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

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

[1]Department of Chemistry, and [2]Department of Biochemistry, Université de Montréal, 2900 Boulevard Edouard-Montpetit, Montréal, Québec, H3T 1J4, Canada

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

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

Corresponding Author:

Joelle N. Pelletier
Department of Chemistry
Université de Montréal
2900 Boulevard Edouard-Montpetit, Montréal, QC
H3T 1J4
Canada
Tel: +1 (514) 343-2124
Fax: +1 (514) 343-7586
E-mail: joelle.pelletier@umontreal.ca

Abstract

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

Introduction

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

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

Enzymes that are used to conjugate proteins are generally limited to using the *N*- or *C*-terminus as the site of modification.[12] The power of these enzymes stems from each enzymatic class having an amino acid recognition sequence that is targeted with high or exclusive selectivity, as long as this sequence is located at a protein terminus. Formylglycine generating enzyme, phosphopantetheinyl transferase, farnesyltransferase, biotin ligase, and lipoic acid ligase are examples of enzymes that catalyze such bioconjugations.[12] As a recent example, formylglycine generating enzyme has been used to construct artificially glycosylated proteins,[13] and DNA-protein conjugates.[14]

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

These successes result from MTG's high promiscuity toward its lysine substrates,[1; 3] with its ability to accept numerous primary amines being a key for the incorporation of diverse chemical functionality, such as bio-orthogonal functional groups to fluorophores. In contrast, MTG's glutamine reactivity is restricted to protein- and peptide-bound glutamine residues. Phage display screening of glutamine-containing peptides has yielded several 'glutamine tags'[20] that were successfully applied to channel MTG's reactivity during protein labeling;[10] we note that this example used a *C*-terminally expressed glutamine recognition tag rather than a reactive glutamine internal to the target protein. Nonetheless, those glutamine-containing sequences are

diverse in composition, revealing no clear pattern in the primary structure surrounding the reactive glutamine.[20]

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

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

GB1 as a probe for investigating MTG's glutamine reactivity and identifying more highly reactive locations for a glutamine residue, towards making MTG a more effective tool for protein conjugation.

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

Results

Design of the Single-Glutamine-Containing GB1 Variants

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

The first mutagenesis step consisted in replacing the sole native glutamine of GB1, Q32, with a structurally similar residue that MTG does not react with, asparagine. The Q32N knock-out served not only as the template for generating future mutants, but as a negative control to verify that no conjugation was occurring at other sites on GB1. This was confirmed by resolving on gel and by high-resolution MS (Table S3). A single glutamine was then introduced at each of the 24 selected locations on the template. We confirmed that all the GB1 variants expressed solubly to similar levels as the native GB1 (Supporting Fig. 1).

Fluorescent MTG Protein Assay

The establishment of a sensitive assay to monitor the efficiency of labeling of the GB1 variants was critical to the success of the study (Fig. 2). We and others previously investigated MTG's ability to accept a variety of primary amines as substrates instead of lysine, and others have exploited this promiscuity as a tool to introduce diverse functionalities into proteins,[29; 1; 15; 30; 3] providing us with considerable flexibility in the choice of our probe. Our standard

methodology for monitoring the products of MTG-catalyzed conjugation has been based on the use of liquid chromatography-mass spectrometry (LC-MS).[31] However, visualizing and quantifying fluorescence output is more rapid and sensitive, and provides a direct means to screening for improved fluorescent protein labeling. To this effect, we recently reported a new class of highly tunable fluorescent compounds that can be readily functionalized to bear a primary amine.[27] These bright fluorescent dyes are characterized by an unusually high excitation-emission differential and are highly soluble in aqueous media, making them good candidates for bioconjugation. Although the primary amine of benzo[*a*]imidazo[2,1,5-*c*,*d*]indolizin-7-ylmethanaminium (**1**) is separated from the bulky, aromatic core by a single methylene, we have demonstrated that MTG can use it as a substrate to label two proteins, α-lactalbumin (α-LA) as well as GB1; while α-LA is well established to be highly reactive with MTG,[32; 33; 24] GB1 had not yet been known to be a substrate prior to our investigation.[27] Here we extend this assay of fluorescent GB1 conjugation to the GB1 variants (Fig. 2).

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

The quantification of fluorescence is described herein according to two properties: selectivity and efficiency (Table 1). Selectivity refers to the degree to which GB1 is labeled in the presence of MTG relative to non-specific binding. When non-specific binding of **1** was

observed, as in the case of native GB1 and some of the GB1 variants (Fig. 3), the selectivity ratio was calculated; the lower the background, the higher the selectivity. Efficiency, instead, compares the fluorescent output of a labeled GB1 variant to that of the labeled native GB1; it expresses the relative reactivity of the glutamine. We observed that selectivity tended to increase as efficiency increased.

Introduction of Glutamine into GB1 Loop Elements

Of the 24 glutamine-displaying GB1 variants prepared, eight of the targeted residues were located on flexible loops (Fig. 3), with at least one mutation being made in each of the four loops present in GB1. Based on the report of higher glutamine reactivity in disordered regions,[24] we anticipated that this subgroup of variants should be the most reactive. While T49Q exhibited good fluorescence following conjugation with **1**, with both selectivity and efficiency well over one order of magnitude higher than native GB1, it was the only strongly improved loop variant. T11Q, located on a different loop, produced a modest increase compared to native GB1; all other loop variants were unreactive, or exhibited the same level of reactivity as the control lacking MTG (indicating non-specific binding of **1** to the GB1 variant). This is particularly surprising when closely observing the location of T49 within the crystal structure. Indeed, T49 is on the same loop as variants D47Q and A48Q; the former was inert to labeling, and A48Q was barely observable. These residues are all within a similar environment, making it difficult to rationalize the drastic difference in reactivity that MTG displays for its glutamine substrate. Similarly, the K10Q variant exhibited no reactivity despite being located beside T11 which, when substituted, was modestly more reactive than native GB1.

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

Relocating Glutamine in the α-Helix of GB1

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

native GB1. We thus demonstrate that the well-structured and tightly packed α-helix of GB1 can harbour highly MTG-reactive glutamines.

Glutamines in the β-Sheet of GB1 Can Also Be Reactive

With the β-sheet being the largest single secondary structure element within GB1, over 40 % of the newly introduced glutamines were located within it. Upon examining the crystal structure, we speculated that many of these mutations would react poorly, particularly those located on the internal β-strands 1 and 4 because they belong to a flat protein surface that does not appear to be complementary to the crevice that forms MTG's active site (Fig. 4).[23] This speculation was invalidated when the most selective and efficient variant was determined to be I6Q, located within β-strand 1. K4Q, the other variant introducing glutamine within β-strand 1, also exhibited high reactivity. T55Q also reacted strongly; that residue is located at the very edge of β-strand 4 and is not as tightly concealed within the structure as is I6Q. T17Q and E19Q lost reactivity, with the remaining four variants exhibiting activities on par with native GB1.

These results are surprising, as MTG has been reported to prefer glutamine-containing regions that are predominantly unstructured,[24] yet in the GB1 framework we observe the highest reactivity in α-helical and β-sheet regions. Therefore, secondary structure (or lack thereof) is not a strong predictor of glutamine reactivity. We attempted to identify patterns in the primary sequence flanking the reactive glutamines but failed to identify any potential markers to predict glutamine reactivity (Fig. 5). We also considered tertiary structure, seeking patterns in surface charge and hydrophobicity (Supporting Figs. 2-4), as well as B-factors (Supporting Fig. 5). No clear sequence or structural pattern or trend was observed amongst the reactive glutamines, making it difficult to predict where MTG will bind.

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

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

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

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

The high reactivity of I6Q GB1 resulted in rapid saturation of the fluorescence signal after the exposure time we had determined to be optimal for quantification of most variants (5 s). To compare the reactivity of MTG and its variants more accurately, we recorded the fluorescent signal after 1 s of exposure, where saturation was not observed (Table 4). Three MTG variants, W69H, Y302G, and Y302H, reacted with I6Q GB1 as well as or better than native GB1. W69H maintained the same level of reactivity as native MTG, where Y302G and Y302H were twice as efficient. Taken together, these results demonstrate that engineered MTG and GB1 variants can be paired to create an effective protein labeling system: it is possible to engineer both the substrate and the catalyst towards higher efficiencies, and these effects of engineering the substrate and the enzyme are cumulative.

In summary, among a library of 24 different glutamine-containing point mutations, covering 43 % of the amino acid sequence, four GB1 variants were observed to be at least 100-fold more reactive towards MTG and four more were at least 10-fold more reactive. Thus, one-third of the glutamines tested in the well-folded, globular GB1 protein provided good substrates for MTG labeling, with the I6Q substitution being the most reactive among all. We were not able to identify any clear trend that MTG displays towards the environment in which the glutamine

residue is located, whether considering the primary sequence flanking the glutamine (Fig. 6) or tertiary structure properties (Supporting Figs. 2-4). We initially expected that loop variants would be the most reactive, as the high flexibility of these elements would make them the most likely candidates to fit into MTG's active-site cleft. However, the loop variants underperformed relative to α-helix or β-sheet variants, leading us to hypothesize that if secondary structure plays a role in MTG's substrate recognition, there are other, more important factors that dominate MTG's glutamine selectivity. If MTG undergoes a significant structural rearrangement upon binding to its glutamine-bearing protein substrate, as does its mammalian TG2 counterpart,[38] then predicting their mode of interaction may require their co-crystallization.

To conclude, through a semi-rational approach, we constructed and improved a protein labeling system in which both the catalyst and substrate were optimized. Enzymatic bioconjugation is advantageous over traditional chemical modifications owing to increased selectivity displayed by enzymes,[12] with MTG's glutamine selectivity being central to its effective application. MTG has the advantage of being independent of an *N*- or *C*-terminal reactive site, which is a requirement for most enzymes used for protein labeling, allowing for increased selection and flexibility of the modification site within the protein target. In addition, as we have demonstrated, MTG labeling activity has the potential to be improved by protein engineering. Point mutations improved the reactivity of the substrate protein, GB1, by over two orders of magnitude, which was further enhanced when coupling with variants of the catalyst, MTG. Through this process, we probed the selectivity MTG displays for its glutamine-bearing protein substrate. Although no clear recognition pattern was observed, we have demonstrated the straightforward engineering of MTG-reactive glutamines in a well-folded domain, suggesting

that other proteins are amenable to similar modification to allow MTG-catalyzed protein labeling.

Acknowledgements

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

Conflicts of Interest

The authors declare no competing financial interests.

Materials and Methods

Materials

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

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

Expression and purification of MTG

MTG was expressed and purified as previously described.[30] Briefly, a 5-mL starter culture of *E. coli* BL21 (DE3) containing the plasmid pET20b-MTG, which expresses a *C*-terminally 6-His-tagged version of MTG, was propagated overnight at 37°C in ZYP-0.8G medium and shaking at 240 rpm. It was used to inoculate 500 mL of auto-inducing ZYP-5052 medium.[40] After 2 h of incubation at 37°C and 240 rpm, the temperature was reduced to 22°C overnight. Cells were collected by centrifugation and resuspended in 50 mM sodium phosphate buffer, 300 mM NaCl, pH 7.5. The cells were lysed using a Constant Systems cell disruptor set at 37 kPSI and cooled to 4°C. After further centrifugation to remove insoluble cellular matter, the inactive form of MTG was incubated with trypsin (1 mg/mL solution, 1:9 ratio of trypsin to MTG, v/v) for the purpose of cleaving its pro-sequence. Activated MTG was purified using a 5-mL His-trap nickel-nitrilotriacetic acid (Ni-NTA) column (GE Healthcare) equilibrated in 50 mM phosphate buffer, pH 7.5, with 300 mM NaCl, and eluted with an imidazole gradient (0 –

250 mM) using an Åtka FPLC (GE Healthcare). After purification, active MTG was dialyzed against 50 mM sodium phosphate buffer, 300 mM NaCl, pH 7.5. The average yield was 25 mg of activated MTG per litre of culture, with ~ 85% purity as estimated by SDS-PAGE and revelation with Coomassie Brilliant Blue stain. Aliquots were snap frozen and stored at -80°C in 15% glycerol.

MTG Mutagenesis

Plasmid pDJ1-3, encoding the open reading frame for MTG, was used as a template for mutagenesis. All mutants were obtained using the rolling circle approach.[41; 42] Following mutagenesis with Phusion® High Fidelity polymerase, the amplified PCR product was treated with FastDigest *Dpn*I before being transformed into *E. coli* BL21 (DE3) for protein expression. Ampicillin (Amp) was used at 100 µg/mL for plasmid maintenance. Sequences were confirmed by DNA sequencing (ABI 3730 DNA sequencer, IRIC Genomic Platform at Université de Montréal). (Table S1).

Expression and purification of native GB1 and variants

A 2-mL starter culture of *E. coli* BL21 (DE3) containing the plasmid pET15b-GB1, or mutagenized plasmids expressing a variant within the same vector, which expresses an *N*-terminally 6x-His-tagged version of GB1, was propagated overnight at 37°C in LB medium containing 100 µg/mL ampicillin and shaking at 240 rpm. A 500 µL volume of the starter culture was used to inoculate 50 mL of LB medium containing 100 µg/mL ampicillin. After 3h of incubation at 37°C and 240 rpm, IPTG was added to a final concentration of 0.2 mM and expression was allowed to proceed for 3 hours. Cells were collected by centrifugation and

resuspended in 2.7 mL of 50 mM sodium phosphate buffer, 300 mM NaCl, pH 7.5. FastBreak™ cell lysis reagent was added to the resuspended cells to a final volume of 3 mL, mixed by inversion, and incubated at room temperature for 10 min. After further centrifugation at 4°C to remove insoluble cellular matter, the clarified lysate was loaded onto 1 mL of nickel-nitrilotriacetic acid (Ni-NTA) resin (GE Healthcare) equilibrated in 50 mM phosphate buffer, pH 7.5, with 300 mM NaCl. The resin was washed with 10 column volumes of the same buffer containing 15 mM imidazole, and eluted in 3 mL using the phosphate buffer containing 250 mM imidazole. After purification, GB1 was dialyzed against 50 mM sodium phosphate buffer, pH 7.5, 300 mM NaCl, 1 mM EDTA. The average yield was 3 mg of GB1 per 50 mL of culture, with ~ 90% purity as estimated by tricine SDS-PAGE[35] and revelation with Coomassie blue stain. Aliquots were snap frozen and stored at -80°C in 20% glycerol.

GB1 Mutagenesis

The pET15b-GB1 plasmid encoding the open reading frame for GB1 was used as a template for mutagenesis. The sequence for glutamine knock-out, Q32N, was generated first from native GB1, and was subsequently used as a template for amplification of all other GB1 mutants. Site overlap extension was used to generate mutant GB1 sequences.[43] The DNA fragments were digested with FastDigest *Nde*I and *Bam*HI restriction enzymes, and religated into pET15b which had been cut with the same enzymes and also dephosphorylated, and transformed in *E. coli* BL21 (DE3).

MTG Activity Assay

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

Fluorescent conjugation assays

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

High resolution mass spectrometry

Aliquots taken at 24 h were centrifuged to remove any insoluble matter, and transferred to LC-MS vials. Undiluted samples (5 μL) were injected without any additional treatment onto

an Aeris peptide XB-C18, 3.6-μm, 150 × 2.1 mm LC column (Phenomenex) and eluted with a 16-minute, 5-50% ACN/H_2O gradient. Masses were detected under positive ionization (ESI) with a Synapt G2S (Q-TOF) triple quadrupole mass detector (Waters).

Differential Scanning Fluorimetry

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

For the T_m calculations, both temperature and fluorescence data were smoothed.[44] The first derivatives dFluo or dTemp were calculated using the cubic spline interpolation. The preliminary maximum was determined to obtain the half-values to the left and right of it. The linear fit for the curve outside the half-values was calculated, followed by the calculation of the average deviation from the fit. If the maximum was below the detection limit (fit value + 3 × deviation), the T_m determination was considered uncertain. The quadratic fit around the maximum was then calculated as follows to obtain T_m. The first derivative of the quadratic fit function (y-value) was set to 0 and the x-axis value (temperature) was resolved. Then, the

average deviation of the curve points around the maximum from the quadratic fit was calculated. If the relative deviation was greater than 5%, the T_m values were rejected if the corresponding maximum was below the detection limit. However, T_m values with a maximum above the detection but a relative deviation greater than 5% were defined as uncertain.

References

1. Rachel NM, Pelletier JN (2013) Biotechnological applications of transglutaminases. Biomolecules 3:870-888.
2. Yokoyama K, Nio N, Kikuchi Y (2004) Properties and applications of microbial transglutaminase. Appl Microbiol Biotechnol 64:447-454.
3. Strop P (2014) Versatility of microbial transglutaminase. Bioconjug Chem 25:855-862.
4. Krall N, da Cruz FP, Boutureira O, Bernardes GJ (2016) Site-selective protein-modification chemistry for basic biology and drug development. Nat Chem 8:103-113.
5. Hoffman RM (2005) The multiple uses of fluorescent proteins to visualize cancer in vivo. Nat Rev Cancer 5:796-806.
6. Kerppola TK (2006) Visualization of molecular interactions by fluorescence complementation. Nat Rev Mol Cell Biol 7:449-456.
7. Chan J, Dodani SC, Chang CJ (2012) Reaction-based small-molecule fluorescent probes for chemoselective bioimaging. Nat Chem 4:973-984.
8. Zhang G, Zheng S, Liu H, Chen PR (2015) Illuminating biological processes through site-specific protein labeling. Chem Soc Rev 44:3405-3417.
9. Kamiya N, Abe H (2011) New fluorescent substrates of microbial transglutaminase and its application to peptide tag-directed covalent protein labeling. Methods Mol Biol 751:81-94.
10. Oteng-Pabi SK, Pardin C, Stoica M, Keillor JW (2014) Site-specific protein labelling and immobilization mediated by microbial transglutaminase. Chem Commun (Camb) 50:6604-6606.
11. Siegmund V, Schmelz S, Dickgiesser S, Beck J, Ebenig A, Fittler H, Frauendorf H, Piater B, Betz UA, Avrutina O, Scrima A, Fuchsbauer HL, Kolmar H (2015) Locked by Design: A Conformationally Constrained Transglutaminase Tag Enables Efficient Site-Specific Conjugation. Angew Chem Int Ed Engl 54:13420-13424.
12. Rashidian M, Dozier JK, Distefano MD (2013) Enzymatic labeling of proteins: techniques and approaches. Bioconjug Chem 24:1277-1294.
13. Liang SI, McFarland JM, Rabuka D, Gartner ZJ (2014) A modular approach for assembling aldehyde-tagged proteins on DNA scaffolds. J Am Chem Soc 136:10850-10853.
14. Smith EL, Giddens JP, Iavarone AT, Godula K, Wang LX, Bertozzi CR (2014) Chemoenzymatic Fc glycosylation via engineered aldehyde tags. Bioconjug Chem 25:788-795.
15. Strop P, Liu SH, Dorywalska M, Delaria K, Dushin RG, Tran TT, Ho WH, Farias S, Casas MG, Abdiche Y, Zhou D, Chandrasekaran R, Samain C, Loo C, Rossi A, Rickert M, Krimm S, Wong T, Chin SM, Yu J, Dilley J, Chaparro-Riggers J, Filzen GF, O'Donnell CJ, Wang F, Myers JS, Pons J, Shelton DL, Rajpal A (2013) Location matters: site of conjugation modulates stability and pharmacokinetics of antibody drug conjugates. Chem Biol 20:161-167.

16. Dennler P, Chiotellis A, Fischer E, Bregeon D, Belmant C, Gauthier L, Lhospice F, Romagne F, Schibli R (2014) Transglutaminase-based chemo-enzymatic conjugation approach yields homogeneous antibody-drug conjugates. Bioconjug Chem 25:569-578.
17. Fontana A, Spolaore B, Mero A, Veronese FM (2008) Site-specific modification and PEGylation of pharmaceutical proteins mediated by transglutaminase. Adv Drug Deliv Rev 60:13-28.
18. Maullu C, Raimondo D, Caboi F, Giorgetti A, Sergi M, Valentini M, Tonon G, Tramontano A (2009) Site-directed enzymatic PEGylation of the human granulocyte colony-stimulating factor. FEBS J 276:6741-6750.
19. Mero A, Spolaore B, Veronese FM, Fontana A (2009) Transglutaminase-mediated PEGylation of proteins: direct identification of the sites of protein modification by mass spectrometry using a novel monodisperse PEG. Bioconjug Chem 20:384-389.
20. Sugimura Y, Yokoyama K, Nio N, Maki M, Hitomi K (2008) Identification of preferred substrate sequences of microbial transglutaminase from Streptomyces mobaraensis using a phage-displayed peptide library. Arch Biochem Biophys 477:379-383.
21. Kashiwagi T, Yokoyama K, Ishikawa K, Ono K, Ejima D, Matsui H, Suzuki E (2002) Crystal structure of microbial transglutaminase from Streptoverticillium mobaraense. J Biol Chem 277:44252-44260.
22. Yang MT, Chang CH, Wang JM, Wu TK, Wang YK, Chang CY, Li TT (2011) Crystal structure and inhibition studies of transglutaminase from Streptomyces mobaraense. J Biol Chem 286:7301-7307.
23. Tagami U, Shimba N, Nakamura M, Yokoyama K, Suzuki E, Hirokawa T (2009) Substrate specificity of microbial transglutaminase as revealed by three-dimensional docking simulation and mutagenesis. Protein Eng Des Sel 22:747-752.
24. Spolaore B, Raboni S, Ramos Molina A, Satwekar A, Damiano N, Fontana A (2012) Local unfolding is required for the site-specific protein modification by transglutaminase. Biochemistry 51:8679-8689.
25. Shea JE, Brooks CL, 3rd (2001) From folding theories to folding proteins: a review and assessment of simulation studies of protein folding and unfolding. Annu Rev Phys Chem 52:499-535.
26. Cheng Y, Patel DJ (2004) An efficient system for small protein expression and refolding. Biochem Biophys Res Commun 317:401-405.
27. Levesque E, Bechara WS, Constantineau-Forget L, Pelletier G, Rachel NM, Pelletier JN, Charette AB (2017) General C-H Arylation Strategy for the Synthesis of Tunable Visible Light Emitting High Stokes Shift Benzo[a]imidazo[2,1,5-c,d]indolizine Fluorophores. J Org Chem.
28. Chica RA, Doucet N, Pelletier JN (2005) Semi-rational approaches to engineering enzyme activity: combining the benefits of directed evolution and rational design. Curr Opin Biotechnol 16:378-384.
29. Ohtsuka T, Sawa A, Kawabata R, Nio N, Motoki M (2000) Substrate specificities of microbial transglutaminase for primary amines. J Agric Food Chem 48:6230-6233.
30. Gundersen MT, Keillor JW, Pelletier JN (2014) Microbial transglutaminase displays broad acyl-acceptor substrate specificity. Appl Microbiol Biotechnol 98:219-230.
31. Rachel NM, Pelletier JN (2016) One-pot peptide and protein conjugation: a combination of enzymatic transamidation and click chemistry. Chem Commun (Camb) 52:2541-2544.
32. Matsumura Y, Chanyongvorakul Y, Kumazawa Y, Ohtsuka T, Mori T (1996) Enhanced susceptibility to transglutaminase reaction of alpha-lactalbumin in the molten globule state. Biochim Biophys Acta 1292:69-76.
33. Nieuwenhuizen WF, Dekker HL, de Koning LJ, Groneveld T, de Koster CG, de Jong GA (2003) Modification of glutamine and lysine residues in holo and apo alpha-lactalbumin with microbial transglutaminase. J Agric Food Chem 51:7132-7139.

34. Oteng-Pabi SK, Keillor JW (2013) Continuous enzyme-coupled assay for microbial transglutaminase activity. Anal Biochem 441:169-173.
35. Schagger H, von Jagow G (1987) Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. Anal Biochem 166:368-379.
36. Niesen FH, Berglund H, Vedadi M (2007) The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. Nat Protoc 2:2212-2221.
37. Folk JE, Cole PW (1966) Mechanism of action of guinea pig liver transglutaminase. I. Purification and properties of the enzyme: identification of a functional cysteine essential for activity. J Biol Chem 241:5518-5525.
38. Pinkas DM, Strop P, Brunger AT, Khosla C (2007) Transglutaminase 2 undergoes a large conformational change upon activation. PLoS biology 5:e327.
39. Marx CK, Hertel TC, Pietzsch M (2007) Soluble expression of a pro-transglutaminase from Streptomyces mobaraensis in Escherichia coli. Enzyme Microb Technol 40:1543-1550.
40. Studier FW (2005) Protein production by auto-induction in high density shaking cultures. Protein Expr Purif 41:207-234.
41. Fire A, Xu SQ (1995) Rolling replication of short DNA circles. Proc Natl Acad Sci U S A 92:4641-4645.
42. Lizardi PM, Huang X, Zhu Z, Bray-Ward P, Thomas DC, Ward DC (1998) Mutation detection and single-molecule counting using isothermal rolling-circle amplification. Nat Genet 19:225-232.
43. Ho SN, Hunt HD, Horton RM, Pullen JK, Pease LR (1989) Site-directed mutagenesis by overlap extension using the polymerase chain reaction. Gene 77:51-59.
44. Savitzky A, Golay MJE (1964) Smoothing + Differentiation of Data by Simplified Least Squares Procedures. Anal Chem 36:1627-1639.

Tables

Secondary Structure	Mutation	Selectivity	Efficiency
α-helix	Native	1.3	1
	Q32N	N.D.	N.D.
Loop	K10Q	N.D.	N.D.
	T11Q	16	4
	V21Q	1.2	-
	D40Q	3.9	-
	E42Q	4.2	-
	D47Q	N.D.	N.D.
	A48Q	1.0	-
	T49Q	53	43
α-helix	A24Q	5.0	4.2
	K28Q	-	-
	V29Q	430	130
	K31Q	270	100
	N35Q	2.5	-
	D36Q	3.7	0.7
β-sheet	K4Q	98	79
	I6Q	190*	180
	E15Q	4.4	3.4
	T17Q	-	-
	E19Q	1.0	-
	T44Q	1.3	1.5
	D46Q	5.0	27
	T51Q	71	46
	T53Q	3.3	2.6
	T55Q	20	160

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

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

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

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

Table 3. Specific activity of MTG variants towards the model dipeptide, ZQG.

Selectivity

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

Efficiency, compared to native MTG

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

Efficiency, compared to native GB1 substrate

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

I6Q GB1 substrate, 1 s exposure

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

Table 4. MTG variant reactivity towards native and I6Q GB1. For the top three tables, fluorescence was quantified after 5 s of exposure, while the last table quantified after 1 s of exposure to prevent saturation of the detector. Selectivity represents the fold-increase of the ratio of fluorescence in the reactions over non-specific fluorescence in the control; the higher the selectivity, the lower the background. Efficiency is the fold-increase of the ratio of fluorescence of the variant relative to native protein and is labeled in increasingly saturated shades of green: 1 to 10-fold, pale green; 10 to 100-fold, bright green; greater than 100-fold, dark green. Signals that were lower than that of native GB1 are represented with a dash (-), and those that could not be accurately quantified due to saturation of the detector are indicated with an asterisk (*). N/A = not available.

Figures

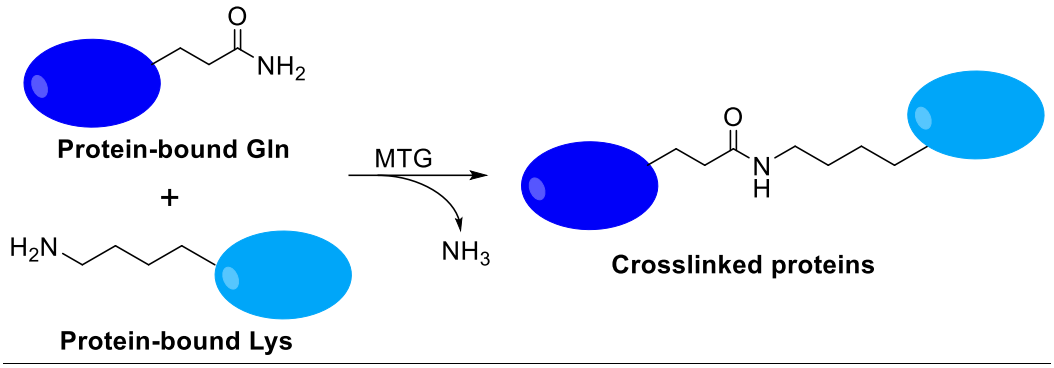

Figure 1. MTG-catalyzed protein crosslinking.

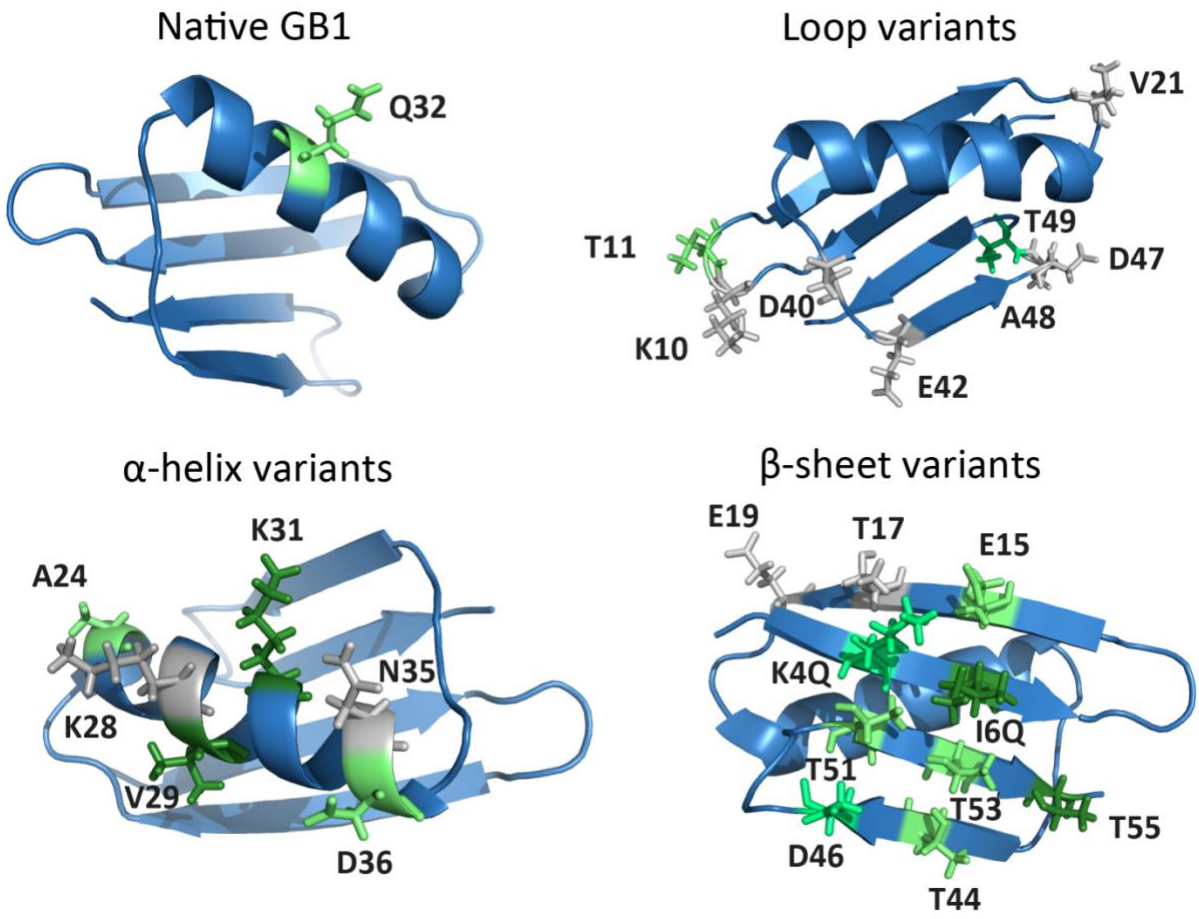

Figure 2. Structures of GB1. Native GB1 (top left), as well as the residues which underwent glutamine substitution; loop variants (top right), helix variants (bottom left), and sheet variants (bottom right). The labeling efficiency of each residue is colored according to the results presented in Table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

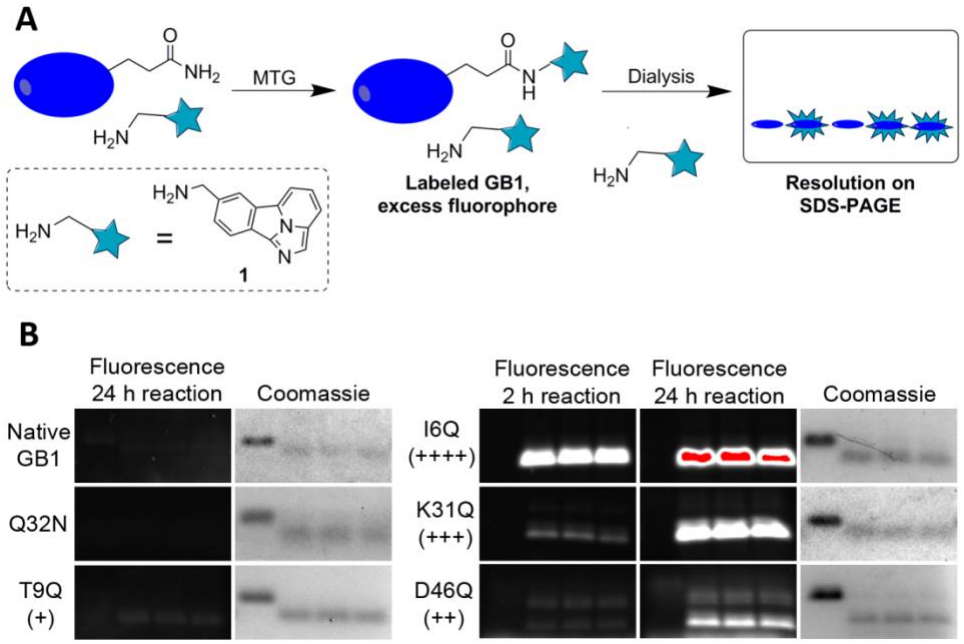

Figure 3. Assay used to conjugate GB1 variants with fluorescent probe **1** and representative SDS-PAGE analysis of fluorescently labeled GB1 variants. A) GB1's single glutamine residue is targeted by MTG, forming an amide bond with the amine-bearing fluorophore. Excess fluorophore is removed by dialysis and reactivity is analyzed using SDS-PAGE. B) Equal quantities of protein were loaded and excited for 5 s using a Cy2 excitation filter prior to Coomassie Brilliant Blue staining. GB1 variants exhibiting low fluorescent conjugation efficiencies were barely visible even after 24 h of reaction time (Native GB1, T9Q; Q32N served as a negative control); for this reason, the 2 h reactions were omitted. Red bands indicate saturation of the detector.

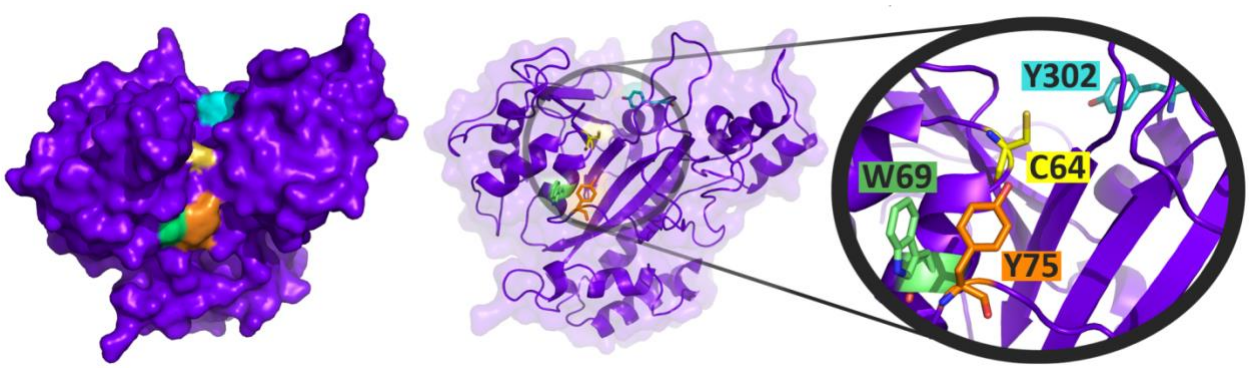

Figure 4. Location of residue substitutions in MTG. Left: A top-down surface view into the active-site crevice, with W69, Y75, and Y302 colored green, orange, and cyan, respectively. Center: cartoon representation with the active site zoomed (right). The catalytic cysteine essential for enzymatic activity, shown in yellow, was not mutated. PDB coordinates 1IU4.

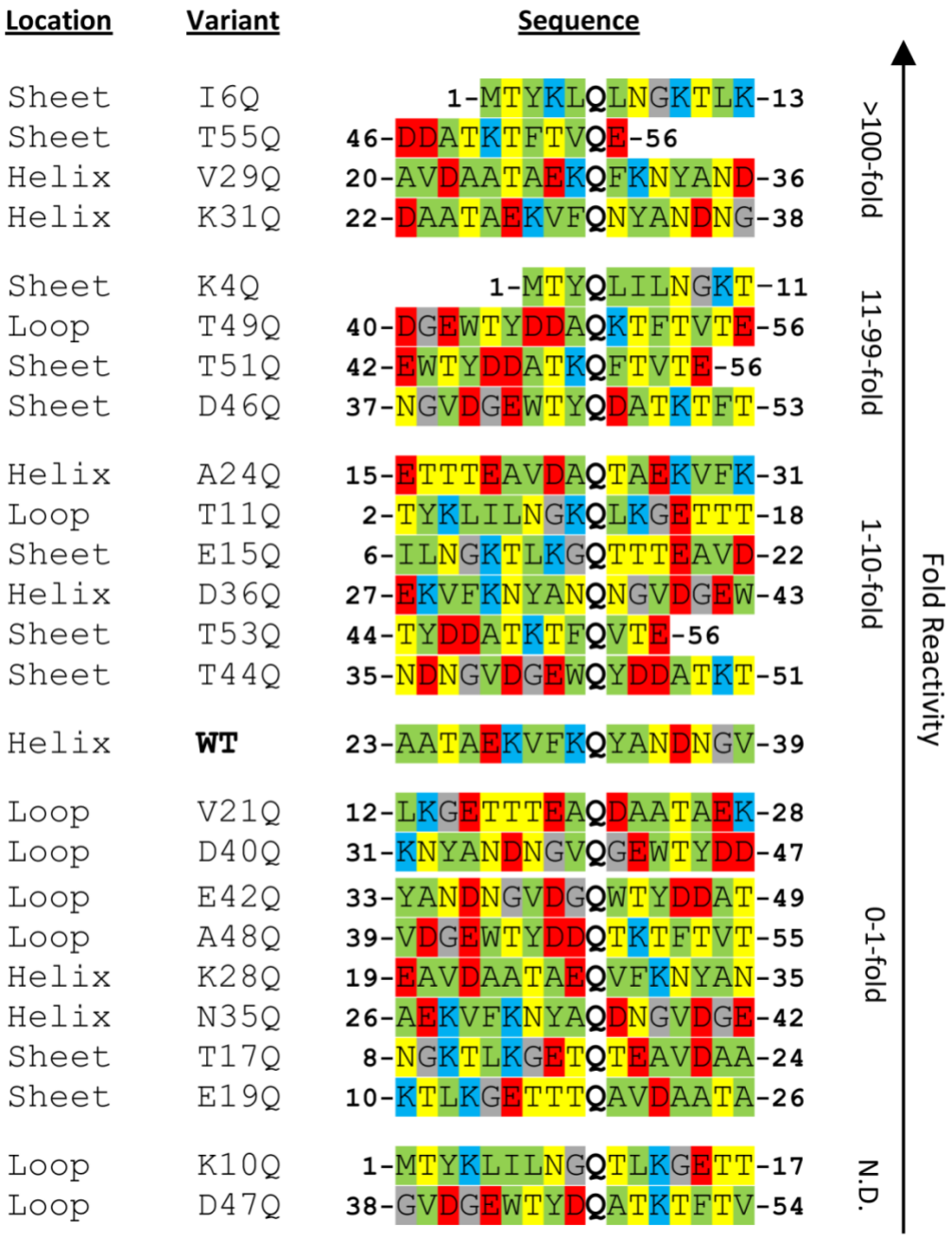

Figure 5. Primary amino acid sequence alignment of GB1 variants, centered on the glutamine residue present in the native or variant GB1s; residue numbering is indicated. Variants are ranked according to their reactivity, with the most reactive variant presented first. Amino acids are colored according to the properties of their side chains: green = hydrophobic; yellow = polar; blue = basic; red = acidic; grey = glycine.