MUTAGENESIS OF MICROBIAL TRANSGLUTAMINASE FOR QTAG-KTAG CONJUGATION

CHAN SOO KHIM

UNIVERSITI SAINS MALAYSIA
2018

MUTAGENESIS OF MICROBIAL TRANSGLUTAMINASE FOR QTAG-KTAG CONJUGATION

by

CHAN SOO KHIM

Thesis submitted in fulfilment of the requirements
for the Degree of
Doctor of Philosophy

May 2018

ACKNOWLEDGEMENT

First and foremost, I would like to express my deepest appreciation to my supervisor, Assoc. Prof. Dr. Lim Theam Soon, for his patient guidance, encouragement and advice provided throughout my time as his student. His guidance has helped me throughout my research and during the writing of this thesis. I am extremely lucky to have a supervisor who cares about my work so much and willing to spend his time to response to my questions and queries so promptly.

Also, my sincere gratitude to my co-supervisor Assoc. Prof. Dr. Choong Yee Siew for her guidance in bioinformatics works. Her guidance and teachings has led me down on a new area which I had never learned before. I would also like to thank all the INFORMM science officers, staffs and my labmates (QiuTing, Chai Fung, Mimie, Miza, Angela, Chia Chiu, and Jing Yi) who always offered me help whenever I need. Not to be forgotten, my sincere thanks to Jia Xin and Yee Ying from the bioinformatics lab who helped me a lot in bioinformatics works.

I would also like to thank USM Fellowship for supporting my study by covering tuition fees and also providing monthly stipend. This work could be completed on time without the concern of financial burden due to this scholarship.

Last but not least, I must express my gratitude to my family and friends for their continued support and encouragement thru all the ups and downs of my research. Completing this work would have been more difficult if they were not there for me.

TABLE OF CONTENTS

CHAPTER 5-CONCLUSION

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

2x YT	Two times yeast extract tryptone
3D	Three-dimensional
ABTS	2,2′-Azino-bis(3-ethylbenzthiazoline-6-sulfonic acid)
ADC	Antibody-drug conjugate
AP	Alkaline phosphatase
Asp	Aspartic acid
bp	Base pair
BSA	Bovine serum albumin
CBZ-Gln-Gly	Benzyloxycarbonyl-L-Glutaminylglycine
CN/DAB	4-chloro-1-naphthol/3,3'-diaminobenzidine tetrahydrochloride
Cys	Cysteine
dH_2O	Distilled water
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
dsDNA	Double-stranded DNA
ECL	Enhanced chemiluminescence
E. coli	*Escherichia coli*
eGFP	Enhanced green fluorescence protein
ELISA	Enzyme-linked immunosorbent assay
Glu	Glutamine
His	Histidine
HRP	Horseradish peroxidase
IPTG	Isopropyl β-D-1-thiogalactopyranoside
KalbTG	*Kutzneria albida* Transglutaminase

Ktag	Lysine tag
Lys	Lysine
mTGase	Microbial transglutaminase
NHS	N-Hydroxysuccinimide
OD	Optical density
OE-PCR	Overlap-extension PCR
o/n	Overnight
PBS	Phosphate-buffered saline
PBST	Phosphate-buffered saline with tween
PCR	Polymerase chain reaction
pNPP	ρ-Nitrophenyl phosphate
Qtag	Glutamine tag
RMSD	Root-mean-square deviation
RT	Room temperature
SAS	Solvent-accessibility surface area
S. mobaraense	*Streptomyces mobaraense*
scFv	Single-chain fragment variable
SDM	Site-directed mutagenesis
SDS-PAGE	SDS-polyacrylamide gel electrophoresis
TAMEP	Transglutaminase activating metalloprotease
TBS	Tris-buffered saline
TBST	Tris-buffered saline with tween
TGase	Transglutaminase
WT	Wild type

LIST OF SYMBOLS

°C	Degree Celcius
%	Percent
g	Gram
h	Hour
g	Gravity force
kDa	KiloDalton
L	Liter
µg	Microgram
µL	Microliter
M	Molar
mg	Milligram
min	Minute
mL	Milliliter
mM	Millimolar
ng	Nanogram
rpm	Revolution per minute
U	Unit of enzyme
v/v	Volume/volume
w/v	Weight/volume

MUTAGENESIS TRANSGLUTAMINASE MIKROBIAL

UNTUK KONJUGASI QTAG-KTAG

ABSTRAK

Penemuan awal transglutaminase mikrobial (mTGase) adalah untuk mengurangkan kos tinggi transglutaminase mamalia dalam aplikasi industri pemprosesan makanan. Enzim ini telah menjadi alatan penting dalam aplikasi bioteknologi disebabkan sifat tahan lasak, ketulenan yang tinggi, tidak memerlukan ion kalsium untuk pengaktifan dan mempunyai kadar tindak balas yang tinggi. Akan tetapi, mTGase mempunyai kekhususan substrat yang rendah disebabkan celah tapak aktif yang lebar pada enzim telah membantutkan aplikasinya dalam bidang bioteknologi. Oleh itu, mutagenesis pada mTGase untuk mendapatkan enzim yang substrat spesifik bagi konjugasi spesifik amat diidamkan. Ketiga-tiga residu yang penting untuk pengecaman substrat (V65, W69, and Y75) dimutasi kepada 20 jenis asid amino untuk pembinaan perpustakaan mutan pertama. Mutasi pada mutan aktif di ketiga-tiga residu digabung kepada semua gabungan yang mungkin untuk membina perpustakaan mutasi bergabung. Mutan MTG 120 yang memperolehi aktiviti spesifik yang tertinggi diguna untuk mengasingkan peptida yang mengandungi glutamin dengan sebuah perpustakaan faj peptida 20-mer NNK. MTG 120 dengan mutasi leusina (L), metionina (M), dan asid glutamik (E) mempamerkan aktiviti spesifik yang paling tinggi. Selain itu, MTG 120 menunjukkan pengurangan yang ketara dalam paut silang secara rawak dengan substrat mTGase seperti albumin serum lembu, dan fosfatase alkali hasilan makmal. Pengikat afiniti tinggi kepada MTG 120 iaitu faj peptida P58 berjaya diasingkan melalui proses '*biopanning*'. Motif RVGQL yang berasal daripada peptida P58 yang diasingkan bercantum dengan fosfatase alkali

hasilan makmal untuk konjugasi dengan scFv antibodi Ubiquitin yang mempunyai tag penta-lisina. Konjugasi antara kedua-dua protein itu berjaya menghasilkan jalur yang bersaiz kira-kira 85 kDa di Blot Barat. Produk konjugasi tersebut juga menunjukkan pengekalan fungsi dengan penetapan kadar immunosorben taut-enzim (ELISA). Mutagenesis mTGase bersamaan dengan pengasingan peptida yang spesifik dengan mutan telah berjaya memperbaiki kekhususan substrat mTGase untuk aplikasi dalam bidang bioteknologi.

MUTAGENESIS OF MICROBIAL TRANSGLUTAMINASE

FOR QTAG-KTAG CONJUGATION

ABSTRACT

The initial discovery of microbial transglutaminase (mTGase) was to soothe the high cost of mammalian transglutaminase in food processing industries. This enzyme has become an important tool in biotechnological applications due to its robustness, availability in high purity, bypass the need of calcium ions for activation and high reaction rate. However, mTGase has low substrate specificity due to its wide active site cleft which impeding its diverse applications in biotechnological use. Therefore, mutagenesis of mTGase to obtain substrate specific enzyme is highly desirable for site-specific conjugations. Three residues (V65, W69, and Y75) that are crucial for substrate recognition were mutated to all 20 amino acids to generate the first mutant library. Mutations at respective sites of the active mutants were then combined at all possible combinations to construct the combinatorial mutant library. MTG 120 mutant with the highest specific activity was used to pan against a 20-mer NNK peptide phage library to isolate glutamine-containing peptide. MTG 120 with mutations of leucine (L), methionine (M), and glutamic acid (E) at residue 65, 69, and 75, respectively, exhibited the highest specific activity. Also, MTG 120 showed significant reduction in random cross-linking with mTGase substrates including bovine serum albumin, and in-house alkaline phosphatase. A high affinity binder against MTG 120, peptide phage P58 has been successfully isolated through biopanning process. The motif RVGQL derived from the isolated peptide P58 was fused with in-house alkaline phosphatase to perform conjugation with penta-lysine tag anti-Ubiquitin scFv. The two proteins were successfully showed conjugation in

Western blot to give a band around 85 kDa. The conjugated product also showed retained of functionality in enzyme-linked immunosorbent assay (ELISA). Mutagenesis of mTGase coupled with selection of mutant specific peptide has successfully improved the substrate specificity of mTGase for applications in biotechnological fields.

CHAPTER 1

INTRODUCTION

1.1 Literature Review

The inherent ability to bring molecules together and allowing them to stay with each other requires many factors to be taken into consideration. The design and modifications required to achieve such as feat is an art form on its own. In the field of medical biotechnology, the target molecules are usually either DNA or proteins. The biomolecules are unique in the way they function, fold and regulate mechanisms in biological systems. However, the advent of recombinant DNA technology has allowed these molecules to serve a greater function via evolutionary reprogramming. The move towards "greener" approaches when dealing with biomolecules is critical in this aspect as harsh modification strategies risk the disruption and destruction of the conjugated biomolecules in terms of stability and function. Therefore protein conjugation is an area of study which is critical to supplement the ever evolving field of medical biotechnology.

1.1.1 Protein Conjugation

Conjugation is a process of joining independent molecules together covalently to achieve a hybrid molecule with combined attributes of each individual components. This is done to attain better functional properties including biocompatibility, biostability, and bioselectivity of the hybrid molecule for various biomedical applications (Koniev and Wagner, 2015). Protein and peptide conjugates especially have been intensively studied due to their high potentials in the biomedical field (van Hest, 2017). Various conjugation strategies have been developed over the years for protein conjugations to generate well-defined conjugates (Boutureira and Bernardes,

2015). Generally, protein conjugation strategies can be divided to two major groups: chemical conjugation, and enzymatic conjugation.

1.1.1 (a) Chemical Conjugation

Chemical protein conjugation methods generally employ lysine (Lys) or cysteine (Cys) residues due to their high nucleophilicity in the side chains (Chalker et al., 2009). Lysine has an accessible primary amine group to act as a good nucleophile for a broad range of organic reactions (Basle et al., 2010). The nucleophilicity of the amine group of Lys residue to react with electrophiles to form a bond with amine group is higher and faster than other amino acids (Brun and Gauzy-Lazo, 2013). Also, the abundance of Lys residues on the protein surface makes them sufficiently exposed or accessible for reaction (Brun and Gauzy-Lazo, 2013, Sesay, 2003). N-hydroxyl-succinimidyl (NHS) ester is the common reagent used in Lys conjugation to form amide bonds between the carboxylic acid and amino group due to its simplicity and availability (Basle et al., 2010).

The Cys residue is another common site targeted for protein conjugation owing to its high nucleophilicity at its thiol (-SH) side chain (Brotzel and Mayr, 2007). At pH levels below 9, amine gets protonated and the thiol group becomes more nucleophilic than amine, resulting in a faster reaction in Cys over Lys residues (Basle et al., 2010). Albeit taking part in most of the functional properties, Cys is low in abundance in proteins which allows better manipulation in conjugation compared to Lys residues (Kim et al., 2015, Koniev and Wagner, 2015). Pre-treatment with dithiothreitol (DTT) is required to free the thiol groups for tagging because thiol groups are often found in disulphide form (Crankshaw and Grant, 2001).

Notwithstanding the promising results of chemical conjugation over the years, many more setbacks regarding this group of conjugation techniques are being reported. For example, Lys conjugation requires high pH values to avoid protonation of the amine group which greatly decrease its reactivity. Protonated amine groups under physiological pH demonstrates no significant nucleophilicity compared to other side chain groups (Koniev and Wagner, 2015). That being said, chemical conjugation might not be ideal for alkaline sensitive proteins (Basle et al., 2010). Often, chemical conjugations cause heterogeneity in products due to the presence of multiple Lys or Cys sites distribute in protein (Dennler et al., 2015b). Presence of heterogeneous products is particularly unfavourable in antibody-drug conjugates (ADCs). This is because it can increases the likelihood for unconjugated antibodies to compete with conjugated antibodies for antigen binding sites which can weaken the therapeutic index (Junutula et al., 2008). Moreover, NHS esters have low specificities due to their reactions with random amino acids to form labile bonds which produce undesired side products (Chih et al., 2011). Despite the widespread use of chemical conjugation, emphasis on production of homogeneous conjugated products has been growing. Especially in the biopharmaceutical industry, payload distribution is exceptionally essential for conjugate stability and efficacy (Sochaj et al., 2015). In order to complement existing chemical conjugation approaches, efforts have been driven to address these problems with alternatives such as enzymatic conjugation.

1.1.1 (b) Enzymatic Conjugation

Evolution of Chemical Biology over the decades has seen to increased use of enzymes to form defined covalent bonds leading to the production of compounds with novel compositions (McFarland and Rabuka, 2015). The advantage of this approach is that further modification in natural amino acid residues for enzymatic conjugation is not required. The introduction of an enzyme recognizable motif with the target protein for co-expression allows conjugation to take place subsequently (Agarwal and Bertozzi, 2015, Appel and Bertozzi, 2014). With the aid of recombinant technology, the required motif could be easily introduced to the target protein by cloning without hassle (Dennler et al., 2015b).

Enzymes like sortase A (srtA), transglutaminase (TGase) and formylglycine-generating enzyme (FGE) have shown promising applications in mediating enzymatic protein conjugation for various industrial applications. TGase is unique in comparison to srtA and FGE as the later two enzymes are capable of recognising a specific motif for conjugation but not TGase. Sortase A (EC 3.4.22.70) is a calcium-dependent transpeptidase which recognises LPXTG motif where 'X' can be any of the naturally occurring amino acids. As shown in Figure 1.1A, the thiol group of srtA (Cys^{184}) attacks the amide bond between threonine (T) and glycine (G) in the LPXTG motif by cleaving it to form thioacyl intermediate.

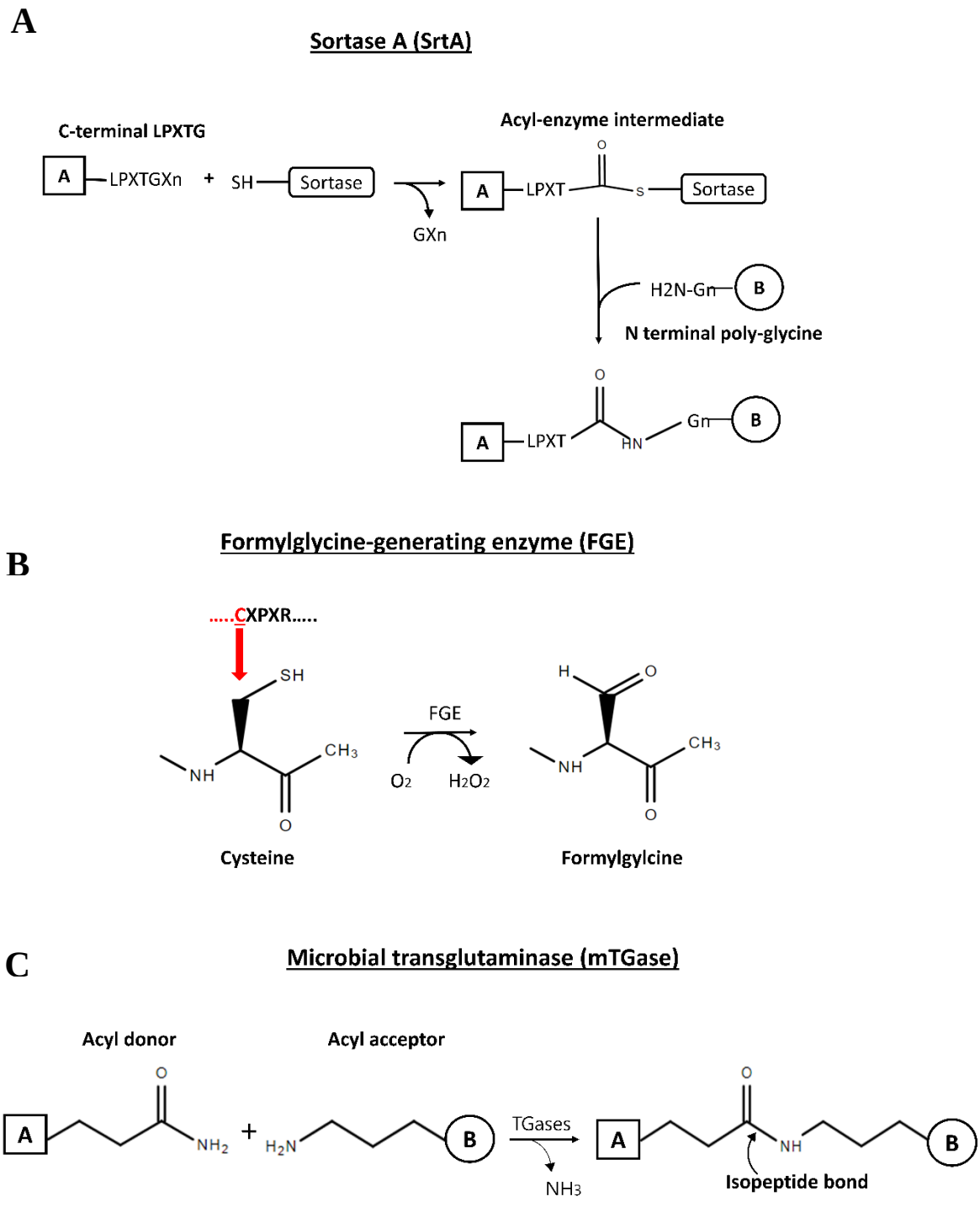

Figure 1.1: Enzymatic conjugation approaches. (A) Sortase A (SrtA). Protein A has LPXTGXn motif near the C-terminal to form intermediate with Sortase A. Then, the acyl-enzyme intermediate was resolved by nucleophile poly-glycine tagged at N – terminal of Protein B to release the enzyme and conjugated product. (B) Formylglycine-generating enzyme (FGE) oxidises the cysteine to formylglycine (FGly). (C) Microbial transglutaminase (mTGase). Isopeptide bond was formed between acyl donor (Protein A) and amine group of acyl acceptor (Protein B) which catalysed by mTGase.

A second wave of nucleophilic attack by the glycine residues in the poly-glycine linker found in the coupling partner releases the intermediate to generate a new amide bond between the LPXTG bound target and poly-glycine tagged partner (Mazmanian et al., 1999, Schumacher et al., 2016). FGE (E.C. 1.8.99.-) on the other hand oxidises a cysteine residue in the conserved sequence of CXPXR found in eukaryotes or serine residue in SXPXR of prokaryotes to an active formylglycine (FGly) (Figure 1.1B). The 'X' in the conserved motifs are recognisable either as serine, threonine, alanine or glycine (Carrico et al., 2007). Introduction of the FGE recognisable motif into a desired protein allows the protein to co-express with the aldehyde tag which can be used for conjugation using aminooxy or hydrazine-functionalised reagents (Rashidian et al., 2013). Different from srtA and FGE, TGase is able to cross-link proteins randomly without the need of a consensus sequence making them extensively studied (Milczek, 2017).

The term transglutaminase (TGase) was first coined in year 1959 to describe the transamidating activity observed in guinea-pig liver (Clarke et al., 1959). Transglutaminases (protein-glutaminase γ-glutamyltransferase EC 2.3.2.13) are transferases that catalyse the transfer of acyl groups between γ-carboxyamide groups in glutamine (Glu) residues and primary ε-amino groups in residues such as lysine (Lys). Glutamine residues act as acyl donors and various primary amines acts as acyl acceptors to form covalent isopeptide bonds between these two residues which results in protein cross-linking (Figure 1.1C) (Griffin et al., 2002, Ohtsuka et al., 2000a).

TGases are widely found in different taxonomic groups such as microorganisms (Ando et al., 1989), plants (Del Duca et al., 2014), animals (Folk and Cole, 1966) and humans (Suedhoff et al., 1990). Generally, TGases require calcium (Ca^{2+}) ions for activation. Upon binding, Ca^{2+} ions function to expose two key residues near the active site which are vital for substrate access (Ahvazi et al., 2002). In animals or humans, TGases are inactive under normal physiological conditions and can only be activated when Ca^{2+} ions bind during the disruption of physiological homoeostatic mechanisms (Griffin et al., 2002). However, the requirement of Ca^{2+} ions for activation in plants are not absolute (Lorand and Graham, 2003, Serafini-Fracassini et al., 1995). On the other hand, microbial transglutaminse (mTGase) is a calcium-independent TGase which was first isolated from *Streptomyces mobaraensis* (formerly known as *Streptoverticillium mobaraense*) by a researcher from Ajinomoto Co., Inc (Ando et al., 1989). Different bacterial strains including *Streptomyces sp*. have been studied for the production of mTGase. However, *Streptomyces sp*. has significantly better yield compared to other strains which makes it the most preferred source in the industry to produce mTGase (Zhang et al., 2009).

1.1.2 Microbial Transglutaminase

1.1.2 (a) Characterisations of mTGase

Recombinant mTGase are commonly expressed in a zymogen form (pro-MTG) to avoid detrimental internal cross-linking of essential cytosolic proteins (Takehana et al., 1994). A 45-residue pro-sequence (DNGAGEETKSYAETYRLTADDVANINALNESAPAASSAGPSFRAP) located at the N-terminal of mTGase is reported to fold into an α-helix adopting a L-shape conformation which covers the enzyme active site as shown in Figure 1.2 (Yang et al., 2011).

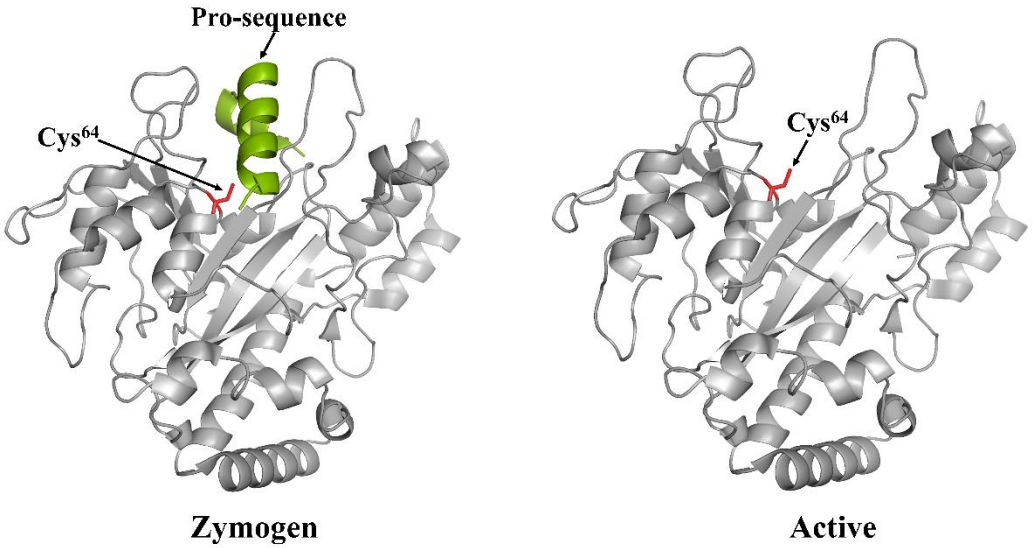

Figure 1.2: Crystal structure of mTGase from *S. mobaraensis*. (Left) PDB ID 3IU0: The 45-residue pro-sequence is fold into an α-helix covering the active site (Cys^{64}) (Yang et al., 2011). (Right) PDB ID 1IU4: Active form of mTGase exposing the active site (Cys^{64}) that is vital for activity (Kashiwagi et al., 2002).

The pro-sequence is vital for enzyme folding and inhibition of enzyme activation within the cells. Removal of the pro-sequence using proteases reveal the active site of the enzyme to initiate the function. This allows for a turn-on / turn-off switch to be introduced via enzyme treatment. Examples of proteases reported to activate pro-MTG include trypsin, dispase, chymotrypsin, transglutaminase activating metalloprotease (TAMEP), thrombin, cathepsin B, and proteinase K (Eder and Fersht, 1995, Pasternack et al., 1998, Marx et al., 2008a).

The complete amino acid sequence of the activated mTGase derived from *Streptomyces sp.* has been characterised to consist of 331 amino acids with a molecular weight of approximately 38 kDa with an isoelectric point of 8.9 (Kanaji et al., 1993). It is reported to be a monomeric, simple protein, with no known glycosylation or lipids (Yokoyama et al., 2004). Also, mTGase exhibits good stability over a wide range of temperature and pH compared to other sources of TGases (Yokoyama et al., 2004). Up till today, two crystal structures of mTGase from *Streptomyces mobaraensis* have been deposited in Protein Data Bank (PDB) including PDB ID: 1U14 (Kashiwagi et al., 2002), and PDB ID: 3IU0 (Yang et al., 2011). Microbial transglutaminase is suggested to have a novel three-dimensional structure due to the fact that no similar proteins to mTGase could be found. Both crystal structures showed mTGase folds into a plate-like shape with a deep cleft at the edge of the molecule which is distinct from factor XIII-like TGases that adopts protease-like catalytic triad structures (Kashiwagi et al., 2002, Yang et al., 2011).

Structurally, a single cysteine (Cys^{64}) residue has been identified in mTGase residing at the bottom cleft of the crystal structure. The Cys^{64} comes in proximity with aspartic acid (Asp^{255}) and histidine (His^{274}) residues to form the active site which superimposes well on a catalytic triad (Kashiwagi et al., 2002). Despite the catalytic

triad of mTGase being able to superimpose well with factor XIII-like TGases, studies showed mTGase are capable of generating higher reaction rates (Shimba et al., 2002). This is associated to the Cys^{64} residue of mTGase which is exposed to surrounding solvent allowing the reaction of mTGase with substrates to promptly occur. In addition to that, the flexibility of the right side wall of the active site cleft reduces the steric hindrance between the substrates and enzyme allowing improved accessibility (Kashiwagi et al., 2002). Another exciting attribute of mTGase is the broad substrate specificity for acyl donors. This unique attribute can be explained by the crystal structure of mTGase which reveals the wide active-site cleft position that accommodates the α-helix pro-sequence. The wide size of the active-site allows for additional flexibility in the active-site to accommodate a less specific substrate (Shimba et al., 2002). This explains the broad specificity of mTGase to react with a large family of acyl donors. This flexibility and efficacy has allowed mTGase to contribute efficiently to many different applications in various industries making it a vital enzyme in industry.

1.1.2 (b) Applications of mTGase

The low substrate specificity of mTGase allows it to effectively cross-link most of the proteins it comes in contact with post-activation. This unique feature of mTGase has been exploited in various industries such as food, textile, and leather processing to improve their functional properties (Mariniello and Porta, 2005). However, bioavailability of isopeptide moieties in native proteins and protein structures will affect the affinity of mTGase to these proteins (Matsumura et al., 1996). In the meat industry, mTGase is used to restructure meat by mixing mTGase with caseinate as a

substituent to salts and phosphates that can be used in conventional meat binding and texturing. When reacted with mTGase, caseinate becomes viscous allowing the meat pieces to stick together. Restructuring meat with different types of meat offer better texture as compared to one-meat product due to the availability of isopeptide bonds in different meats (Zhang et al., 2009). In dairy products and soy proteins, mTGase is applied to improve characteristics such as gelling strength and viscosity (Gan et al., 2009b, Kuraishi et al., 2001). In wool and textile industries, mTGase is used to recover the damage of wool caused by proteolytic treatment of protease in biological detergents. Also, it is used to enhance the tensile strength of wool fibers. In leather processing industries, mTGase is applied in the filling step where the voids in the fibers of leather can be sealed (Tesfaw and Assefa, 2014).

The early discovery of mTGases was meant to replace the animal derived TGases in food industries. Since then, mTGase has surged on to become an impeccable tool in the biotechnology field. Many publications showed the utilization of mTGase in conjugation processes for biotechnological applications (Dennler et al., 2014, Lin and Ting, 2006, Spolaore et al., 2016, Strop et al., 22 Nov 2012). A recent publication summarized the research and biotechnological applications of mTGase for the conjugation of proteins with proteins, DNA, polymers, and radioisotopes (Strop, 2014).

Conjugation of different moieties together using mTGase mostly depends on the glutamine-tag (Qtag) and lysine-tag (Ktag) since mTGase does not show any preference towards specific peptide sequences. As long as both lysine (K) and glutamine (Q) residues are accessible to mTGase, they could serve as potential substrates for protein conjugation to occur (Coussons et al., 1992). A variety of Qtag DNA conjugation with Ktag protein by mTGase has been reported over the years (Takahara et al., 2013, Tominaga et al., 2007). Also, various other molecules could be

conjugated with proteins like antibodies simply by the introduction of a Ktag and Qtag which are recognized by mTGase (Kamiya et al., 2003a, Lee et al., 2013, Lin and Ting, 2006). Qtag single-chain fragment variable (scFv) was successfully fused to a Ktag enhanced yellow fluorescent protein (EYFP) to yield a heterodimer (scFv-EYFP fusion protein) product by mTGase. Validation of this fusion protein with fluorescent immunoassay (FIA) confirmed the functionality of both proteins was intact post-conjugation (Kamiya et al., 2003a). In a separate study, mTGase conjugation demonstrated higher specificity compared to chemical modifications for immobilisation of Ktag alkaline phosphatase onto casein-coated polyacrylic resin. This technique created a stable functional protein array which is crucial for biotechnological applications (Tominaga et al., 2004a). The versatility of mTGase in conjugation was challenged when myc-tag (EQKLISEEDL) was exploited as the acyl donor to conjugate various moieties including antibodies, biotin, fluorescent dye, radioisotopes, chemical functional group and surface. The group successfully conjugated all tested functionalities which could be potentially used for various applications (Dennler et al., 2015a).

The ability of mTGase to conjugate proteins easily has been harnessed in the biopharmaceutical industry to produce antibody drug conjugates (ADCs) (Dorywalska et al., 2015, Farias et al., 2014, Strop et al., 2016). However, researchers observed mTGase only recognize glutamine residues (Q^{295}) in constant regions of aglycosylated or deglycosylated antibodies but not those in glycosylated antibodies (Jeger et al., 2010, Mindt et al., 2007, Strop et al., 22 Nov 2012). Presence of N-glycan at N^{297} hinders the transamination of mTGase due to steric hindrance. Removal of N-glycan at N^{297} using Peptide-N-Glycosidase (PNGase F) allows Q^{295} more accessibility to mTGase and enhances the flexibility of the loop where the residue is located (Jeger et al., 2010).

The work was extended to conjugate different moieties including antimitotic toxin monomethyl auristatin E (MMAE) and radioactive substrates (^{67}Ga and ^{89}Zr) to antibodies using the similar strategy (Jeger et al., 2010, Dennler et al., 2014). Later, researchers developed a glutamine tag (LLQG) that can be engineered at desired locations of the antibody after performing a systematic scan of the constant domains. The introduction of amine containing linkers would allow mTGase to conjugate various probes and drugs including fluorophores and potent tubulin inhibitor (MMAD) to antibodies with similar efficiencies across all antibody subtypes (Strop et al., 2013). Despite the low substrate specificity of mTGase, applications of mTGase in various fields have conferred the advantages of mTGase based conjugations.

1.1.2 (c) Advantages of mTGases Based Conjugations

Microbial transglutaminase provides a convenient approach for protein conjugation without major modifications to the natural amino acid residues. Peptide tags can be cloned directly with the protein of interest for co-expression prior to conjugation (Agarwal and Bertozzi, 2015, Appel and Bertozzi, 2014). Moreover, mTGase recognisable motifs are usually short (about 5-6 amino acids) which will not interfere with the expression, purification and function of the proteins. Complications regarding immunogenicity triggered by peptides could also be minimized (Hagemeyer et al., 2015). As mentioned earlier, mTGase provides a promising alternative for conjugation because it does not interfere with antigen binding. Also, the mild conjugation condition of mTGase helps to reserve the intact functionalities of the conjugated proteins (Kamiya and Mori, 2015, Swee et al., 2013, Wu et al., 2009). This is an added advantage over chemical conjugation methods as the use of glutaraldehyde

has been reported to impair antigen conformation when used at a higher percentages (Chan and Lim, 2016). However, this is not a concern when conjugating proteins using mTGase (Strop, 2014).

Conjugation of mTGase is irreversible unlike srtA which is also another favourite workhorse for protein conjugation. Despite its high specificity, the reversible reaction of srtA requires the need for a higher enzyme concentration for efficient conjugation (Chen et al., 2011, Rashidian et al., 2013). Other advantages of mTGase is the lower cost of production (Yurimoto et al., 2004) and its ability to avoid red pigmentation in food products unlike animal derived TGases. Red pigmentation of animal TGases causes detrimental changes to the appearance of the product making it unsuitable for sale (Yokoyama et al., 2004, Zhang et al., 2009). Furthermore, mTGase exhibits a higher activity which is crucial for large scale production and enzymatic conjugation (Steffen et al., 2017). Importantly, mTGase have broad substrate specificity, calcium independence, and is stable over a wide range of temperature and pH making it ideal for the conjugation of various substrates (Kuraishi et al., 2001, Zhu et al., 1995). The broad substrate specificity of mTGases is very useful in food and textile industries to catalyse random amide bond formation (Gundersen et al., 2014). However, it is a huge challenge when site-specific conjugation is desired since mTGases randomly cross-links with non-target substrates (Steffen et al., 2017, Strop, 2014). In order to overcome these issues, various approaches including mutagenesis and phage display biopanning were then taken to address the substrate preference of mTGases with the hopes of improving its specificity and functionality (Sugimura et al., 2008, Yokoyama et al., 2010).

1.1.3 Enzyme Engineering

1.1.3 (a) Strategy and Techniques for Enzyme Engineering

In nature, enzymes naturally aid in sustaining the development of organisms in a natural environment. Therefore, natural enzymes usually are not equipped with adequate properties to be utilised for biotechnological applications (Fernandez-Gacio et al., 2003). The generation of enzymes with desired properties like thermostability, bioactivity, and specificity has ushered in the development of enzyme engineering which serves as a powerful tool to improve enzyme performance (Denard et al., 2015, Dalby, 2011). Generally, strategies for enzyme engineering are divided as the rational method, random method (directed evolution) and the semi-rational method which is a combination of both random and rational approaches (Steiner and Schwab, 2012). In the rational method, replacement of amino acids at precise locations are designed based on the knowledge of protein structure and mechanism information (Steiner and Schwab, 2012). In comparison with the rational method, similar information and knowledge of the protein are not required in the random method approach. The approach employs a random process such as error-prone PCR to create a mutant library (Chen, 2001). Both the rational and random methods have their own pros and cons but yet they can work to complement each other. The rational method requires the generation of a smaller library size, hence less effort and time is required for screening. The random method on the other hand introduces mutations randomly without bias yielding a large mutant library which would require the need to screen large numbers of mutants (Tachioka et al., 2016). Meanwhile the semi-rational approach combines the advantages from both rational and random methods requiring a smaller size library based on protein biochemical and structural data (Steiner and Schwab, 2012) where specific residue sites are determined ‘rationally’ and then randomized to all 20 amino

acids (McLachlan et al., 2009). The library is then screened to yield the desired mutants.

Site-directed mutagenesis (SDM) is commonly used for rational method amino acid substitution, addition, or deletion in the template DNA sequence (Urban et al., 1997). Overlap-extension PCR (OE-PCR) is often described as the approach to perform SDM to introduce desired mutations. In OE-PCR, two separate PCRs are performed to generate two separate DNA fragments with similar overlapping ends by using 4 primers (Figure 1.3).

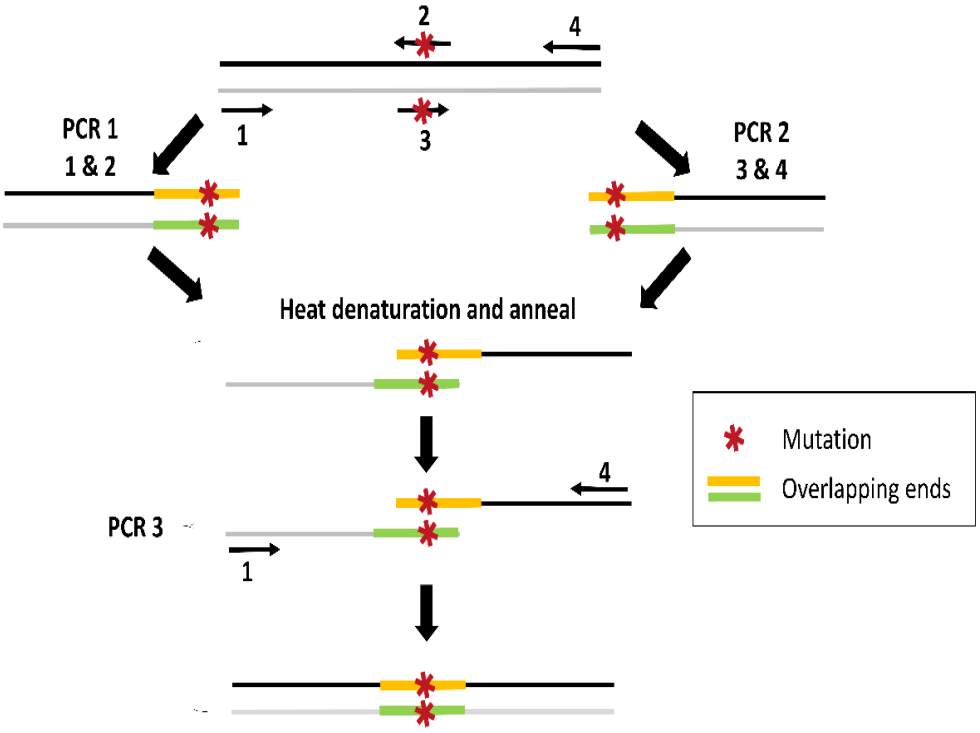

Figure 1.3: Overlap-extension PCR (OE-PCR) for site-directed mutagenesis. Primer 1 and 4 is MTG outer forward and reverse primer respectively. Primer 2 and 3 is respective reverse and forward primer carrying desired mutation.

The desired mutations are introduced at this stage. Later, these two DNA fragments are heat denatured and joined via PCR to generate the whole DNA fragment with a set of outermost primers (Antikainen and Martin, 2005). This technique is vital in enzyme engineering to generate mutants with altered amino acid sequences in order to improve enzyme functional attributes (Urban et al., 1997, Xiao et al., 2007).

1.1.3 (b) Engineering of mTGase

Enzyme engineering is widely used to improve or alter the functional properties of mTGase over the past few years particularly for industrial purposes. Few publications have successfully showed improved mTGase properties including substrate specificity (Sugimura et al., 2008), thermostability (Marx et al., 2008b), activity (Yokoyama et al., 2010), and yield (Rickert et al., 2016) via various strategies.

Both rational and random methods have been applied to screen for improved activity of mTGase by Yokoyama and his group. They started off with the rational mutagenesis by computing the percentage of solvent-accessibility surface area (SAS) of each amino acid in the activated mTGase. Forty amino acid residues with high values of SAS percentage was selected and subjected to SDM. Thirty two mutants exhibited a higher specificity than the wild type. For random mutagenesis, error prone PCR was applied to randomly mutate the whole region of the mTGase gene. Ten clones with higher specificity than the wild type were successfully isolated. Mutant S199A isolated from the random method demonstrated the highest specific activity, which was 1.7 times higher than the wild type (Yokoyama et al., 2010).

Another group of researchers performed random mutagenesis on mTGase to generate thermostable and heat-sensitive variants for cross-linking at higher

temperatures. They observed all single point mutations which altered thermal properties of wild type mTGase were located at the N-terminal. One variant (S2P) demonstrated 270 % increase in the half-life after the serine residue near to the N-terminal was mutated to proline (Marx et al., 2008b). In a separate case, mutagenesis on the pro-domain region was reported to affect the yield of soluble activated mTGase in *E. coli* cytoplasm. The rational approach was applied to perform alanine scanning on the pro-domain to identify mutants which are able to retain its chaperone function yet capable to be activated at a designated temperature. Alanine mutation at the pro-domain region with an insertion of 3C protease expression successfully generated yields and enzymatic activities similar to wild type mTGase from *S. mobarensis* (Rickert et al., 2016).

To reveal the substrate specificity of mTGase, researchers simulated a published three-dimensional docking model (PDB ID: 1IU4) with a peptide substrate, N-Benzyloxycarbonyl-L-Glutaminylglycine (CBZ-Gln-Gly). Residues in the vicinity of Cys^{64} which interact frequently with CBZ-Gln-Gly were replaced with alanine by SDM. They observed significant reduction in mTGase activities after those sites were mutated to alanine (A). Also, they found that residues such as valine (V), leucine (L), isoleucine (I), phenylalanine (F), tyrosine (T), or tryptophan (W) in the vicinity of the glutamine (Q) residue are ideal to influence substrate recognition by forming hydrophobic interaction or Π-Π interaction (Tagami et al., 2009).

Mutagenesis has been successfully applied in altering enzymes to introduce novel or improved properties. However, this approach often limits the success of designed mutations due to the basis of structure-function relationship. Therefore, the ability to carry out directed evolution of enzymes will provide a more systematic and high-throughput approach to enzyme engineering. This will also involve the creation of

mutant libraries and screening for desired the target within these libraries in a high-throughput manner. Therefore, display systems have been employed as an attractive alternative to generate a collection of variants with novel properties which can be screened and isolated simultaneously (Fernandez-Gacio et al., 2003).

1.1.4 Phage Display Technology

Various display systems have been applied successfully to evolve enzymes including phage display (Fernandez-Gacio et al., 2003), ribosome display (Amstutz et al., 2002), mRNA display (Seelig, 2011), and bacterial display (Kim et al., 2000). Phage display technology is known as one of the best display systems for *in vitro* selection because it can handle large-sized libraries which increases the efficiency of screening for targets (Fernandez-Gacio et al., 2003).

Phage display technology was first introduced by George Smith in 1985 to display peptides on the surface of phage particles as a fusion to the coat proteins (Smith, 1985). Presentation of peptides or proteins on phage surfaces allows for a direct physical linkage between the phenotype and genotype as shown in Figure 1.4, making the identification of isolated particles convenient (Strachan et al., 2002, Garet et al., 2010).

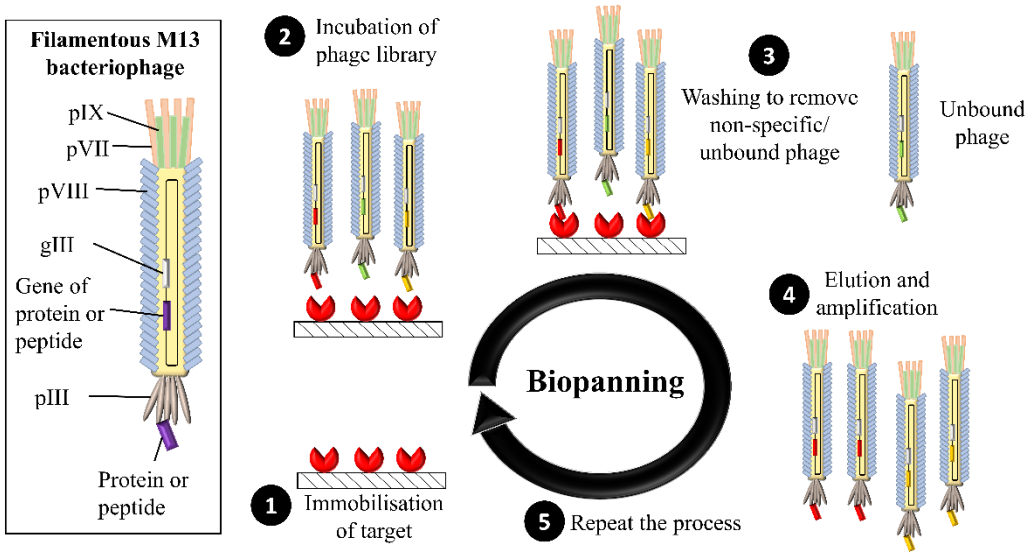

Figure 1.4: Structure of filamentous M13 bacteriophage and phage display biopanning. Biopanning process has four major steps. (1) Immobilisation of target on solid surface. (2) Incubation of phage library with the coated target. (3) Washing step to remove non-specific/unbound phage. (4) Elution and amplification of bound phage. (5) The process is repeated by incubating the amplified phage with target.

Filamentous phages (f1, fd, M13) are commonly used in phage display systems to infect *E. coli* strains carrying the F' episome due to their non-lytic characteristics (Tikunova and Morozova, 2009). Proteins or peptides are commonly displayed on phage surface by fusion to phage coat proteins pVIII (p8) and pIII (p3) (Hust and Dübel, 2005, Vodnik et al., 2011). Even so, fusion to coat protein pVII (p7) (Kwaśnikowski et al., 2005, Løset et al., 2011b) and pIX (p9) (Løset et al., 2011a, Løset et al., 2011b) have also been reported. However, p3 is the preferred coat protein for display because it is the largest coat protein with 406 amino acids and able to display large proteins for selection (Zhao et al., 2016).

The selection procedure via phage display technology to obtain specific proteins or peptides is known as biopanning which involves repetitive cycles of binding, washing, elution, and amplification of isolated phages (Figure 1.4). Generally, three to five rounds of biopanning are required to enrich high affinity predominant population. This allows clonal selection to define a concentrated pool of a particular population of clones with specific characteristics. Stringency of selection could be adjusted by varying the number of wash steps, condition of wash solution, incubation time of phages or concentration of target protein. After the last round of biopanning, selected phage clones are then identified with DNA sequencing since the genotype and phenotype are directly linked (Bazan et al., 2012, Lim and Chan, 2016).

Previously, phage display was applied to display enzymes for selecting desired enzyme presenting phage clones based on catalytic activity (Fujita et al., 2005, Soumillion et al., 1994). However, the large molecular weight of enzymes often limits their maximum concentration in solution which results in long measurement times for low activities and background problem. Moreover, exportation of the bulky enzymes through the cytoplasmic membrane introduces another set of challenges. This is due

to the inefficiency of phage particles to present proteins that are large in size. In addition to that, selection for phage particles presenting enzymes based on catalytic activity is much more difficult than affinity based discrimination (Fernandez-Gacio et al., 2003). Therefore, these display formats are not widely used in enzyme engineering. Instead, phage display was applied successfully to select preferred substrates for enzymes.

Sugimura and his team identified preferred substrate sequences of mTGase via peptide phage display. The phage peptide library was first incubated with biotinylated cadaverine and mTGase. Then, samples were passed through an avidin column to isolate the desired peptide phages which are bound to the biotinylated cadaverine. The isolated peptide presenting phages were then proceeded with rounds of selection to enrich the desired peptide presenting phages. The group managed to enrich peptides with a consensus sequence which is different from mammalian TGases (Sugimura et al., 2008). An almost similar strategy was adopted to screen for preferred substrate for mTGase derived from *Kutzneria albida* (KalbTG). A 5-mer peptide library was constructed by using all combinations of 18 natural amino acids excluding cysteine and methionine or dimer, repeat of similar amino acid. Both N- and C-terminals of the 5-mer peptides were flanked by a 3 amino acid linker. Peptide arrays were performed by incubating the peptide library with mTGase together with biotinylated cadaverine which was then detected with streptavidin labelled fluorescent dye. Motifs for both acyl donor and acyl acceptor were successfully identified and applied in protein labelling. Computational approach was later applied to superimpose KalbTG to mTGase derived from *S. mobaraense* to predict the substrate binding mode (Steffen et al., 2017).

1.1.5 Computational Modelling for Protein

Functional characterization of proteins based on sequence information is usually facilitated via three-dimensional (3D) structure to provide crucial structure-function information of a particular protein of interest. These structures can either be determined via experimental methods or computational methods. In experimental methods, X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy are the two most common approaches to determine 3D structures of proteins at high-resolution (Webb and Sali, 2014, Madhusudhan et al., 2005). In the computational approach, the 3D structure of the protein of interest could be predicted via *ab inito* modelling or via comparative modelling by alignment with known protein structures (Baker and Sali, 2001, Fiser et al., 2002). The former is essential when similar structures are not available or unidentified, then models are built from scratch based on the laws of physics (Madhusudhan et al., 2005). Comparative modelling on the other hand predicts 3D structure of the protein (target) by alignment to proteins (template) with known structures (Webb and Sali, 2014).

There are a few factors that needs to be considered when it comes to the selection of suitable templates for modelling. Firstly, the purpose of modelling has to be determined. For example, if a protein-ligand model is to be constructed, templates with similar ligands should be chosen. Next, sequence and subfamily similarity between target and template should be high to enable the selection of a template closest to the target to be modelled. Also, the environment between template and target including pH, ligands, solvent, quaternary interaction should be considered. Last but not least is the resolution and residual/reliability factor (R factor) of crystallographic structure. Higher resolution template should be used if two templates have comparable similar sequences with the target (Eswar et al., 2006).

Currently, MODELLER is one of the most popular choice for protein structure modelling experiments. Target sequences are inputs to be modelled based on template structures and MODELLER will automatically calculate a model without any user intervention (Eswar et al., 2006). Till today, two crystal structures of mTGase from *S. mobaraensis* (PDB ID: 1IU4; PDB ID: 3IU0) obtained by X-ray diffraction have been deposited in Protein Data Bank (PDB) (Kashiwagi et al., 2002, Yang et al., 2011). Both these crystal structures have provided valuable information for better understanding of mTGase especially in comparative modelling when property alteration is desired. Comparative model of mTGase from *S. mobaraensis* has not yet been reported so far. However, with the increase in number of crystal structures deposited in PDB, accuracy of comparative modelling is also improved to bridge the gap between known sequences with 3D model (Eswar et al., 2006, Fiser, 2010).