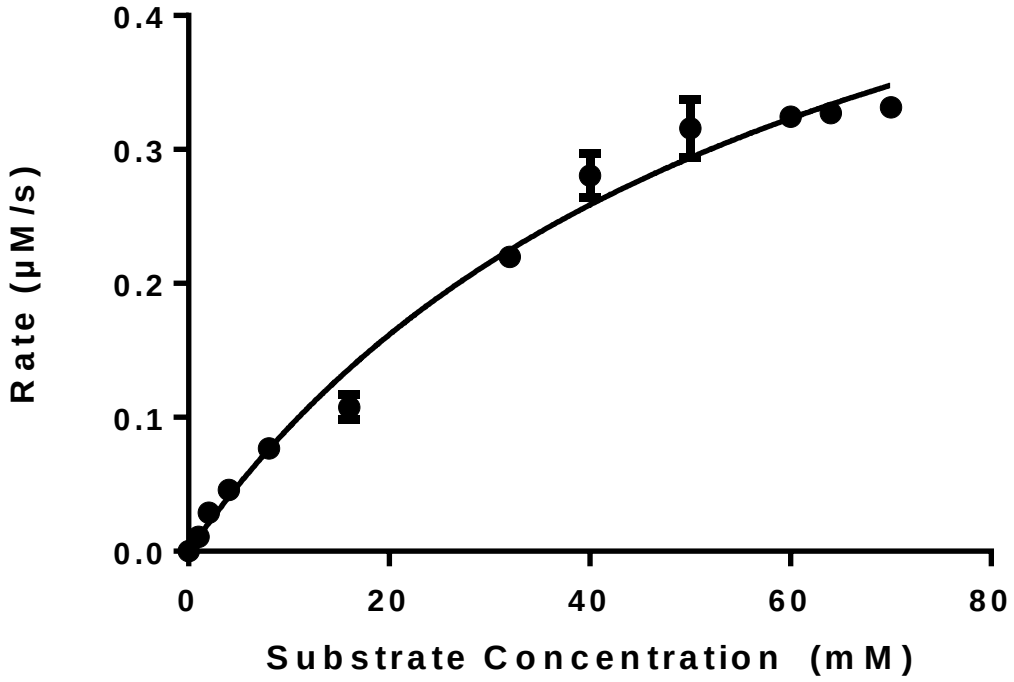


Figure 2.7 Michaelis-Menten plot for mTG activity of CBz-Gln-Gly. mTG activity measured by the coupled-GDH assay, using varied concentrations of CBz-Gln-Gly as acyl-donor substrate (0.125-68 mM) in the presence of 10 mM Gly-OMe as acyl-acceptor substrate, carried out in triplicate using 0.2 U of enzyme at pH 7.2 and 37ºC carried out in 96-well microplate reader using a 96-well microplate reader. Error bars represent standard error of the mean values. The solid line through the data points was fit to the hyperbolic Michaelis-Menten equation, giving kinetic parameters of V_{max}= 0.546 mM/s and K_M = 58.9 mM.

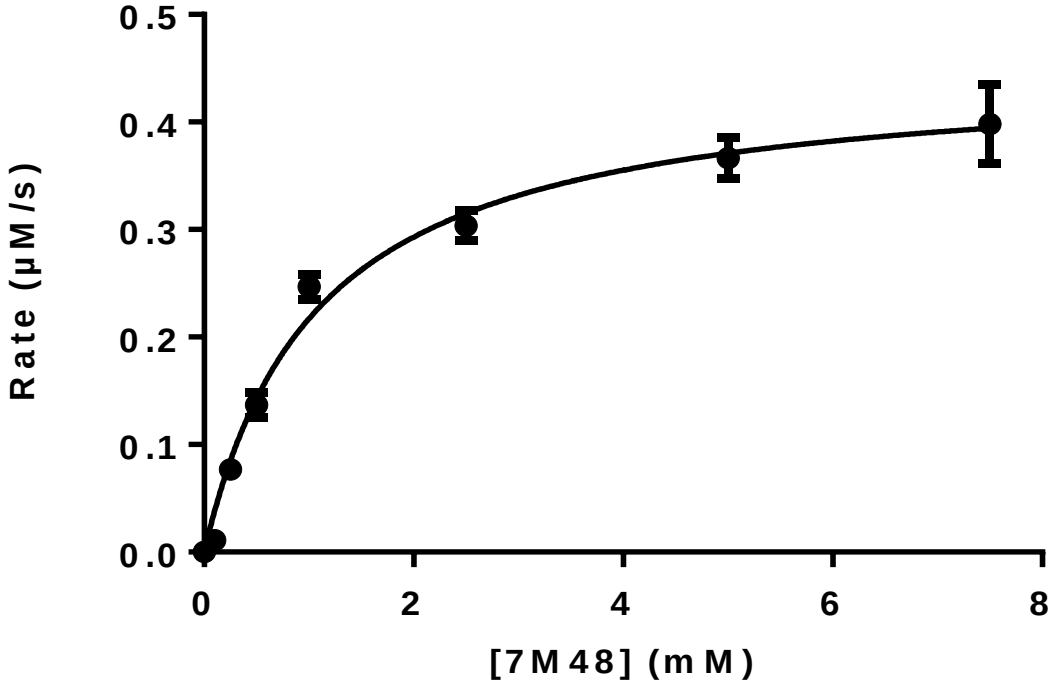


Figure 2.8 Michaelis-Menten plot for mTG activity of 7M48. mTG activity measured by the coupled-GDH assay, using varied concentrations of 7M48 as acyl-donor substrate (0.125-7.5 mM) in the presence of 10 mM Gly-OMe as acyl-acceptor substrate, carried out in triplicate using 0.2 U of enzyme at pH 7.2 and 37°C carried out in 96-well microplate reader. Error bars represent standard error of the mean values. The solid line through the data points was fit to the hyperbolic Michaelis-Menten equation, giving kinetic parameters of V_{max}= 0.435 mM/s and K_M = 3.07 mM.

High-reactivity substrate peptide sequences discovered in this way should prove useful in biocatalysis and bioconjugation applications. For example, polymeric materials comprising a high-reactivity peptide sequence could be modified using mTG as a catalyst for the synthetic process. Alternatively, the genetically-encodable peptide sequence could be fused to a protein of interest, transforming it into a specific target for mTG-mediated labelling *in cellulo*.

Table 2.3 Kinetic evaluation of transamidation and deamidation of mTG peptides. Kinetic values obtained from transamidation and deamidation (hydrolysis), mediated by mTG, of high reactivity peptides and Cbz-Gln-Gly (acquired by both GDH assay and hydroxamate assay). Results obtained were performed in triplicate, using a 96-well microplate reader. Reported errors are stand errors of the mean values.

Substrate	K_M **(mM)**	k_{cat} **(s^{-1})**	k_{cat} **/** K_M **(mM^{-1} s^{-1})**
Transamidation [a]			
Cbz-Gln-Gly	58.9 ± 2.4 (55.1± 3.3) [b]	(1.37 ± 0.19) x 10^3 (1.05 ± 0.13) x 10^3 [b]	(0.23 ± 0.04) × 10^2 (0.19 ± 0.02) × 10^2 [b]
7M42	15.3± 1.4	(1.07 ± 0.07) x 10^3	(0.69 ± 0.03) × 10^2
7M48	3.07 ± 0.70	(1.09 ± 0.25) x 10^3	(3.54 ± 0.19) × 10^2
Deamidation			
Cbz-Gln-Gly	1.610± 0.103	60.3 ± 13.1	(0.38 ± 0.07) × 10^2
7M42	0.245± 0.034	103.6 ± 5.3	(4.18 ± 0.54) × 10^2
7M48	0.157± 0.028	90.4 ± 7.5	(5.73 ± 0.71) × 10^2

[a] Measured in the presence of 10mM Gly-OMe as acceptor substrate.

[b] Values in were measured using the hydroxamate assay, for comparison.

2.4 Conclusions

In conclusion, we have developed and optimized an assay that can be used to evaluate the activity and kinetic parameters of substrates for microbial transglutaminase. Through the optimization, it has been proven that pH range in which the assay is functional will allow for the detection of peptide substrate at physiological pH as well as several units outside this range. Given that the final envisioned application for mTG is as a biocatalyst, it is important to note that not only is the enzyme functional, but so is the activity assay, and over a wide range. The limit of detection also presented a favourable result; showing detection down to 0.02 U highlights how sensitive this assay can be. This will provide added value when enzyme quantities are limited.

As mentioned previously, the value of a continuous assay cannot be overstated, especially when it permits the detection of a broad range of substrates, not even limiting itself to peptides. Using this assay, two acyl-donor substrates (7M42 and 7M48) were identified that demonstrated significantly greater reactivity for mTG relative to the commonly used substrate Cbz-Gln-Gly. The considerable improvement in k_{cat} / K_M (mM^{-1} s^{-1}) clearly demonstrates that better substrates can been identified.

2.5 Perspectives

It is worth noting that the discovery of a better mTG substrate than Cbz-Gln-Gly is not the only added value of this work. The ability to genetically encode the high-reactivity sequences 7M42 and 7M48 increases their value as substrates. Given the current field of enzyme facilitated bioconjugation, it is critical that any tags used for labelling can be expressed endogenously. This prevents the tagging system from being limited to *in vitro* applications. These substrates demonstrate the potential for the development of bio-catalytic and bio-conjugation applications mediated by mTG and this will be explored in the next chapter.

2.6 Experimental Section

2.6.1 Expression and purification of mTG

The plasmid pDJ1-3 was prepared by Professor Markus Pietsch and kindly provided by Professor J. Pelletier (Université de Montréal). pDJ1-3 encodes the pro-enzyme of mTG from *Streptomyces mobaraensis*, inserted between the *Nde*I and *Xho*I restriction sites of the vector pET20b[94]. mTG was expressed and purified according to a previously published protocol[95] with the following amendments: The plasmid encoding for mTG bearing a *C*-terminal hexa-histidine tag was transformed in to *E. coli* BL21-Gold (DE3) in the presence of 100 µg/mL ampicillin. mTG was expressed according to the auto-induction protocol[96]. The culture was centrifuged at 3000 × g for 30 min at 4 ºC; the supernatant was discarded and the pellet was resuspended in 50 mM phosphate buffer pH 8.0, with 300 mM NaCl. Cells were disrupted by sonication over ice (three cycles of 30-s pulse at 20% intensity / 1 min pause) using a Branson sonicator. mTG was then activated by cleavage of the pro-enzyme alpha sequence through incubation in a 1:10 ratio (w/v) of trypsin (1 mg/mL) to unpurified mTG for 75 min, at 37 °C. The activated mTG was purified using a 1 mL His-trap Ni-NTA column (GE Healthcare) equilibrated in 50 mM phosphate buffer pH 8.0, with 300 mM NaCl, and eluted with an imidazole gradient (0-140 mM), on an Åkta FPLC (GE Healthcare). The purified, activated mTG was dialyzed against 50 mM phosphate buffer pH 8.0. The average yield was 75 mg of activated mTG per litre of culture, with greater than 80% purity as estimated through evaluation of 10% SDS-PAGE followed by staining with Coomassie blue. Aliquots were snap-frozen and stored at -80°C in 15% glycerol.

2.6.2 Peptide Synthesis

Heptamer peptides were synthesized using an automated peptide synthesizer (Liberty CEM Microwave Peptide Synthesizer) by conventional Fmoc-based SPPS. Synthesis was carried out on a scale of 200 mg of Knorr-amide resin with a loading capacity of 0.93 mmol/g (0.25 mmol scale). Each round of peptide elongation consists of: A) Fmoc deprotection with 20% piperidine in DMF, 5 min, three times, B) washing with DMF, C) coupling with HBTU, DIEA, NMP and the appropriate amino acid in DMF, 11 min, D) washing with DMF. The molar ratio of the amino groups on the resin/building block/HBTU was 1/4/3.6. When the peptide was complete, acetylation at the native N-terminus (to protect primary amine) was carried out using 20% pyridine in DMF. Peptide was then cleaved from the resin using 90/5/5 TFA/DCM/thianisole for 120 minutes. After cleavage, the resin was filtered and washed, three times under vacuum using 1:1 cylcohexane:acetone. The peptide was dissolved in a minimal amount of acetone and precipitated from the solution using cold Et_2O, with sonication. The precipitated peptide was then recovered through centrifugation at 1100 × g for 30 min.

Pentamer peptides were synthesized using traditional solid phase peptide synthesis techniques. Synthesis was carried out on a 1 g scale of Knorr-amide resin. Each round of peptide elongation consists of: A) Fmoc deprotection with 20% piperidine in DMF, 10 min, B) washing with DMF, C) coupling with HOBT, DIPEA, EDC and the appropriate amino acid in DMF, 30 min, two times D) washing with DMF. The molar ratio of the amino groups on the resin/building block/HOBT was 1/2.3/3. Peptide cleavage and precipitation as per peptides synthesized via peptide synthesizer.

2.6.3 HPLC Analysis

HPLC was performed on a Waters 2996 instrument equipped with a photodiode array detector. UV absorbance of peptides was measured at 254 nm and 274 nm. Analytical HPLC was performed on a Waters C18 column (5 µm, 3.8 mm × 150 mm) with a run time of 30 min, a flow rate of 0.5 mL/min, and a linear gradient of 50–95% of 0.1% formic acid (FA) in methanol (eluent B) in 0.1% aqueous FA (eluent A).

2.6.4 Hydroxamate Assay

The previously reported hydroxamate assay[88,97]⁹⁷was adapted to use with mTG as follows: 0.75 mL of 0.2 M acyl-donor substrate was mixed with 0.25 mL of 2 M hydroxylamine, 0.25 mL of 0.02 M EDTA, and 1 mL of 1.0 M Tris-acetate and adjusted to pH 6. A 400-µL aliquot of this solution was incubated with 100 µL of enzyme solution (at varying concentrations) at 37 ºC for 10 minutes. After incubation, 500 µL of 2.0 M $FeCl_3$•6 H_2O, 0.3 M trichloroacetic acid and 0.8 M HCl was added. The solution was centrifuged for 5 min at 10000 × g (to clear any precipitated protein) and the absorbance of the supernatant fraction was measured at 525 nm. Activity was determined, using the extinction coefficient of Cbz-L-glutamyl-(gamma-hydroxylamine)-glycine, determined to be 340 M^{-1} for a one-centimetre path length at 525 nm[97], using authentic synthetic material [98].

2.6.5 Glutamate Dehydrogenase Coupled-Enzyme Assay

The optimized assay was performed using 200 mM MOPS buffer (pH 7.2), 1 mM EDTA, 500 μM NADH, 10 mM ketoglutarate, 2 units glutamate dehydrogenase (GDH), varying concentrations of acyl-donor substrate (1 to 66 mM) and 10 mM of glycine methyl ester as an acyl-acceptor substrate[99], in a final volume of 180 μL. The reaction solution was pre-incubated for 5 min at 37 °C prior to initiation of the reaction by the addition of 0.10 units mTG (in a volume of 20 μL) to give a final volume of 200 μL. For the corresponding blank solution, the mTG solution was replaced with water. The assays were performed in a 96-well plate, and the decrease in absorbance due to the oxidation of NADH was followed against a blank at 340 nm in a Synergy HT Multi-Mode Microplate Reader thermostated at 37°C. After a brief lag during which sufficient ammonia is produced to saturate GDH, linear slopes of absorbance versus time were measured. This lag phase is intrinsic to all continuous enzyme-coupled assays, for which it is difficult to measure true initial rates. However, the brevity of this lag in this optimized assay allows for the accurate measurement of approximate initial rates, which are still a valid indicator of substrate turnover. These approximate initial rates were corrected for path length and blank using the instrument software.

Chapter Three: **Site-Specific Labelling and Immobilization, Mediated by mTG**

The majority of the work presented in this chapter was published in:

Oteng-Pabi, S. K., et. al. Site-specific labelling and immobilization, mediated by microbial transglutaminase. *Chemical Communications* **50,** 6604-6606 (2014).

3.1 Introduction

This chapter presents the development of an assay used to elucidate kinetic parameters for potential high reactivity peptide substrates. Identifying substrates for mTG is crucial to the development of mTG as a bio-conjugation tool. mTG, similar to its labelling counterparts, must be validated for its ability to facilitate the covalent attachment of fluorogens with a POI. A proof-of-principle (POP) experiment should therefore include the labelling of a protein, as well as a demonstration of the versatility of the technique, high specificity and low background reaction. These factors are crucial to the development of an effective enzyme for bio-conjugation.

3.1.1 Identification of advantages associated with mTG

Defining the role of genes is crucial to identification of their function. Although this knowledge is inscribed in DNA, reading genetic material is not the most effective path to elucidating this information[100]. It is more realistic to monitor the effects the translated genes will have and map their interactions to determine the functional role of a gene. As mentioned in chapter 1, there are a multitude of methods that can be used to chemically modified proteins to monitor their function[101]. Generally, these methods result in chemical labelling of a specific functional group. However, chemically labelling limits the scope in which protein modification can be utilized. To combat this issue, enzymatic site-specific labelling of a residue was introduced[12,13,102,103]. The facilitation of protein labelling by enzymes led to labelling in live cells, where information on expression, localization and trafficking can be expanded upon. This critical information made the site-specific labelling of proteins relevant and led to the discovery of many potential enzymes that can facilitate bio-conjugation[12,104].

Several enzymes have been discovered and/or engineered to accommodate this demand: including farnesyl transferase[32], biotin ligase[37], myristyl transferase[105], formylglycine- generating enzyme[19], sortase[106], lipoic acid ligase[107] and tissue transglutaminase (TG2)[48]. Although they have their advantages and disadvantages, transglutaminase appears to have the most potential as a tool for selective labelling of a POI. Transglutaminase recognizes a minimal 5-7 amino acid peptide tag, unlike other enzymes that require l3-15 residue tags, such as ligase and formylglycine-generating enzyme. Furthermore, transglutaminases are able to incorporate a wide molecular scope of labels onto a POI while farnesyl transferase, myristyl transferase and lipoic acid ligase are limited to relative large substrates. Additionally, transglutaminases have the added benefit of catalytically-efficient labeling and relative stability (over the duration of a labelling experiment).

Transglutaminase appears to have the most potential as a tool for protein labelling, specifically when minimal perturbation of the POI is desired. Natively, transglutaminases catalyze the cross-linking of proteins and/or peptides, through a transamidation reaction between glutamine and lysine residues in wild-type sequences (**Figure 3.1**). This activity has been showcased throughout the family of transglutaminase enzymes, specifically Transglutaminase 2[62]. Transglutaminase 2 is a relatively ubiquitous enzyme, causing its dysfunction to be implicated in numerous diseases and being the target of many studies. Unfortunately, its cellular application is limited due to its relative large size, 76 kDa, its calciun-dependence for activation and its broad specificity.

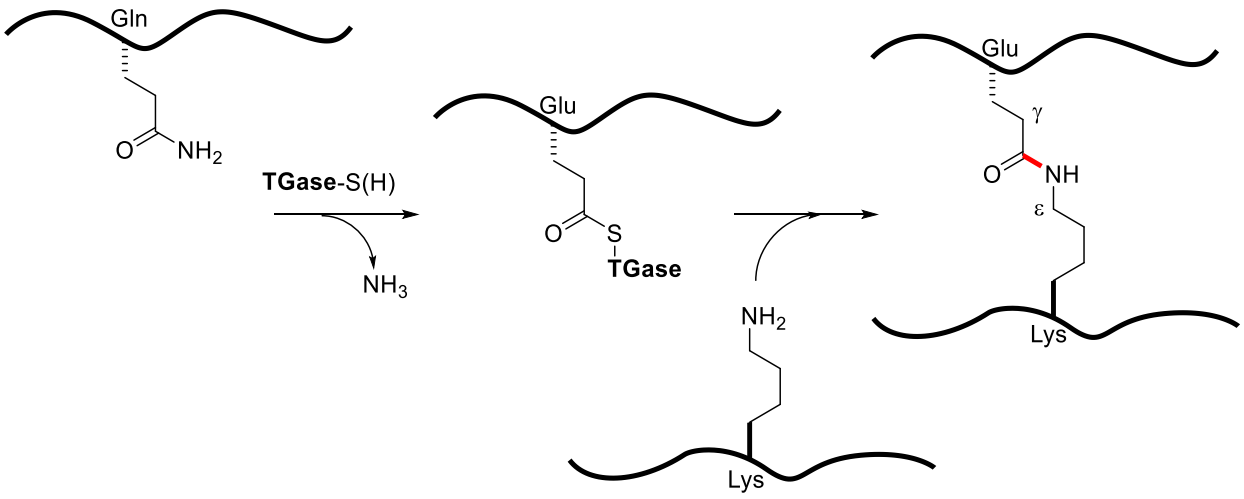


Figure 3.1 Transglutaminase (TGase)-mediated protein cross-linking. Transamidation between protein-bound Gln and Lys residues leads to the formation of γ-glutamyl-ε-lysyl isopeptide bonds (red).

Bacterial transglutaminases have been developed for different applications[49], including an inexpensive way to cross-link proteins/peptides in the food industry[64]. However, research into microbial transglutaminase (mTG) from *Streptoverticillium mobaraense*, may also be the answer to the limitations of protein labelling with Transglutaminase 2[72]. mTG is a smaller, 38 kDa, calcium-independent enzyme[66,108]. mTG cross-links proteins through the same Gln/Lys acyl transfer reaction as the mammalian enzymes[48,109].

3.1.2 Previous work associated with mTG peptide discovery

Two discoveries regarding the substrate specificity of mTG have enabled its application as a labelling tool. Firstly, minimal peptide sequences (5-7 peptide) have been developed that act as genetically encodable Glutamine substrates for mTG ('Q- tags')[72,108]. The short tags fulfill the requirement to have minimal perturbation to the POI. With this goal, it would be logical to

ensure all modifications to a wild-type protein are kept to a minimum to replicate wild-type behaviour as much as possible. Previous work by Hitomi and co-workers used phage- displayed peptide libraries to discover sequences that serve as efficient Gln-donor substrates for mTG, while maintaining orthogonality with mammalian transglutaminases[72,73]. This work proved to be the basis in which mTG tags were developed. In Chapter 2 Hitomi's sequences were reduced to heptamers and pentamers to identify their specificity. These tags can now be genetically fused to test proteins, with the hope of retaining their efficient reactivity with mTG.

Secondly, propargylamine has been shown to be a surprisingly efficient amine substrate for mTG[110]. Through subsequent copper-catalyzed azide- alkyne cycloaddition (CuAAC), the modification of small glutamine peptides[15] and a native protein is demonstrated. The potential of these findings can be introduced in a protein-labelling method that takes advantage of the specificity of mTG Q-tag recognition with the scope of 'click' chemistry, thereby adding a wide range of unnatural functionality to a POI, through selective and site-specific post-translational modification. Herein, we demonstrate two proof-of-principle applications of mTG-mediated protein labelling, in the site-specific modification of a Q-tagged protein for 1) fluorescent labelling and 2) immobilization.

3.2 Spacer length optimization between high reactivity tag and POI

Prior to demonstrating the capacity of mTG as a tool for bio-conjugation, optimization of the affinity peptides as genetic tag was required. Although kinetic parameters for 7M38/42/48 had already been established, it was not known whether these tags could be recognized on the termini of the protein. Merely attaching the tag to the termini of a protein may not sufficient, given the folding and post-translation modifications proteins undergo as the reach their mature state. It is also worth noting that even if the proteins are successful expressed, the genetic tags may not be able to reach the active site of mTG. To address this issue, maltose-binding protein (MBP) was selected for optimization. The relative ease of expression and stability of MBP made it a clear practical option for protein modification. Furthermore, MBP has been established as a fusion protein[111]; therefore, its termini can be manipulated without any significant consequences to the solubility of the protein.

7M38/42/48 were introduced to the multiple cloning site (MCS) of MBP, located at the C-terminus of the protein. To address concerns over the accessibility of the tag to mTG, a short spacer sequence (GSS) was placed upstream of the tag. The MBP Q-tagged variants were then tested to determine the necessity for a spacer, ideally where a common theme can be identified, depending on the spacer used. To elucidate this information, the GDH- assay was used to measure the transamidation of mTG where the MBP-Q-tagged variants and glycine-methyl-ester used as substrates. Glycine methyl ester was already proven to be an effective mTG acceptor substrate[110], whereas MBP-Qtag had not previously been demonstrated to act as a substrate for mTG.

Table 3.1 Initial rates of mTG-mediated transamidation using the GDH-coupled assay. 0.8 µM of Q-tagged MBP test proteins and 10 mM glycine methyl ester was reacted with 0.1 U of mTG. Initial rates were tested over 20 min using a plate reader.

MBP-Qtag	Initial Rates **(µM/s)**
7M38	n.d.
7M42	n.d.
7M48	n.d.
GSS-7M38	n.d.
GSS-7M42	$(0.1 \pm 0.5) \times 10^{-2}$
GSS-7M48	$(0.7 \pm 0.1) \times 10^{-2}$
$(GSS)_2$-7M38	$(4.9 \pm 0.9) \times 10^{-2}$
$(GSS)_2$-7M42	$(20.5 \pm 0.8) \times 10^{-2}$
$(GSS)_2$-7M48	$(37.2 \pm 0.3) \times 10^{-2}$
MBP	n.d.

Given the promiscuous nature of mTG, selectivity is an ongoing concern, and must be accounted for if mTG is to be used to label proteins. The presence of glutamine residues, specifically on the surface of a protein, may limit the effectiveness of mTG. MBP has 9 glutamine residues that could potential react with the enzyme, leading to non-specific transamidation. As a control, untagged- MBP was also tested for transamidation, with the intent that the absence of a mTG specific tag will prevent any transamidation above background from occurring.

A positive trend can be seen in this initial optimization *in vitro* of mTG-facilitated transamidation of a protein. Namely, the greater the spacer length, the more favourable the reactivity for the tag is for mTG. When no spacer is incorporated into the termini of the protein, there appears to be little to no reactivity with mTG. As spacer length increases, so does the reactivity of mTG for the POI. When spacer lengths are normalized, the specific reactivity of the affinity tag takes precedent, resulting in 7M48 yielding the highest initial rate.

As mentioned previously, untagged MBP was used as a negative control for mTG specificity. No reaction was detected when this protein was used as a substrate for mTG. Regardless of the presence of glutamines on the surface of MBP, mTG showed no measurable affinity for these residues. It can be postulated that although mTG is relative promiscuous in nature, it still requires sequences that flank the reactive glutamine to favour recognition and substrate binding.

3.3 Fluorescent labelling of MBP-Qtag, mediated by mTG

To demonstrate fluorescent labelling, a test protein was first prepared by sub-cloning into the multiple cloning site (MCS) of Maltose Binding Protein (MBP) which is located at the C-terminus of the protein, a spacer sequence (GSSGSS), followed by one of three Q-tags: 7M38 (YPMQGWF), 7M42 (WELQRPY), or 7M48 (WALQRPH). The MBP-Q tag variants were then expressed and purified according to established protocols[112] (see Experimental Section). Once purified, the proteins were incubated at 37 ºC at a final concentration of 0.02 mg/mL in the presence of 1 U of mTG in a reaction mixture containing 2 mM propargylamine, 200 mM MOPS (pH 7.2) and 1 mM EDTA at a total volume of 500 µL. After 2 h, 1 mL of a 1:1 water:tert-butanol solution of 10 mM of 2-azido-1-*N*-dansylethylamine (see Experimental Section for synthesis), 0.1% copper sulphate and 0.01 % sodium ascorbate was added. The ensuing 'click' reaction proceeded overnight at 4 ºC. After the reaction, labelled MBP was washed repeatedly over a membrane having a 14-kDa molecular weight cut-off, effectively removing any unbound dansylamino-ethylazide.

The labelled proteins were then analysed by SDS-PAGE. As shown in in **Figure 3.3**, MBP was effectively fluorescently labelled when bearing the Q-tags 7M48 and 7M42. However, when Q-

tagged with 7M38, a sequence known to have relatively weaker reactivity for mTG[108], little fluorescent labelling was observed, as expected. Qualitatively the fluorescence labelling between 7M48 and 7M42 tags appears to be identical; however, this is not surprising given that the reaction was allowed to run for an extended amount of time. If more stringent reaction conditions were selected (less incubation time with mTG), the 7M48 variant would be expected to show much greater relative fluorescent labelling.

As a control, labelling of less accessible tags was also tested, with the belief that inaccessibility would result in lower reactivity. **Figure 3.2** shows the results from an attempt to label the MBP variants having a shorter spacer or no spacer between the protein and the tag. This experiment provided interesting results. It was assumed that without a spacer, no propargylation was expected; however, some detectable labelling was expected when the GSS spacer was included in the tag sequence. The measured rates suggest that the relatively slow reaction with GSS-7M42/48 would not translate to effective protein labelling within the set experimental parameters. As a negative control, untagged MBP was also subjected to the same labelling experiment. Although MBP contains nine native Gln residues, no fluorescent labelling was observed. This clearly demonstrates that non-specific labelling of intrinsically unreactive Gln residues is negligible, illustrating the distinctive affinity conferred by select Q-tag sequences.

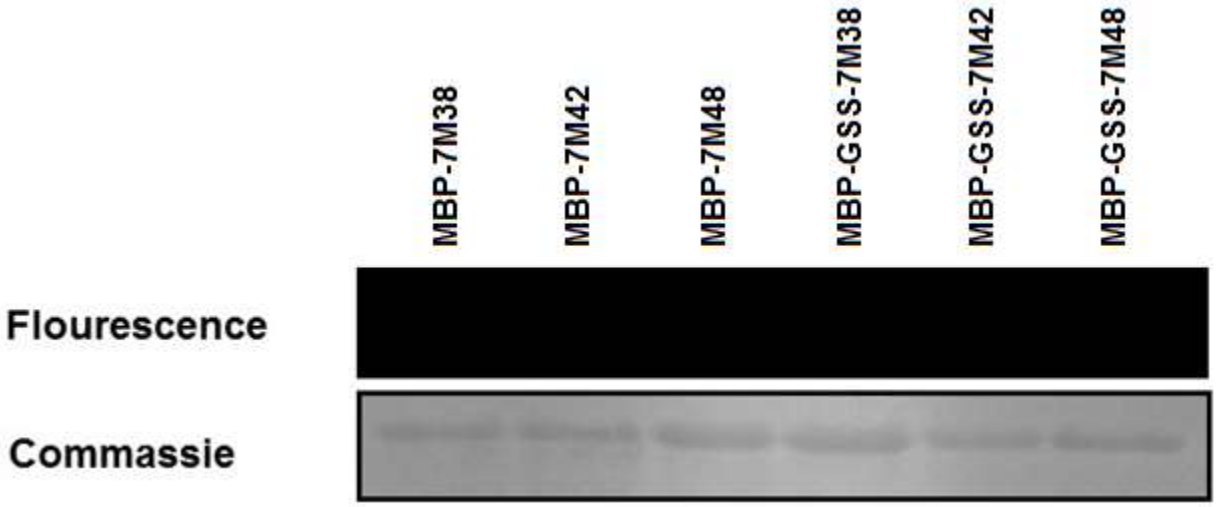


Figure 3.2 mTG-mediated propargylation of MBP bearing high-reactivity Q-tags. Following propargylation, test proteins with no spacer or GSS spacer and mTG recognition tags were fluorescently labeled with dansyl-ethylazide.

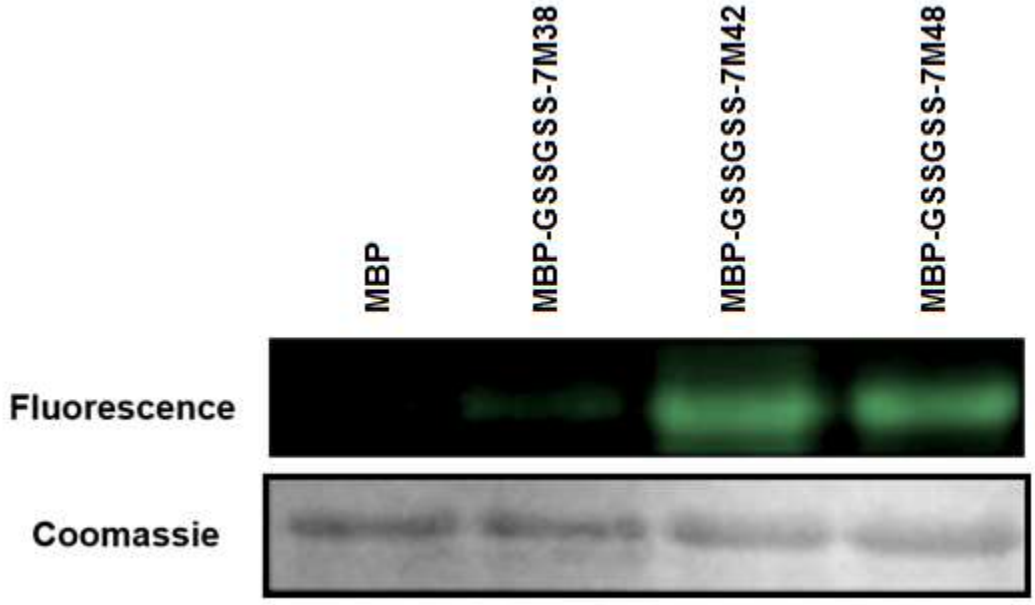


Figure 3.3 mTG-mediated propargylation of MBP bearing high-reactivity Q-tags. Following propargylation, test proteins bearing a GSSGSS spacer and mTG recognition tag were fluorescently labelled with dansyl-ethylazide.

To quantitatively analyse these results, the reactive Q-tagged MBPs were then characterized as Gln-donor substrates for mTG, using the continuous enzyme-coupled assay[113]. Comparison of the specificity constants (k_{cat}/K_M) for the proteins with propargylamine demonstrate that MBP-7M48 is at least ten times more efficient than the other Q-tagged proteins (**Table 3.2**). This add credibility to our previous findings[108], and confirms the 7M48 sequence as a high-reactivity peptide tag for mTG-facilitated protein labelling.

Table 3.2 Specificity constants of mTG-mediated transamidation. The GDH assay was used to measure reactivity of mTG over a protein concentration range between 0.2-0.8 µM Q-tagged MBP and 1 mM propargylamine. Using the Michaelis-Menten equation, initial rates of reaction were plotted against protein concentration and the slope of the linear portion was used to determine the specificity constant.

MBP-Q-tag	k_{cat}/K_M **(µM^{-1} s^{-1})**
MBP-7M38	$(1.3 \pm 0.3) \times 10^{-2}$
MBP-7M42	$(9.7 \pm 0.6) \times 10^{-2}$
MBP-7M48	$(6.8 \pm 0.1) \times 10^{-1}$

3.4 Immobilization, facilitated by mTG, of mRuby2 to azide-functionalized beads (nanoparticles)

In the second proof-of-principle experiment, the aim was the immobilization of a POI. Protein immobilization is used extensively in molecular biology[114] as well as in variety of commercial purposes[115,116]. The ability to immobilize an enzyme onto a solid support is of critical importance, specifically in bio-catalysis. To validate mTG as a tool for immobilization, a GSSGSS spacer and the 7M48 Q-tag was sub-cloned onto the C-terminus of the highly photostable monomeric red fluorescent protein mRuby2. The resulting variant, hereby known as mRuby2-7M48, was then expressed and purified (see Experimental Section) and diluted to 0.01 mg/mL prior to mTG-mediated propargylation. mTG-mediated propargylation was carried out in a similar fashion to MBP-Q tag variants. Following transamidation and washing off of non-covalently bound substrates, the propargylated test protein was incubated with 10 mg of magnetized azide-functionalized nanoparticles (TurboBeads®), in the presence of 0.1% copper sulphate and 0.01 % sodium ascorbate, in 1 mL of a 1:1 water:tert-butanol solution. After overnight reaction at 4 ºC, the functionalized beads were washed and characterized. Since the beads that were covalently bound to mRuby2 were magnetic, they were held to the wall of a microcentrifuge tube with a magnet as they were washed.

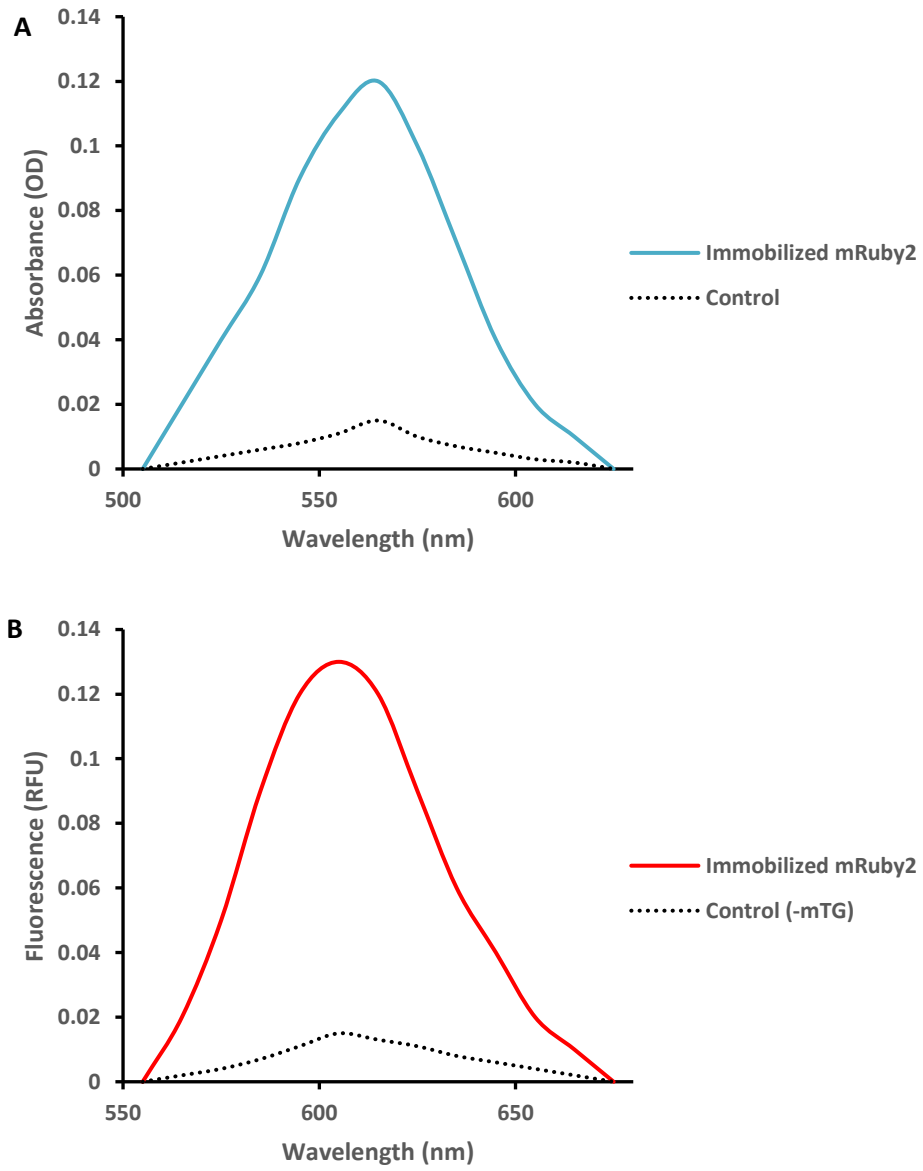


Figure 3.4 Spectral analysis of immobilized mRuby2. Test protein mRuby2-7M48 was propargylated in the presence and absence (negative control) of mTG, then 'clicked' onto azide-functionalized beads. The absorption spectra (A) and fluorescence emission spectra (B) of the treated beads are shown after washing. Spectra were recorded in a 96-well microtiter plate reader.

Spectral analysis in a microtiter plate reader was used to verify that after vigorous washes, beads were now covalently bound to immobilized mRuby2-7M48. As shown in **Figure 3.4**, the labelled beads showed significant absorbance and fluorescence, characteristic of the red fluorescent protein. To ensure that this was not a result of non-specific binding, samples were washed repeatedly to remove any non-covalently bound mRuby2. Due to the time spent incubated with the protein and their tendency to aggregate at low temperatures[117], effective washes were critical. After washes, spectral analysis yielded similarly significant fluorescence at 605 nm, corresponding with mRuby2. As a negative control, the experiment was repeated without the mTG-mediated transamidation. The beads of the negative control did show some minor levels of fluorescence prior to washing. However, fluorescence was diminished with repeated (3x) washing, suggesting it was a result of non-specific binding of mRuby2. Regardless of the washes, there was always clear difference between the control and transamidated product's relative fluorescence. Comparison of the spectra unambiguously confirms the efficacy of the mTG-facilitated immobilization. Additional qualitative evidence in support of this conclusion was obtained by fluorescence microscopy (Experimental Section).

3.5 Conclusions

Through the optimization and demonstration of mTG-mediated propargylation, it is clear that this enzyme can be utilized for bio-conjugation. Once a genetically encodable tag is attached with a spacer having sufficient length to allow the recognition sequence to reach the active site of mTG, successful transamidation has been shown. It was observed that as the spacer length between a POI and an reactivity tag is increased, the relative reactivity of mTG increases proportionately. A spacer exceeding 6 amino acids (GSSGSS) was not attempted, given the desire to minimize the modification of a POI for any bio-conjugation applications.

Furthermore, a large concern in mTG's promiscuity was proven to be a non-issue for the MBP test protein. Untagged MBP was found to have no detectable reactivity according to the GDH-coupled assay. Given the available glutamines on the surface of MBP, this suggests a preference for mTG to recognize glutamines specifically flanked by a recognition sequence. However, given that this is one validated system, it is vital to support mTG-mediated labelling with controls. Later work with mRuby2 added greater confidence to mTG preference for recognition sequences rather than randomly flanked glutamines for reactivity.

Once the POI's were optimized for cross-linking, experiments that clearly demonstrated the effectiveness of mTG were carried out. The efficiency of site-specific propargylation using the 7M48 Q-tag sequence, genetically fused to a POI, was verified. Propargylation was purposefully selected as the first example of transamidation because the subsequent 'click' chemistry that is now available can be used to functionalize the POI with a broad range of azide derivatives. This highlights the versatility of mTG for tagging proteins, not simply for a variety of labelling purposes but also for immobilization. The negligible background reactivity of the Gln residue,

the site-selectivity of the mTG-mediated propargylation of a select Q-tag and the rich diversity of the subsequent bioorthogonal CuAAC 'click' reaction[15] all illustrate the potential for the extension of this method to *in cellulo* labelling.

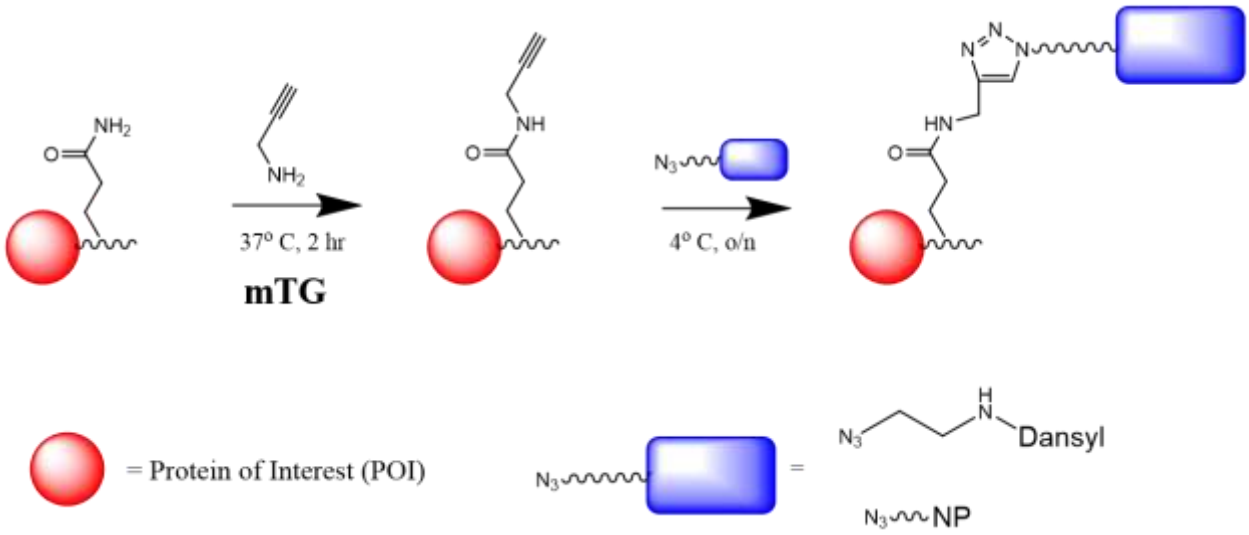


Figure 3.5 Schematic representation mTG labelling. POI are genetically fused with peptide tags, followed by mTG-mediated propargylation of their reactive Gln residues. The propargylated proteins were subjected to copper-assisted azide-alkyne cycloaddition to demonstrate either fluorescent labelling or immobilization.

3.6 Perspectives

Given the relative success demonstrated when utilizing mTG as a tool for bio-conjugation, there are still limitations that must be addressed. *In vitro* and cell surfaces labelling is well within the range of possibility for this enzyme and should be explored to greater extents with the current tools that are available. However, limiting this enzyme to bio-catalytic applications does not take full advantage of this enzyme's potential. Expanding the work with transglutaminase to include *in cellulo* labelling would be the logical next step. Unfortunately, the need to express mTG as a zymogen introduces additional limitations that must be addressed first. A realistic solution in which a TGase is expressed directly in its active form would make that TGase the definitive tool for bio-conjugation. This extension of TGase-mediated labelling will be discussed in the following chapters.

3.7 Experimental Section

3.7.1 Cloning and expression of MBP-Qtag proteins

Expression and purification of tagged MBP variants was performed using established molecular biological techniques[112]. For amplification of MBP-7M38/42/48, forward and reverse primers were synthesized by IDT and used in Touchdown PCR with a pMAL-c5X as a template:

Table 3.3 Oligonucleotides used for cloning the MBP-Qtag variants

MBP_Fw_Bg1 II	5'- GAT CGC TGT TGA AGC GTT ATC GCT GAT TTA TAA CAA AGA TCT GCT GCC GAA CCC GCC AAA - 3'
MBP_7M38_ Rv_EcoRI	5'- GTC GAG AAT TCA AAC CAT CCT TGC ATT GGA TAG ATA TCG CGG CCG CCC ATG GAC ATA TGT GAA ATC CTT CCC TC - 3'
MBP_7M42_ Rv_EcoRI	5'- GTC GAC GAA TTC ATA TGG ACG TTG TAA TTC CCA GAT ATC GCG GCC GCC CAT GGA CAT ATG TGA AAT CCT TCC CTC - 3'
MBP_7M48_ Rv_EcoRI	5' – GTC GAC GAA TTC ATG TGG ACG TTG TAA TGC CCA GAT ATC GCG GCC GCC CAT GGA CAT ATG TGA AAT CCT TCC CTC – 3'
MBP_GSS_7 M38_Rv_Eco RI	5' – GTC GAC GAA TTC AAA CCA TCC TTG CAT TGG ATA GGA GGA GCC GAT ATC GCG GCC GCC CAT GGA CAT ATG TGA AAT CCT TCC CTC – 3'
MBP_GSS_7	5' – GTC GAC GAA TTC ATA TGG ACG TTG TAA TTC GGA GGA

M42_Rv_Eco RI	GCC CCA GAT ATC GCG GCC GCC CAT GGA CAT ATG TGA AAT CCT TCC CTC – 3’
MBP_GSS_7 M48_Rv_Eco RI	5’ – GTC GAC GAA TTC ATG TGG ACG TTG TAA TGC CCA GGA GGA GCC GAT ATC GCG GCC GCC CAT GGA CAT ATG TGA AAT CCT TCC CTC – 3’
MBP_GSSGS S_7M38_Rv	5’ -GTC GAC GAA TTC AAA CCA TCC TTG CAT TGG ATA GGA GGA GCC GGA GGA GCC GAT ATC GCG GCC GCC CAT GGA CAT ATG TGA AAT – 3’
MBP_GSSGS S_7M42_Rv_ EcoRI	5’ – GTC GAC GAA TTC ATA TGG ACG TTG TAA TTC GGA GGA GCC GGA GGA GCC CCA GAT ATC GCG GCC GCC CAT GGA CAT ATG TGA AAT CCT – 3’
MBP_GSSGS S_7M48_Rv_ EcoRI	5’ – GTC GAC GAA TTC ATG TGG ACG TTG TAA TGC CCA GGA GGA GCC GGA GGA GCC GAT ATC GCG GCC GCC CAT GGA CAT ATG TGA AAT CCT – 3’

The reaction was performed in a BioRad® thermal cycler. After amplification, digestion and ligation into pMAL-c5X vector were performed. Inclusion of the Q-tag was verified through DNA sequencing. The constructed vector was expressed in *Escherichia coli* BL21-Gold (DE3) in the presence of 100 mg/mL ampicillin. Purification of maltose binding protein has been established and the protocol is summarized as followed[118]. A 1-mL sample of MBP pre-culture was used to inoculate a culture of 500 mL LB medium + 0.2% glucose. The culture was incubated at 37 ºC until it reached an OD600 of 0.5, followed by the addition of IPTG to a final concentration of 0.1 mM for 2 h. Cells were harvested through centrifugation (4000 g for 20

min). Cells were re-suspended in buffer (20 mM Tris-HCl, 200 mM NaCl 1 mM EDTA, pH 7.4). Suspended pellets were lysed by sonication (3 cycles of 1 min pulses, at a 20% intensity/1 min pause between cycles) using a Branson sonicator. After lysis, the lysate was centrifuged at 20000 g for 20 min, followed by filtration using 0.45-µm filters. A 2-mL volume of amylose resin was added to the column, followed by washing with buffer for 5 column volumes. Once the resin was conditioned, the lysate was poured onto the column with resin and incubated at 4 ºC for 2 h while shaking. Using gravity chromatography, the flow-through was collected, followed by 3 washes (using 2 column volumes per wash). MBP-Q was eluted with 1 column elution buffer (20 mM Tris-HCl, 200 mM NaCl 1 mM EDTA, 10 mM Maltose, pH 7.4). Protein concentration was quantified using Bradford protein assay.

3.7.2 Cloning and expression of RFP-Qtag proteins

Expression and purification of tagged mRuby2-7M48 was performed using established molecular biological techniques. For amplification of mRuby2-7M48, forward and reverse primers were synthesized by IDT and used for Touchdown PCR with a pCDNA3- mRuby2 as a template:

Table 3.4 Oligonucleotides used for cloning the mRuby2-Qtag variants

mRuby2_Fw_BamHI	5’- AAGGATCCGATGGTGTCTAAGGGCGAAGAGCTG- 3’
mRuby2_GSSGSS_7M48 _Rv_EcoRI	5’- AGAGAATTCTTACTTGTACAGCTCGTCATGTGGTCTTTGT

	AATGCCCAACTACTTCC ACTACTTCC – 3’

The reaction was performed in a BioRad® thermal cycler. After amplification, digestion and ligation into pET11a vector were performed. Inclusion of the Q-tag was verified through DNA sequencing. The constructed vector was expressed in *Escherichia coli* BL21-Gold (DE3) in the presence of 100 mg/mL ampicillin. Purification of maltose binding protein has been established and the protocol is summarized as followed. A 1-mL aliquot of pET11a-mRuby2-Q preculture was used to inoculate a culture of 500 mL TB medium. The culture was incubated at 37 ºC until it reached an OD600 of 0.5, followed by the addition of IPTG to a final concentration of 0.1 mM for 3 h. Cells were harvested through centrifugation (3000 g for 30 min). Cells were then resuspended in buffer (50 mM phosphate buffer, 250 mM NaCl 1 mM EDTA, pH 8). Suspended pellets were then lysed by sonication (3 cycles of 2 min pulses, at a 20% intensity/1 min pause between cycles) using a Branson sonicator. After lysis, the lysate was centrifuged at 18000 g for 30 min, followed by filtration using 0.45-µm filters. The mRuby2 lysate was purified using a 1-ml His Trap Ni-NTA column (GE Healthcare) equilibrated in 50 mM phosphate buffer (pH 8.0) with 300 mM NaCl and eluted with an imidazole gradient (0–140 mM) on an Åkta FPLC instrument (GE Healthcare). The purified mRuby2 was dialyzed against 50 mM phosphate buffer (pH 8.0). Protein concentration was quantified using Bradford protein assay

3.7.3 Synthesis of dansylethylazide

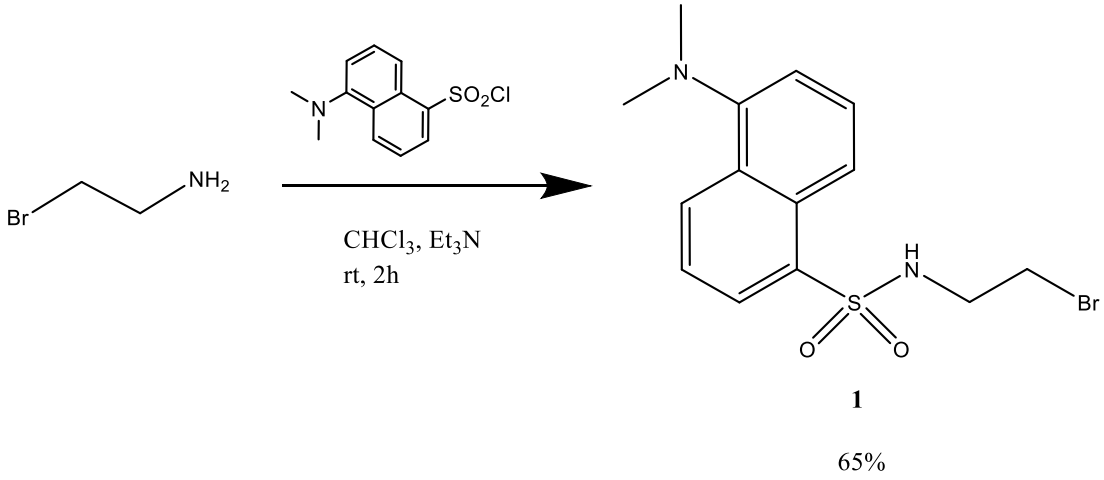


Figure 3.6 Synthetic scheme of 2-Azido-1-*N*-dansylethylamine

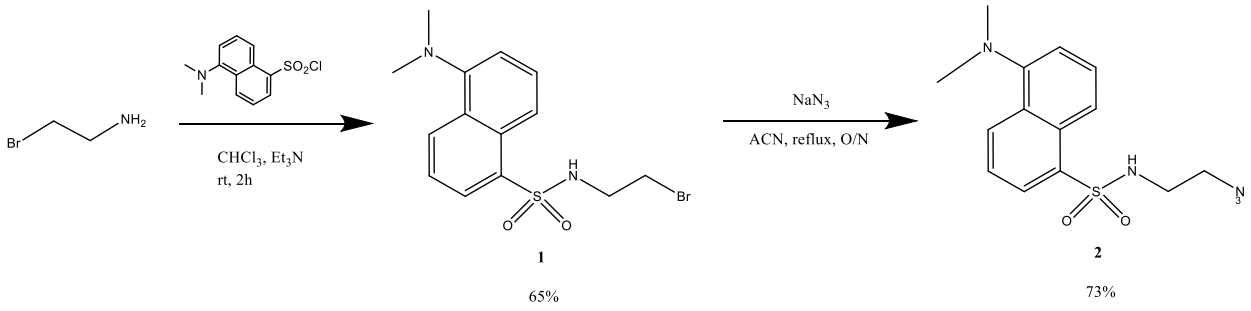


Figure 3.7 Synthetic scheme of 2-bromo-1-*N*-dansylethylamine

2-bromo-1-*N*-dansylethylamine (1): A solution of 2-bromoethylamine (0.21 g, 1.04 mmol, 1.1 eq) in $CHCl_3$ (5 mL) was added to Et_3N (0.35 ml, 2.5 mmol, 2.5 eq). While stirring, a solution of dansyl chloride (0.24g, 0.91 mmol, 1.0 eq) in $CHCl_3$ (1 mL) was added dropwise. The reaction mixture was stirred at room temperature for 2.5 h. The solvent was then evaporated and a yellow oil was obtained and dissolved in EtOAc. This was washed three times with a solution of HCl (1M) in water, three times with a saturated solution of $NaHCO_3$ and three times with a saturated solution of NaCl. The organic phase was dried over $MgSO_4$ and evaporated to give 1, as a yellow oil (231 mg, 0.65 mmol, 65%).

^{1}H NMR ($CDCl_3$, 300 MHz) : δ (ppm) 8.54 (m, 1H); 7.25 (m, 2H); 7.54 (m, 2H); 7.18 (m, 1H); 5.47 (s l, 1H); 3.27 (s, 4H); 2.86 (s, 6H); **^{13}C NMR ($CDCl_3$, 75 MHz):** δ (ppm) 152.0; 134.5; 130.8; 129.9; 129.8; 128.6; 128.4; 123.1; 118.6; 115.3; 45.5; 44.6; 31.5; 27.9

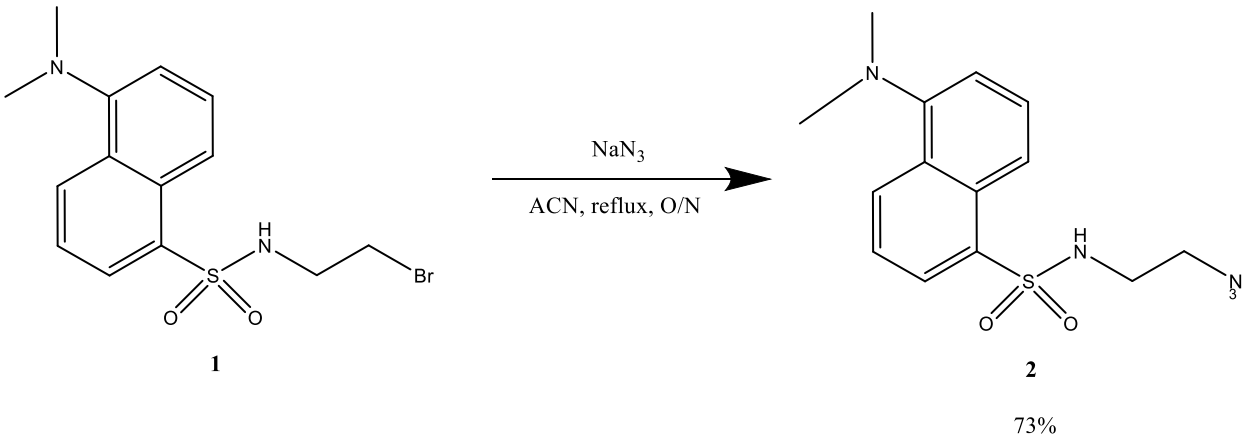
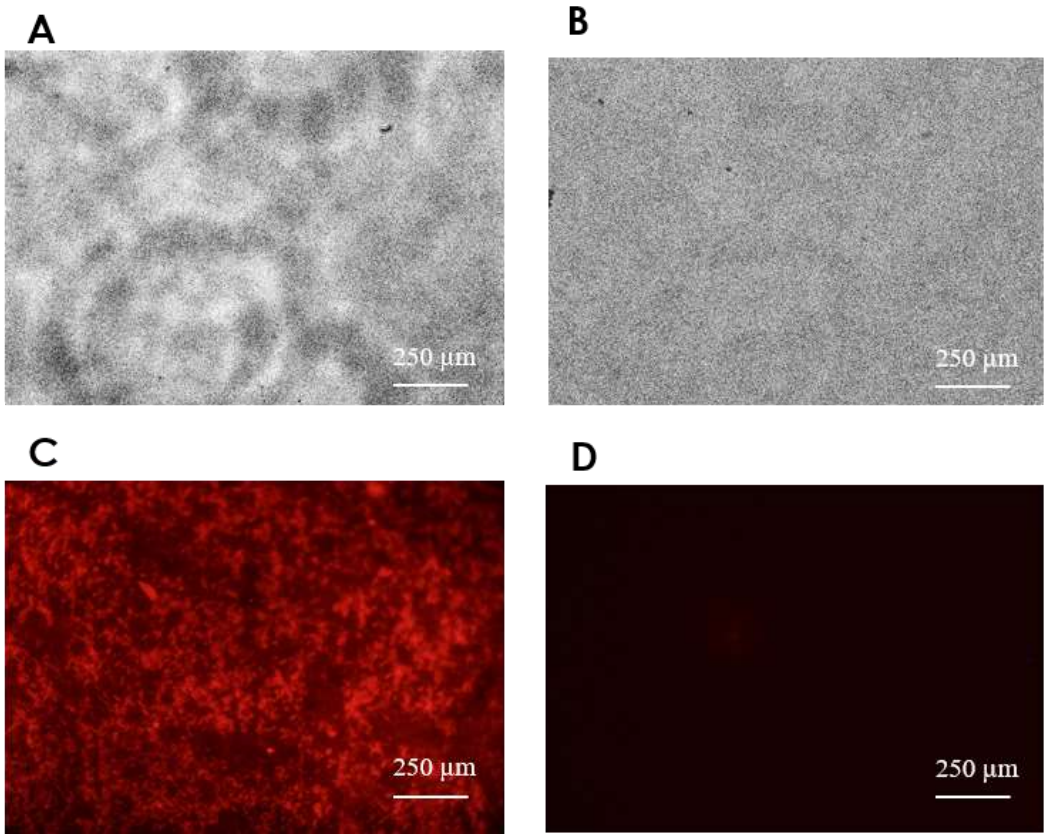


Figure 3.8 Synthetic scheme of 2-Azido-1-*N*-dansylethylamine

2-Azido-1-*N*-dansylethylamine (2): The yellow oil 2 (1.14 g, 3.19 mmol, 1.0 eq), was mixed with ACN (50 mL) followed by the addition of NaN_3 (0.84 g, 12.96 mmol, 5.0 eq) while stirring. The reaction mixture was heated to reflux overnight. After evaporation of the solvent, the aqueous phase was extracted with EtOAc. The organic phase was washed three times with water and once with a saturated solution of NaCl. Thereafter, the solution was dried over MgSO4 and evaporated to dryness. The oil obtained was purified by flash chromatography on silica gel with Hexane/ EtOAc (7:3) to give 2 as a yellow oil (750 mg, 2.34 mmol, 73%).

^{1}H NMR ($CDCl_3$, 300 MHz): δ (ppm) 8.57 (d, 1H, *J*=8.49 Hz); 8.26 (d, 2H, *J*=7.65 Hz); 7.60 (t, 1H, *J*=7.86 Hz); 7.53 (t, 1H, *J*=8.49 Hz); 7.21 (d, 1H, *J*=7.32); 4.96 (t, 1H, *J*=6.66 Hz); 3.31 (t, 2H, *J*=5.49 Hz); 3.06 (q, 2H, *J*=5.52, 6.15 Hz); 2.89 (s, 6H) **^{13}C NMR (CDCl3, 75 MHz):** δ (ppm) 130.8; 129.6; 128.7; 123.1; 118.4; 115.4; 50.9; 45.4; 42.4. **IR:** 2104 cm^{-1} .

3.7.4 Imaging of mRuby2-conjugated Nanoparticles



Imaging of magnetic, azide-functionalized nanoparticles (30 nm/particle), conjugated through cylco-addition to propargylated mRuby2-Qtag. Images were taken with a confocal fluorescence microscope using the appropriate filters at a magnification of 40X. A) Bright field imaging of mRuby2-Qtag, conjugated to nanoparticles through mTG-mediated propargylation of the protein and B) control experiment performed in the absence of mTG. C) Dark field imaging of mRuby2-Q-tag, immobilized to nanoparticle through mTG-mediated propargylation of the protein and D) control experiment performed in the absence of mTG.

Chapter Four: **Identification and characterization of *Bacillus subtilis* transglutaminase substrates**

4.1 Introduction

In this chapter, we shift our focus from the application of mTG as a tool for bio-labelling to the characterization of bTG, a lesser-known homologue in the transglutaminase family. At this stage of the research, the limitations of mTG have resulted in the necessity to discover an alternative labelling strategy, ideally while maintaining the positive attributes of the transglutaminase family.

4.1.1 Disadvantages of mTG

As a tool for bio-conjugation, microbial transglutaminase shows great promise. In the preceding chapters, the advantages to mTG in comparison to other bio-labelling enzymes are discussed in detail. The relative small size (38 kDa), efficient catalytic turnover, broad substrate specificity and stability of mTG suggest that this enzyme is among the best suited for bio-catalysis[119]. The work in previous chapters helps to determine the potential of this enzyme, specifically as a bio-catalyst and highlight its capacity for *in vitro* labelling.

However, mTG has one significant limitation that mutes its functionality, specifically for *in cellulo* labelling. Natively, mTG is derived from *Streptomyces mobaraensis* in its pro-peptide-enzyme form. The inclusion of the pro-peptide is hypothesized to facilitate the transfer of the enzyme through the cytoplasm of a cell and into its medium[120]. More importantly, analysis of the structure of mTG suggests that the N-terminal pro-peptide sequence shields the active site of mTG, rendering the enzyme inactive while the peptide is bound. Given the promiscuity of mTG, it is hypothesized that the active form of the enzyme is potentially toxic to the host. Its ability to cross-link proteins may prove to be disastrous within a cell, making inactive expression

necessary[121]. To activate the original zymogen, *S. mobaraensis* also secretes two proteases that are responsible for the cleavage of the N-terminal pro-peptide[122]. Extensive research has gone into expressing mTG in yeast and *E. coli*; however, at the beginning of this research, expression of the enzyme in its active form was not feasible. Although this had no bearing on the initial uses of mTG (cross-linking meat by-products in the food industry[63,64,123]), as the focus of the field shifted to bio-catalysis/conjugation, the requirement for mTG to be expressed in its active form became a limitation.

Trypsin and dispase are the proteases of choice for the non-native cleavage of the mTG pro-peptide[124]; these proteases efficiently and quickly activate mTG. However, as the purification process summarizes, trypsin's lack of specificity may lead to degradation of mTG over time[125]. Therefore, removal of the protease as quickly as possible through ion-exchanges is critical. This additional purification step does not hinder any *in vitro* applications of mTG; however, it does present a new problem when attempting to label within a cell. Expressing a protease within a cell may lead to a variety of complications due to the protease's lack of specificity over time. Unfortunately, without trypsin, there is no guarantee that mTG will be activated, limiting the scope of enzymatic labelling with mTG.

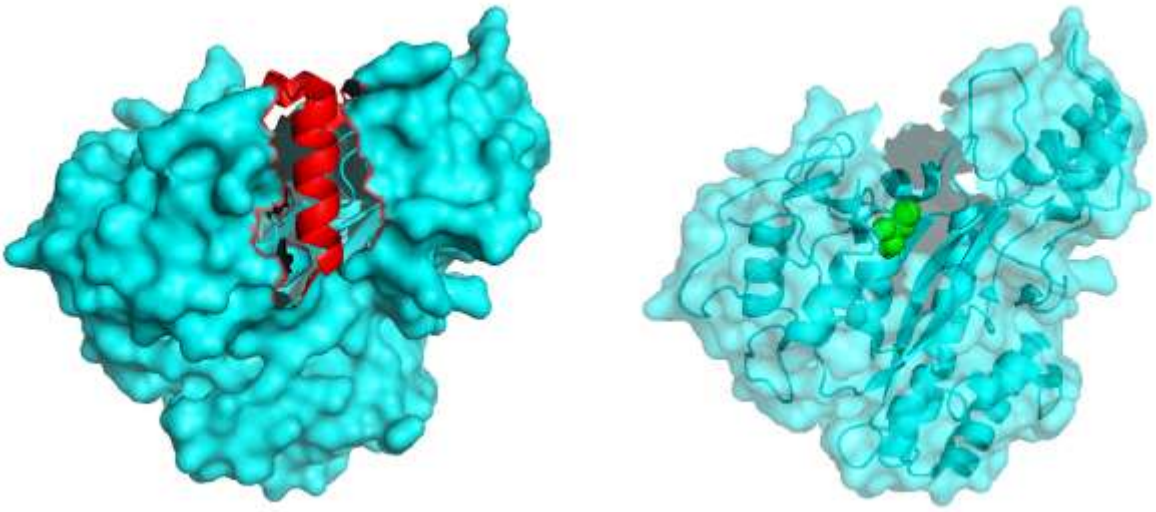


Figure 4.1 Crystal structure of mTG. PDB:3IU0. (Left) Surface representation of mTG zymogen with active site covered by bound α-helix pro-peptide (red). (Right) Surface representation of active mTG, exposing the active site cysteine (green) after cleavage of pro-peptide.

4.1.2 Solutions to mTG limitations

To mitigate the issues presented by mTG's activation, there are realistic solutions that can be explored. The proteolysis of a zymogen enzyme is the main deterrent to *in cellulo* modification of POI. Due to the inability to control trypsin-mediated activation, there is an increased risk of protein degradation throughout a cell when the protease is over-expressed. However, there are other proteases that exhibit much greater precision and may be better suited to cleave mTG after expression. Factor Xa is a relatively high-fidelity protease[126], in part because of its expanded recognition site: trypsin generally cleaves on the carboxyl side of arginine or lysine, Factor Xa cleaves specifically after the recognition site isoleucine-glutamate/aspartate-glycine-arginine[127,128]. By exchanging the trypsin for a less promiscuous protease, the concerns over non-specific proteolysis are significantly reduced. This may then allow mTG to be expressed as a zymogen, which results in the greatest yield, and still be activated within a cell. Unfortunately,

this strategy still depends on the co-expression of a protease. Given the potential for non-specific protease activity, specifically as time increases, there is a strong likelihood for unforeseen effects on the cell. One of the initial qualities discussed for a bio-labelling system was minimal perturbation to a system; this cannot be guaranteed in the presence of a protease. Furthermore, proteases are derived from eukaryotic and prokaryotic cells[129,130], which may mean the expression of the protease could lead to over-expression in certain cells types. This could also have unpredictable results for any *in cellulo* applications of the enzyme labelling strategy.

As previously mentioned, at the beginning of this research, there was no published work on *E. coli*-based recombinant active mTG expression. However, midway through this work, seminal research from Salis and co-workers resulted in the engineering of mTG towards expression in its active form[121]. The group's objective was to obtain active mTG in large quantities for *in vitro* applications such as food processing and labeling of therapeutic proteins for improved drug delivery[36,131,132]. This strategy rectifies the limitations previously held by mTG while maintaining its advantages. If proteolytic activation of mTG is required, *in cellulo* applications using mTG are limited. Although engineering the enzyme is a reasonable solution to address this issue, there are still questions about relative expression and activity levels that may not be addressed[121]. mTG activated via proteolysis yields enzyme at approximately 30 U/mg while recombinant actively expressed mTG's activity is half that at 15 U/mg[121,133]. As a bio-catalyst, improved activity is generally the goal.

Removal of the pro-peptide sequence seems to be the most effective future approach to attempt for in-cell labelling of a POI. However, given the inherent drawbacks of mTG, we thought it would be beneficial to consider another enzyme, ideally one where the advantages of mTG are preserved, but the activation of the enzyme is not required. There are members of the

transglutaminase family that do not required proteolytic activation; unfortunately, most of these enzymes are mammalian[134]. This would introduce a variety of issues that mTG was able to circumvent in the first place. Alternatively, utilizing another bacterial transglutaminase, preferably expressed in its active form, would be the ideal solution to the transglutaminase issue. A lesser known bacterial transglutaminase, derived from *Bacillus subtilis*, has been shown to have similar attributes to mTG and does not require activation[135]. Substituting this enzyme for any *in cellulo* bio-labelling applications maybe the best option for enzyme-facilitated protein modification.

4.2 Bacillus subtilis transglutaminase

Bacillus subtilis transglutaminase, bTG, is derived from the *Bacillus subtilis* and was discovered in 1996[136]. It is implicated in the protection of *B. subtilis* via the cross-linking of multiple coat proteins on the surface of a spore[136]. This multiprotein layer serves as a coat of resistance against lytic enzymes and noxious chemicals, thereby enabling normal germination[137]. Studies have identified bTG expression levels to be directly proportional to life cycle of the cell, with bTG expression increasing as cells begin to form spores[137]. Therefore, bTG is characterized as part of a defense mechanism, specifically in adverse conditions. This native functionality suggests that the enzyme is robust in nature and may be a practical mTG replacement if tailored correctly[137,138].

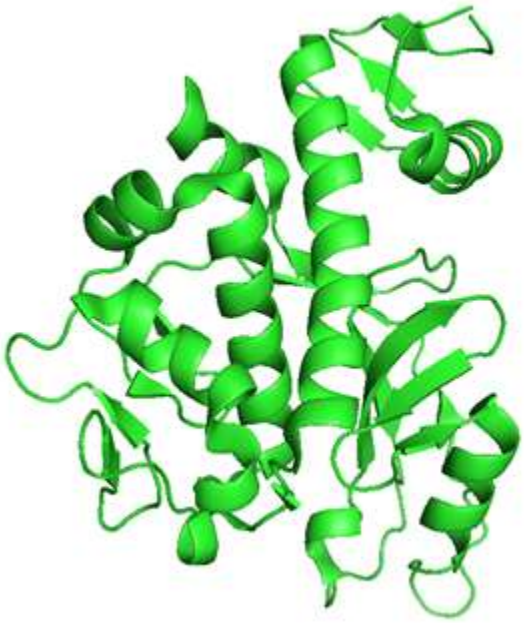


Figure 4.2 Crystal structure of bTG. Cartoon representation of bTG crystal structure (PDB:4P8I).

At 28 kDa, bTG is smaller than mTG and shows little structural homology to its bacterial homologue. There are additional characteristics that differentiate the two bacterial enzymes, including optimal temperature and pH (60 ºC and pH 8.2 for bTG/ 50 ºC and pH 6-7 for mTG)[139,140]. Although there is not much similarity, we hypothesized that the new enzyme maintains an equally broad substrate scope; additionally, bTG's native environment suggest it is a robust enzyme that may be well-suited for bio-catalysis[44,138]. It should be noted that the over-expression of bTG in its mature form does have a toxic effect on bacterial hosts, hindering cell growth and resulting in relatively low yields of bTG[138]. Studies to express bTG in heterologous strains have been successfully executed; however, this defeats the purpose of this research, since active expression is desired.

4.2.1 Reaction Mechanism

To date, there has been limited research on bTG, apart from its native functionality[136,138]. However, recent work by the Fernandes and co-workers resulted in the first crystal structure of bTG[141]. This confirmed the hypothesis (based on sequence alignments) that there was limited structural homology between mTG and bTG[141]. This also limits the relevance of mTG-based specificity to this new enzyme. Substrate scope and reaction mechanism must be determined for bTG independently of previous TGases. Although bTG does not bear any resemblance to other TGases, it does show similarities to other proteases, specifically papain proteases[142]. The Cys-His catalytic core of papain is largely conserved when superimposed with bTG (asparagine in papain replaced with aspartate in bTG)[141,142]. This structural overlap allowed for the proposal of a catalytic mechanism in which the Cys-His-Asp within the assumed catalytic core mediates the transamidation reaction. As shown in **Figure 4.3**, the active site facilitates the nucleophilic attack on the glutamine carbonyl, leading to the formation of a tetrahedral intermediate. During the collapse of the tetrahedral intermediate, an equivalent of ammonia is released, resulting in an acyl-enzyme intermediate. Lysine is then deprotonated by the catalytic histidine during the nucleophilic attack on the acyl-enzyme intermediate. An additional tetrahedral intermediate is formed and collapses, regenerating the catalytic core and yielding a cross-linked product.

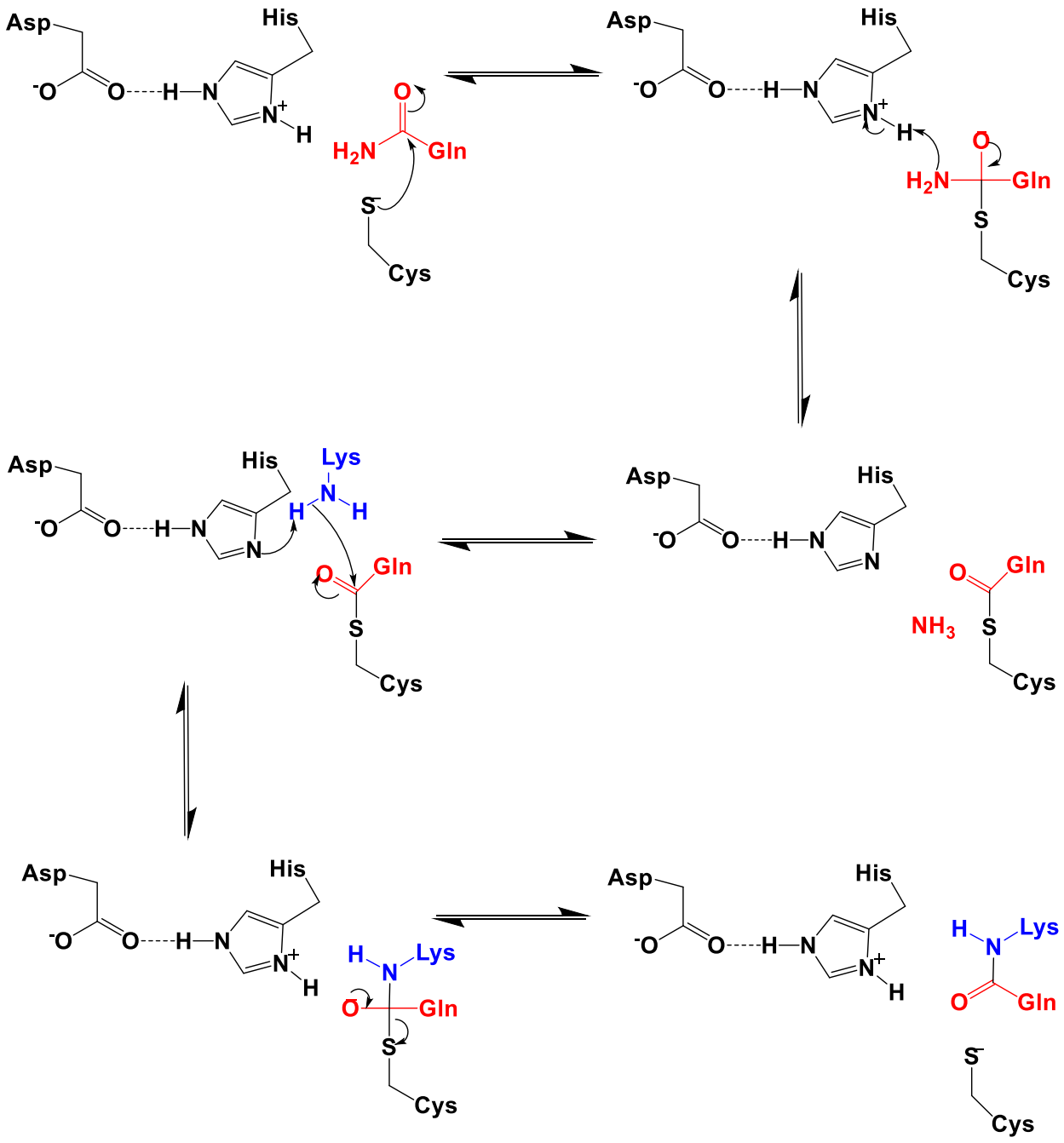


Figure 4.3 Proposed reaction mechanism of mTG. Reaction mechanism of transamidation, based on the catalytic core similarity between bTG and papain proteases. In this scheme, Gln (in red) and Lys (in blue) represent residues bound to individual proteins.

4.2.2 Validation of proposed bTG 'high reactivity' peptides

To utilize bTG in any bio-labelling strategy, the identification of a high reactivity peptide is critical. Unlike mTG, there are few examples in the literature of peptides that confer high reactivity for bTG. Lee and co-workers published preliminary results on proposed "high-reactivity" peptides[143] in the first attempt to study bTG substrate specificity. Three peptides were tested for bTG reactivity based on their relative affinity for various TGases. The sequences TQ1 (RTQPA) and TQ2 (RLQQP) were designed as bTG donor substrates, based on mTG specificity studies; the authors hoped that these peptides would show affinity for bTG as well[143,144]. As a control, the "Q peptide" (PQPQLPYPQPQLPY from gluten), a known substrate for TG2[145], was also tested. Reactivity was assessed by analyzing conversion over time. Q and K substrates were incubated with bTG for 40 min. After this time, HPLC analysis of each reaction solution was used to determine how much transamidation product was formed, relative to unreacted substrates.

Table 4.1 Evaluation of relative TGase reactivity by substrate conversion (%). Solutions of 200 µM of Q-substrates were reacted with 200 µM of hexa-lysine over 40 min at 37 °C. The assay was performed in triplicates; conversion averages and SD are shown[143,144].

Enzyme	Q substrate	Conversion (%)	SD (%)
TG2 from guinea pig	Q peptide	95.7	± 1.2
	TQ1	13.0	± 2.6
	TQ2	38.3	± 1.5
mTG from *Streptomyces mobaraensis*	Q peptide	50.7	± 4.0
	TQ1	93.0	± 2.0
	TQ2	52.7	± 2.1
bTG from *Bacillus Subtilis*	Q peptide	75.6	± 5.9
	TQ1	90.4	± 0.9
	TQ2	58.4	± 2.0

Over the course of 40 min, both TQ1 and TQ2 demonstrated varying levels of reactivity with all three TGases that were tested. The results shown in **Table 4.1** suggest that both tags have some affinity for bTG. Unfortunately, this information does not indicate whether these tags have an increased reactivity for bTG over other TGases. Furthermore, the lack of selectivity may limit the application of these tags in mammalian cells, where TG2 is endogenously expressed[146]. Although conversion rates varied depending on the TGase tested, it is difficult to extrapolate information about reactivity or selectivity based solely on these results. Therefore, we concluded that to identify the sequence determinants that confer reactivity for bTG, a more thorough approach was needed.

The GDH-coupled assay has been successfully applied to the measurement of TG2 and mTG activity[108,147], since it is based on the release of ammonia, which is common to all such transamidation reactions. Therefore, this assay could also be adapted to monitor bTG activity, as a suitable method to determine substrate reactivity. Previous optimization of the assay for mTG minimizes the necessity to repeat those steps for bTG. Both bacterial TGases are functional over a similar pH range and do not have any detrimental effect on glutamate dehydrogenase. Both bTG peptides proposed by Lee (TQ1 and TQ2) were synthesized to identify whether bTG showed significant reactivity. Using the GDH-assay, initial rates were measured over the course of 20 min at varying concentrations of peptide substrate and a constant concentration of glycine-methyl ester. Given the limited information of substrate reactivity, glycine-methyl ester was selected as a starting point for bTG acceptor substrate. This experiment was carried out for both bTG and mTG to determine the relative reactivity of each substrate.

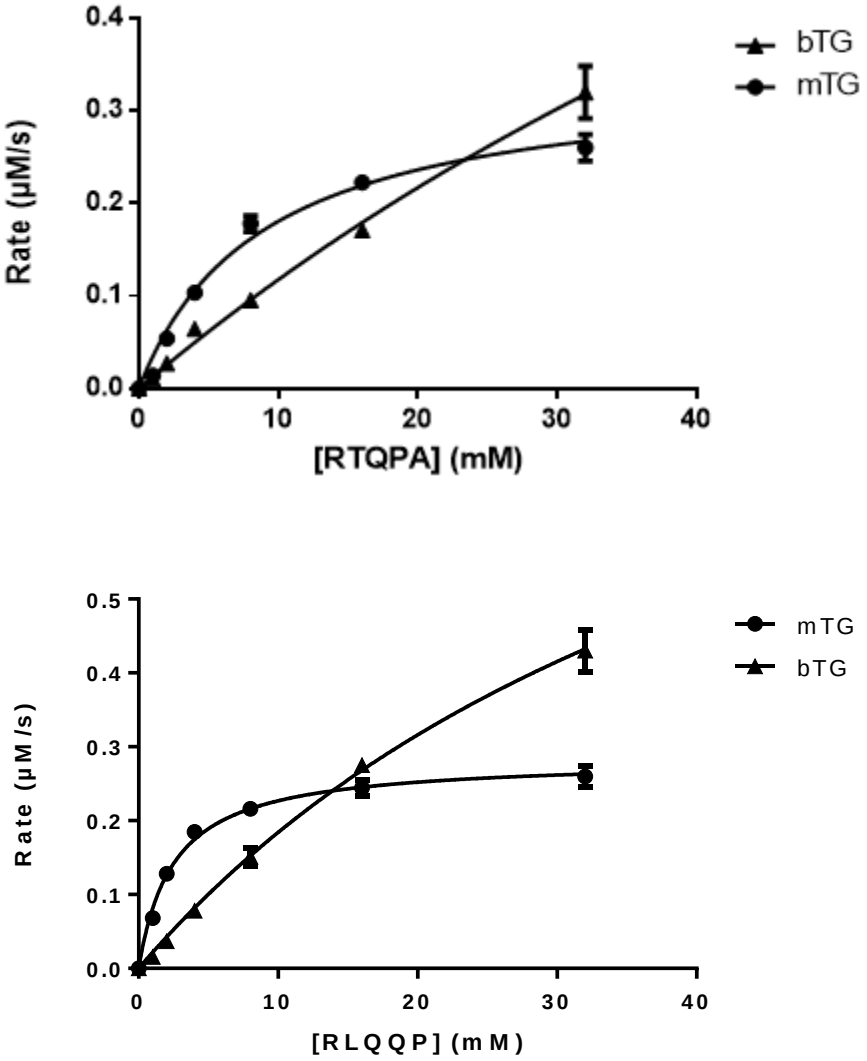


Figure 4.4 Michaelis-Menten plot of mTG and bTG activity measured by the coupled-GDH assay. Using varied concentrations of RTQPA (top) and RLQQP (bottom) as acyl-donor substrate (1-32 mM) in the presence of 10 mM Gly-OMe as acyl-acceptor substrate, GDH-coupled activity assays were carried out in triplicate using a 96-well microplate reader. 0.2 units of mTG and bTG were used. Error bars represent standard error of the mean values. The solid line through the data points represents fitting to the hyperbolic Michaelis-Menten equation.

Figure 4.4 unambiguously shows that at concentrations up to the solubility of the donor substrate, kinetic parameters such as K_M and V_{max} cannot be determined for bTG. This would suggest that both RLQQP and RTQPA are relatively low-reactivity substrates for bTG. RTQPA appeared to approach a plateau at the highest concentration of 32 mM; however, working at such high concentrations is not practical, especially for *in cellulo* labelling. A minimal curvature in transamidation rate was observed when RLQQP was tested with bTG. Interestingly, both peptides showed markedly improved results in the presence of mTG. RLQQP plateaued well within the saturation concentrations of the peptide. Furthermore, the apparent K_M of RLQQP was approximately the same as 7M48 (WALQRPH) at 3 mM. However, the presence of two glutamines within the tag is assumed to have an enhanced effect on reactivity as there are multiple opportunities for catalysis on the same peptide. In summation, kinetic characterization of the bTG peptides suggests that they do not confer high reactivity. It should be noted that the origin of these tags was the reactivity of their sequences with mTG; therefore, poor bTG reactivity may not be surprising. More work must be done to identify a bTG tag that can be used for bio-labelling applications.

4.3 *In vitro* FRET based assay

As discussed previously, limited research has been conducted on bTG, and even less on bTG's substrate specificity. Due to the dearth of information on the scope of bTG affinity, primary research to study peptide substrate scope is necessary, in order to advance labelling applications with this enzyme. For other TGases, phage-displayed libraries were designed that allowed screening of an array of donor peptides with amino acids at different positions, relative to a

reactive glutamine[72,73,144]. This high-throughput approach was an effective method to identify potential high-reactivity tags. The strategy takes advantage of the enzyme's ability to incorporate a fluorescent amine, typically dansyl cadaverine, into a glutamine peptide substrate[72]. The ability to relate a visual signal with a transamidation event is what makes this technique viable for screening peptide's TGase reactivity. This concept can be utilized in an alternative approach where screening donor substrates in a high-throughput manner leads to the identification of peptides that bTG shows relatively high reactivity towards. Herein we present a method developed to identify peptides that exhibit high reactivity for bTG.

4.3.1 Protein-Protein FRET-based assay

As discussed in chapter 1, fluorescent proteins are used throughout molecular biology, mainly for tracking proteins of interest. The intrinsic fluorescence of these proteins and the ability to monitor them using a multitude of diagnostic tools affords the opportunity to use them in a screening strategy. Palmer has reviewed many strategies focussed on how FP's can be adapted as a method of detection[148] for in-cell imaging. Through identification of FP's with ideal spectral characteristics, monitoring the effects of protein-protein interactions becomes much more simplistic. Lee took advantage of bTG and its ability to cross-link peptides, utilizing FRET as a proximity-based indicator. A method was created to determine whether two FRET partners are covalently attached to each other. Encoding potential bTG tag sequences to the termini of fluorescent proteins permits the detection of bTG activity by the observation of a FRET signal[143]. As Lee used this methodology to scan the substrate scope of bTG acceptor peptides, our research concurrently used a similar approach to probe the donor substrate of bTG.

Figure 4.5 gives a pictorial representation of protein-protein FRET-based monitoring of transamidation. In the presence of a transglutaminase, the Q and K tag are cross-linked, resulting in a protein-protein fusion. The fusion product can be detected by exciting the FRET donor; due to the proximity of the two FPs, an energy transfer results in excitation of the FRET acceptor and its subsequent emission. The FRET- based determination of transamidation can be expanded upon in a variety of ways. For the purposes of this study, it can be modified into a high-throughput assay. The incorporation of randomized residues flanking the reactive glutamine defines an experiment to identify peptide sequences that exhibit affinity for bTG. However, before screening for peptides that are favoured by bTG, the assay must be optimized.

Figure 4.5 Visual representation of FRET-based peptide screening assay. Cartoon representation of the conjugation of mRuby2-Q-tag and Clover-K-tag, in the presence of TGase, resulting in a cross-linked product. Due to the spectral overlap of mRuby2 and Clover, when in close proximity, excitation of Clover leads to FRET, and red emission by mRuby2.

4.3.2 Optimization of FRET-based assay

The success of the FRET-based assay is reliant on the spectral overlap of the two FRET partners. To address this concern, mRuby2 and Clover were selected as FRET partners since they have been optimized for maximal FRET signal in previous studies by Lin and co-workers[149]. This work allowed us to focus our optimization efforts on transamidation rather than on the viability of FRET. **Figure 4.6** highlights the spectral overlap of Clover and mRuby2 and demonstrates why the two proteins yield a high FRET signal to FRET donor ratio. The fluorescent proteins have been engineered so that the emission band of Clover coincides with the excitation band of mRuby2. Additionally, minimized overlap between the excitation bands of the FRET donor and the FRET acceptor is key. Overlap would lead to direct excitation of the acceptor, resulting in a false positive of FRET; mRuby2 and Clover show minimal overlap in this respect. Due to the large overlap of Clover emission/mRuby2 excitation and minimal overlap of Clover excitation/mRuby2 excitation, the two FPs are effective FRET partners.

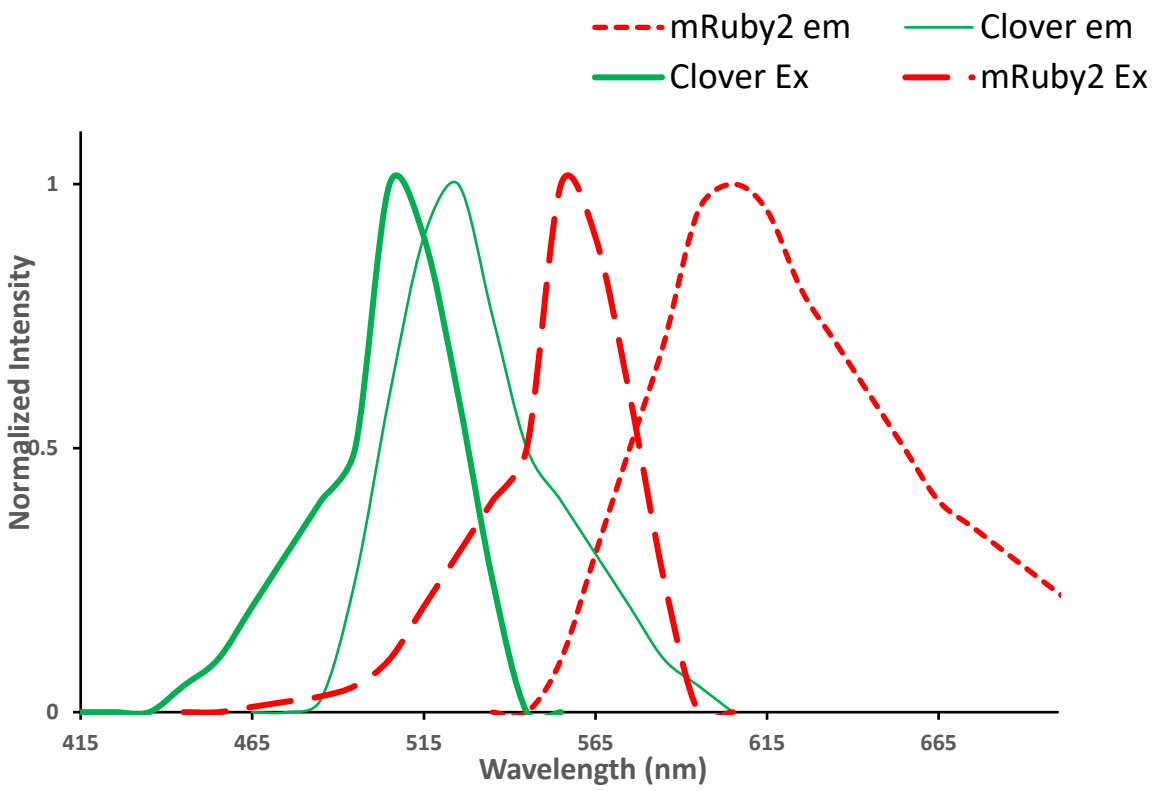


Figure 4.6 Assessment of FRET reporters. Normalized intensity for excitation and emission of mRuby2 and Clover. Significant overlap of Clover emission and mRuby2 excitation is critical for success of FRET signal.

Before the screening experiment could be carried out, it was necessary to determine what a positive result would look like. As a control for this experiment, an mRuby2-Clover fusion protein was designed that mimicked the spacer length between the expected mRuby2-Clover transamidation product. Given that all designed tagged proteins include a GSSGSS spacer and that reactive glutamine/lysines were to be placed four residues into a tag, the fusion product was designed with a 19-residue spacer between the two FPs. The fusion control acts as the best possible result for the FRET assay, setting the bar for what a positive result should resemble. It is evident in **Figure 4.7** that the positive control gives a strong FRET signal (excitation in the green, emission in the red) but also shows significant green emission. This would suggest that

the proximity and orientation of the FPs is not perfect; however, it is very difficult to observe 100% energy transfer.

Although the FPs selected have minimal excitation overlap, there is still a concern that the limited direct excitation of mRuby2 may result in false positives that can distract from the true results of the experiment. To address this issue, we explored the excitation of Clover at a lower wavelength, below its maximum. This would reduce the overall signal transferred, but it would eliminate any direct excitation of mRuby2, reducing false positive results. This result is shown in Figure 4.7 where shifting the excitation of Clover reduces the overall signal of red emission but ensures the signal is based on FRET and not direct excitation of mRuby2.

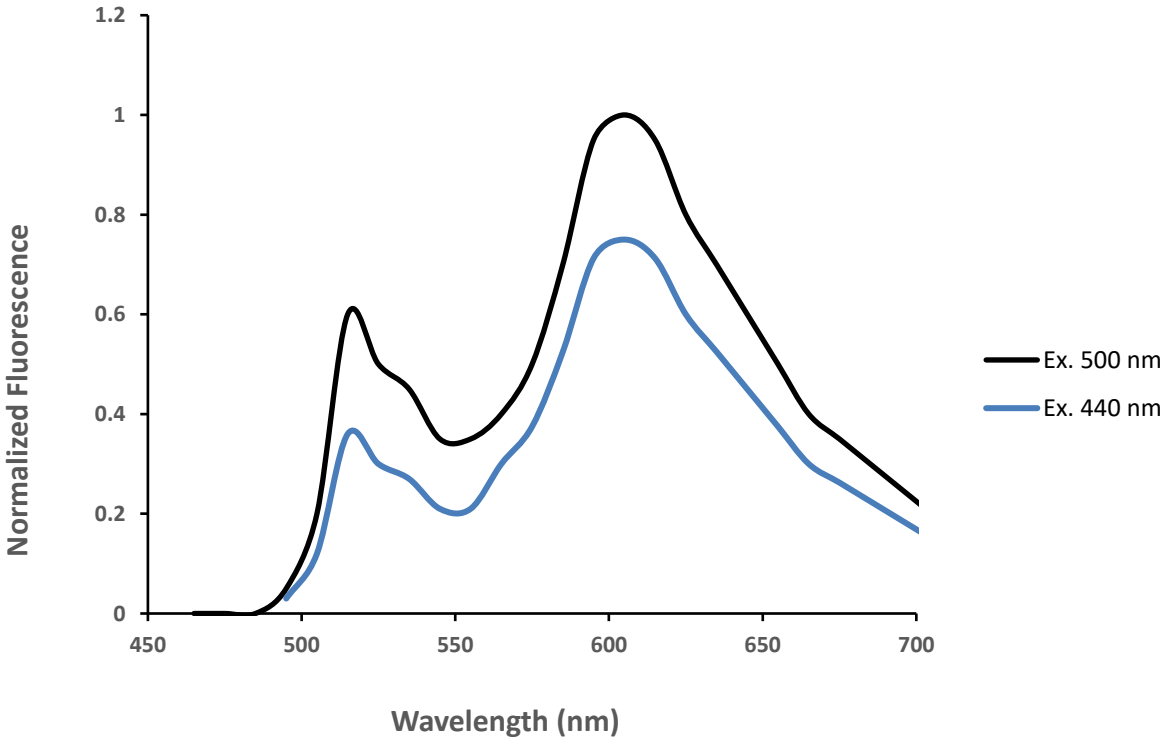


Figure 4.7 FRET Calibration. Emission spectra of mRuby2-Clover fusion at different excitation wavelengths. Excitation of the fusion protein at the Clover excitation maximum, 500 nm, resulted in greater red and green emission. Excitation at the front end of Clover's excitation band reduced the overall emission but still resulted in significant red emission due to FRET.

4.3.3 In vitro FRET-based assay

Before attempting to screen for a high reactivity bTG tag, we reasoned it would be beneficial to complete the assay with a well-known enzyme. Given how little is understood about bTG relative to other TGases, a trial conducted with mTG would help eliminate uncertainty during the optimization process. As a trial, mTG transamidation was monitored through the detection of FRET over time between mRuby2-7M48 and Clover-6K, where 6K represents a hexa-lysine tag. The hexa-lysine tag was originally used by Lee in the study that identified TQ1 and TQ2 as potential high reactivity bTG tags[144].

The FRET assay was monitored over 24 hours through emission spectrum scans with excitation at 440 nm. As the reaction proceeded (**Figure 4.8**), there was a noticeable decrease in fluorescence intensity in the green region of the visible spectrum and an increase in the red. Given the composition of the reaction mixture, it was evident that as the transamidation product accumulated, the FRET signal increased. The decrease in green fluorescence was not proportional to red fluorescence; however, this was expected since it was already established that perfect energy transfer was not attainable. Quantitatively, the transamidation product's FRET signal was significantly attenuated, relatively to the fusion control. **Table 4.2** highlights the difference in intensity after 24 hours; the transamidation product's intensity was 14381 RFU, or four-fold less fluorescent than that of the fusion protein control. Although the experiment was not able to replicate the spectral readout of the mRuby2-Clover fusion, successful monitoring of the transamidation of two tagged fluorescent proteins was achieved.

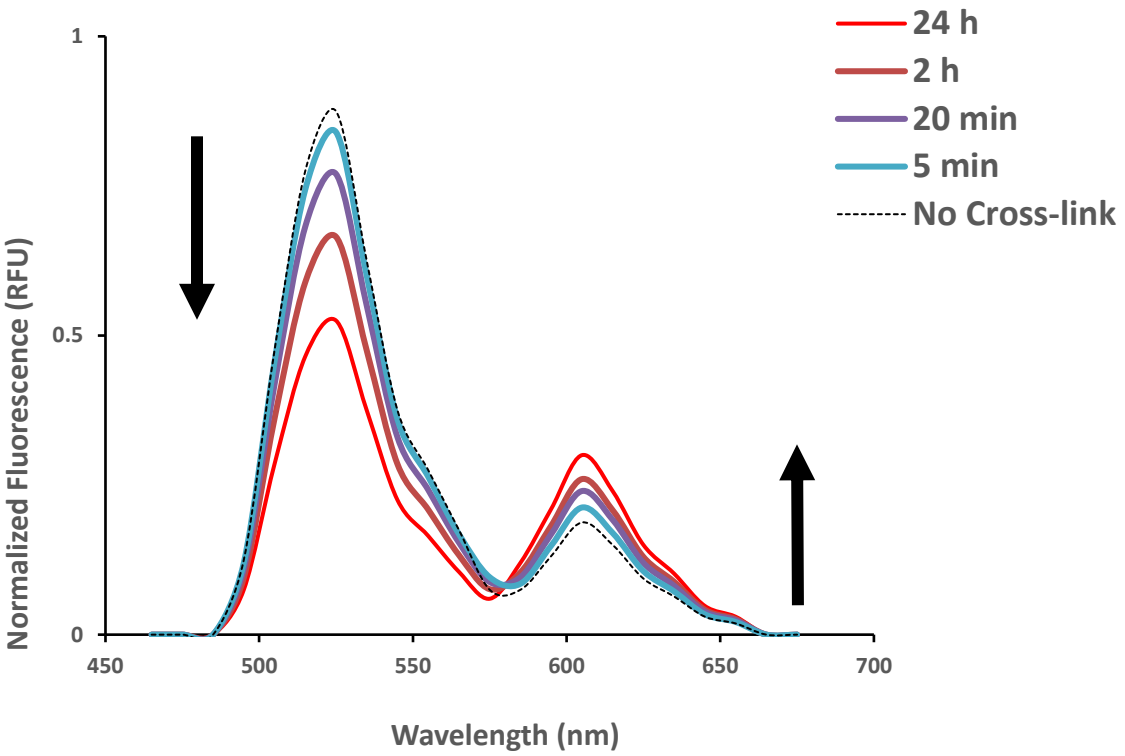


Figure 4.8 *in vitro* optimization of FRET assay using mTG. Emission spectrum scan of 0.8 µM mRuby2-7M48 coupled with 0.8 µM Clover-6K by 0.2 U of mTG. Experiment was performed at pH 7.2 and 37°C. Samples were excited at 440 nm and scanned over a range of 450-700 nm at multiple time points (5 min to 24 h).

Once the assay was established with mTG and proven successful, repeating the experiment with bTG was the next step. Under the same reaction conditions, the experiment was repeated, replacing mTG with bTG. **Figure 4.9** shows a similar decrease in green fluorescence and increase in red fluorescence over time. This is indicative of FRET-signalled transamidation and is the first assay designed to monitor transamidation by bTG. It should be noted that the rate of change of FRET emission was slower with bTG than mTG. This can be explained by pH and tag reactivity. As discussed previously, bTG's optimal pH of 8.2 is 1 unit above the pH range of the assay[139]. bTG has been shown to still have activity at physiological pH; however, a diminished

activity was predictable. Secondly, the tag used for assay optimization, 7M48, does not confer high reactivity for bTG. **Figure 4.10** shows the Michaelis-Menten plot of 7M48 reacted with mTG and bTG. mTG clearly has greater reactivity for 7M48 than bTG, according to the hyperbolic fitting of the initial rate data. Since the substrate used in the FRET assay has rather low reactivity for bTG, it is understandable that the reaction would not necessarily result in a strong FRET signal. When the bTG FRET signal is quantitatively compared to the fusion protein or the mTG transamidation product, it is evident that the tagged fluorescent proteins were not as reactive with bTG. The bTG FRET signal, after 24 hours, was 18 times less intense than the fusion product and 4 times less intense than the mTG FRET signal. Regardless of the relatively poorer results obtained with bTG, it is noteworthy that we were successful at monitoring transamidation through FRET. This convinced us the assay could be applied to screen a library of peptides to identify high reactivity tags. Ideally, the results obtained during the screen would yield a peptide tag that has greater reactivity than 7M48.

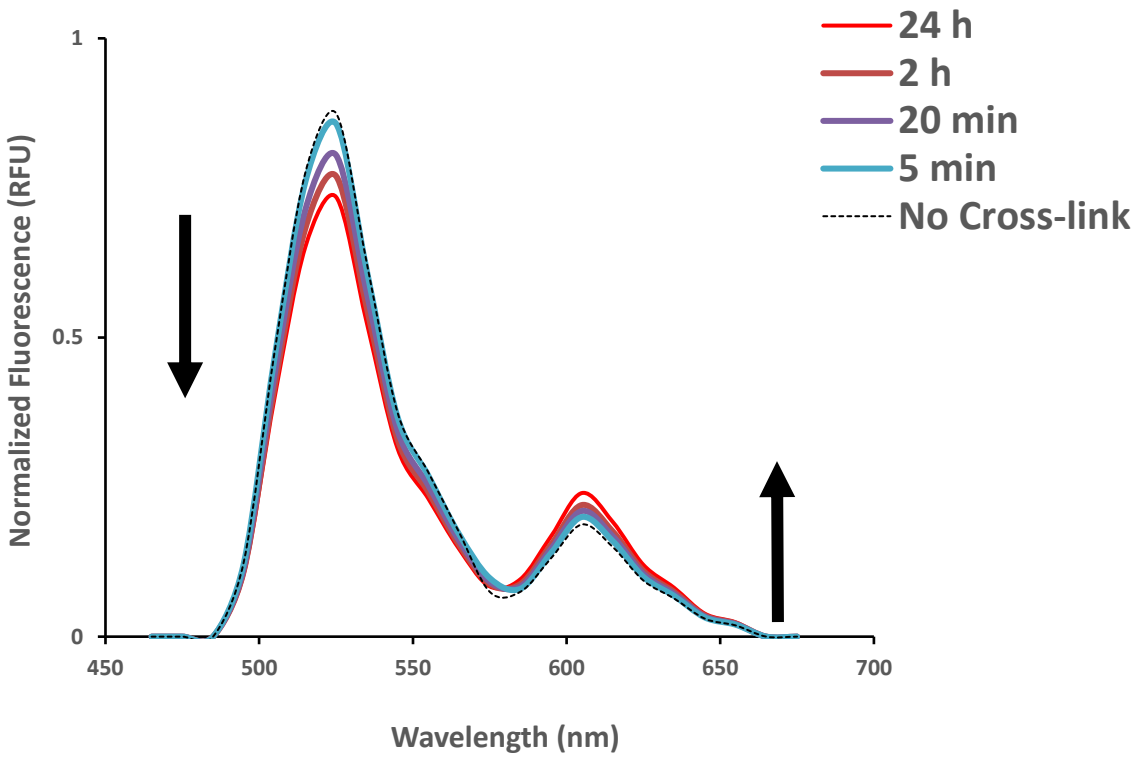


Figure 4.9 *in vitro* optimization of FRET assay using bTG. Emission spectrum scan of bTG-mediated conjugation 0.8 µM of mRuby2-7M48 with 0.8 µM of Clover-6K and 0.2 U of bTG. Experiment was performed at pH 7.2 and 37°C. Samples were excited at 440 nm and scanned over a range of 450-700 nm at multiple time points (5 min to 24 h).

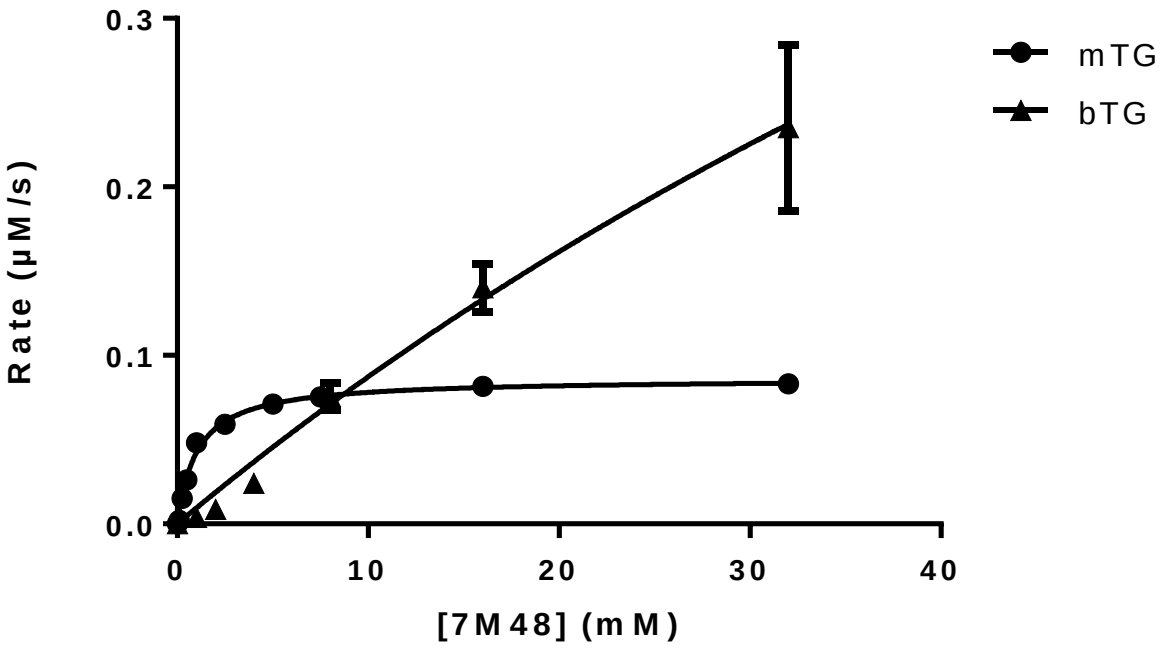


Figure 4.10 Michaelis-Menten plot of mTG and bTG activity measured by the coupled-GDH assay. using varied concentrations of 7M48 as acyl-donor substrate (1-32 mM) in the presence of 10 mM Gly-OMe as acyl-acceptor substrate, 0.2 U of enzyme at pH 7.2 and 37ºC carried out in triplicate using a 96-well microplate reader. Error bars represent standard error of the mean values. The solid line through the data points was fit to the hyperbolic Michaelis-Menten equation.

Table 4.2 Quantitative analysis of FRET signal. Fluorescent intensity of FRET-based assay after 24 h of tagged fluorescent proteins incubation with mTG or bTG. FRET signal of transamidation product is compared to designed mRuby2-Clover fusion.

FRET Pair	Fluorescence Intensity at 600 nm after 24 h
mRuby2-7M48/ Clover-6K	***mTG***
	14381
	bTG
	3377
mRuby2-Clover Fusion	63219

4.3.4 Designing peptide library for FRET-assay screen

After establishing the parameters of the screen, designing the peptide library was the final step before running the assay. Although comprehensive studies on peptide length have not been established for bTG, it is known that for mTG, heptapeptides maintain the desired minimal tag size without sacrificing reactivity[150]. Furthermore, 7M48 has greater reactivity than the pentamer TQ1 and TQ2 peptides suggested in the literature. Given the limited information available, designing a heptamer library appeared to be a logical starting point for the assay. The size of the peptide library that can be screened is dependent on how exhaustive the screening tools available are. In the design of a heptamer tag, varying the 6 positions flanking the reactive glutamine with all 64 codons, covering all 20 amino acids, would result in library with at least 68 x 10^9 codons. To ensure 90% coverage of all possible sequence variants, 1 x 10^{11} colonies would need to be

screened[151,152]. This is within the capacity of flow cytometry used in this experiment. However, it is not necessary to screen every possible codon when a representative library will provide valuable information about substrate affinity. Furthermore, a smaller library is amenable to other diagnostic methods. As shown previously, the *in vitro* assay is functional and could be modified to screen a representative library.

Alternatively, the 'NMT' codon is a degenerate codon that represents a variety of base pairs, coding for a range of amino acid types. N represents all four base pairs, M represents Thymine/Guanine and T signifies Thymine. This combination of base pairs results in 8 codons that code for 8 amino acids: alanine (hydrophobic), asparagine (polar/uncharged), aspartate (negative/charged), histidine (positive/charged), proline (hydrophobic), serine (polar uncharged), threonine (polar/uncharged) and tyrosine (aromatic). This restricted library represents all major amino acid types while keeping the library at a manageable size. Varying 6 positions with the NMT codon would result in a 262,144-codon library. To ensure 90% coverage of all possible peptide variants, 603,607 colonies would require screening. For these reasons, a representative donor substrate library was designed using NMT degenerate codons at the 6 positions flanking a glutamine residue. Given the demonstrated reactivity of the hexalysine tag with bTG, the '6K' tag was chosen as the acceptor substrate in the screening experiment to identify a high-reactivity Q-tag substrate.

4.4 *In cellulo* FRET-based assay screen

4.4.1 Expected FACS results

Once conditions for performing and monitoring *in vitro* FRET-based transamidation were established, *in cellulo* monitoring of transamidation was the next objective. This presented a different set of challenges to overcome. Specifically, the expression and visualization of three separate proteins within a cell is a daunting task. First, it is necessary to predict what the potential positive and negative results would be and how to interpret this information. As **figure 4.11** indicates, there are three results possible, two of which are the same phenotypically. If successful co-expression occurs, and the reaction commences as designed, some cells may undergo a transamidation reaction, resulting in a cross-linked product and a FRET signal under the appropriate excitation. If transamidation does not occur, there will not be a FRET signal, and the blue excitation will yield green emission. Green emission will also occur if bTG is not expressed. As a control, the fusion protein will also be expressed to calibrate a positive FRET signal for the screening platform.

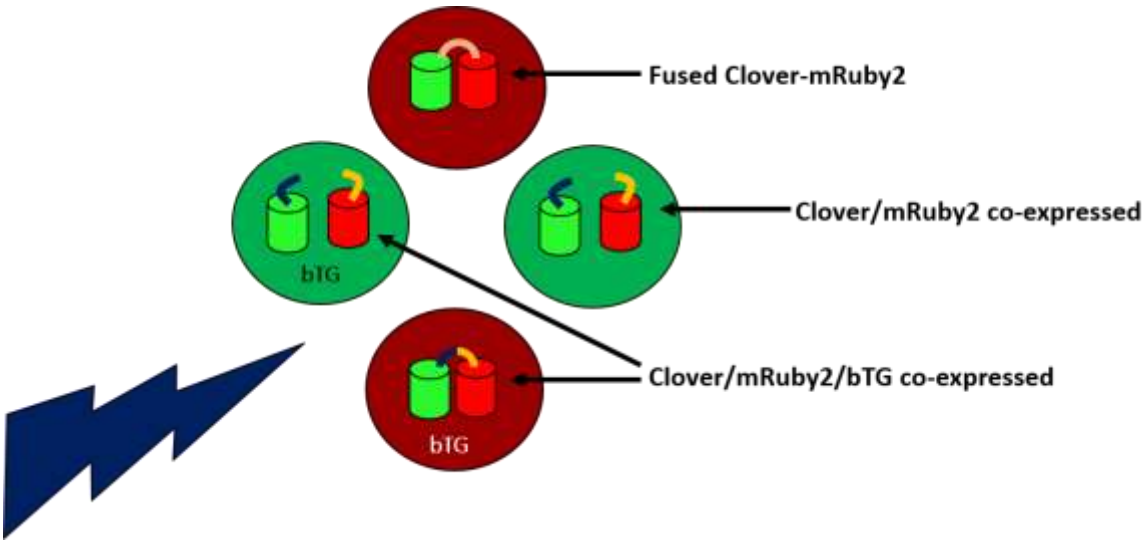


Figure 4.11 Possible outcomes of the *in cellulo* FRET-based assay. Three possible results when cells are induced to co-express bTG, mRuby2-Q and Clover-K: 1) Successful co-expression of three proteins, resulting in positive FRET signal. 2) Successful co-expression; however, bTG lacks reactivity for tags, resulting in no FRET signal. 3) Unsuccessful co-expression of any one of the three proteins, resulting in no FRET signal. As a control, a fused mRuby2-Clover was also designed and expressed, to calibrate the expected positive FRET signal.

4.4.2 FRET-based selection strategy

Figure 4.12 summarizes the expression and screening strategy used to identify potential-high reactivity tags. To ensure consistency in fluorescent protein expression levels, a duet vector was used for the expression of both mRuby2-Q tag and Clover-6K. Plasmids for fluorescent proteins and bTG were expressed sequentially, establishing a stable cell line between the bTG and fluorescent protein transformation. Details on the expression can be found in the Experimental Section.

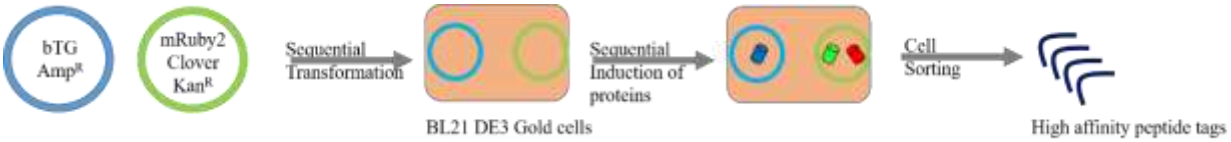


Figure 4.12 Strategy for *in cellulo* expression and FRET-based selection of mRuby2-tagged proteins. Pictorial representation of FRET-assay selection of high reactivity peptide candidates. Plasmids coding for bTG and tagged fluorescent protein substrates are sequentially transformed into BL21 (DE3) Gold cells. After successful transformation, expression of fluorescent proteins is induced and the chromophore is given time to mature; this is followed by the induction of bTG. Once bTG is expressed, cells are analyzed and sorted based on their FRET signal. Sorted cells are collected, plated and sequenced to identify successful peptide tags.

4.4.3 In cellulo FRET assay excitation channels and controls

Unfortunately, the FACS (fluorescence activated cell sorting) instrument does not have the same range or flexibility of excitation wavelengths as the plate reader that was used for assay optimization. Due to this limitation, excitation at 440 nm was not possible. In the sorting experiment, cells experiment were excited using the 488 nm channel, the highest energy channel that allows monitoring of emission in the green and red channels. As **Figure 4.7** shows, excitation at 488 nm unfortunately leads to direct excitation of mRuby2, and increases the opportunity for false positive readings. Therefore, controls that allow the separation of direct excitation and FRET of mRuby2 must be established, in order to properly identify a FRET signal due to transamidation. Before sorting the peptide library, controls were therefore performed to identify positive and negative readings. In total, five different controls were analyzed: Clover alone, mRuby2 alone, the mRuby2-Clover fusion protein, mRuby2-7M48 + Clover-6K and mRuby2-7M48 + Clover-6K + bTG.

4.4.4 FRET-based assay screen control: Clover

Sorting cells expressing only Clover helps identify ineffective FRET. Since this protein alone is unable to undergo transamidation, the cells expressing it should maintain green emission when excited at 488 nm. The representative FACS plots in **Figure 4.13** show Clover FP, behaving as predicted. The first plot (A) determines what events are actual bacteria vs other sub-micron particulates, based on the size and granularity of the event. Although the samples were filtered before cell-sorting, sub-micron particulates can still make their way into the FACS, increasing background and noise during analysis. The second plot (B) compares excitation/emission channels of 488 nm/620 nm to 488 nm/513 nm. 488 nm/620 nm represents the "FRET" channel, since a Stokes shift of that magnitude would require an energy transfer. The 488 nm/513 nm combination represents green emission, indicative of a green fluorescent protein. The third plot (C) compares the combinations of 561 nm/ 614 nm with 488 nm/ 513 nm, or red excitation and emission vs. green excitation and emission. In plots (B) and (C), cells are only found along the y-axis, where green excitation and emission is expected. This control assures us of the lack of red emission due to FRET when Clover is the only FP expressed.

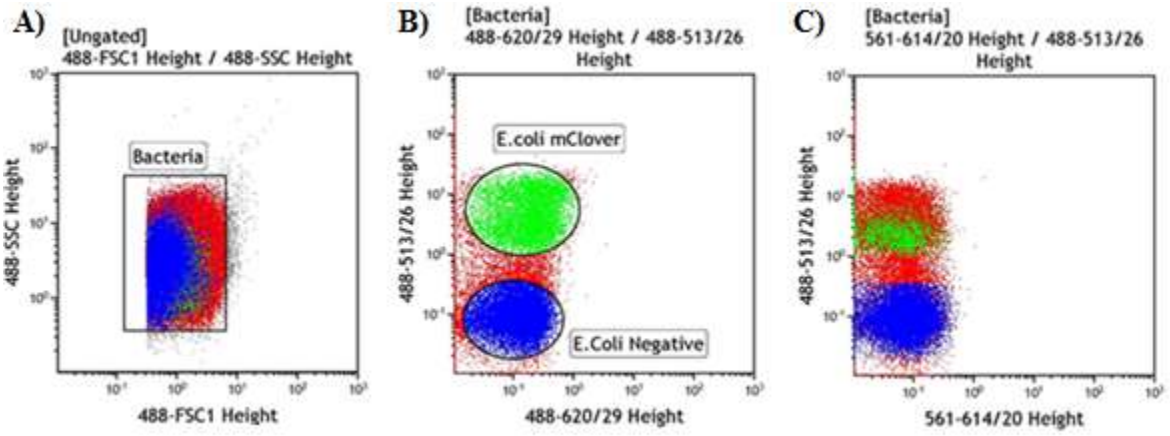


Figure 4.13 Flow cytometry analysis of cells expressing Clover fluorescent protein. FACS plots of BL21(DE3) Gold cells expressing the Clover fluorescent protein. A) Ungated forward and side scatter, to establish bacterial populations vs. other submicron particulate. B) Gated for bacteria, excitation at 488 nm, emission at 620 nm vs excitation at 488, emission at 513 nm. C) Gated for bacteria, excitation at 561 nm, emission at 614 nm vs excitation at 488 nm, emission at 513 nm. Gated cells are colour-coded to identify distributions of populations from plot to plot (Blue= total bacteria population, green= Clover+ cells, blue= clover- cells)

4.4.5 FRET-based assay screen control: mRuby2

Similarly to the Clover FP cells, mRuby2 was also expressed individually and gated to identify what to expect during the sorting of the library. The same channels used to gate and sort Clover FP were used for mRuby2 FP and the results were noticeably different. **Figure 4.14** plot B) shows a population of cells along the x-axis, indicating what may be assumed to be FRET. However, the excitation band of mRuby2 reaches 488 nm, resulting in direct excitation of mRuby2 and not necessarily FRET. This false positive must be monitored throughout the experiment, and a method to distinguish between FRET and direct red excitation must be developed. Plot C) gives the expected result of red excitation and emission along the x-axis for mRuby2 cells.

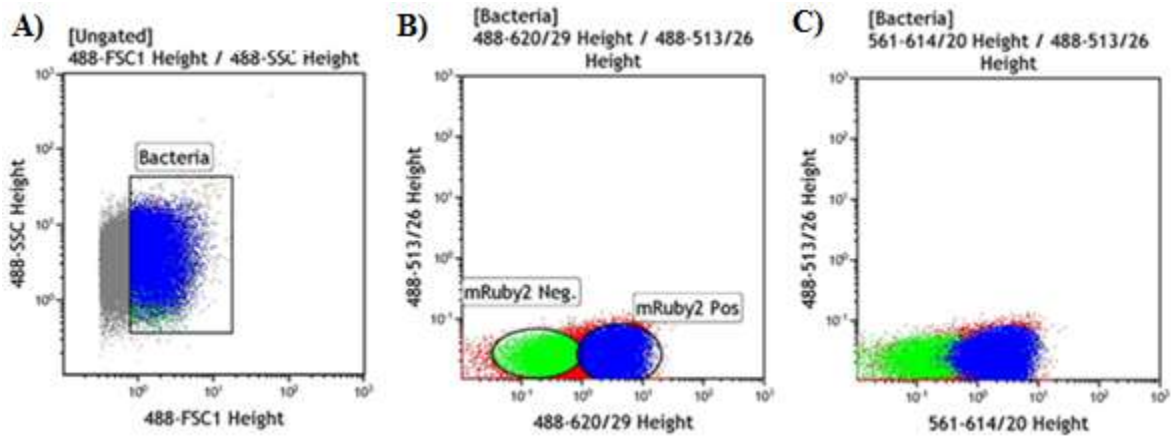


Figure 4.14 Flow cytometry analysis of cells expressing mRuby2 fluorescent protein. FACS plots of BL21(DE3) Gold cells expressing the mRuby2 fluorescent protein. A) Ungated forward and side scatter, to establish bacterial populations vs. other submicron particulate. B) Gated for bacteria, excitation at 488 nm, emission at 620 nm vs. excitation at 488, emission at 513 nm. C) Gated for bacteria, excitation at 561 nm, emission at 614 nm vs. excitation at 488 nm, emission at 513 nm. Gated cells are colour-coded to identify distributions of populations from plot to plot.

4.4.6 FRET-based assay screen control: mRuby2-Clover fusion

The fusion protein is expected to have very different results than the individually expressed fluorescent proteins. In our previous analysis of the fusion protein using the plate reader, it was clear that 100% FRET efficiency was not possible. Looking back at **Figure 4.7**, FRET occurs upon excitation at 440 nm; however, significant green emission is still observed. During optimization, a wavelength of 440 nm was used to excite the fusion protein, meaning that red fluorescence was based solely on FRET and not direct excitation. This assumption cannot be made for excitation at 488 nm when using FACS, due to excitation limitations. Therefore, additional analysis was needed to be confident in the identification of FRET-positive cells. **Figure 4.4.6.** plot B) highlights the population scattering of a fusion protein where green emission is compared to red emission possibly due to FRET. Populations can be found along the

x- and y-axes, indicating cells emitting red fluorescence or green fluorescence, respectively. More importantly, a new population in the top right quadrant of the plot is observed; this area represents cells that are positive for green and red fluorescence due to FRET. As was the case when the emission spectrum was analyzed, the fusion protein has two maxima, in the green and red. These cells are believed to undergo FRET when excited at 488 nm. However, a strategy is still required to differentiate cells demonstrating FRET from cells that simply demonstrate red emission due to direct excitation of mRuby2. We can assume the population along the x-axis represents cells that are undergoing direct excitation. This population is identical to the mRuby2 control of **Figure 4.14** plot B). To verify that the FRET-positive cells are distinct from the mRuby2 cells, a plot comparing 561 nm/614 nm vs. 488 nm/620 nm (direct red excitation vs. "FRET" red excitation) was prepared. FRET-positive cells should represent a distinct population shifted towards a higher value down the 488 nm/620 nm axis. On the other hand, FRET-negative cells should be slightly shifted to a higher value down the 561 nm/614 nm axis, since the 561-nm excitation channel is closer to the excitation maximum of mRuby2 than the 488-nm channel. Plot C) clearly shows two distinct populations, suggesting the cells can be sorted into FRET-positive (FRET+) and FRET-negative (FRET-) groups. It should be noted that the observation of cells that apparently express only mRuby2 could be due to a translation issue as the expression of the large fusion protein may result in a Clover folding/maturation failure (mRuby2 is N-terminal protein in fusion).

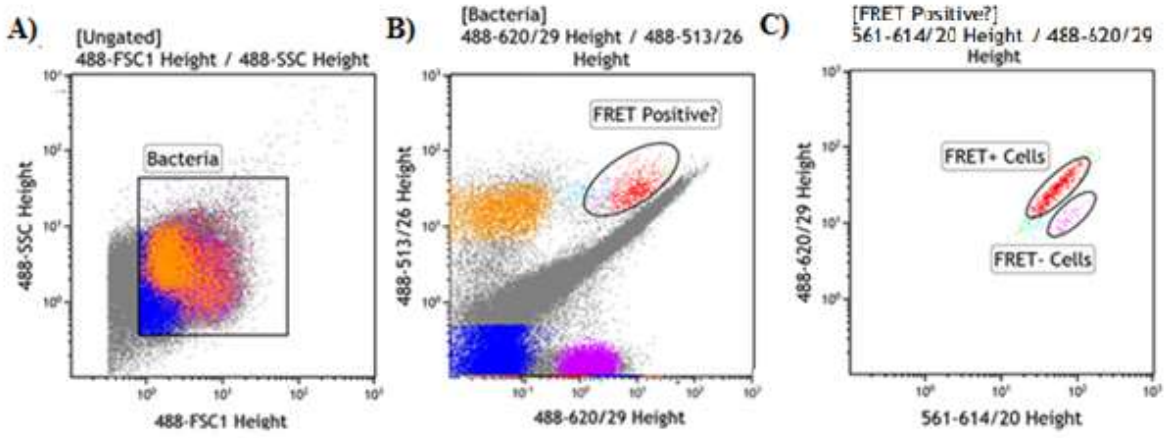


Figure 4.15 Flow cytometry analysis of cells expressing mRuby2-Clover fusion fluorescent protein. FACS plots of BL21(DE3) Gold cells expressing the mRuby2-Clover fusion fluorescent protein. A) Ungated forward and side scatter, to establish bacterial populations vs. other submicron particulate. B) Gated for bacteria, excitation at 488 nm, emission at 620 nm vs. excitation at 488, emission at 513 nm. C) Gated for FRET+ cells, excitation at 561 nm, emission at 614 nm vs. excitation at 488 nm, emission at 620 nm. Gated cells are colour-coded to identify distributions of populations from plot to plot (red= FRET+ cells).

4.4.7 FRET-based assay screen control: mRuby2-7M48 + Clover-6K

Co-expression of mRuby2 and Clover in the absence of bTG represents a negative control; without bTG, FRET should not occur. Based on the analysis of the fusion protein, "FRET" red emission vs. green emission should show a population of cells in the top right quadrant when FRET has occurred. **Figure 4.16** plot B) shows minimal events in this region, suggesting that the FRET has not taken place. Without bTG to cross-link mRuby2 and Clover, the proximity necessary for FRET cannot be achieved. There does appear to be a small population in this quadrant; however, this is most likely random collisions of FPs that leads to a background FRET signal.

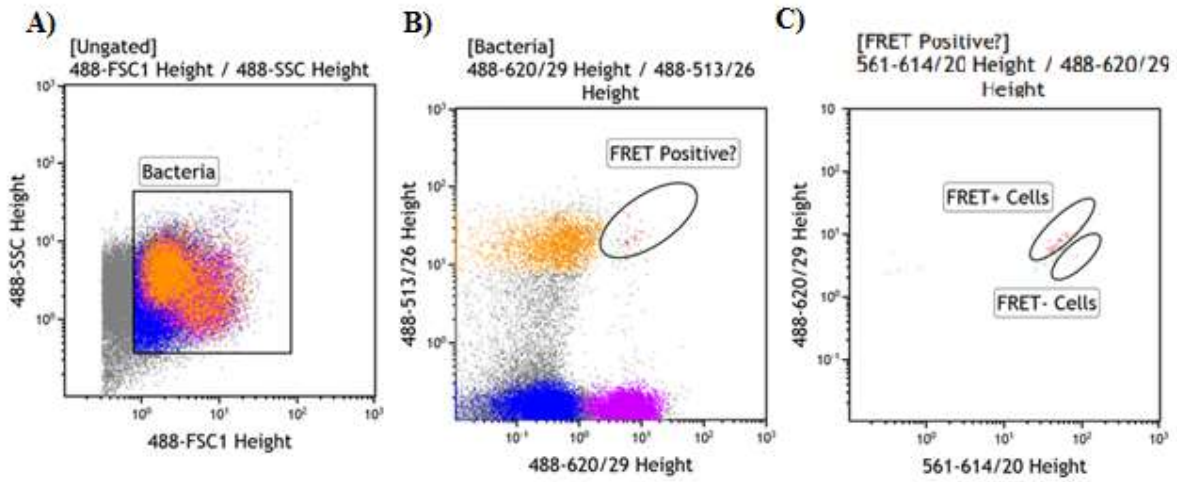


Figure 4.16 Flow cytometry analysis of cells expressing mRuby2 + Clover fluorescent proteins. FACS plots of BL21(DE3) Gold cells expressing mRuby2 + Clover fluorescent proteins, cultured for three hours after maturation of fluorescent proteins. A) Ungated forward and side scatter, to establish bacterial populations vs. other submicron particulate. B) Gated for bacteria, excitation at 488 nm, emission at 620 nm vs. excitation at 488, emission at 513 nm. C) Gated for FRET+ cells, excitation at 561 nm, emission at 614 nm vs. excitation at 488 nm, emission at 620 nm. Gated cells are colour-coded to identify distributions of populations from plot to plot.

4.4.8 FRET-based assay screen control: mRuby2-7M48 + Clover-6K + bTG

Given the demonstrated reactivity of 7M48 and 6K tags with bTG, mRuby2 and Clover were tagged with each, respectively, to serve as a positive control. A result similar to the fusion protein was expected: a population in the upper right quadrant when observing green emission vs. “FRET” red emission. **Figure 4.17** verifies the successful, in cell, transamidation of mRuby2-7M48 and Clover-6K. In the presence of bTG, transamidation of proteins led to a population of FRET+ cells, distinct from cells only demonstrating red fluorescence directly from excitation of mRuby2.

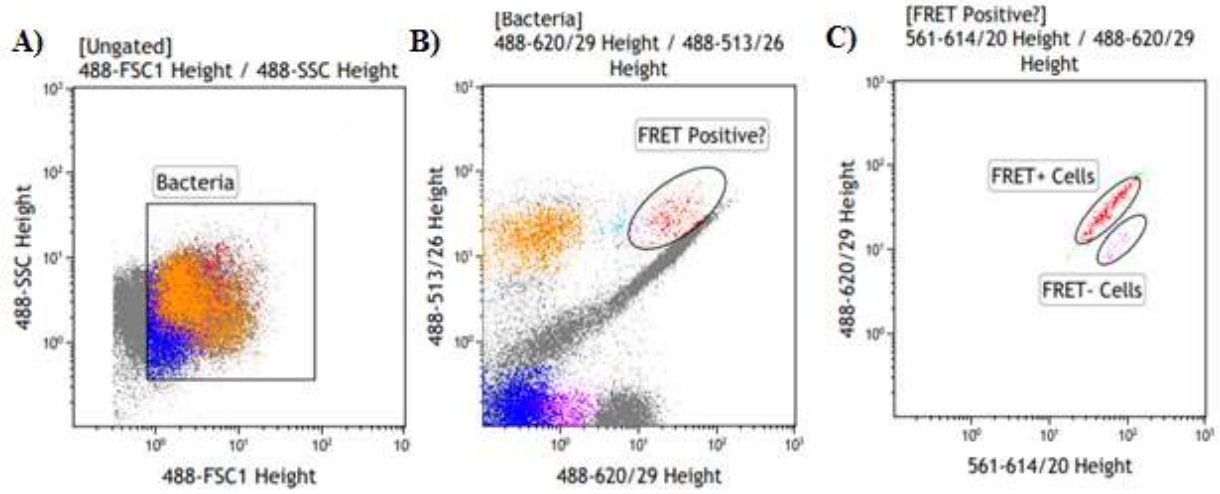


Figure 4.17 Flow cytometry analysis of cells expressing mRuby2 + Clover + bTG. FACS plots of BL21(DE3) Gold cells expressing mRuby2 + Clover + bTG. A) Ungated forward and side scatter, to establish bacterial populations vs. other submicron particulate. B) Gated for bacteria, excitation at 488 nm, emission at 620 nm vs. excitation at 488, emission at 513 nm. C) Gated for FRET+ cells, excitation at 561 nm, emission at 614 nm vs. excitation at 488 nm, emission at 620 nm. Gated cells are colour-coded to identify distributions of populations from plot to plot. bTG expression was induced for 3 h prior to analysis.

These controls allowed us to understand what to expect during the subsequent library screening experiments. Namely, using the gating parameters identified in the control experiments, cells in which transamidation occurs, demonstrating a genuine FRET signal, will be identified, collected and analyzed further.

For the peptide library screen, mRuby2-7M48 was replaced with mRuby2-6NMT-Q, representing the XXX-Q-XXX heptamers library that was cloned onto the C-terminus of mRuby2. After induction of bTG, sorting analysis was performed periodically, over a 3-hour period, at 15 min, 30 min, 60 min, 120 min and 180 min. It was expected that there would be an increase in FRET+ cells as time progressed.

4.4.9 FRET-based assay screen results: 15 min

After 15 min, no events were observed in the designated FRET+ area, as shown in **Figure 4.15** plot E). As a control, cells not expressing bTG were analyzed to ensure that any background FRET signal would not give rise to a false positive. As plot B) shows, no cells were found in the FRET+ region for cells not expressing bTG, either. The most reactive peptides should undergo transamidation the fastest. Ideally, the screen would result in FRET+ cells at the earliest point of analysis; however, more time was needed to observe a FRET+ cells.

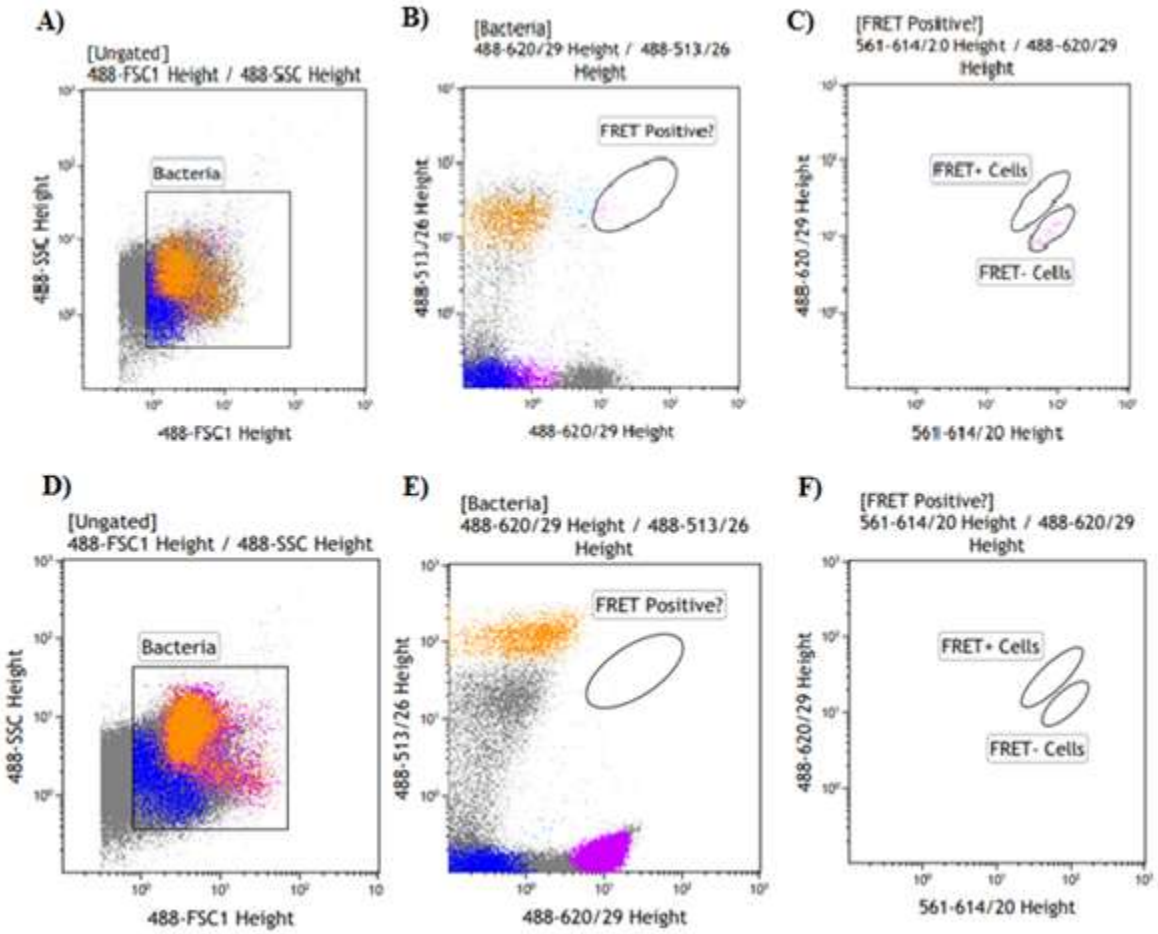


Figure 4.18 Flow cytometry analysis of cells expressing peptide library. FACS plots of BL21(DE3) Gold cells expressing mRuby2-NMT + Clover-6K , with and without bTG, 15 min after 3h bTG induction. A/D) Ungated forward and side scatter, to establish bacterial populations vs. other submicron particulate. B/E) Gated for bacteria, excitation at 488 nm, emission at 620 nm vs. excitation at 488, emission at 513 nm. C/F) Gated for FRET+ cells, excitation at 561 nm, emission at 614 nm vs. excitation at 488 nm, emission at 620 nm. Plots A-C are of cells that do not contain bTG, plots D-F are of cells in which bTG has been induced. Gated cells are colour-coded to identify distributions of populations from plot to plot.

4.4.10 FRET-based assay screen results: 30 min

Similar results were observed when the cells were analyzed 30 min after induction period, with no FRET+ cells identified at this point in the experiment.

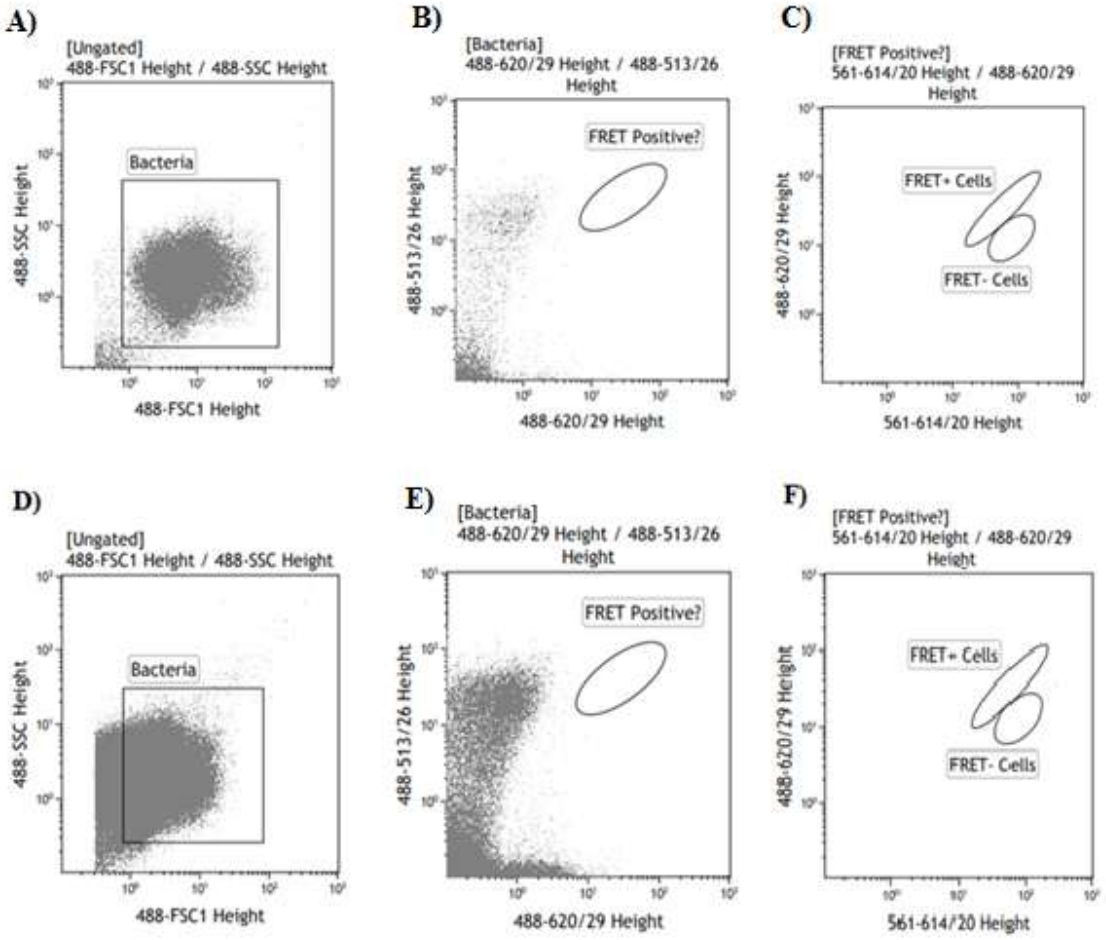


Figure 4.4.10 Flow cytometry analysis of cells expressing peptide library. FACS plots of BL21(DE3) Gold cells expressing mRuby2-NMT + Clover-6K with/without bTG, 30 min after bTG induction. A/D) Ungated forward and side scatter, to establish bacterial populations vs. other submicron particulate. B/E) Gated for bacteria, excitation at 488 nm, emission at 620 nm vs. excitation at 488, emission at 513 nm. C/F) Gated for FRET+ cells, excitation at 561 nm, emission at 614 nm vs. excitation at 488 nm, emission at 620 nm. Plots A-C are of cells that do not contain bTG, whereas plots D-F are of cells in which bTG expression has been induced.

4.4.11 FRET-based assay screen results: 60 min

At 60 min, a shift in the bacterial population can clearly be observed. **Figure 4.19** plot E) shows a small population of events in the top right quadrant, the area defined to indicate FRET+ cells. To validate the increase in FRET signal was due to transamidation, plot B) shows no events in the same region, meaning that the signal was not due to background FRET. This population of cells was sorted, collected and cultured for further analysis; however, the reduced size of the population led to difficulty in harvesting these cells.

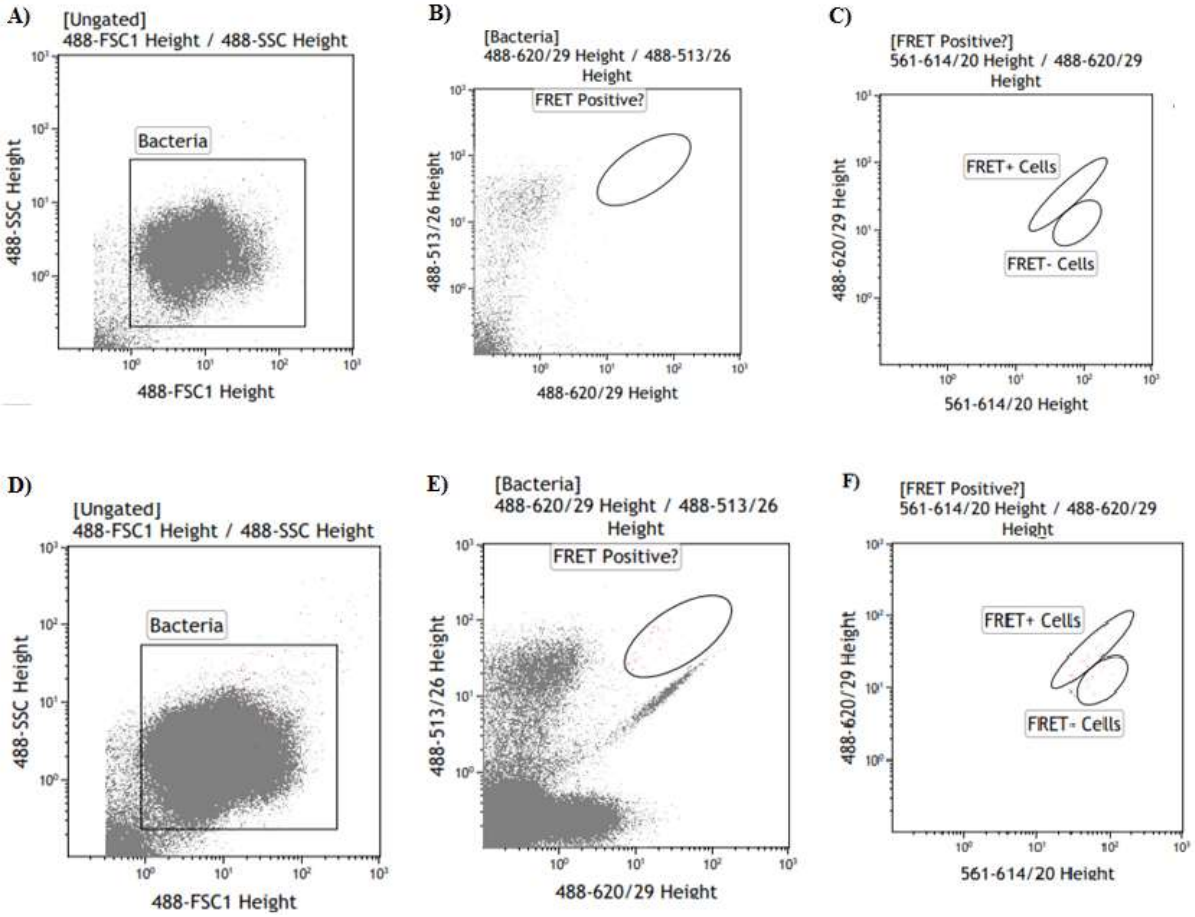


Figure 4.19 Flow cytometry analysis of cells expressing peptide library. FACS plots of BL21(DE3) Gold cells expressing mRuby2-NMT + Clover-6K with/without bTG 60 min after induction period. A/D) Ungated forward and side scatter, to establish bacterial populations vs. other submicron particulate. B/E) Gated for bacteria, excitation at 488 nm, emission at 620 nm vs. excitation at 488, emission at 513 nm. C/F) Gated for FRET+ cells, excitation at 561 nm, emission at 614 nm vs. excitation at 488 nm, emission at 620 nm. Plots A-C are of cells that do not contain bTG, whereas plots D-F show cells in which bTG expression has been induced for 3 h prior to analysis. Gated cells are colour-coded to identify the distribution of populations from plot to plot.

4.4.12 FRET-based assay screen results: 120 min

The results observed at 60 min after induction period suggested that as time elapsed, there was an increase in the number of FRET+ cells. This would suggest that the assay was successfully observing the increase in the concentration of transamidation product within the cell. The increase in the FRET+ population can be visualized by comparing **Figure 4.20** to **Figure 4.19** Although the increase is small and not immediately clear in plot B), in plot F), which focuses on FRET+ cells, a greater number of cells in the desired population can clearly be detected. These cells were also collected for further analysis.

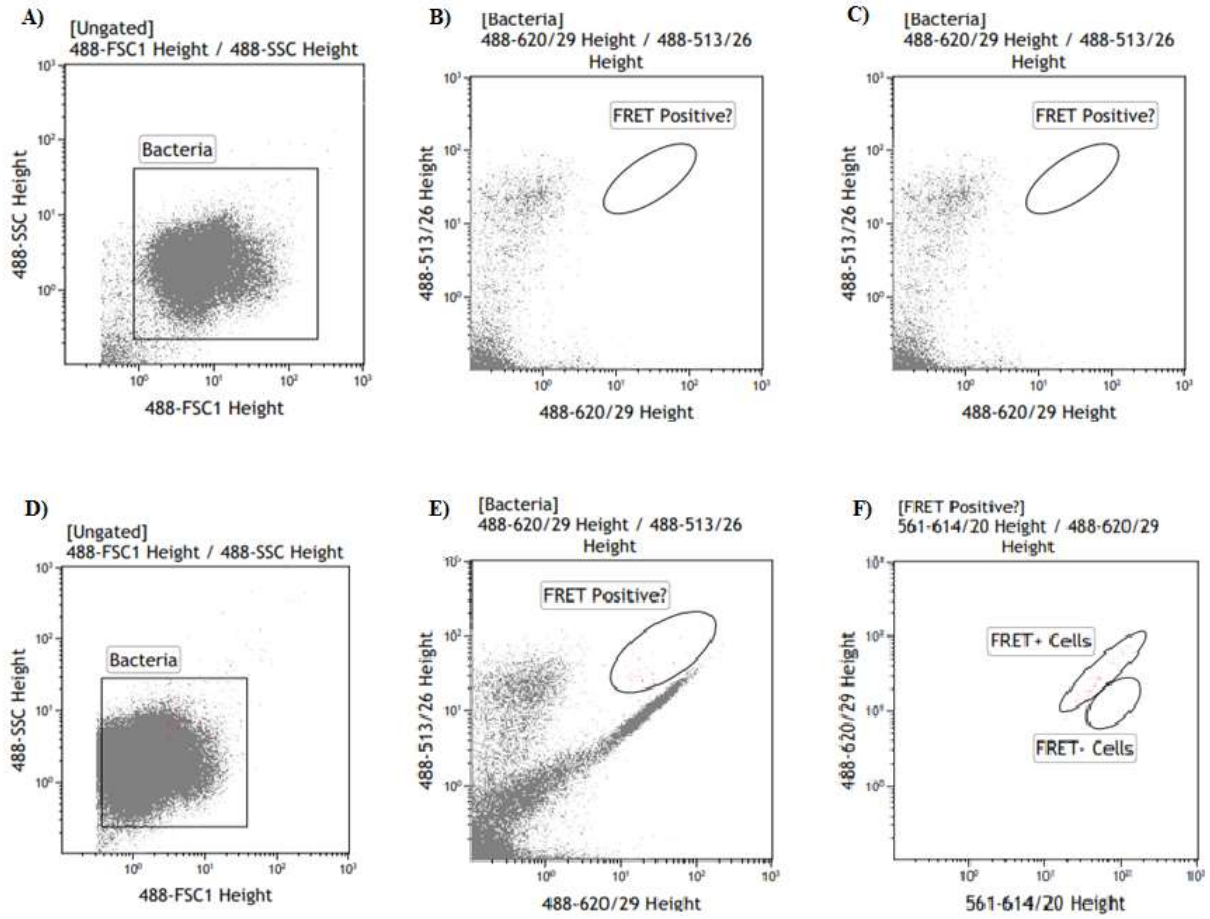


Figure 4.20 Flow cytometry analysis of cells expressing peptide library. FACS plots of BL21(DE3) Gold cells expressing mRuby2-NMT + Clover-6K with/without bTG 120 min after induction period. A/D) Ungated forward and side scatter, to establish bacterial populations vs. other submicron particulate. B/E) Gated for bacteria, excitation at 488 nm, emission at 620 nm vs. excitation at 488, emission at 513 nm. C/F) Gated for FRET+ cells, excitation at 561 nm, emission at 614 nm vs. excitation at 488 nm, emission at 620 nm. Plots A-C are of cells that do not contain bTG, whereas plots D-F show cells in which bTG expression has been induced for 3 h prior to analysis. Gated cells are colour-coded to identify the distribution of populations from plot to plot.

4.4.13 FRET-based assay screen results: 180 min

The final analysis was performed after 180 min incubation. **Figure 4.21** clearly shows further increase in cells demonstrating FRET. The population of FRET+ cells at 3 h is the largest observed in this experiment. As the controls and optimization experiments demonstrated,

transamidation of the tagged fluorescent proteins will lead to an elevation in FRET signal, over time. However, with the goal being identification of peptides that confer high reactivity for bTG, the tags identified at the later time points in the experiment are more likely to include lower reactivity sequences than the population of peptides that were reactive after only 60 min. For this reason, the cells of the 3-h time point were collected, but not analyzed; in the case that inconclusive results were found with earlier cells, the 3 h cells could be used to provide more information on reactivity.

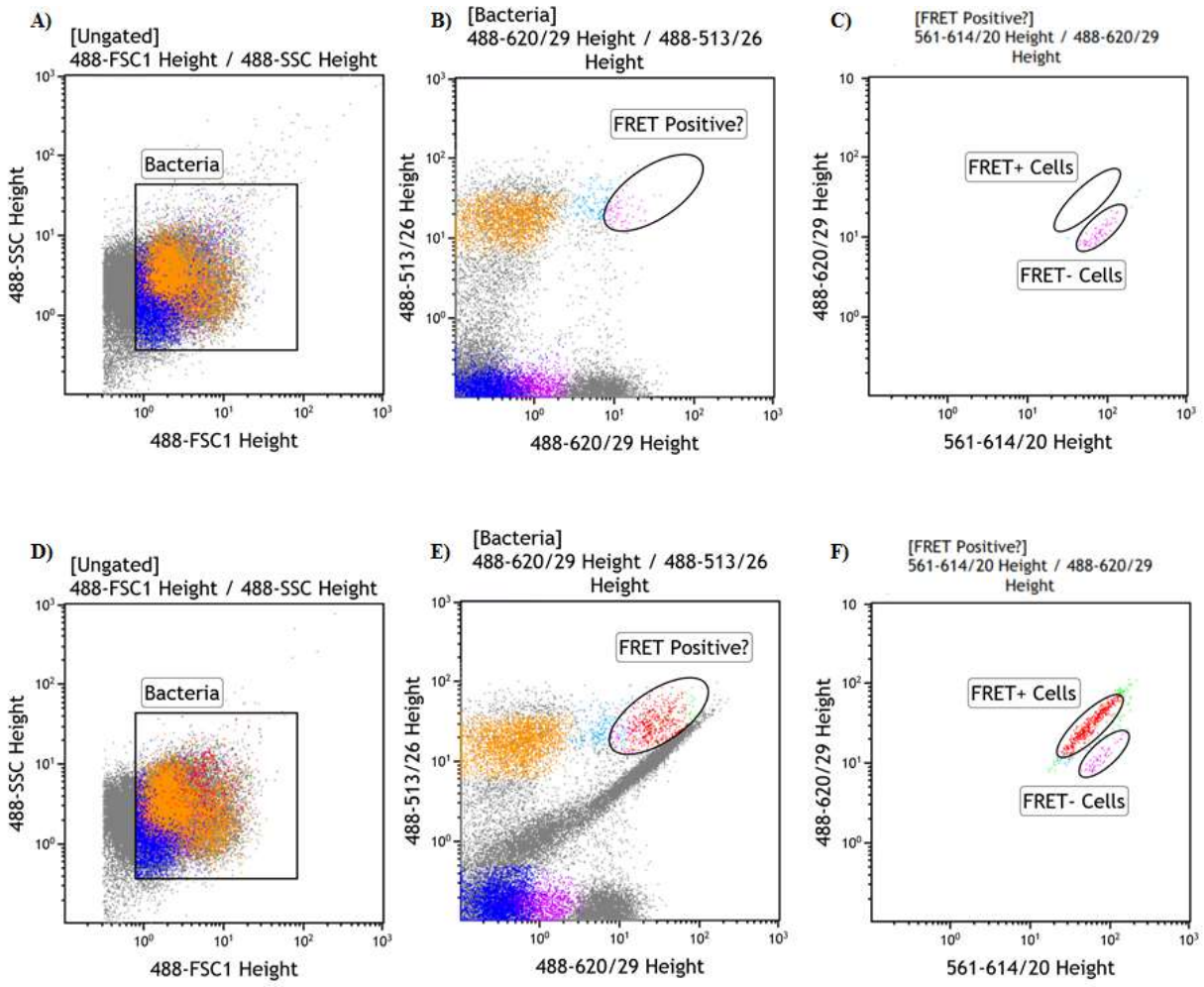


Figure 4.21 Flow cytometry analysis of cells expressing peptide library. FACS plots of BL21(DE3) Gold cells expressing mRuby2-NMT + Clover-6K with/without bTG 180 min after induction period. A/D) Ungated forward and side scatter, to establish bacterial populations vs. other submicron particulate. B/E) Gated for bacteria, excitation at 488 nm, emission at 620 nm vs. excitation at 488, emission at 513 nm. C/F) Gated for FRET+ cells, excitation at 561 nm, emission at 614 nm vs. excitation at 488 nm, emission at 620 nm. Plots A-C are of cells that do not contain bTG, whereas plots D-F show cells in which bTG expression has been induced for 3 h prior to analysis. Gated cells are colour-coded to identify the distribution of populations from plot to plot.

4.5 Analysis of NMT-peptide library

As described above, FRET+ cells were observed at 60 min, 120 min and 180 min; these cells were sorted and collected in approximately 1 mL of PBS each. Cells were then transferred to 3 mL of LB containing 100 µg/mL of ampicillin, of which 200 µL was plated immediately on LB agar, and the remaining culture was incubated overnight at 37 ^{o}C and plated the following day. Incubation with growth medium was important in order to help the cells recover after stress, specifically cells from the 60- and 120-min analyses, which did not grow as well otherwise. This strategy led to successful growth of colonies for 120 min and 180 min FRET+ cells; however, the 60-min cells did not grow and were not pursued any further. When working with flow cytometry, dense populations are more likely to result in successful sorting. Sorting is based on a set of criteria established based on desired results, if the results are too stringent, no cells may qualify. If the parameters are too loose, all cells (and other particulates) will count as a successful sort. The plot at 60 min suggested tat there were FRET+ cells; however, there were so few sorting events that it is possible that no cells were included in the sorted fractions, and that if they were, there were not enough to collect and to culture.

4.5.1 Identification of potential high reactivity bTG peptides

As previously stated, the goal of this screen was to identify the most reactive, most reactive peptides; therefore, we were initially interested in characterizing the positive hits from 120 min rather than from 180 min. Of the colonies grown from the 120-min FRET+ population, 94 were picked and grown up in a 96-well plate (200 µL LB +100 µg/mL ampicillin, 10% glycerol

solution overnight at 37°C). DNA extraction and sequencing was performed by *Eurofins Genomics.*

Of the 94 samples sent for sequencing, 78 were successfully sequenced. By way of reminder, the NMT degenerate codon codes for 8 possible amino acids: alanine, asparagine, aspartate, histidine, proline, serine, threonine and tyrosine. The glutamine residue is fixed at the centre of every heptamer tag sequence, flanked by three amino acid residues on either side. We did not characterize the transformant library, so we are unable to confirm the diversity of our peptide pool. However, if we assume the pool was diverse, of the 78 sequenced peptides, over 50% had a tyrosine at position 1. This suggests that bTG may have an affinity for aromatic amino acids within the tunnel where its active site cysteine can be found. Furthermore, 19% of samples contained the sequence H-Q-A at positions 3-4-5. Slightly less commonly, 13% of samples presented H-Q-P instead. Although there is a structural commonality between the two trends (proline and alanine both being hydrophobic), the conformational flexibility afforded to a peptide with alanine over proline can result in a very different binding mode and affinity. Fourteen percent of samples contained H-Y at positions 6, 7 and 9% of samples contained Y-A-H at positions 1,2,3. **Figure 4.22** depicts a protein logo of the sequencing results, a tool used to visualize the frequency at which an amino acid appears in a set of sequences. Structurally, the most common types of clones to appear at each position are: 1- aromatic, 2- aliphatic/aromatic, 3- basic, 4- glutamine, 5- aliphatic, 6- basic and 7- aromatic/aliphatic.

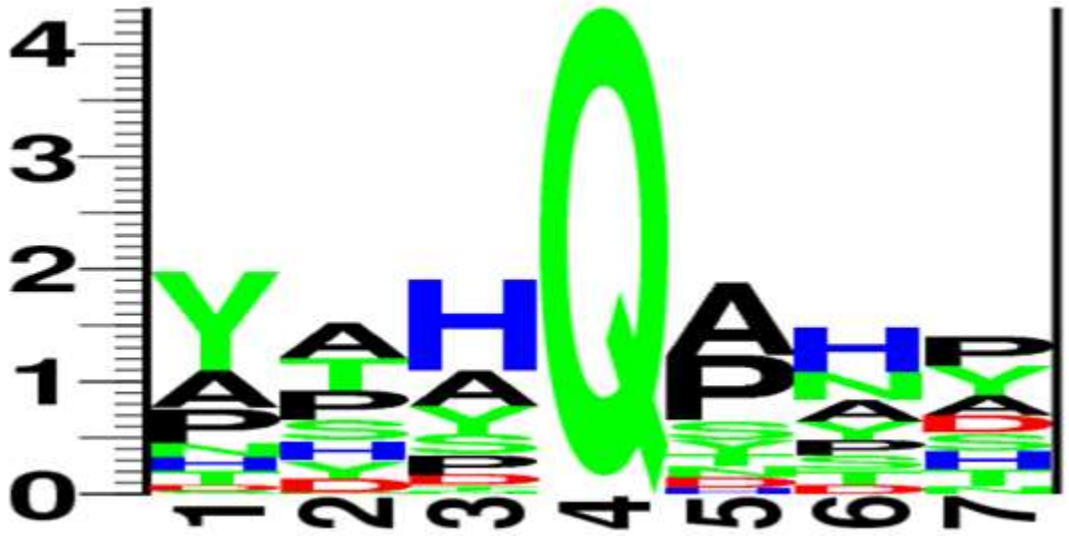


Figure 4.22 Protein logo of peptide library sequencing results. Protein logo representing the sequencing results for the mRuby2-NMT library at 120 min. Cells detected based on observation of FRET due to transamidation facilitated by bTG. FRET observed after excitation at 488 nm, emission at 620 nm. (http://rth.dk/resources/plogo/).

Within the 78 samples successfully sequenced, four included the most common structural trends discussed above:

Sequences of Interest
YAHQAHY
YPHQPHY
YPHQAHY
YAHQPHY

Given the similarity of these sequences and the rate at which they underwent successful transamidation (120-min time point, less than the 180-min time point of the M748 control), we chose these tag sequences for subsequent kinetic characterization. YAHQAHY is the most

interesting sequence because it is the only sequence that was observed three times, while no other candidate was found more than once.

4.5.2 Characterization of potential high reactivity peptides

Based on the peptide library screen results, four peptides of interest were selected as potentially having high reactivity with bTG. As a control, two other peptides (YAHQAAH and YSHQAHY) were selected that did not display all the trends possessed by the four candidates but may still prove to confer some level of affinity for bTG.

These proteins were then used as substrates in the GDH-coupled assay to determine the relative reactivity of their Q-tags (**Table 4.3**). Specificity constants were also measured using mTG and TG2 to also determine the TGase-selectivity of the different sequences.

Table 4.3 Specificity constants of TGase-mediated transamidation. 0.2-0.8 µM mRuby2-Qtag test proteins and 5 mM Gly-OMe were reacted with 0.1 U TGase. Initial rates were measured over 20 min using the GDH-coupled assay. Rates were fit to the Michaels-Menten equation, using Graphpad.

POI-Q	bTG	mTG	TG2
	k_{cat}/K_M (10^{-2} µM^{-1} s^{-1})		
mRuby2-YAHQAHY	32 ± 5	56 ± 5	11 ± 1
mRuby2-YAHQPHY	21 ± 1	7 ± 0.3	7 ± 0.3
mRuby2-YPHQPHY	8 ± 4	5 ± 1	14 ± 8
mRuby2-YPHQAHY	14 ± 3	22 ± 5	2 ± 0.9
mRuby2-YSHQAHY	7 ± 2	0.3 ± 0.06	12 ± 3
mRuby2-YAHQAAH	3 ± 0.8	49 ± 3	3 ± 1
mRuby2-7M48	12 ± 2	68 ± 7	15 ± 4
MBP-RTQPA	5 ± 0.6	11 ± 5	0.5 ± 0.07
MBP-RLQQP	2 ± 0.1	10 ± 3	2 ± 0.8

The peptide that appeared most frequently in the screen also showed the highest reactivity with bTG. mRuby2-YAHQAHY was 6 times more reactive than MBP-RTQPA, the previously reported "high reactivity" tag for bTG. mRuby2-YAHQPHY also demonstrated relatively high reactivity in comparison to previously discovered tags[143]. At 21 x10^{-2} µM^{-1} s^{-1}, YAHQPHY is 10-fold more efficient than RLQQP, which contains two potentially reactive glutamines. The improved kinetic efficiency of these tags validates the success of the peptide library screen and

firmly establishes YAHQAHY and YAHQPHY as a great starting point for the study of bTG substrate specificity.

As an additional consideration, specificity constants were tested for mTG and TG2 as well. With the end goal of harnessing bTG for *in cellulo* protein labelling, it is important that the selected tag is selective for bTG, and shows limited reactivity with TG2, a ubiquitous enzyme within mammalian organisms[153]. **Table 4.3** shows the efficiency constants of the peptide candidates with TG2; most of the peptide sequences show varying levels of reactivity. TG2 showed the greatest efficiency with 7M48, YPHQPHY and YSHQAHY. This could pose a problem for in-cell labelling studies using mTG or bTG. However, both 7M48 and YAHQAHY are significantly more efficient with their intended TGase (mTG and bTG, respectively) than they are with TG2.

Table 4.4 Selectivity of bTG peptides for TG2. Specificity constants of peptides found to have relatively high bTG reactivity were compared to the specificity constants of those same peptides, in their reaction with TG2. The ratio indicates how much more efficient the peptides are for bTG than TG2.

POI-Q	$bTG_{kcat/Km}$ / $TG2_{kcat/Km}$
mRuby2-YAHQAHY	3
mRuby2-YAHQPHY	3
mRuby2-YPHQPHY	0.6
mRuby2-YPHQAHY	7
mRuby2-YSHQAHY	0.6
mRuby2-YAHQAAH	1
mRuby2-7M48	0.8
MBP-RTQPA	10
MBP-RLQQP	1

Table 4.5 compares the relative reactivities of the identified peptides with bTG to TG2. When the ratio of reactivity for bTG:TG2 is analysed, YAHQAHY and YAHQPHY both possess enhanced reactivity for bTG over TG2 (both with a ratio of 3). However, RTQPA (TQ1) shows the greatest selectivity, with a relative reactivity of 10. This does not minimize TQ1's relatively poor absolute reactivity. Considering the absolute reactivity as well as the bTG:TG2 relative reactivity, YAHQAHY is the best tag moving forward.

4.6 Conclusion

In the past, the minimal research on the substrate specificity of bTG has limited its potential as a bio-labelling tool. Unlike other TGases, no comprehensive peptide library has been screened to better understanding its substrate affinity. To address these shortcomings, a focussed heptapeptide library was screened with the intent of identifying high reactivity peptide candidates. To increase the flexibility of the library, 8 representative amino acid residues were screened at all 6 flanking positions, rather than all 20 possible amino acids. This significantly limited the size of the library but still allowed the identification of the most structurally suitable peptides as bTG substrates. To screen the library, a FRET-based assay was developed where Q and K tags were genetically encoded onto the termini of FRET-paired fluorescent proteins. This allowed the detection of transamidation by monitoring increased red fluorescence due to FRET. Controls were used to demonstrate that the assay was effective, and the assay was then adapted to screen an array of Q-tags rather than a single glutamine peptide. The screening results potentially suggest specific trends such as, aromatic amino acids at the both ends of the peptide and flexible aliphatic/basic residues within the sequence. Without screening for library diversity, we cannot be confident in the selection results. However, this does represent an effective screening method for future libraries. The screen also provided a sub-set of specific peptide sequences that corresponded to the structural similarities of the observed trends. Kinetic characterization of those specific peptides led to identifying YAHQAHY as the most reactive peptide for bTG to date. It possesses elevated catalytic efficiency over TQ1 and TQ2, the only other peptides screened for bTG reactivity in the literature[143]. This peptide sequence also shows promising, although not perfect, selectivity for bTG over TG2.

4.7 Perspectives

The identification of YAHQAHY shows that improved reactivity with bTG is attainable. This result was especially valuable given the limited size of the peptide library that was screened. Although the representative library resulted in an improved tag, not all 20 amino acids were varied each of the 6 flanking positions. Therefore, it is more than likely that a tag of greater reactivity can be identified with a larger library. Regardless, the trends observed from the NMT library provide greater insight into bTG affinity. This information can be used when attempting to design an improved peptide through more focused alterations to the recognition sequence.

The success of the library also makes the screening of a K-tag more intriguing. With so little information on the substrate specificity of bTG, a K-tagged peptide library would provide invaluable information bTG. Repeating the experiment, with a K-tag library and YAHQAHY as the constant could result in a K-tag tuned for reactivity with the newly discovered Q-tag.

To discover a Q- and K- tag that are most reactive with one another, a screen of both libraries could be executed. Varying both recognition tags allows the identification of a tag pair that are most suited to react with one another. With the end goal of improved protein labelling, this strategy may be the best option to identify high-reactivity peptides.

4.8 Experimental Section

4.8.1 Cloning of bTG

Two cloning strategies were attempted to obtain transglutaminase from *Bacillus subtilis*. Forward and reverse primers were synthesized by IDT. Primers were designed that cloned the *tgI* gene from the genomic DNA of *B. subtilis*. Using Touchdown PCR, the *tgI* gene was amplified into either pBAD24 or pMAL-c5X vectors. When cloned into pBAD24, additional cloning was necessary to include a C-terminal His-tag.

Table 4.5 Oligonucleotides used for cloning bTG from *Bacillus subtilis*

pBAD_bTG_Fw_NdeI	5’ - CGC CAT ATG ATG ATT ATT GTA TCA GGA CAA T - 3’
pBAD_bTG_Rv_NcoI	5’ - CCC CCA TGG TTA ATG GCG GAC GAT GCG G - 3’
pBAD_bTG_6His_Rv_NcoI	5’ - CCC CCA TGG TTA ATG ATG ATG ATG ATG ATG ATG GCG GAC GAT GCG G - 3’
MBP_6His_bTG_Fw_NdeI	5’ - GGG CTA CATATG CAT CAT CAT CAT CAT CAT AAC ATG ATT ATT GTA TCA GGA CAA TTG CTC - 3’
MBP_bTG_Rv_BamHI	5’ - CGA TGG GGATCC GGC TTT TTA GCG GAC GAT GCG - 3’

The reaction was performed in a BioRad® thermal cycler. After amplification, the gene was digested and ligated into the pBAD24 vector. The constructed vector was then transformed into chemically competent *Escherichia coli* BL21-Gold (DE3) or BL21(A1) in the presence of 100 µg/mL ampicillin by heat shock. Gene inclusion and reading frame were verified through sequencing.

4.8.2 Expression and Purification of bTG

The pBAD24-bTG plasmid encoding for bTG bearing a *C*-terminal hexa-histidine tag was transformed in to *E. coli* BL21(A1) in the presence of 100 µg/mL ampicillin. A 5-mL pre-culture sample was grown overnight and used to inoculate a 1-L culture that was grown to an OD of 0.3 before 0.5% L-arabinose induction for 2 h. The culture was centrifuged at 3000 × g for 30 min at 4 ºC; the supernatant was discarded and the pellet was resuspended in 40 mM phosphate buffer pH 8.0, with 250 mM NaCl. Cells were disrupted by sonication over ice (three cycles of 30-s pulse at 20% intensity / 1 min pause). The solution was transferred to centrifuge bottles and centrifuged for at 18000 x g for 30 min at 4 ºC. During centrifugation, 1 mL of Ni-NTA resin was equilibrated in 40 mM phosphate buffer pH 8.0, with 300 mM NaCl. After centrifugation of the lysate, the supernatant was transferred to a column of Ni-NTA resin and incubated for 1 h while gently rotating at 4 ºC. After incubation, the flow through was discarded, followed by two column-volume washes and elution of bTG using 40 mM phosphate buffer pH 8.0, with 300 mM NaCl and 250 mM imidazole. Eluted bTG was dialyzed against 50 mM phosphate buffer pH 8.0. The average yield was 7 mg of activated bTG per litre of culture, with greater than 70% purity as estimated through evaluation of 10% SDS-PAGE (Figure 4.23) followed by staining with Coomassie blue. Aliquots were snap-frozen and stored at -80°C in 15% glycerol.

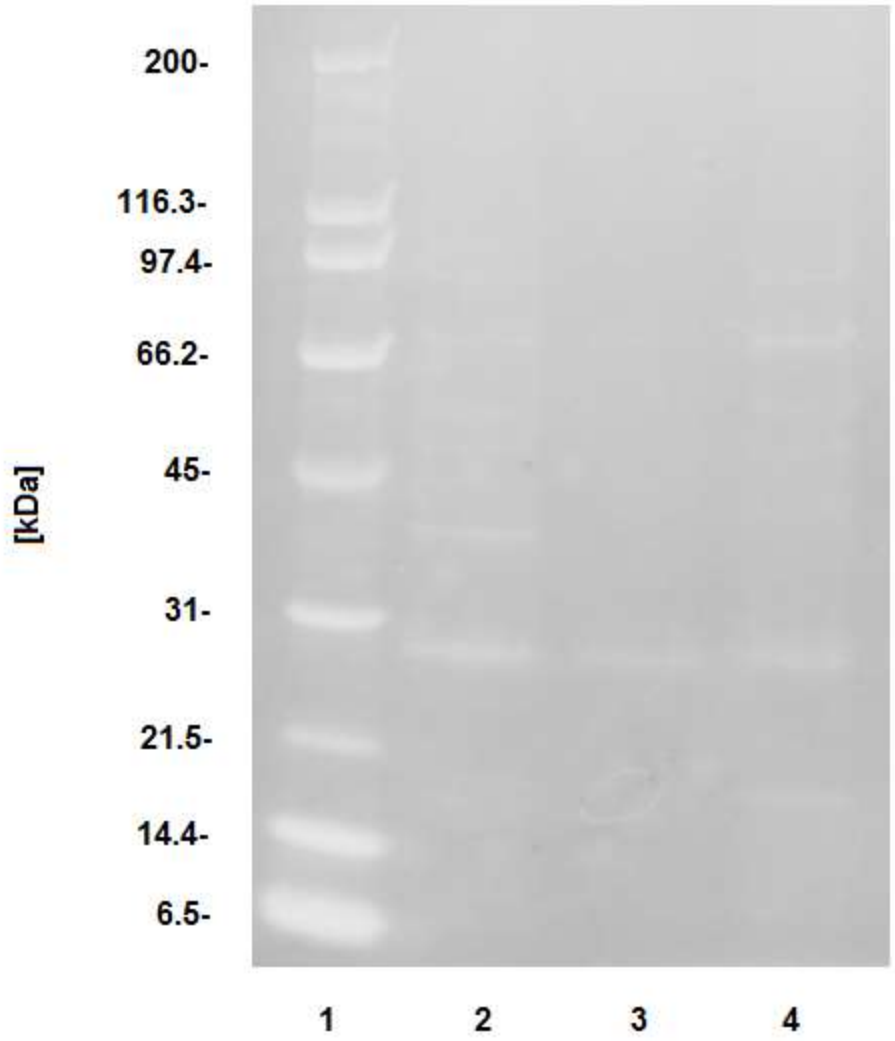


Figure 4.23 SDS-PAGE, coomassie stained gel of bTG, expressed in Bl21(DE3) Gold from pBAD24-bTG plasmid. 1-ladder, 4- flow through, 3-wash and 2- elution (expected band, 28 kDa). 20 ul of samples were loaded in each lane. Gel was run over 2 h at 100 V.

The pMAL-c5X-bTG plasmid encoding a bTG fusion protein was transformed into BL21 (DE3) Gold and expressed and purified according to the previously established maltose binding protein protocol summarized in Chapter 3. The average yield was 35 mg of activated bTG per litre of culture, with greater than 70% purity as estimated through evaluation of 10% SDS-PAGE

(Figure 4.24) followed by staining with Coomassie blue. Aliquots were snap-frozen and stored at -80°C in 15% glycerol.

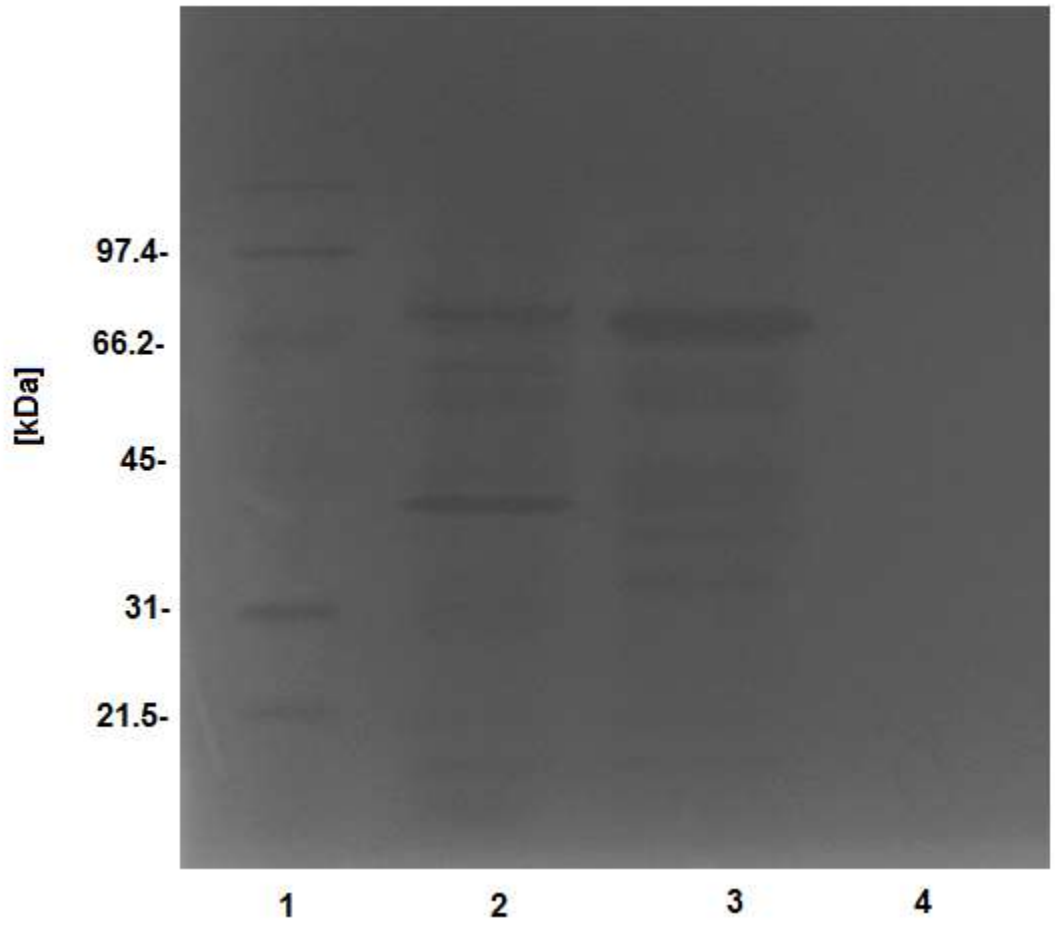


Figure 4.24 SDS-PAGE, coomassie stained gel of bTG, expressed in Bl21(DE3) Gold from pBAD24-bTG plasmid. 1-ladder, 2- flow through, 2- elution (expected band, 70 kDa) and 3-wash. 20 ul of samples were loaded in each lane. Gel was run over 2.5 h at 80 V.

4.8.3 Cloning and expression of mRuby2 and Clover in pACYCduet-1

Plasmids pCDNA3.1-mRuby2 and pCDNA3.1-Clover, coding for mRuby2 and Clover respectively, were purchased from Addgene. To maximize expression levels, both FPs were fused to maltose binding protein (MBP), with the hopes of improved solubility.

Table 4.6 Oligonucleotides used for cloning of Clover-6K

pCDNA3_Clover_Fw _NdeI	5’ - AGT GTG CTG GAA TTC CAT ATG GTG - 3’
pCDNA3_Clover_Rv _NdeI	5’ - TAT AGA CAT ATG ATA GGG CCC TCT AGA CTT- 3’
pMAL_Fw_ MfeI	5’ - CGA GCA ATT GAC CAA CAA GGA CCA TAG ATT ATG - 3’
pMAL_Clover_ spacer_Rv_NdeI	5’ - TAT AGA CAT ATG GGA GGA GCC GGA GGA GCC GGG CCC TCT AGA - 3’
pMAL_Clover_ spacer_6K_Rv_ NdeI	5’ - TAT AGA CAT ATG TTT TTT TTT TTT TTT TTT GGA GGA GCC GGA - 3’
pACYC_MBP_ Clover_Fw_Bg1II	5’ – AAA AGA TCT C ATG AAA ATC GAA GAA - 3’

pACYC_MBP_ Clover6K_Rv_EcoRV	5’ - TAT AGA GAT ATC TTT TTT TTT TTT - 3’

Table 4.7 Oligonucleotides used for cloning of mRuby2-7M48

pCDNA3_mRuby 2_Fw_NdeI	5’ - GCC GCC ACC CAT ATG GTG CGG GGT- 3’
pCDNA3_mRuby 2_Rv_NcoI	5’ – TTT CCA TGG GGC ACA GTC GAG GCT GAT - 3’
pMAL_Fw_ MfeI	5’ - CGA GCA ATT GAC CAA CAA GGA CCA TAG ATT ATG - 3’
pMAL_mRuby2_ spacer_Rv_NcoI	5’ – AGA CCA TGG GGA GGA GCC GGA GGA GCC GGC ACA GTC GAG- 3’
pMAL_mRuby2_ spacer_7M48_Rv _NcoI	5’ –TAT AGA CCA TGG ATG TGG ACG TTG TAA TGC CCA GGA GGA GCC GGA- 3’
pACYC_MBP_ mRuby2_Fw_Ba mHI	5’ - AGA TCT C G GGA TCC ATG AAA ATC GAA GAA - 3’
pACYC_MBP_ mRuby2_7M48_ Rv_EcoRI	5’ –TAT AGA GAA TTC ATG TGG ACG TTG TAA TGC - 3’

Table 4.8 Oligonucleotides used for cloning of mRuby2-NMT

pCDNA3_mRuby 2_Fw_NdeI	5’ - GCC GCC ACC CAT ATG GTG CGG GGT- 3’
pCDNA3_mRuby 2_Rv_NcoI	5’ – TTT CCA TGG GGC ACA GTC GAG GCT GAT - 3’
pMAL_Fw_ MfeI	5’ - CGA GCA ATT GAC CAA CAA GGA CCA TAG ATT ATG - 3’
pMAL_mRuby2_ spacer_Rv_NcoI	5’ – AGA CCA TGG GGA GGA GCC GGA GGA GCC GGC ACA GTC GAG- 3’
pACYC_MBP_ mRuby2_Fw_Ba mHI	5’ - AGA TCT C G GGA TCC ATG AAA ATC GAA GAA - 3’
pACYC_MBP_ mRuby2_spacer_ Rv_EcoRI	5’ –TAT AGA GAA TTC GGA GGA GCC GGA GGA GCC - 3’
pACYC_MBP_ mRuby2_spacer_ NMT_Rv_EcoRI	5’- TAT AGA GAA TTC **NMT NMT NMT** CTG **NMT NMT NMT** GGA GGA GCC GGA GGA - 3’

-**NMT** represents degenerate codon, where N=all four base pairs, M= adenine and cytosine, T=thymine

Forward and reverse primers were synthesized by IDT. Primers were designed to sub-clone the FP gene from pCDNA3.1 to the MCS of pMAL-c5X. FPs were sub-cloned out of pCDNA 3.1 template using PCR and ligated into pMAL-c5X MCS to create MBP-FP fusion protein. After sequencing confirmation, GSS-GSS spacer was introduced to the C-teminal of each FP using similar cloning techniques. Following sequencing of MBP-FP-GSSGSS fusion protein, the recognition tag (7M48, 6K or NMT library) were introduced to the termini of each fusion protein.

For the NMT library, the degenerate codon primers were used to introduce the peptides onto the termini of the FP gene. Due to the length of the reverse primers, touchdown PCR was necessary to amplify the gene of interest After amplification of the MBP-mRuby2-GSSGSS-NMT gene, the gene was digested and ligated in the pMAL-c5X vector with BamHI and EcoRI. The ligated product was then transformed into chemically competent *E. coli*-bTG cells but not plated. Cells were used to inoculate 10 ml LB pre-culture. MBP fusion proteins were expressed per previously established protocols. Successful expression of FPs confirmed via SDS-PAGE (Figure 4.25) and spectroscopy.

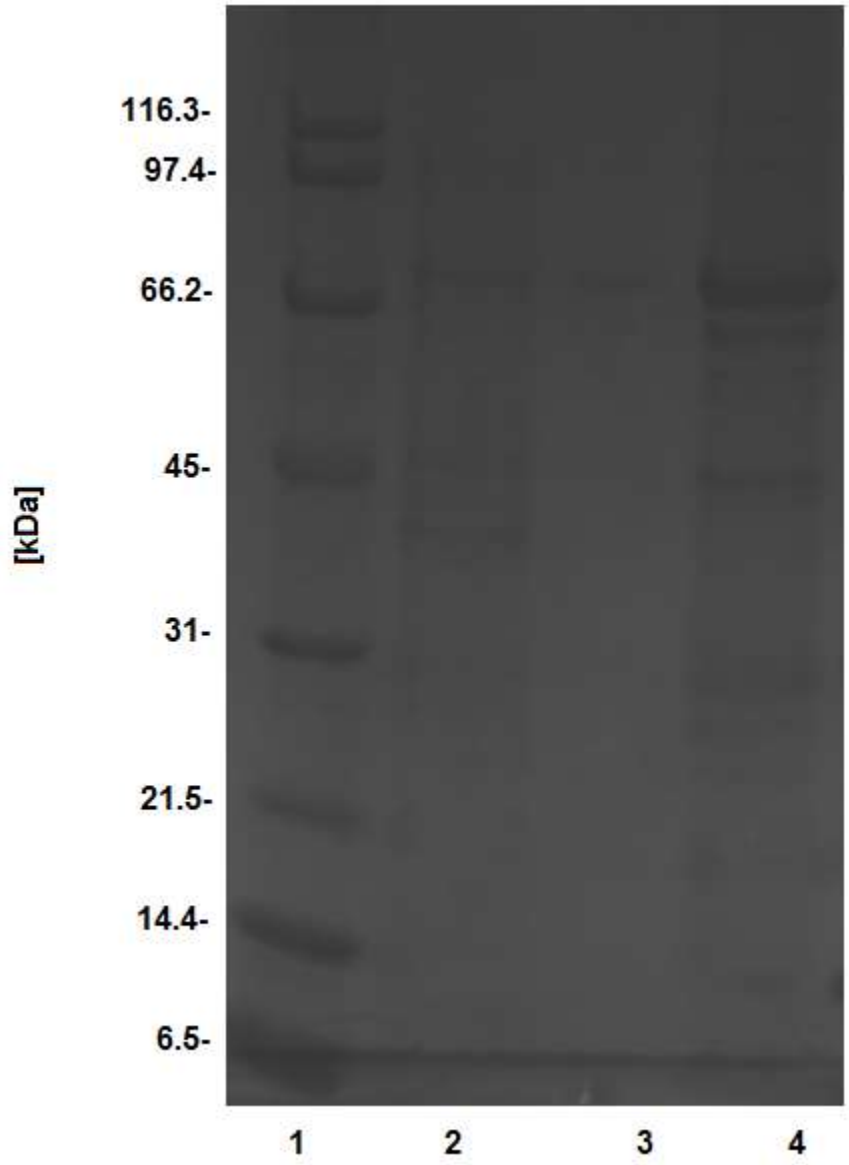


Figure 4.25 SDS-PAGE, coomassie stained gel of mRuby2, expressed in Bl21(DE3) Gold from pACYC-duet1-mRuby2-7M48/Clover-6K plasmid. 1-ladder, 2- flow through, 2- wash and 3-elution (expected band, 69 kDa). 20 ul of samples were loaded in each lane. Gel was run over 2 h at 100 V.

4.8.4 FRET-based assay optimization

After the expression and purification of tagged fluorescent proteins and mTG/bTG, 0.8 µM of mRuby2-7M48 and Clover-6K were incubated in 1 mL reaction mixture, buffered at pH 7.2 with 40 mM MOPS. The sample was excited at 440 nm and scanned from 440-700 nm to identify the baseline of emission before cross-linking had occurred. This was followed by the addition of 1 U of mTG into the reaction mixture. Spectral scans were run on the reaction over the next 24 h using a plate reader. 200-µL samples were scanned at 37^{o}C with no shaking.

4.8.5 In cellulo expression of mRuby2, Clover and bTG

Prior to sorting, 100-mL cultures were inoculated with a 5-mL pre-culture. Cultures were grown to an OD of 0.3 at 37^{o}C. At this point, an IPTG induction (final concentration 400 µM) was carried out for 3 hours at 28 ^{o}C. After the induction period, cells were lightly pelleted (2000 rpm, 30 min) and re-suspended in 100 mL of fresh LB. Cells were then incubated at 4 ^{o}C, for 24-36 hours (to allow for the slow maturation of fluorescent proteins, specifically mRuby2). Afterwards, cultures were incubated at 28 ^{o}C prior to induction of bTG expression using 5% arabinose for 3 hrs. Expression in sub-optimal conditions necessitated the increase in arabinose concentration for induction. After 3 h induction, 5-mL aliquots were taken at different time points (15 min, 30 min, 1 h, 2 h and 3 h) and cooled to 4 ^{o}C to halt transamidation activity prior to flow cytometry.

To validate the co-expression of the three proteins as well as the effective transamidation of the substrates, a gel was run on the reaction culture containing mRuby2-7M48, Clover-6K and bTG,

expressed within *E. coli* after 24 h. As a control, a culture that did not undergo arabinose induction was also analysed.

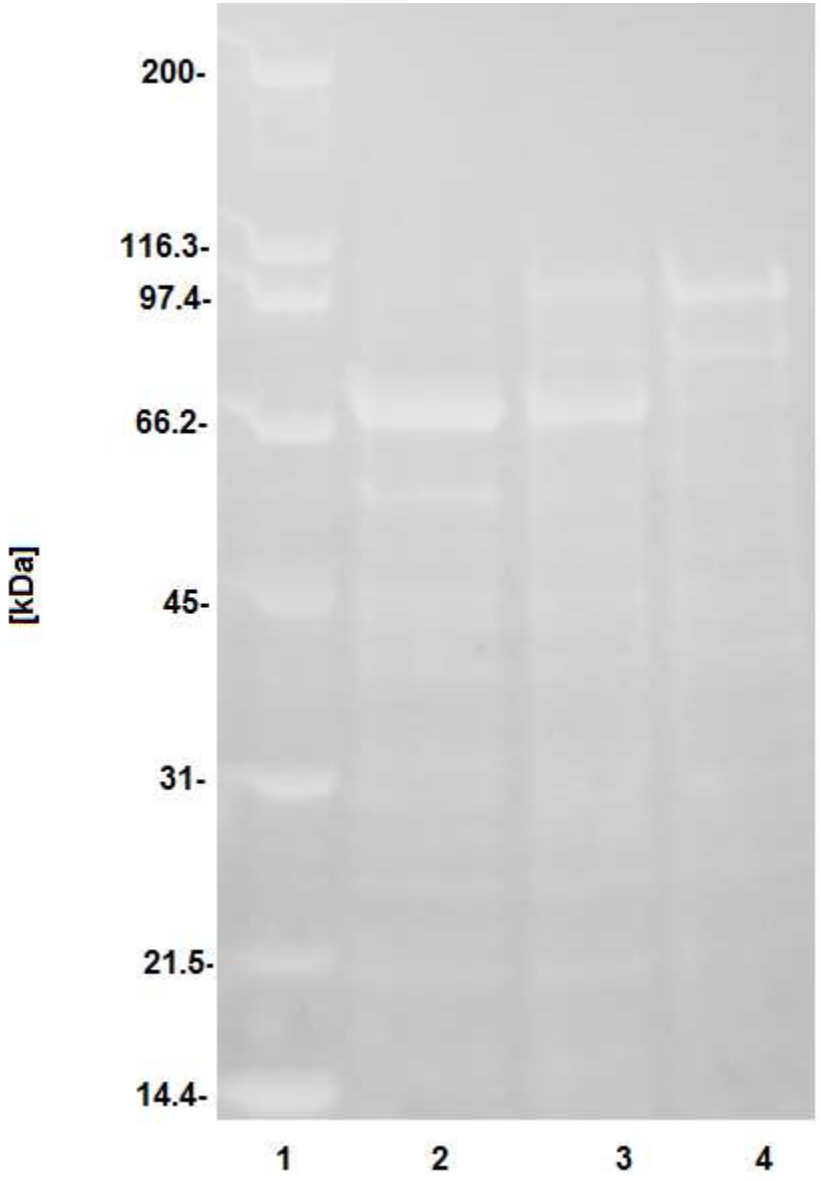


Figure 4.26 SDS-PAGE, coomassie stained gel of co-expressed mRuby2-7M48/Clover-6K/bTG, in Bl21(DE3) Gold from pACYC-duet1-mRuby2-7M48/Clover-6K and pBAD24-bTG plasmids. 1- ladder, **2**- mRub2-7M48/Clover6K lysate no arabinose induction (expected band, 69 kDa), **3**- mRub2-7M48/Clover6K + bTG lysate (expected band, 96 kDa for crosslinked product, 69 for unreacted FP and 28 kDa for bTG) and **4**- mRuby2-Clover fusion protein lysate

(expected band, 96 kDa). 20 ul of samples were loaded in each lane. Cell cultures co-expressing proteins were lysed after 3 h. Gel was run over 2 h at 100 V.

The experiment was repeated, using MBP-bTG rather than bTG given the higher expression levels when bTG is fused to MBP.

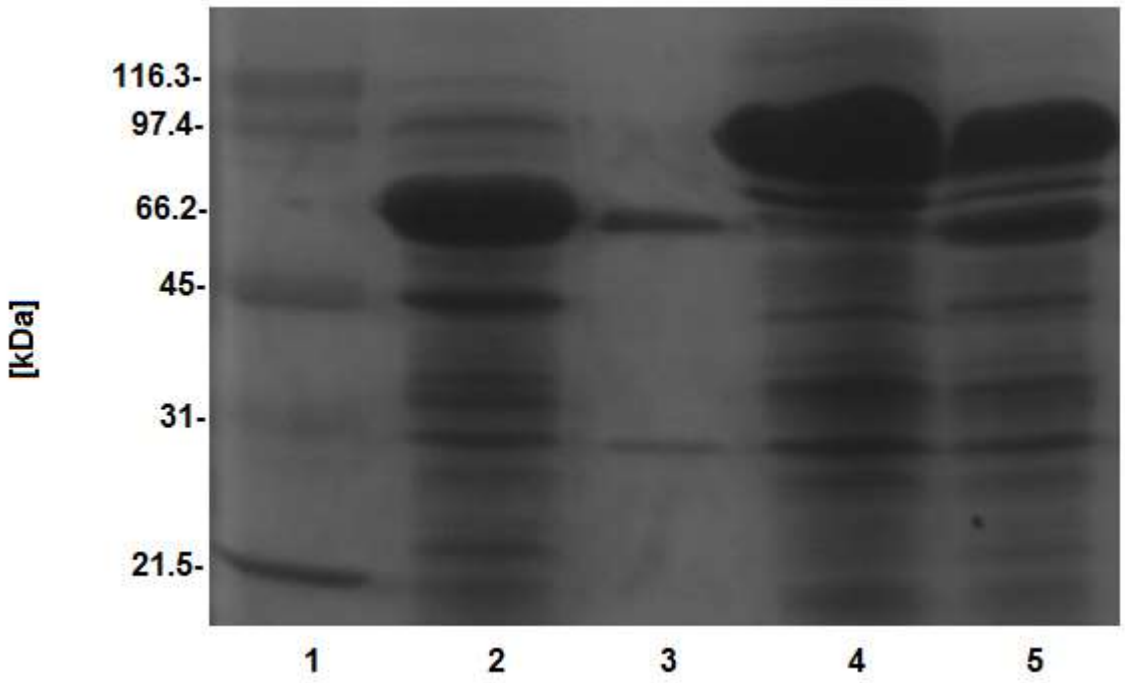


Figure 4.27 SDS-PAGE, coomassie stained gel of co-expressed mRuby2-7M48/Clover-6K/bTG, in Bl21(DE3) Gold from pACYC-duet1-mRuby2-7M48/Clover-6K and pBAD24-MBP-bTG plasmids. **1**- ladder, **2**- mRub2-7M48/Clover6K lysate no arabinose induction (expected band, 69 kDa), **3**- purified MBP-bTG, **4**- mRuby2-Clover fusion protein lysate (expected band, 96 kDa) and **5**- mRub2-7M48/Clover6K + MBP-bTG lysate (expected band, 96 kDa for crosslinked product, 69 for unreacted FP and 28 kDa for bTG). 40 ul of samples were loaded in each lane. Cell cultures co-expressing proteins were lysed after 3 h. Gel was run over 3 h at 80 V.

4.8.6 Expression and purification of Q-tagged MBP-mRuby2 variants from library

Pre-cultures of cell-stocks from selected FRET+ cells were made using 5 mL LB+amp. MBP expression protocols detailed in chapter 3 were followed to express MBP fusion variants. To separate the mRuby2 from the Clover, affinity chromatography was used to purify MBP-mRuby2-Q. Only the MBP-mRuby2-Q proteins possess a Hex-His tag that can be used for additional purification. The lysate was purified using 1 mL of Ni-NTA resin, equilibrated in 50 mM phosphate buffer (pH 8.0) with 300 mM NaCl and eluted with an imidazole gradient (50 mM phosphate buffer, pH 8.0, with 300 mM NaCl 140 mM imidazole) on a gravity column. The purified activated MBP-mRuby2-Q was dialyzed against 50 mM phosphate buffer (pH 8.0). Protein concentration was quantified using Bradford protein assay

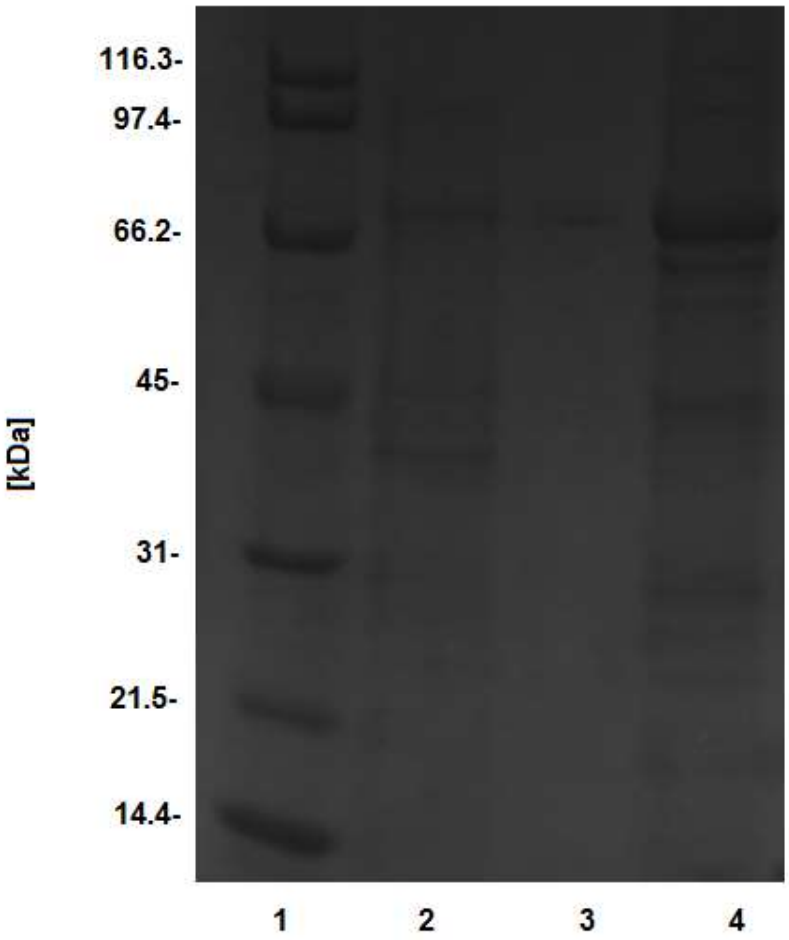


Figure 4.28 SDS-PAGE, coomassie stained gel of mRuby2, expressed in Bl21(DE3) Gold from pACYC-duet1-mRuby2-7M48/Clover-6K plasmid. 1-ladder, 2- flow through, 2- wash and 3-elution (expected band, 69 kDa). 20 ul of samples were loaded in each lane. Gel was run over 2 h at 100 V.

4.8.7 Sorting Results

For this experiment, a *Beckman Coulter MoFlo Astrios EQ* cell sorter was used. 2200 events were sorted per second.

Table 4.9 Sequencing results for peptide library of FRET+ cells, screen at 120 min.

	1	2	3	4	5	6	7	8	9	10	11	12
A	X	Y-T-A-Q-A-H-Y	**Y-A-H-Q-A-H-Y**	T-A-Y-Q-P-N-S	T-T-S-Q-D-N-A	X	H-D-H-Q-S-N-P	Y-H-P-Q-P-Y-A	A-H-H-Q-N-S-H	H-A-A-Q-D-T-Y	Y-P-Y-Q-A-N-D	Y-S-H-Q-A-H-Y
B	A-S-H-Q-Y-T-N	Y-A-Y-Q-N-A-T	X	P-T-H-Q-P-H-Y	P-D-A-Q-P-D-H	P-H-S-Q-A-S-T	P-S-H-Q-T-H-Y	Y-T-P-Q-A-P-P	Y-P-S-Q-A-A-S	X	Y-A-D-Q-S-Y-D	H-T-H-Q-P-T-A
C	T-P-H-Q-A-P-T	**Y-A-H-Q-A-H-P**	P-S-T-Q-A-A-S	P-T-A-Q-P-H-H	A-H-H-Q-P-S-H	**Y-A-H-Q-A-H-Y**	A-D-H-Q-A-T-P	X	Y-T-A-Q-N-A-D	**Y-P-H-Q-P-H-Y**	Y-P-H-Q-A-H-A	X
D	A-D-H-Q-A-Y-P	T-Y-A-Q-P-N-D	**Y-P-H-Q-A-H-A**	X	Y-A-H-Q-T-N-T	Y-D-H-Q-P-H-Y	A-P-Y-Q-Y-N-S	A-S-H-Q-A-Y-H	N-H-S-Q-T-H-P	P-T-H-Q-A-N-P	A-Y-T-Q-P-D-D	P-A-P-Q-A-Y-S
E	X	S-T-P-Q-H-S-D	Y-S-A-Q-H-N-A	Y-D-A-Q-T-H-H	P-Y-P-Q-P-Y-P	N-T-Y-Q-P-H-P	Y-Y-S-Q-S-A-Y	N-Y-H-Q-P-P-P	H-P-A-Q-A-S-P	X	X	X
F	N-Y-A-Q-S-P-N	N-A-Y-Q-A-S-S	A-T-H-Q-P-A-P	P-P-S-Q-A-N-S	P-S-Y-Q-P-Y-S	T-H-A-Q-A-N-P	X	Y-T-N-Q-S-H-T	A-A-Y-Q-A-P-P	Y-P-S-Q-P-Y-D	Y-N-H-Q-P-S-A	Y-A-H-Q-A-A-H
G	P-H-D-Q-T-T-N	X	X	Y-S-A-Q-P-A-H	Y-T-H-Q-P-D-T	A-P-P-Q-Y-Y-D	D-T-H-Q-N-H-Y	Y-A-S-Q-Y-A-Y	**Y-A-H-Q-P-H-Y**	A-P-A-Q-P-P-S	A-P-Y-Q-P-H-A	H-A-H-Q-A-H-Y
H	D-Y-H-Q-D-N-P	**Y-A-H-Q-A-H-T**	Y-S-Y-Q-D-A-H	Y-H-P-Q-S-D-N	Y-Y-H-Q-A-T-A	X	Y-S-D-Q-Y-P-D	X	Y-H-H-Q-P-T-A	Y-T-D-Q-S-N-T	Seq. Control	Seq. Control

Chapter Five: **Site-Specific Protein Labelling Mediated by bTG**

5.1 Introduction

This chapter evaluates *Bacillus subtilis* transglutaminase as a tool for bio-conjugation. Through the kinetic evaluation of potential high-reactivity peptide substrates, tags have been identified that allow the covalent modification of POI's. Through the adaptation of the glutamate dehydrogenase assay, the validation of peptide or small molecule acceptors with bTG helps determine the potential of the enzyme for any labelling strategy. After validation, a POP experiment to demonstrate the capabilities of bTG for labelling while simultaneously highlighting its advantages is necessary. Based on these experiments, bTG stands out as a versatile tool for all forms of protein labelling.

5.1.1 Drawbacks to mTG

In the previous chapter, the limitations of mTG facilitated labelling were discussed in detail. Although the enzyme maintains a variety of beneficial properties, there are still shortcomings that limit its use as bio-conjugation tool[121]. Most notably, the required proteolytic activation of mTG restricts the environments in which the enzyme can be utilized. Applications in the food industry and various other *in vitro* bio-catalytic implementations of the enzyme are unaffected by this limitation[150,154]; however, active expression of the biocatalyst is necessary for any projects that involve in-cell labelling of a POI. Due to these factors, the majority of research involving mTG focuses on *in vitro* or cell surface labelling, where the purified enzyme can be added exogenously[82]; regardless, versatile conjugation strategies based on mTG have been identified[87]. Furthermore, mTG's well-known broad substrate specificity may lead to the modification of additional reactive glutamines and lysines within a POI. The opportunity for excessive cross-

linking counteracts the benefits attributed to mTG mediated bio-conjugation. Although there are not many examples of this occurrence in literature, a bias towards successful studies with limited reactive glutamines may explain this. In this study, mTG controls were run to identify any reactive glutamines prior to inclusion of Q-tag to a POI. This step is crucial for all labelling strategies that include proteins.

5.1.2 Advantages to bTG labelling

In response to the disadvantages of mTG, another bacterial transglutaminase was studied. In chapter 4, bTG was examined as a potential protein-labelling replacement. By its native function, bTG is expressed in its active form, eliminating the biggest drawback to the use of mTG for labelling. Taking advantage of this enzyme presents an opportunity for the co-expression of bTG along with a Q-tagged POI. With the intention of labelling a protein inside a cell, bTG can function at a level previously unseen by mTG. Reviewing other labelling strategies shows that *in cellulo* protein-conjugation is common and cannot be bypassed[155]. Ligases, sortases[156] and FlAsH probes[157] have all been used within living cells to monitor the progression of POI's[34]. This level of protein detection was not an option with mTG-based labelling, but bTG does have this potential. Before attempting to address bTG-facilitated *in cellulo* labelling, a POP experiment that highlights *in vitro* labelling is required. Similarly to mTG, taking advantage of the Q-tag specificity to modify a POI is critical to the advancement of bTG protein labelling. Herein we demonstrate proof-of-principle applications of bTG in the site-specific modification of POIs.

5.2 Fluorescent labelling of POI-Qtag, mediated by bTG

To showcase the labelling potential of bTG, peptide sequences with high k_{cat}/K_M values were introduced to the C-terminus of mRuby2. Thanks to the peptide library screen used to identify bTG peptide substrates, additional cloning was not required. YAHQAHY, YAHQPHY, YPHQAHY and YPHQPHY (hereby known as BQ1, BQ2, BQ3 and BQ4, respectively) tags were selected to demonstrate the labelling capabilities of bTG.

5.2.1 Characterization of bTG substrates

After expression and purification of the most reactive Q-tagged mRuby2 variants, the tagged proteins were validated as substrates for bTG using the glutamate dehydrogenase-coupled assay. In Chapter 4, the relative efficiencies of peptide tags were tested, with the tags acting as glutamine donors and glycine methyl ester as the acceptor substrate. This information proved to be very valuable when assessing reactivity of the newly discovered tags. However, it does not provide direct information about the effectiveness of tagging with bTG. To elucidate this information, screening biologically relevant acceptors substrates was required. By comparing the kinetic parameters of acceptors that can translate to a conjugation strategy, a clearer idea on the effectiveness of the tags can be identified.

The reactivity of these tagged proteins with different acceptors substrates was compared to identify which tags were best suited for labelling. Two acceptor substrates were screened with tags BQ1-4, namely propargyl amine[150] and dansyl-cadaverine[158]. Both acceptor substrates are biologically relevant to a variety of tagging strategies. As demonstrated in chapter 3, introducing a clickable handle onto a POI introduces a multitude of orthogonal chemical diversity that can be

used for labelling or immobilization of a protein. **Table 5.1** shows the initial rates measured for the reaction of mRuby2-Q proteins with propargyl amine, in the presence of bTG, monitored by the GDH-coupled assay. Similar to the results found when glycine methyl ester was used as an acceptor, BQ1 exhibited the greatest reactivity in the presence of propargyl amine. It should be noted that the initial rate of the best known bTG tag is significantly lower than that of mTG reacting with its most effective tag. mRuby2-7M48, in the presence of mTG had an initial rate of $(37.2 \pm 0.3) \times 10^{-2}$ μM/s, three-fold higher than the rate of mRuby2-BQ1, when incubated with bTG and propargylamine. Although the rate is reduced, it is still bio-catalytically relevant and should not impede any labelling experiments. Furthermore, the muted relative rate of transamidation may prove to be beneficial for in cell labelling. As discussed in chapter 4, the promiscuity of bacterial transglutaminases presents a unique problem for bio-conjugation, especially in-cell labelling[138].. The relative rates of transamidation of the bacterial transglutaminases and their effect on non-specific cross-linking warrants further research.

Table 5.1 Initial rates of bTG-mediated transamidation POI-Q and propargyl amine, using the GDH-coupled assay. 0.8 µM of Q-tagged test proteins and 10 mM propargyl amine were reacted with 0.1 U of bTG. Initial rates were tested over 20 min using a plate reader.

Q-tagged test protein	Initial Rates **(µM/s)**
mRuby2-BQ1	$(12.1 \pm 0.2) \times 10^{-2}$
mRuby2-BQ2	$(8.7 \pm 0.5) \times 10^{-2}$
mRuby2-BQ3	$(0.5 \pm 0.8) \times 10^{-2}$
mRuby2-BQ4	$(5.5 \pm 0.9) \times 10^{-2}$
mRuby2	n.d.
MBP-TQ1	$(0.9 \pm 0.4) \times 10^{-2}$
MBP-TQ2	$(0.2 \pm 0.9) \times 10^{-2}$
mRuby2-7M48*	$(37.2 \pm 0.3) \times 10^{-2}$

*transamidation mediated with mTG, for comparison

As a control, untagged mRuby2 was incubated with bTG and propargyl amine; no reactivity was observed, indicating that without the high-reactivity Q-tag introduced to the terminus of the test protein, bTG showed no reactivity. Given the promiscuity of bacterial transglutaminases and the potential for excessive cross-linking, this result bodes well for bTG-facilitated labelling with fluorescent proteins. Given that mRuby2 does not contain a reactive glutamine, it can be reasoned that the family of RFPs, with identical structure aside from a few point mutations, is also not reactive with bTG. Due to the frequency in which fluorescent proteins are used for labelling projects, a lack of reactivity with red fluorescent proteins is an important result.

Incorporation of a chemical handle introduces versatility to any labelling project; however, direct fluorescent labelling is still the most common method for monitoring a POI[159]. To this date, bTG had not been used to facilitate the covalent addition of a fluorescent molecule onto a protein. Although a more comprehensive study of the acceptor substrate scope is needed, lysine is known

to be the native acceptor substrate for the enzyme[160]. Therefore, it would be reasonable to predict that a lysine mimics would provide comparable levels of reactivity with bTG. Dansyl-cadaverine contains a primary amine, followed by a 5-carbon spacer leading up to the fluorophore. The alkyl chain acts as an ideal substitute when compared to lysine. **Figure 5.1** demonstrates the space between the primary amine and the rest of the molecule for both dansyl-cadaverine and lysine. When entering the active site of TGase, the amines are unencumbered by the rest of the molecule, promoting efficient catalytic turnover. The similarity between the molecules has been utilized by TGases in various other labelling projects and could most likely be adapted by bTG as well.

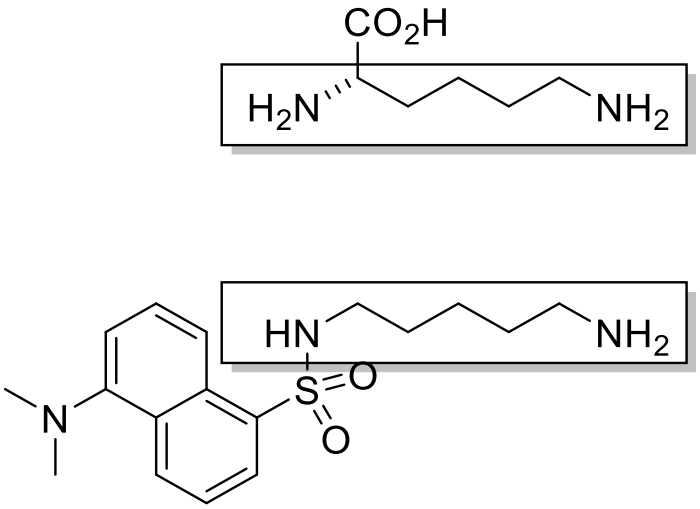


Figure 5.1 Comparison of dansyl-cadaverine and lysine. Five carbon linkers allow primary amines to be used as lysine mimics when reacting with TGase.

To identify whether dansyl-cadaverine serves as a substrate, the GDH-coupled assay was used to measure the initial rates of transamidation between Q-tagged mRuby2 and dansyl-cadaverine. **Table 5.2** compares the initial rates of transamidation for the fluorophore and the mRuby2-Q. As was the case with other acceptor substrates, BQ1 maintains the highest rate of reactivity when compared to other bTG peptides screened. This included previously identified TQ1 and TQ2; the rate of BQ1 being 4-5 times faster than those reported in the literature.

mTG has been shown to not only react with tags on the terminus of a protein but also with internal tags[161]. Reactivity with internal tags not only adds versatility to the labelling strategy but also confirms that mTG affinity is more specific than initially assumed. That is, if an internal glutamine is not reactive, it is due to the residues flanking it, not inaccessibility of the residue itself. Screening for this attribute in bTG may present an additional function that provides added value to the enzyme. To identify the recognition of an internal tag, the 7M48 recognition sequence was placed between MBP and mRuby2. The MBP-7M48-mRuby2 fusion protein was then screened for reactivity with dansyl-cadaverine using the GDH-coupled assay. **Table 5.2** shows that in the presence of an acceptor substrate, the enzyme-coupled assay successfully monitored the transamidation of MBP-7M48-mRuby2. This marks the first time bTG was used to successfully facilitate transamidation with an internal tag.

Both propargyl amine and dansyl-cadaverine act as efficient glutamine acceptor substrates for bTG. Whether the substrate is a small propargyl amine or the larger lysine-mimic, dansyl-cadaverine, bTG catalyzes transamidation with no limitations. The diversity of substrate recognized by the enzyme may prove to be advantageous as a wider range of chemical tools are identified for bio-conjugation strategies.

Table 5.2 Initial rates of bTG-mediated transamidation of POI-Q and dansyl cadaverine, using the GDH-coupled assay. 0.8 µM of Q-tagged mRuby2 test proteins and 5 mM dansyl cadaverine were reacted with 0.1 U of bTG. over 20 min using a plate reader.

POI-Qtag	Initial Rates **(µM/s)**
mRuby2-BQ1	$(20.4 \pm 0.9) \times 10^{-2}$
mRuby2-BQ2	$(15.1 \pm 0.7) \times 10^{-2}$
mRuby2-BQ3	$(3.2 \pm 0.9) \times 10^{-2}$
mRuby2-BQ4	$(11.2 \pm 0.4) \times 10^{-2}$
mRuby2	n.d.
MBP-TQ1	$(4.2 \pm 0.2) \times 10^{-2}$
MBP-TQ2	$(6.2 \pm 0.5) \times 10^{-2}$
MBP-7M48-mRuby2	$(5.9 \pm 0.2) \times 10^{-2}$

5.2.2 Fluorescent labelling of POI-Q tag, mediated by mTG

After quantitative confirmation of the acceptor substrates reactivity with bTG, a proof-of-principle experiment is required to visualize the fluorescent labelling of a POI. As was the case with mTG, introduction of recognition tag to the terminus of a protein converts it into a substrate for bTG. 0.02 mg/mL of Q-tagged POI's were incubated at 37 ºC in the presence of 5 mM dansyl-cadaverine and 0.1 U bTG. The reaction mixture was incubated for 3 h, followed by a series of buffer exchanges to remove non-covalently bound dansyl cadaverine. Unlike propargylation and subsequent click chemistry, vigorous washes were not necessary to remove non-covalently bound dansyl-cadaverine. The extended incubation of test proteins with dansyl-azide may explain the necessity for additional wash steps due to non-specific binding.

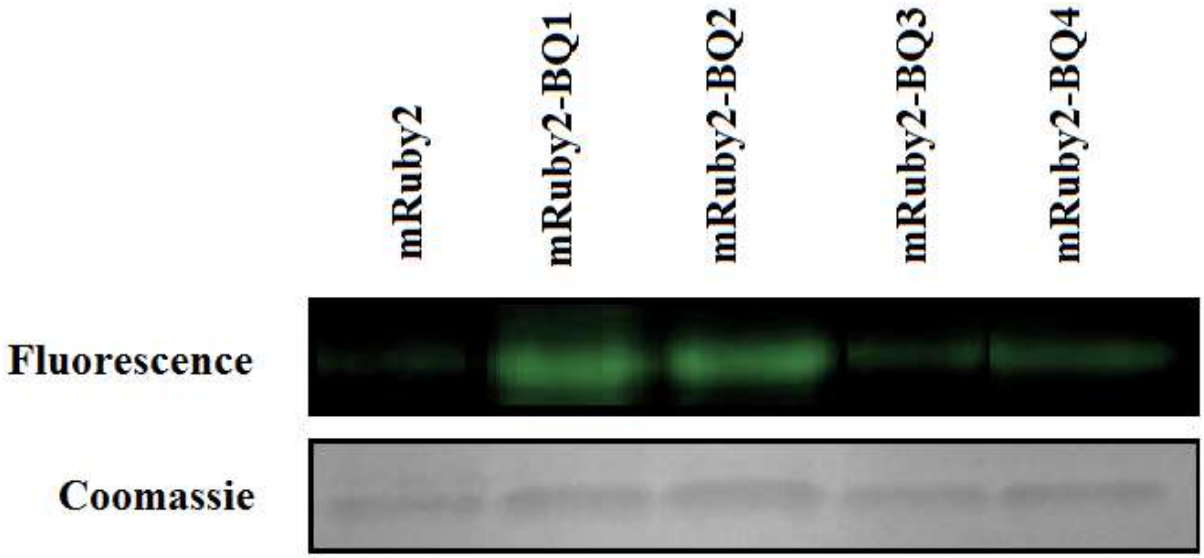


Figure 5.2 bTG-mediated fluorescent labelling of mRuby2 bearing high-reactivity Q-tags. Test proteins bearing bTG recognition tags were fluorescently labelled with dansyl-cadaverine through bTG-mediated transamidation. SDS-PAGE gels of test proteins were run, followed by irradiation to visualize any fluorescent bands. After fluorescent visualization, Coomassie staining were performed to confirm the presence of protein bands.

Figure 5.2 shows the fluorescent labelling of POI-Q, mediated by bTG. bTG successfully cross-linked dansyl-cadaverine to the terminus of a POI; this is the first time this enzyme has been used to label a protein. As expected, a significantly more fluorescent band was observed for mRuby2 test proteins bearing BQ1 and BQ2 tags on their terminus. Given the increased rate of reaction when these peptide sequences are used, and the elevated specificity constants they show, regardless of the acceptor substrate, BQ1 and BQ2 can be considered to be high-reactivity donor peptide tags. They have been proven to be effective, in several different labelling strategies. As a control, fluorescent labelling was attempted using untagged-mRuby2. Fortunately, there was no perceived fluorescence of the protein, indicating that no bio-conjugation occurred, or it was at such low levels that it was not observable.

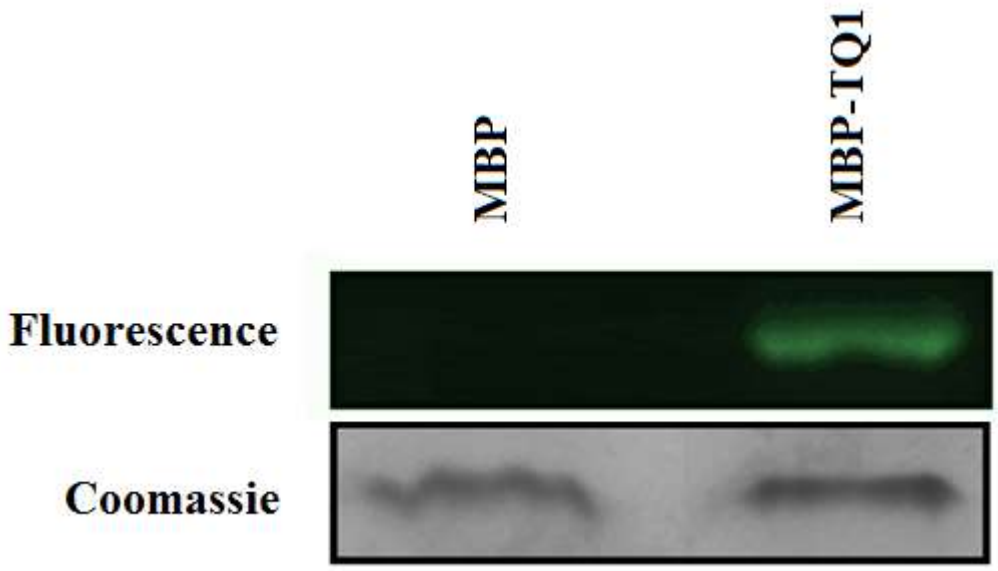


Figure 5.3 bTG-mediated fluorescent labelling of MBP bearing high-reactivity Q-tags. Test proteins bearing the TQ1 tag published by Lee was fluorescently labelled with dansyl-cadaverine through bTG-mediated transamidation. As a control, labelling of untagged MBP was also attempted. SDS-PAGE gels of test proteins were run, followed by irradiation to visualize any fluorescent bands. After fluorescent visualization, Coomassie staining was performed to confirm the presence of protein bands.

It should be noted that **Table 5.2** shows muted reactivity for TQ1 and TQ2 in comparison to BQ-series of peptides. This difference suggests that site-specific labelling with TQ1/2 may be too difficult for TQ tags, within the labelling parameters that were set for bTG (3 h, 37^{o}C). To determine whether the tag could be used for bio-labelling, MBP-TQ1 reacted with dansyl-cadaverine, in the presence of bTG. TQ2 was not tested for fluorescent labelling; two glutamines within the recognition sequence makes identifying which glutamine undergoes transamidation ambiguous. **Figure 5.3** shows that regardless of the reduced reactivity for TQ1, per the enzyme-coupled assay, bTG was still able to directly label the test protein. Fluorescence intensity was not observed over time; therefore, the less reactive tag was still able to label the protein within the time frame given. To better understand labelling of TQ1 vs. BQ- peptides, monitoring

fluorescence at different time points is a logical follow-up experiment. An untagged MBP was run as a control, exhibiting no fluorescence in the presence of bTG, indicating no transamidation without the recognition tag.

Notably, MBP is a second test protein that does not contain any glutamines that are reactive with bTG. This result is positive but it does not preclude the necessity to run this control whenever bTG or any other TGase is being used for labelling. The promiscuity of the enzyme must always be considered when developing a labelling strategy with transglutaminases.

5.3 Conclusions

After diligent optimization and characterization of high-reactivity peptides, bTG-mediated fluorescent labelling was established. By fusing two different tags onto MBP, two test proteins were prepared and subsequently confirmed to be bTG substrates using the previously established GDH-coupled assay. The versatility of the assay enabled the verification of previously identified bTG peptides, BQ1 and BQ2, as significantly more reactive in comparison to TQ1 and TQ2. Two commonly used acceptor substrates, propargyl amine and dansyl-cadaverine were shown to function as acyl-acceptor substrates for the BQ1 and BQ2 tags. This was an important result, due to the utility of those two small molecules in a variety of labelling strategies. This opens the door to direct protein labelling or the incorporation of a chemical handle that permits a variety of orthogonal chemistry to occur. Once characterization was complete, the visualization of direct protein labelling was performed. **Figure 5.2** demonstrates the first time bTG was used to directly fuse a fluorophore onto the terminus of a POI. This opens the door to a variety of possibilities involving bTG-mediated labelling. The added benefit of bTG's active expression pushes the scope of functionality to places that could not be reached with mTG at the start of this study. Additionally, bTG did not show any reactivity when incubated with two different untagged test proteins, namely mRuby2 and MBP. This does not mean bTG is not promiscuous; however, it does speak to the value of the Gln-recognition sequence for conferring bTG affinity.

5.4 Perspectives

The often-discussed *in cellulo* POP experiment is the final step in the validation of bTG as a complete bio-labelling tool. With *in vitro* bio-conjugation fully established and optimized, demonstrating that bTG can be used in-cell is all that remains. As discussed throughout the chapter, the biggest challenge to successful *in cellulo* labelling is the potential promiscuity of bTG. Although bTG appears to be less catalytically efficient than mTG, it is still a robust enzyme, potentially capable of non-specific cross-linking within a cell. This possibility may have toxic effects and prevent the enzyme from enjoying the same flexibility it maintains for *in vitro* labelling. Bacterial hosts are known to undergo significant cross-linkage over time when expressing bTG, but it is unknown whether a similar effect will be had on mammalian cells. Controls must be run to identify whether bTG expression decreases mammalian cell survival before *in cellulo* labelling can be tested.

Separate from *in cellulo* labelling, very little is known about the acceptor substrate scope of the enzyme. This work showed that glycine methyl ester, propargyl amine and dansyl-cadaverine all demonstrate reactivity with bTG. However, there are a wide range of small molecules that bTG may have affinity for. Given the lack of homology between the bacterial transglutaminases studied in this research, it would be unwise to assume they maintain the same substrate specificity. A study to elucidate information of substrate scope would provide much needed information about how to most effectively take advantage of bTG.

5.5 Experimental Section

5.5.1 Cloning of MBP-7M48-mRuby2

As was the case when introducing 7M48 tag to the MCS of MBP, forward and reverse primers were synthesized by IDT and used in Touchdown PCR with a pMAL-c5X as a template:

Table 5.3 Oligonucleotides used for cloning MBP-7M48-mRuby2

MBP_Fw_NdeI	5'- ATC GAG GGA AGG ATT TCA CAT ATG - 3'
MBP_spacer_Rv_EcoRV	5' - TCC GTC GAC GAT ATC GGA GGA GCC GGA GGA GCC - 3'
MBP_spacer_7M48_Rv_ EcoRV	5' - TTT TTT GAT ATC ATG TGG ACG TTG TAA TGC CCA GGA GGA GCC GGA - 3'

Two PCR reactions were performed to introduce the spacer and 7M48 tag, sequentially in a BioRad® thermal cycler. Once amplified, the tagged MBP was digested and ligated into the pMAL-c5X vector MCS. Inclusion of the Q-tag was verified through DNA sequencing after successful chemical transformation using DH5α cells. Due to the abundance of restriction sites down stream of the tagged-MBP within pMAL-c5X, additional PCR was not needed to introduce the mRuby2 gene to the C-termni of 7M48. Instead, mRuby2 was digested from another pMAL-c5X plasmid using BamHI/EcoRI and inserted into the MCS, following the 7M48 tag. The constructed vector was expressed in *E, coli* BL21-Gold (DE3) in the presence of 100 mg/mL ampicillin. Expression and purification followed the established protocols in Chapter 3.

5.5.2 Fluorescent labelling of Q-tagged test proteins

Multiple 0.02 mg/ml Q-tagged test proteins (mRuby2-BQ1/2/3/4 and MBP-TQ1) and untagged controls (mRuby2 and MBP) were incubated with 5 mM dansyl-cadaverine and 0.1 U of bTG were incubated in 1 mL of reaction mixture, buffered at a pH of 7.2 with 40 mM MOPS, 5 mM sodium acetate, 1 mM EDTA. The reaction was run at 37 ^{o}C for 3 h with no shaking. After time had elapsed, the samples underwent five rounds of buffer exchange using Amicon Filters with a cut off of 14 kDa. After washes were complete, a 10% SDS-PAGE gel was run at 80 V for 2 hrs. Once the gel was complete, it was irradiated using a transilluminator to visualize the dansyl group which is excited at 340 nm and emits at 470 nm[162]. After the gel's fluorescence was observed, it was stained with Coomassie blue to identify proteins and ensure the fluorescence band corresponds to the test protein.

Chapter Six: **Conclusions and Future Directions**

The research in this thesis focused on the development of two enzymes as tools to facilitate protein labelling. Four main objectives were presented that would shape the success or failure of the overall project. In each chapter, results based on the stated objective were discussed at length. In this last chapter, the research completed will be summarized and evaluated based on the initial goals of the work. Additionally, the future perspectives of the research will be discussed, as it will provide a guideline for the direction of the project.

6.1 Objective 1: Development of enzyme-coupled assay to monitor mTG activity

6.1.1 Achieved results and conclusions

With the guidance of previous work from the Keillor group, the GDH-coupled assay was adapted to evaluate the activity of microbial transglutaminase. This assay includes the variability of screening any donor and acceptor substrates, permitting that the enzymatic reaction results in the release of ammonia. During optimization, important parameters were set, including the pH range of assay effectiveness (6-9) and mTG's limit of detection (0.02 U). Overall, the assay provides an efficient, low-cost method to enzyme activity detection that is tunable to high-throughput screening.

The adaptation of this method goes beyond assay development; specific peptidic sequences required validation as high-reactivity peptide substrates. CBz-Gln-Gly was the reported industry standard for mTG glutamine-donor substrates. However, its low-reactivity suggested that an improved substrate could be identified. Furthermore, to use mTG for bio-conjugation, a peptide

substrate was more advantageous. Through the work of Hitomi, contracted peptide substrates identified in a phage-display library were characterized using the enzyme-coupled assay. Of the heptamers peptides identified, **7M48** was selected as it demonstrated the greatest reactivity for mTG. This discovery validates the assay and provides a genetically encodable tag that facilitates protein labelling.

6.1.2 Future work for the assay and the peptide tag

Due to the flexibility of the assay, and the broad substrate specificity of mTG, additional work can be done to screen new substrates. An extensive study on the acceptor substrate versatility has already been performed[110]; however, there is not much known about the donor substrates. With the help of the assay, a variety of small molecule substrates can be quickly screened for mTG affinity. Will mTG be responsive to the substitution of an ester bond cleavage, rather than an amide bond? How will the steric bulk of the parts of a molecule that flank a glutamine residue affect its reactivity? What other amide bonds can mTG react with? There are numerous donor candidates that can be tested that will provide invaluable information about the extent to which native mTG reactivity can be pushed.

The assay can also be used to screen mTG variants constructed to modify its substrate scope. mTG mutations have been limited to modifying or eliminating the pro-peptide with the intention of altering expression and activity[68]. To date, there has been no work on engineering the active site of the enzyme to alter its binding pocket and subsequent specificity. The assay presents an efficient method to identify successful mutants while minimizing the consumables used.

7M48 was shown to have the greatest reactivity for mTG of the candidates selected; however, its K_M value was in the high micromolar range. Additional optimization of the Q-tag peptide could

be done to improve the reactivity of the recognition sequence. Modification of residues within 7M48 that maintain the structural environment of the peptide or follow trends highlighted in the peptide screen may enhance the tag's reactivity. Improved reactivity can only help to increase the effectiveness of the tag when used for protein labelling.

6.2 Objective 2: Site-specific labelling and immobilization, mediated by mTG

6.2.1 . Achieved results and conclusions

With the identification of high-reactivity peptide substrates, demonstrating the capacity of mTG as bio-catalyst was crucial. The minimal peptide tags were already known to have elevated reactivity for mTG, in comparison to CBz-Gln-Gly; however, they had not been used to site-specifically incorporate a molecule onto a test protein. Peptides sequences were successfully encoded to the termini of MBP and screened for reactivity with the enzyme. The promiscuity of mTG necessitates controls that identify any non-specific transamidation. Potentially reactive glutamines on the surface of MBP could invalidate the site-specific claim of the labelling strategy. However, when untagged MBP was tested, no reactivity was found, indicating that the tag is necessary to facilitate mTG activity. After optimization of the spacer length was completed to make the peptide tags more accessible, proteins successfully had propargyl amine incorporated onto their mTG recognition tag. This molecule was selected, not simply for its high reactivity[110,150] for mTG but also for the chemical diversity it introduces to a POI through click chemistry. After propargylation, different chemical anchors can be introduced to a protein; this led to fluorescent labelling of MBP as well as the immobilization of mRuby2 onto a nanoparticle.

6.2.2 Future work for in vitro labelling

The *in vitro* modification of a test protein was shown in this research; however, more POP experiments can be done to fully display the range of mTG as bio-conjugation tool. mTG can facilitate cell-surface labelling[48], antibody-drug conjugation[163,164] and polymer formation[165] to highlight its versatility. With the use of the 7M48 tag, conjugation can be more efficient then ever before while introducing the intrinsic benefits of a peptide tag rather than a small molecule tag.

The expansion of mTG research to facilitate *in cellulo* labelling is the final progression in mTG-mediated bio-conjugation. Due to the limitations of mTG expression, this was not seen as a realistic option at the onset of these studies. However, with the development of a mature mTG expression strategy, the enzyme can now be optimized for in-cell labelling. As an alternative, a better-suited enzyme could be characterized and optimized for covalent modification. This strategy was selected when the research shifted to bTG-mediated labelling.

6.3 Objective 3: Identification and characterization of bTG peptide substrates

6.3.1 Achieved results and conclusions

The limited research on bTG in comparison to mTG set the initial starting point for this work. A lack of information on specificity of the enzyme as well as no structural homology between bTG and other bacterial TGases meant limited inferences could be made based on other homologs. A peptide library was designed to better understand substrate affinity; this would also provide insight into the similarities of affinity with other TGases. Library design focused on limiting the

number of residues screened while maintaining a set of amino acids that were structurally representative of all 20 possible amino acids. Eight representative amino acids were screened at all 6 positions flanking the fixed reactive glutamine. The focus of the library ensured that any screening method selected would be adaptable to the number of candidates designed. As a screening method, a FRET-based assay was designed that took advantage of bTG's capacity to cross-link reactive glutamines/lysines and the optimized FRET pairing of mRuby2 and Clover FPs. Q- and K- tags were encoded to the termini of mRuby2 and Clover, respectively. This permitted the detection of transamidation by the observation of red fluorescence due to FRET. Once optimized, a Q-tag library was successfully screened, with FRET+ cells being sorted using FACS. The selected Q-tags, YAHQAHY specifically, showed elevated catalytic efficiency with bTG in comparison to TQ1 and TQ2, the only other sequences tested for bTG reactivity. These results are sufficient starting point to understand bTG reactivity.

6.3.2 Future work

The desire for versatility of the library limited the range of amino acids the library would cover. Designing an NMT degenerate codon library resulted in 262,144 different codons and 603607 cells needing to be screened to cover 90% of the desired possible sequences variants. This seems like a large library; however, the speed of a cell sorter reduces the library size concerns. At 2200 events/second, a much larger library can be screened in relative short amount of time. As mentioned before, **BQ1** is a promising starting point for bTG peptide reactivity; however, a library that screened all 20 possible amino acids at every flanking positions would yield more comprehensive results.

Aside from a larger Q library, screening a K library may also be of interest. As said before, so little is known about the acceptor substrate affinity of bTG that it may prove to be significantly different than other TGases. With the groundwork being set with the Q library, repeating the screen with a fixed Q-tag and varied K-tags may provide useful insight into the substrate affinity of the enzyme. A more optimistic screen may involve screening a Q and K library simultaneously. By screening the libraries separately, the results are biased towards the sequence most reactive to the fixed Q or K tag being screened against. By introducing both libraries simultaneously, the greatest Q/K tag pair would be selected, which is the desired outcome for a protein-labelling project. Future work is dependent on the size of library the cell sorter can realistic analyze; within reason, this method can be adapted to identify precisely what trends are important for bTG affinity.

6.4 Objective 4: Identification and characterization of bTG peptide substrates

6.4.1 Achieved results and conclusions

Following the identification of peptides that confer improved reactivity with bTG, protein labelling was successful established. Two test proteins, encoded with bTG recognition tags, underwent transamidation in the presence of bTG and previously un-tested acceptor substrates. Not only was improved bTG reactivity observed when BQ1 and BQ2 were compared to TQ1 and TQ2, but propargyl amine and dansyl-cadaverine were shown to have considerable reactivity with the enzyme. This was the first time that bTG was shown to react with the two acceptor substrates selected. Given the value and versatility present by the two acceptor substrates, bTG shows great potential as a tool for bio-conjugation. Furthermore, the direct incorporation of

dansyl-cadaverine to the termni of a test protein marks the first time bTG was used to fluorescently label a protein. This result will act as the starting point for bTG-mediated labelling and conjugation strategies.

6.4.2 Future work

As a member of the TGase family, certain characteristics are afforded to bTG. Bacterial TGases have been proven to possess wider substrate scope then their mammalian homologs; however, most of this work is based on the study of mTG. Comprehensive analysis on the substrate scope of bTG will provide much greater insight into the capabilities and catalytic potential of the enzyme. Identifying whether bTG can catalyze amide bond formation as well as isopeptide formation may result in a new pathway for the enzyme and peptide chemistry.

The impetus for shifting this research to bTG was due to its potential for labelling a POI within a living cell. As discussed in detail, the active expression of the enzyme makes in-cell expression and subsequent bio-conjugation a realistic option. However, extensive optimization will be required before this potential can be realized. Future work that identifies non-specific activity that bTG may possess within a cell will help to determine how feasible this labelling method is. No research has studied the effect of bTG expression on mammalian hosts; however, it has suggested[68] that bacterial TGases have a toxic effect on bacterial hosts, specifically when expressed in active form. Comprehensive studies to determine what level of expression can be tolerated by different host cells will be beneficial when attempting to label a POI-Q that is endogenously expressed. The detection of POI's within living cells is the final objective of this work.

Successful labelling in living cells is incumbent on overcoming the promiscuity of bTG. If unattainable with the native form of the enzyme, alternatives must be made. Enzyme engineering may be considered to reduce the substrate scope of the enzyme. The broad substrate range of bacterial TGases has been useful for most bio-catalytic applications; however, for protein labelling, this may prove to be a hindrance. A labelling strategy that is truly specific for its target is crucial to its success.

References

(1) Duan, G.; Walther, D. *PLoS Comput. Biol.* **2015**, *11* (2), 1–23.

(2) Spicer, C. D.; Davis, B. G. *Nat. Commun.* **2014**, *5*, 4740.

(3) Giepmans, B. N. G.; Adams, S. R.; Ellisman, M. H.; Tsien, R. Y. *Science (80-.).* **2006**, *312* (5771), 217–224.

(4) Lichtman, J. W.; Conchello, J. A. *Nat. Methods* **2005**, *2* (12), 910–919.

(5) Terai, T.; Nagano, T. *Pflugers Arch. Eur. J. Physiol.* **2013**, *465* (3), 347–359.

(6) Romanini, D. W.; Cornish, V. W. *Nat. Chem.* **2012**, *4* (4), 248–250.

(7) Chalfie, M.; Sciences, B. *Photochem. Photobiol.* **1995**, *62* (4), 651–656.

(8) Jensen, E. C. *Anat. Rec.* **2012**, *295* (12), 2031–2036.

(9) Zhang, J.; Campbell, R. E.; Ting, A. Y.; Tsien, R. *Nat. Rev. Mol. Cell Biol.* **2002**, *3* (12), 906–918.

(10) Højlys-Larsen, K. B.; Sørensen, K. K.; Jensen, K. J.; Gammeltoft, S. *Mol. Biosyst.* **2012**, *8* (5), 1452–1460.

(11) Crivat, G.; Taraska, J. W. *Trends Biotechnol.* **2012**, *30* (1), 8–16.

(12) Rashidian, M.; Dozier, J. K.; Distefano, M. D. *Bioconjug. Chem.* **2013**, *24* (8), 1277–1294.

(13) Takaoka, Y.; Ojida, A.; Hamachi, I. *Angew. Chem. Int. Ed. Engl.* **2013**, 2–21.

(14) Van Berkel, S. S.; Van Eldijk, M. B.; Van Hest, J. C. M. *Angew. Chemie - Int. Ed.* **2011**, *50* (38), 8806–8827.

(15) Li, L.; Zhang, Z. *Molecules* **2016**, *21* (10), 1–22.

(16) Nikić, I.; Kang, J. H.; Girona, G. E.; Aramburu, I. V.; Lemke, E. a. *Nat. Protoc.* **2015**, *10* (5), 780–791.

(17) Gong, Y.; Pan, L. *Tetrahedron Lett.* **2015**, *56* (17), 2123–2132.

(18) Rush, J. S.; Bertozzi, C. R. *J. Am. Chem. Soc.* **2008**, *130* (37), 12240–12241.

(19) Wu, P.; Shui, W.; Carlson, B. L.; Hu, N.; Rabuka, D.; Lee, J.; Bertozzi, C. R. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (9), 3000–3005.

(20) Rabuka, D.; Rush, J. S.; deHart, G. W.; Wu, P.; Bertozzi, C. R. *Nat. Protoc.* **2012**, *7* (6), 1052–1067.

(21) Yin, J.; Liu, F.; Li, X.; Walsh, C. T. *J. Am. Chem. Soc.* **2004**, *126* (25), 7754–7755.

(22) Zhou, Z.; Cironi, P.; Lin, A. J.; Xu, Y.; Hrvatin, S.; Golan, D. E.; Silver, P. A.; Walsh, C. T.; Yin, J. *ACS Chem. Biol.* **2007**, *2* (5), 337–346.

(23) Clarke, K. M.; Mercer, A. C.; La Clair, J. J.; Burkart, M. D. *J. Am. Chem. Soc.* **2005**, *127* (32), 11234–11235.

(24) Quadri, L. E. N.; Weinreb, P. H.; Lei, M.; Nakano, M. M.; Zuber, P.; Walsh, C. T. *Biochemistry* **1998**, *37* (6), 1585–1595.

(25) Worthington, A. S.; Burkart, M. D. *Org. Biomol. Chem.* **2006**, *4* (1), 44–46.

(26) Mazmanian, S. K.; Liu, G.; Ton-That, H.; Schneewind, O.; Gold, H. S.; Moellering, R. C.; Hiramatsu, K.; Sieradzki, K.; Roberts, R. B.; Haber, S. W.; Tomasz, A.; Strominger, J. L.; Walsh, C. T.; Navarre, W. W.; Schneewind, O.; Uhlén, M.; Guss, B.; Nilsson, B.; Götz, F.; Lindberg, M.; Navarre, W. W.; Schneewind, O.; Ton-That, H.; Faull, K. F.; Schneewind, O.; Navarre, W. W.; Ton-That, H.; Faull, K. F.; Schneewind, O.; Schneewind, O.; Fowler, A.; Faull, K. F.; Schneewind, O.; Mihaylova-Petkov, D.; Model, P.; Schneewind, O.; Model, P.; Fischetti, V. A.; Ton-That, H.; Labischinski, H.; Berger-

Bächi, B.; Schneewind, O.; Berger-Bächi, B.; Schindler, C. A.; Schuhardt, V. T.; Kopp, U.; Roos, M.; Wecke, J.; Labischinski, H.; Boothby, D.; Daneo-Moore, L.; Shockman, G. D.; Navarre, W. W.; Ton-That, H.; Faull, K. F.; Schneewind, O. *Science (80-.).* **1999**, *285* (5428), 760–763.

(27) Tsukiji, S.; Nagamune, T. *ChemBioChem* **2009**, *10* (5), 787–798.

(28) Popp, M. W.; Antos, J. M.; Grotenbreg, G. M.; Spooner, E.; Ploegh, H. L. *Nat. Chem. Biol.* **2007**, *3* (11), 707–708.

(29) Duckworth, B. P.; Xu, J.; Taton, T. A.; Guo, A.; Distefano, M. D. *Bioconjug. Chem.* **2006**, *17* (4), 967–974.

(30) Kho, Y.; Kim, S. C.; Jiang, C.; Barma, D.; Kwon, S. W.; Cheng, J.; Jaunbergs, J.; Weinbaum, C.; Tamanoi, F.; Falck, J.; Zhao, Y. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101* (34), 12479–12484.

(31) Weinrich, D.; Lin, P. C.; Jonkheijm, P.; Nguyen, U. T. T.; Schröder, H.; Niemeyer, C. M.; Alexandrov, K.; Goody, R.; Waldmann, H. *Angew. Chemie - Int. Ed.* **2010**, *49* (7), 1252–1257.

(32) Duckworth, B. P.; Zhang, Z.; Hosokawa, A.; Distefano, M. D. *ChemBioChem* **2007**, *8* (1), 98–105.

(33) Xu, J.; DeGraw, A. J.; Duckworth, B. P.; Lenevich, S.; Tann, C. M.; Jenson, E. C.; Gruber, S. J.; Barany, G.; Distefano, M. D. *Chem. Biol. Drug Des.* **2006**, *68* (2), 85–96.

(34) Rashidian, M.; Mahmoodi, M. M.; Shah, R.; Dozier, J. K.; Wagner, C. R.; Distefano, M. D. *Bioconjug. Chem.* **2013**, *24* (3), 333–342.

(35) Dirksen, A.; Hackeng, T. M.; Dawson, P. E. *Angew. Chemie - Int. Ed.* **2006**, *45* (45), 7581–7584.

(36) Algar, W. R.; Prasuhn, D. E.; Stewart, M. H.; Jennings, T. L.; Blanco-Canosa, J. B.; Dawson, P. E.; Medintz, I. L. *Bioconjug. Chem.* **2011**, *22* (5), 825–858.

(37) Howarth, M.; Takao, K.; Hayashi, Y.; Ting, A. Y. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (21), 7583–7588.

(38) Slavoff, S. A.; Chen, I.; Choi, Y. A.; Ting, A. Y. *J. Am. Chem. Soc.* **2008**, *130* (4), 1160–1162.

(39) Fernández-Suárez, M.; Baruah, H.; Martínez-Hernández, L.; Xie, K. T.; Baskin, J. M.; Bertozzi, C. R.; Ting, A. Y. *Nat. Biotechnol.* **2007**, *25* (12), 1483–1487.

(40) Farazi, T. A.; Waksman, G.; Gordon, J. I. *J. Biol. Chem.* **2001**, *276* (43), 39501–39504.

(41) Price, H. P.; Menon, M. R.; Panethymitaki, C.; Goulding, D.; McKean, P. G.; Smith, D. F. *J. Biol. Chem.* **2003**, *278* (9), 7206–7214.

(42) Wright, M. H.; Heal, W. P.; Mann, D. J.; Tate, E. W. *J. Chem. Biol.* **2010**, *3* (1), 19–35.

(43) Mero, a; Spolaore, B.; Pasut, G.; Veronese, F. M.; Fontana, a. **2009**, 384–389.

(44) Jeger, S.; Zimmermann, K.; Blanc, A.; Gr??nberg, J.; Honer, M.; Hunziker, P.; Struthers, H.; Schibli, R. *Angew. Chemie - Int. Ed.* **2010**, *49* (51), 9995–9997.

(45) Keillor, J. W.; Apperley, K. Y. P.; Akbar, A. *Trends Pharmacol. Sci.* **2015**, *36* (1), 32–40.

(46) Mironov, G. G.; Clouthier, C. M.; Akbar, A.; Keillor, J. W.; Berezovski, M. V. *Nat. Chem. Biol.* **2016**, *5* (September).

(47) Kerr, C.; Szmacinski, H.; Fisher, M.; Nance, B.; Lakowicz, J. R.; Akbar, A.; Keillor, J. W.; Toth, E. A.; Weber, D. J.; Eckert, R. L. *Oncogene* **20016**, *accepted* (May), 1–10.

(48) Lin, C. W.; Ting, A. Y. *J. Am. Chem. Soc.* **2006**, *128* (14), 4542–4543.

(49) Abe, H.; Goto, M.; Kamiya, N. *Chem. - A Eur. J.* **2011**, *17* (50), 14004–14008.

(50) Rabuka, D. **2011**, *14* (6), 790–796.

(51) De Graaf, A. J.; Kooijman, M.; Hennink, W. E.; Mastrobattista, E. *Bioconjug. Chem.* **2009**, *20* (7), 1281–1295.

(52) Tanaka, T.; Kamiya, N.; Nagamune, T. *FEBS Lett.* **2005**, *579* (10), 2092–2096.

(53) Kanda, T.; Sullivan, K. F.; Wahl, G. M. *Curr. Biol.* **1998**, *8* (7), 377–385.

(54) Beld, J.; Sonnenschein, E. C.; Vickery, C. R.; Noel, J. P.; Burkart, M. D. **2015**, *31* (1), 61–108.

(55) Ohtsuka, T.; Masafumi, O.; Noriki, N.; Motoki, M. 2000, pp 2608–2613.

(56) Rainer, H. B.; Bode-b, S. M. **2001**, No. 2, 79–99.

(57) Caja, S.; Mäki, M.; Kaukinen, K.; Lindfors, K. *Cell. Mol. Immunol.* **2011**, *8* (2), 103–109.

(58) Fisher, M. L.; Keillor, J. W.; Xu, W.; Eckert, R. L.; Kerr, C. *Mol. Cancer Res.* **2015**, *13* (7), 1083–1094.

(59) Mastroberardino, P. G.; Piacentini, M. *Erasmus* **2011**, *268* (5), 419–431.

(60) Wilhelmus, M. M. M.; de Jager, M.; Smit, A. B.; van der Loo, R. J.; Drukarch, B. *Sci. Rep.* **2016**, *6* (October 2015), 20569.

(61) Andringa, G.; Lam, K. Y.; Chegary, M.; Wang, X.; Chase, T. N.; Bennett, M. C.; Branch, E. T.; Neurosciences, B. R. *Control* **2004**, No. 3.

(62) 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. *Can. J. Chem.* **2008**, *86* (4), 271–276.

(63) Kieliszek, M.; Misiewicz, A. *Folia Microbiol. (Praha).* **2014**, *59* (3), 241–250.

(64) Mirzaei, M. *Food Eng.* **2011**, *9*, 2000–2004.

(65) Zotzel, J.; Keller, P.; Fuchsbauer, H.-L. *Eur. J. Biochem.* **2003**, *270* (15), 3214–3222.

(66) Kashiwagi, T.; Yokoyama, K. ichi; Ishikawa, K.; Ono, K.; Ejima, D.; Matsui, H.; Suzuki, E. ichiro. *J. Biol. Chem.* **2002**, *277* (46), 44252–44260.

(67) Rodrigues-Lima, F.; Delomenie, C.; Goodfellow, G. H.; Grant, D. M.; Dupret, J. M. *Biochem J* **2001**, *356* (Pt 2), 327–334.

(68) YURIMOTO, H.; YAMANE, M.; KIKUCHI, Y.; MATSUI, H.; KATO, N.; SAKAI, Y. *Biosci. Biotechnol. Biochem.* **2004**, *68* (10), 2058–2069.

(69) Rachel, N. M.; Pelletier, J. N. *Biomolecules* **2013**, *3*, 870–888.

(70) Kamiya, N.; Abe, H.; Goto, M.; Tsuji, Y.; Jikuya, H. *Org. Biomol. Chem.* **2009**, *7* (17), 3407–3412.

(71) Yang, M.-T.; Chang, C.-H.; Wang, J. M.; Wu, T. K.; Wang, Y.-K.; Chang, C.-Y.; Li, T. T. *J. Biol. Chem.* **2011**, *286* (9), 7301–7307.

(72) Sugimura, Y.; Yokoyama, K.; Nio, N.; Maki, M.; Hitomi, K. *Arch. Biochem. Biophys.* **2008**, *477* (2), 379–383.

(73) Hitomi, K.; Kitamura, M.; Sugimura, Y. *Amino Acids* **2009**, *36* (4), 619–624.

(74) Sugimura, Y.; Hosono, M.; Wada, F.; Yoshimura, T.; Maki, M.; Hitomi, K. *J. Biol. Chem.* **2006**, *281* (26), 17699–17706.

(75) Ohtsuka, T.; Ota, M.; Nio, N.; Motoki, M. *Bioscience biotechnology and biochemistry*. J-Stage 2000, pp 2608–2613.

(76) Grenard, P.; Bresson-Hadni, S.; El Alaoui, S.; Chevallier, M.; Vuitton, D. a; Ricard-Blum, S. *J. Hepatol.* **2001**, *35* (3), 367–375.

(77) Takahashi, N.; Takahashi, Y.; Putnam, F. W. *Proc. Natl. Acad. Sci. U. S. A.* **1986**, *83* (21), 8019–8023.

(78) Wan, X. H.; Lee, E. H.; Koh, H. J.; Song, J.; Kim, E. K.; Kim, C. Y.; Lee, J. B.; Kim, S.-Y.; Yao, K.; Lee, J. H. *Br. J. Ophthalmol.* **2002**, *86* (11), 1293–1298.

(79) Salter, N. W.; Ande, S. R.; Nguyen, H. K.; Nyomba, B. L. G.; Mishra, S. *J. Mol. Endocrinol.* **2012**, *48* (3), 203–216.

(80) Grosso, H.; Mouradian, M. M. *Pharmacol. Ther.* **2012**, *133* (3), 392–410.

(81) Chhabra, A.; Verma, A.; Mehta, K. *Anticancer Res.* **2009**, *29* (6), 1909–1919.

(82) Kumazawa, Y.; Ohtsuka, T.; Ninomiya, D.; Seguro, K. *Biosci. Biotechnol. Biochem.* **1997**, *61* (7), 1086–1090.

(83) Aluko, R. E.; Yada, R. Y. *J. Sci. Food Agric.* **1999**, *290* (June 1998), 286–290.

(84) Kanaji, T.; Ozaki, H.; Takao, T.; Kawajiri, H.; Ide, H.; Motoki, M.; Yasutsugu, S. *J. Biol. Chem.* **1993**, *268* (16), 11565–11572.

(85) Piredda, L.; Farrace, M. G.; Lo Bello, M.; Malorni, W.; Melino, G.; Petruzzelli, R.; Piacentini, M. *FASEB J.* **1999**, *13* (2), 355–364.

(86) Shi, Y.-G.; Qian, L.; Zhang, N.; Han, C.-R.; Liu, Y.; Zhang, Y.-F.; Ma, Y.-Q. *Molecules* **2011**, *16* (12), 10046–10058.

(87) Kressler, R. G.; Ma, K.; Besheer, A.; Hertel, T. C. *J. Pharm. Sci.* **2009**, *98* (4420), 4420–4428.

(88) Folk, J. E.; Cole, P. W. *J. Biol. Chem.* **1966**, *241* (23), 5518–5525.

(89) Lorand, L.; Campbell, L.; Robertson, B. *Biochemistry* **1972**, No. 11, 434–438.

(90) Day, N.; Keillor, J. W. *Anal. Biochem.* **1999**, *274* (1), 141–144.

(91) Qiao, S.; Piper, J.; Haraldsen, G.; Fleckenstein, B.; Molberg, Ø.; Øynebråten, I.; Khosla, C.; Sollid, L. M. *J. Immunol.* **2005**, *174* (3), 1657–1663.

(92) Mechanism, G. D.; Leblanc, A.; Gravel, C.; Labelle, J.; Keillor, J. W. *Biochemistry* **2001**, *40* (6), 8335–8342.

(93) Meiying, Z.; Guocheng, D.; Jian, C. *Enzyme Microb. Technol.* **2002**, *31* (4), 477–481.

(94) Marx, C. K.; Hertel, T. C.; Pietzsch, M. *Enzyme Microb. Technol.* **2007**, *40* (6), 1543–1550.

(95) Gundersen, M. T.; Keillor, J. W.; Pelletier, J. N. *Appl. Microbiol. Biotechnol.* **2013**, in press.

(96) Studier, F. W. *Protein Expr. Purif.* **2005**, *41* (1), 207–234.

(97) Folk, J. E.; Chung, S. I. L. *Methods Enzymol.* **1985**, *113* (1980), 358–375.

(98) Gross, M.; Folk, J. E. *J. Biol. Chem.* **1974**, *249* (10), 3021–3025.

(99) Ohtsuka, T.; Sawa, a; Kawabata, R.; Nio, N.; Motoki, M. *J. Agric. Food Chem.* **2000**, *48* (12), 6230–6233.

(100) Liang, F.; Holt, I.; Pertea, G.; Karamycheva, S.; Salzberg, S. L.; Quackenbush, J. *Nat. Genet.* **2000**, *25* (2), 239–240.

(101) Iwakura, M.; Nakamura, D.; Takenawa, T.; Mitsuishi, Y. *Protein Eng.* **2001**, *14* (8), 583–589.

(102) Tobergte, D. R.; Curtis, S. *J. Chem. Inf. Model.* **2013**, *53* (9), 1689–1699.

(103) Soh, N. *Sensors* **2008**, *8* (2), 1004–1024.

(104) Sunbul, M.; Yin, J. *Org. Biomol. Chem.* **2009**, *7* (17), 3361–3371.

(105) Heal, W. P.; Wickramasinghe, S. R.; Leatherbarrow, R. J.; Tate, E. W. *Org. Biomol. Chem.* **2008**, *6* (13), 2308.

(106) Strijbis, K.; Spooner, E.; Ploegh, H. L. *Traffic* **2012**, *13* (6), 780–789.

(107) Uttamapinant, C.; White, K. a.; Baruah, H.; Thompson, S.; Fernández-suárez, M.;

Puthenveetil, S.; Ting, a. Y. *Pnas* **2010**, *107*, 10914–10919.

(108) Oteng-Pabi, S. K.; Keillor, J. W. *Anal. Biochem.* **2013**, *441* (2), 169–173.

(109) Odii, B. O.; Coussons, P. *Sci. World J.* **2014**, *2014* (Table 1), 7–9.

(110) Gundersen, M. T.; Keillor, J. W.; Pelletier, J. N. *Appl. Microbiol. Biotechnol.* **2013**.

(111) Silhavy, T. J.; Bassford, P. J. J.; Beckwith, J. R. *J. Bacteriol.* **1979**, *139* (July), 19–31.

(112) Riggs, P. *Curr. Protoc. Mol. Biol.* **2001**, *Chapter 16* (1994), Unit16.6.

(113) Oteng-Pabi, S. K.; Keillor, J. W. *Anal. Biochem.* **2013**.

(114) Kim, D.; Herr, A. E. *Biomicrofluidics* **2013**, *7* (4), 1–47.

(115) Scouten, W. H.; Luong, J. H. T.; Stephen Brown, R. *Trends Biotechnol.* **1995**, *13* (5), 178–185.

(116) Rusmini, F.; Zhong, Z.; Feijen, J. *Biomacromolecules* **2007**, *8* (6), 1775–1789.

(117) Shemiakina, I. I.; Ermakova, G. V; Cranfill, P. J.; Baird, M. A.; Evans, R. A.; Souslova, E. A.; Staroverov, D. B.; Gorokhovatsky, A. Y.; Putintseva, E. V; Gorodnicheva, T. V; Chepurnykh, T. V; Strukova, L.; Lukyanov, S.; Zaraisky, A. G.; Davidson, M. W.; Chudakov, D. M.; Shcherbo, D. *Nat. Commun.* **2012**, *3*, 1204.

(118) Weiss, J. B.; Ray, P. H.; Bassford, P. J. *Proc. Natl. Acad. Sci. U. S. A.* **1988**, *85* (23), 8978–8982.

(119) Zhu, Y.; Rinzema, A.; Tramper, J.; Bol, J. *Appl Microbiol Biotechnol* **1995**, *44*, 277–282.

(120) Liu, S.; Zhang, D.; Wang, M.; Cui, W.; Chen, K.; Du, G.; Chen, J.; Zhou, Z. *Microb. Cell Fact.* **2011**, *10* (1), 1–7.

(121) Salis, B.; Spinetti, G.; Scaramuzza, S.; Bossi, M.; Saccani Jotti, G.; Tonon, G.; Crobu, D.; Schrepfer, R. *BMC Biotechnol.* **2015**, *15* (1), 84.

(122) Noda, S.; Miyazaki, T.; Tanaka, T.; Chiaki, O.; Kondo, A. *Biochem. Eng. J.* **2013**, *74*, 76–80.

(123) Zhu, Y.; Tramper, J. *Trends Biotechnol.* **2008**, *26* (10), 559–565.

(124) Pasternack, R.; Dorsch, S.; Otterbach, J. T.; Robenek, I. R.; Wolf, S.; Fuchsbauer, H. L. *Eur. J. Biochem.* **1998**, *257* (3), 570–576.

(125) Krieger, M.; Kay, L. M.; Stroud, R. M. *Journal of Molecular Biology*. 1974.

(126) Eaton, D.; Rodriguez, H.; Vehar, G. a. *Biochemistry* **1986**, *25* (2), 505–512.

(127) Bjork, I.; Jackson, C. M.; Jornvall, H.; Lavine, K. K.; Nordling, K.; Salsgiver, W. J. *J. Biol. Chem.* **1982**, *257* (5), 2406–2411.

(128) Coughlin, S. R. *Nature* **2000**, *407* (6801), 258–264.

(129) Tripathi, L. P.; Sowdhamini, R. *BMC Genomics* **2008**, *9*, 549.

(130) Huete-Pérez, J. A.; Engel, J. C.; Brinen, L. S.; Mottram, J. C.; McKerrow, J. H. *J. Biol. Chem.* **1999**, *274* (23), 16249–16256.

(131) Sato, H.; Hayashi, E.; Yamada, N.; Yatagai, M.; Takahara, Y. *Bioconjug. Chem.* **2001**, *12* (5), 701–710.

(132) Chudasama, V.; Maruani, A.; Caddick, S. *Nat. Chem.* **2016**, *8* (2), 114–119.

(133) Marx, C. K.; Hertel, T. C.; Pietzsch, M. *Enzyme Microb. Technol.* **2007**, *40* (6), 1543–1550.

(134) Lorand, L.; Graham, R. M. *Nat. Rev. Mol. Cell Biol.* **2003**, *4* (2), 140–156.

(135) Kobayashi, K.; Hashiguchi, K.; Yokozeki, K.; Yamanaka, S. *Biosci Biotechnol Biochem*. 1998, pp 1109–1114.

(136) Mcpherson, D. C.; Kim, H.; Hahn, M.; Wang, R.; Grabowski, P.; Eichenberger, P.; Driks, A. *J. Bacteriol.* **2005**, *187* (24), 8278–8290.

(137) Plácido, D.; Fernandes, C. G.; Isidro, A.; Carrondo, M. A.; Henriques, A. O.; Archer, M. *Protein Expr. Purif.* **2008**, *59* (1), 1–8.

(138) Liu, Y.; Lin, S.; Zhang, X.; Liu, X.; Wang, J.; Lu, F. *J. Ind. Microbiol. Biotechnol.* **2014**, *41* (8), 1227–1235.

(139) Ragkousi, K.; Setlow, P. *J. Bacteriol.* **2004**, *186* (17), 5567–5575.

(140) Agyare, K. K.; Damodaran, S. *J. Agric. Food Chem.* **2010**, *58* (3), 1946–1953.

(141) Fernandes, C. G.; Plácido, D.; Lousa, D.; Brito, J. A.; Isidro, A.; Soares, C. M.; Pohl, J.; Carrondo, M. A.; Archer, M.; Henriques, A. O. *Biochemistry* **2015**, *54* (37), 5723–5734.

(142) Brömme, D. *Curr. Protoc. Protein Sci.* **2001**, *Chapter 21*, Unit 21.2.

(143) Lee, J.-H.; Song, E.; Lee, S.-G.; Kim, B.-G. *Biotechnol. Bioeng.* **2013**, *110* (11), 2865–2873.

(144) Lee, J.-H.; Song, C.; Kim, D.-H.; Park, I.-H.; Lee, S.-G.; Lee, Y.-S.; Kim, B.-G. *Biotechnol. Bioeng.* **2013**, *110* (2), 353–362.

(145) DØrum, S.; Arntzen, M. O.; Qiao, S. W.; Holm, A.; Koehler, C. J.; Thiede, B.; Sollid, L. M.; Fleckenstein, B. *PLoS One* **2010**, *5* (11).

(146) Van Den Akker, J.; Van Weert, A.; Afink, G.; Bakker, E. N. T. P.; Van Der Pol, E.; Böing, A. N.; Nieuwland, R.; VanBavel, E. *Amino Acids* **2012**, *42* (2–3), 961–973.

(147) Cornelius Klöcka, Xi Jina, Kihang Choia, Chaitan Khoslaa, B.; Madridc, P. B.; Andrew; Spencerd, Brian C. Raimundoe, Paul Boardmane, Guido Lanzae, and J. H. G. *Bioorg Med Chem Lett.* **2011**, *21* (9), 2692.

(148) Dean, K. M.; Palmer, A. E. *Nat. Chem. Biol.* **2014**, *10* (7), 512–523.

(149) Lam, A. J.; St-pierre, F.; Gong, Y.; Marshall, J. D.; Cranfill, P. J.; Baird, M. a; Mckeown, M. R.; Wiedenmann, J.; Davidson, M. W.; Schnitzer, M. J.; Tsien, R. Y.; Lin, M. Z. *Nat. Methods* **2012**, *9* (10), 1–8.

(150) Oteng-pabi, S. K.; Pardin, C.; Keillor, J. W.; Pabi, S. K. O. **2013**, No. November.

(151) Wu, C.-H.; Liu, I.-J.; Lu, R.-M.; Wu, H.-C. *J. Biomed. Sci.* **2016**, *23* (1), 8.

(152) Szardenings, M. *J. Recept. Signal Transduct. Res.* **2003**, *23* (4), 307–349.

(153) Halim, D.; Caron, K.; Keillor, J. W. *Bioorganic Med. Chem. Lett.* **2007**, *17* (2), 305–308.

(154) Lerner, A.; Matthias, T. *Int. J. Celiac Dis.* **2015**, *3* (1), 1–6.

(155) Progatzky, F.; Dallman, M. J.; Lo Celso, C. *Interface Focus* **2013**, *3* (3), 20130001.

(156) Spirig, T.; Weiner, E. M.; Clubb, R. T. *Mol. Microbiol.* **2011**, *82* (5), 1044–1059.

(157) Fujii, T.; Shindo, Y.; Hotta, K.; Citterio, D.; Nishiyama, S.; Suzuki, K.; Oka, K. *J. Am. Chem. Soc.* **2014**, 1–12.

(158) Dutton, a; Singer, S. J. *Proc. Natl. Acad. Sci. U. S. A.* **1975**, *72* (7), 2568–2571.

(159) Suzuki, T.; Matsuzaki, T.; Hagiwara, H.; Aoki, T.; Takata, K. *Acta Histochem. Cytochem.* **2007**, *40* (5), 131–137.

(160) Monroe, A.; Setlow, P. *J. Bacteriol.* **2006**, *188* (21), 7609–7616.

(161) Strop, P. *Bioconjug. Chem.* **2014**, *25* (5), 855–862.

(162) Holmes-Farley, S. R.; Whitesides, G. M. *Langmuir* **1986**, *2* (3), 266–281.

(163) Josten, A.; Haalck, L.; Spener, F.; Meusel, M. *J. Immunol. Methods* **2000**, *240* (1), 47–54.

(164) Dennler, P.; Chiotellis, A.; Fischer, E.; Brégeon, D.; Belmant, C.; Gauthier, L.; Lhospice, F.; Romagne, F.; Schibli, R. *Bioconjug. Chem.* **2014**, *25* (3), 569–578.

(165) Lim, E.-K.; Haam, S.; Lee, K.; Huh, Y.-M. *Bioconjugation protocols*; 2011; Vol. 751.

List of Publications

Oteng-Pabi, S.K., Cloutier, C. and Keillor, J. W. Identification and characterization of bTG substrates using FRET-based assay. (manuscript in preparation to be submitted shortly)

Oteng-Pabi, S. K., Pardin, C., Stoica, M. and Keillor J. W. Site-specific labelling and immobilization, mediated by microbial transglutaminase. *Chemical Communications* **50,** 6604-6606 (2014).

Oteng-Pabi, S.K. and Keillor, J. W. Continuous enzyme-coupled assay for microbial transglutaminase activity. *Analytical Biochemistry* **441,** 169-173 (2013).