

## Accepted Manuscript

Title: Specificity of Transglutaminase-Catalyzed Peptide Synthesis

Author: Antony D. St-Jacques Natalie M. Rachel Dan R. Curry Steve M.F.G. Gillet Christopher M. Clouthier Jeffrey W. Keillor Joelle N. Pelletier Roberto A. Chica



PII: S1381-1177(15)30101-6  
DOI: <http://dx.doi.org/doi:10.1016/j.molcatb.2015.11.009>  
Reference: MOLCAB 3273

To appear in: *Journal of Molecular Catalysis B: Enzymatic*

Received date: 23-7-2015  
Revised date: 23-10-2015  
Accepted date: 3-11-2015

Please cite this article as: Antony D.St-Jacques, Natalie M.Rachel, Dan R.Curry, Steve M.F.G.Gillet, Christopher M.Clouthier, Jeffrey W.Keillor, Joelle N.Pelletier, Roberto A.Chica, Specificity of Transglutaminase-Catalyzed Peptide Synthesis, Journal of Molecular Catalysis B: Enzymatic <http://dx.doi.org/10.1016/j.molcatb.2015.11.009>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

## Specificity of Transglutaminase-Catalyzed Peptide Synthesis

Antony D. St-Jacques<sup>a,b</sup>, Natalie M. Rachel<sup>c,d,e</sup>, Dan R. Curry<sup>a,b</sup>, Steve M.F.G. Gillet<sup>c</sup>,  
Christopher M. Clouthier<sup>a,b</sup>, Jeffrey W. Keillor<sup>a,b</sup>, Joelle N. Pelletier<sup>c,d,e</sup>, Roberto A. Chica<sup>a,b,d\*</sup>  
rchica@uottawa.ca

<sup>a</sup>Department of Chemistry and Biomolecular Sciences, University of Ottawa, Ottawa, Ontario,  
K1N 6N5, Canada

<sup>b</sup>Centre for Catalysis Research and Innovation, University of Ottawa, Ottawa, Ontario, K1N  
6N5, Canada

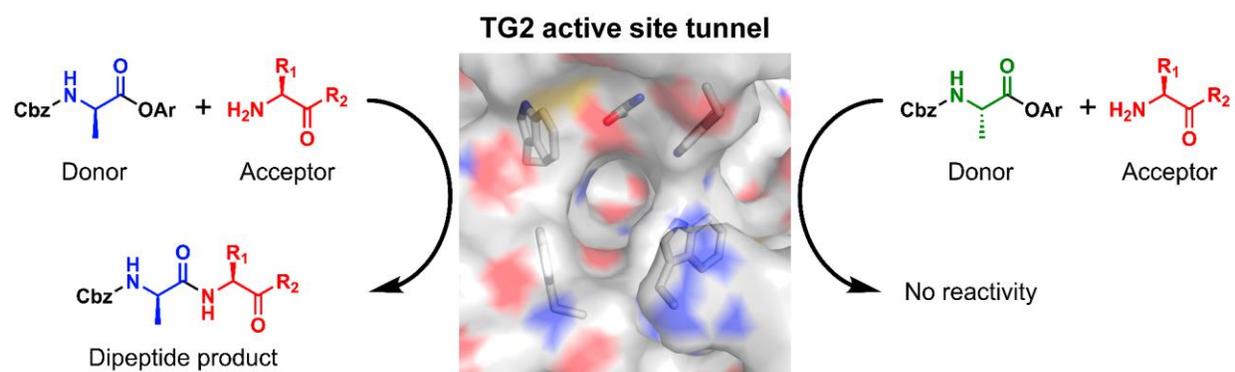
<sup>c</sup>Département de Chimie, Université de Montréal, Montréal, Québec, H3T 1J4, Canada

<sup>d</sup>PROTEO Network, Université Laval, Québec, Québec, G1V 0A6, Canada

<sup>e</sup>Center for Green Chemistry and Catalysis, Montréal, Québec, H3A 0B8, Canada

\*Corresponding author at: Department of Chemistry and Biomolecular Sciences, University of  
Ottawa, 10 Marie Curie, Ottawa, Ontario, Canada, K1N 6N5

## Graphical Abstract



**Highlights**

1. Wild-type TG2 can catalyze synthesis of Cbz-protected dipeptides.
2. Donor substrate specificity is narrow.
3. Aromatic ester derivatives of Gly and D-Ala serve as donor substrates.
4. Hydrophobic cavity formed by Trp332 and Phe334 dictates substrate specificity.
5. Protein engineering based on our computational models may expand substrate scope.

**Abstract**

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

**Abbreviations**

7HC: 7-hydroxycoumarin

Cbz: carbobenzyloxy

TG2: tissue transglutaminase

MD: molecular dynamics

GABA:  $\gamma$ -aminobutyric acid

gTG2: guinea pig liver transglutaminase

NTA: nitrilotriacetic acid

MOPS: 3-(*N*-morpholino)propanesulfonic acid

EDTA: ethylenediaminetetraacetic acid

DMF: *N,N*-dimethylformamide

HRMS: high-resolution mass spectrometry

ESI: electrospray ionization

FAB: fast atom bombardment

LC-MS: liquid chromatography-mass spectrometry

DMSO: dimethyl sulfoxide

HPLC: high-performance liquid chromatography

MOE: molecular operating environment

Boc: *tert*-butyloxycarbonyl

**Keywords:** TG2; enzyme kinetics; LC-MS; biocatalysis; molecular modeling

## 1. Introduction

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

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

intermediate can be hydrolyzed, transforming the acyl-donor glutamine residue into glutamate and regenerating the free enzyme [7].

TG2 enzymes exhibit broad specificity towards the acyl-acceptor substrate [8]. Although the native acyl-acceptor substrate is generally a lysine-containing protein or peptide, many non-natural primary amines, such as glycinamide [9, 10], and anilines, such as *N,N*-dimethyl-1,4-phenylenediamine [11], can also react. However, amines containing free carboxylic acid groups, such as free amino acids, do not act as substrates [10]. On the other hand, TG2 displays narrow specificity for its acyl-donor substrates. The side chain of a protein or peptide-bound L-Gln residue is the native substrate while the side chain of the similar amino acid L-Asn is not reactive [9]. In addition to amides,  $\gamma$ -glutamyl aromatic ester derivatives of L-Glu, such as *N*-carbobenzyloxy-L-glutamyl( $\gamma$ -*p*-nitrophenyl ester)glycine (Fig. 1A), have also been shown to be acyl-donor substrates of TG2 and are used to measure the enzyme's activity [12]. However, secondary amide derivatives of L-Gln, such as *N*- $\gamma$ -methyl-L-glutamine or anilides, are not substrates of TG2 [13]: the  $\gamma$ -carboxamide group of L-Gln is the only known amide that is an acyl-donor substrate of TG2.

We and others previously demonstrated that TG2 could use a novel class of acyl-donor substrates that are neither L-Gln nor L-Glu derivatives [14, 15]. Namely, 4-(*N*-carbobenzyloxyglycylamino)-butyric acid-coumarin-7-yl ester (Cbz-Gly-GABA-7HC) and 4-(*N*-carbobenzyloxyphenylalanylamino)-butyric acid-coumarin-7-yl ester (Cbz-Phe-GABA-7HC) (Fig. 1B) can react with TG2 to release 7-hydroxycoumarin (7HC), resulting in a fluorescence increase that makes these compounds useful for quantifying TG2 reaction rates. The scaffolds of these substrates, based on known irreversible inhibitors of TG2 [16, 17], differ from L-Glu aromatic ester acyl-donor substrates of TG2 in that the reactive ester function is located on the

main chain of the peptide analogue, rather than on the side chain. As a result, they give rise to products that do not contain a  $\gamma$ -glutamyl- $\epsilon$ -lysyl isopeptide bond. An analogue in which the 7HC leaving group is attached directly to the glycine residue carboxylate group, *N*-carbobenzyloxyglycyl-coumarin-7-yl ester (Cbz-Gly-7HC, Fig. 1C), is also a donor substrate of TG2 [18]. Significantly, the reaction of this substrate with an acceptor amine substrate would result in the formation of a peptide-like  $\alpha$ -amide bond (Scheme 1). These results illustrate that specificity for acyl-donor substrates with aromatic ester functions is broader than had previously been supposed and demonstrate that the enzyme can generate products with novel scaffolds.

In this study, we investigate guinea pig liver TG2 (gTG2)-catalyzed peptide bond formation between the Cbz-Gly-7HC donor substrate in combination with various amino acid-derived acceptors. Using LC-MS analysis of the reaction products, we demonstrate that the enzyme is able to react directly with the  $\alpha$ -carboxyl group of Cbz-Gly-7HC to form Gly-Xaa dipeptide products, confirming that its natural transamidation activity can be co-opted for peptide synthesis. Additionally, we explore the substrate specificity of the enzyme in peptide synthesis by measuring its reactivity toward a variety of potential acyl-donor substrates having an aromatic ester function on the  $\alpha$ -carboxyl group of various amino acids. We observed that the aromatic ester of Gly is an efficient acyl-donor substrate; the aromatic ester of D-Ala is also reactive though to a lesser extent, and that of L-Ala showed no detectable reactivity.

To elucidate how the stereochemical configuration of the side-chain of alanine-containing donor substrates affects gTG2 catalytic efficiency, we used a computational strategy combining computational protein design and molecular dynamics simulations to model the binding modes of donors in the gTG2 active site. We show that the inability of gTG2 to efficiently catalyze peptide synthesis from donors other than Cbz-Gly-7HC results from the narrow substrate binding

tunnel, which prevents bulkier donors to adopt a catalytically productive binding mode. Our observations pave the way to future protein engineering efforts to expand the substrate scope of gTG2 in peptide synthesis, which may lead to useful biocatalysts for the synthesis of desirable bioactive molecules.

## 2. Materials and Methods

### 2.1 Materials

All reagents used were of the highest available purity. Lysozyme, 7HC, *N*-acetyl-L-lysine methyl ester hydrochloride (*N*-AcLysOMe), *N*-carbobenzyloxy-L-glutaminyglycine (Cbz-L-Gln-Gly), glycineamide (GlyNH<sub>2</sub>) and L-leucine methyl ester (LeuOMe) hydrochlorides were purchased from Sigma-Aldrich. L-Alanineamide hydrochloride (AlaNH<sub>2</sub>) was purchased from Novabiochem (Mississauga, ON). Ni-NTA agarose resin was purchased from Qiagen (Mississauga, ON). Restriction enzymes and DNA-modifying enzymes were from New England Biolabs. All aqueous solutions were prepared using water purified with a Millipore BioCell system.

### 2.2 Synthesis of donor substrates

#### 2.2.1. Synthesis of Cbz-Gly-7HC and Cbz-L-Ala-7HC

The synthesis of Cbz-Gly-7HC was based on a previously reported protocol [14]. Namely, 0.2 g (1 mmol) of Cbz-Gly and 0.4 g (2.5 mmol) of 7HC were dissolved in 10 mL of ethyl acetate. Then, 0.22 mL (0.2 g, 2 mmol) of *N*-methylmorpholine and 0.8 mL (0.63 g, 5 mmol) of *N,N*-diisopropylcarbodiimide were added with stirring at room temperature. Stirring was continued until the complete disappearance of Cbz-Gly, as followed by thin layer

chromatography (ethyl acetate). The reaction mixture was then washed once with 1 M NaOH, three times with 0.1 M NaOH, 3 times with 0.1 M HCl, once with saturated NaHCO<sub>3</sub>, and once with brine. The organic phase was then dried over MgSO<sub>4</sub>, filtered and evaporated under reduced pressure. The resulting residue was purified by silica gel chromatography (ethyl acetate) to remove traces of diisopropylurea, giving the desired ester in 70 % yield (0.25 g). Cbz-L-Ala-7HC was synthesized according to a similar protocol.

Cbz-Gly-7HC. <sup>1</sup>H NMR (300 MHz, CDCl<sub>3</sub>): δ 4.30 (2H, d), 5.18 (2H, s), 5.33 (1H, s), 6.44 (1H, d), 7.08 (1H, d), 7.10 (1H, s), 7.37 (5H, m), 7.51 (1H, d), 7.70 (1H, d). <sup>13</sup>C NMR (75 MHz, CDCl<sub>3</sub>): δ 168.5, 160.6, 156.7, 154.9, 152.9, 143.1, 136.3, 129.0, 128.9, 128.6, 128.5, 118.4, 117.2, 116.6, 110.5, 67.7, 43.2. HRMS (FAB) calculated for C<sub>19</sub>H<sub>16</sub>NO<sub>6</sub> ([M<sup>+</sup>H]<sup>+</sup>): 354.0972, found 354.0968.

Cbz-L-Ala-7HC. <sup>1</sup>H NMR (300 MHz, CDCl<sub>3</sub>): δ 1.63 (3H, d), 4.62 (1H, m), 5.18 (2H, s), 5.33 (1H, d), 6.42 (1H, d), 7.07 (1H, d), 7.09 (1H, s), 7.39 (5H, m), 7.51 (1H, d), 7.70 (1H, d). <sup>13</sup>C NMR (75 MHz, CDCl<sub>3</sub>): δ 171.5, 160.5, 156.0, 154.9, 153.1, 143.1, 136.4, 129.0, 128.9, 128.6, 128.5, 118.4, 117.2, 116.6, 110.5, 67.5, 50.2, 18.5. HRMS (FAB) calculated for C<sub>20</sub>H<sub>18</sub>NO<sub>6</sub> ([M<sup>+</sup>H]<sup>+</sup>): 368.1129, found 368.1118.

### 2.2.2. Synthesis of Cbz-D-Ala-7HC

The synthesis of Cbz-D-Ala-7HC followed the protocol employed for Cbz-L-Ala-7HC. Equimolar amounts of Cbz-D-Ala (4 mmol, 0.89 g) and 7HC (4 mmol, 0.65 g) were dissolved in 15 mL of dichloromethane at room temperature. To the stirring solution, 0.38 mL (4.4 mmol) of *N*-methylmorpholine and 0.82 mL (8 mmol) of *N,N*-diisopropylcarbodiimide were added. The consumption of Cbz-D-Ala was monitored by thin-layer chromatography. Upon completion, the

reaction mixture was washed successively with 0.1 M NaOH, 0.1 M HCl, saturated sodium bicarbonate, and brine. The organic phase was dried over MgSO<sub>4</sub>, filtered, and concentrated under reduced pressure. The crude product was purified *via* flash column using chloroform/methanol (9:1), affording the titular compound in 66 % yield (0.59 g).

Cbz-D-Ala-7HC. <sup>1</sup>H NMR (CDCl<sub>3</sub>, 400 MHz): δ 1.41 (d, 3H), 4.25 (q, 1H), 5.10 (s, 2H), 6.63 (d, 1H), 7.37 (m, 5H), 7.55 (br s, 1H), 7.71 (d, 1H), 7.84 (d, 1H). <sup>13</sup>C NMR (CDCl<sub>3</sub>, 100 MHz): δ 17.6, 53.5, 67.4, 110.1, 115.4, 116.1, 118.7, 127.9, 128.1, 129.7, 136.5, 146.3, 156.6, 156.9, 157.1, 161.8, 169.3. HRMS (ESI) Calculated for C<sub>20</sub>H<sub>17</sub>NO<sub>6</sub>: 367.1056. Found: 367.1060.

### 2.2.3. Synthesis of Cbz-GlyNH<sub>2</sub>

The synthesis of Cbz-GlyNH<sub>2</sub> was adapted from a previously reported protocol [19]. Glycinamide hydrochloride (18 mmol, 2.00 g) was dissolved in water (60 mL) and acetone (8 mL), prior to the addition of Na<sub>2</sub>CO<sub>3</sub> (54 mmol, 5.7 g) and NaHCO<sub>3</sub> (18 mmol, 1.5 g). Benzyl chloroformate (22 mmol, 3.20 mL) was added dropwise to the stirring solution over the course of 30 minutes. The resulting mixture was stirred for 3 hours at room temperature, after which the products were isolated by washing with diethyl ether (50 mL). The protected product was precipitated out of solution by the slow addition of 0.1 M HCl. The precipitate was filtered and subsequently dried *in vacuo* to afford a white solid in 86% yield (3.24 g).

<sup>1</sup>H NMR (CDCl<sub>3</sub>, 400 MHz): δ 3.87 (s, 2H), 5.07 (s, 2H), 7.26 (m, 5H), 7.35 (s, 2H), 7.95 (s, 1H). <sup>13</sup>C NMR (CDCl<sub>3</sub>, 100 MHz): δ 45.1, 67.3, 126.2, 126.9, 128.9, 136.3, 156.8, 170.1. HRMS (ESI) calculated for C<sub>10</sub>H<sub>12</sub>N<sub>2</sub>O<sub>3</sub>: 208.0848. Found: 208.0851.

### 2.3 Overexpression and purification of gTG2

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

#### 2.4 Specific activity

The hydroxamate assay [9] was used to quantify gTG2 activity. Briefly, gTG2 was incubated at 37 °C for 10 minutes with 30 mM Cbz-L-Gln-Gly and 100 mM hydroxylamine in 200 mM Tris-acetate buffer (pH 7.0) containing 5 mM CaCl<sub>2</sub> and 1 mM EDTA. The reaction was quenched with a solution containing 2.0 M ferric chloride hexahydrate, 0.3 M trichloroacetic acid, and 0.8 M HCl. The mixture was vortexed and left at room temperature for 10 minutes before measuring absorbance at 525 nm. One unit (U) of gTG2 produces 1 μmol of L-glutamic acid γ-monohydroxamate per minute at 37 °C.

#### 2.5 Kinetic assays

All assays were performed in triplicate. The following solutions were prepared: a standard stock buffer solution (100 mM MOPS buffer pH 7.0, 5 mM CaCl<sub>2</sub>, and 0.05 mM EDTA), a 2 mM (Cbz-Gly-7HC, Cbz-L-Ala-7HC) or 100 mM (Cbz-D-Ala-7HC) solution of acyl-donor substrate in *N,N*-dimethylformamide (DMF), and a 200 mM solution of acyl-acceptor

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

## 2.6 LC-MS

Reaction mixtures containing 150  $\mu$ M of ester donor substrate (Cbz-Gly-7HC, Cbz-L-Ala-7HC, or Cbz-D-Ala-7HC) or 20 mM of amide donor substrate (Cbz-GlyNH<sub>2</sub>) and 50 mM of acceptor substrate (GlyNH<sub>2</sub>, AlaNH<sub>2</sub>, or LeuOMe) were prepared in a buffer composed of 100 mM MOPS pH 7.0, 5 mM CaCl<sub>2</sub>, 0.05 mM EDTA, and 5 % DMF. The pH of each substrate mixture was verified with indicator paper prior to the addition of enzyme. The reaction was initiated upon the addition of 0.1 U/mL of gTG2 (or an equivalent volume of buffer for reactions

run in the absence of enzyme) in a final volume of 2 mL. Reactions were incubated at 37 °C for up to 20 min. Control reactions without gTG2 or without amine acceptor were run for each combination of substrate mixtures. Experiments and controls were performed in triplicate.

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

### *2.7 Homology modeling*

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

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

### 2.8 Construction of acyl-enzyme intermediates

Using MOE, the catalytic Cys residue (Cys277) on the gTG2 homology model was acetylated. The carbonyl moiety of the acetyl group was then manually oriented *via* dihedral angle rotation to form a hydrogen bond with the indole nitrogen of Trp241, which has been shown to be essential for catalytic activity, presumably by stabilizing the transition state [29]. Following energy minimization as described above (root-mean-square gradient below 0.1 kcal mol<sup>-1</sup> Å<sup>-1</sup>) to optimize H-bonding between the acetyl group and Trp241, the acetylated Cys277 residue was extracted from the gTG2 structure and used as a template to build a Cys residue acylated with the Cbz-glycyl moiety in MOE. The added Cbz and Gly atoms were then energy minimized *in vacuo* by conjugate gradient minimization to a root-mean-square gradient below 0.1 kcal mol<sup>-1</sup> Å<sup>-1</sup> in order to refine bond lengths and angles. The resulting minimized acylated Cys residue was then used to generate rotamers via the introduction of the following dihedrals: O1-C1-C2-N1, 90 ± 20° and -90 ± 20°; C1-C2-N1-C3, 180 ± 20°; C2-N1-C3-O3, 180°; N1-C3-

O3-C4,  $120 \pm 20^\circ$ ,  $180 \pm 20^\circ$ , and  $-120 \pm 20^\circ$ ; C3-O3-C4-C5,  $60 \pm 20^\circ$ ,  $180 \pm 20^\circ$ , and  $-60 \pm 20^\circ$ ; O3-C4-C5-C6A,  $0 \pm 20^\circ$ ,  $60 \pm 20^\circ$ , and  $120 \pm 20^\circ$  (see Fig. S1 for atom names). The internal energy of the resulting 13,122 rotamers was evaluated *in vacuo* using MOE with the previously described force field, and only rotamers whose energy was within 10 kcal/mol from the lowest energy rotamer were included in the final rotamer library. A similar procedure was utilized to prepare a rotamer library for Cys277 acylated with the Cbz-D-alanyl moiety.

To build the gTG2 acyl-enzyme intermediates, computational protein design was performed using the fast and accurate side-chain topology and energy refinement (FASTER) algorithm as implemented in PHOENIX [30-32]. Rotamers for acylated Cys277 and surrounding residues (positions Gln169, Trp241, Asn243, Tyr245, Met252, Gln276, Trp278, Phe316, Arg317, Met330, Trp332, Asn333, Phe334, His335, and Cys336) were optimized on the fixed backbone of the gTG2 homology model. The backbone independent Dunbrack rotamer library with expansions of  $\pm 1$  standard deviation around  $\chi_1$  and  $\chi_2$  [33] was used to model side-chain conformations. A four-term potential energy function consisting of a van der Waals term from the Dreiding II force field with atomic radii scaled by 0.9 [34], a direction sensitive hydrogen-bond term with well depth at 8.0 kcal/mol [35], an electrostatic energy term modelled using Coulomb's law with a distance dependent dielectric of 10, and an occlusion-based solvation penalty term [31] were used to evaluate rotamer combinations. The lowest energy acyl-enzyme intermediate structure obtained from each donor was retained for further analysis.

## 2.9 Molecular dynamics

For generation of molecular dynamics (MD) trajectories, structures of the gTG2 acyl-enzyme intermediates prepared as described in Section 2.8 were used as templates. The thioester

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

### 3. Results

#### 3.1 Wild-type gTG2 can form peptide bonds

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

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

We also determined the kinetic parameters for the gTG2-catalyzed transamidation reaction of acyl-donor substrate Cbz-Gly-7HC with the widely used acyl-acceptor substrate *N*-AcLysOMe. As previously observed with other amino acid derivatives [18], we confirmed that the rate of release of 7HC from Cbz-Gly-7HC increased in the presence of *N*-AcLysOMe relative to the rate of hydrolysis. This result suggests that gTG2 catalyzes the formation of a covalent bond between the  $\alpha$ -carboxyl group of Gly in Cbz-Gly-7HC and the  $\epsilon$ -amino group of Lys in *N*-AcLysOMe, whose apparent  $K_M$  and  $k_{\text{cat}}$  values were determined to be  $4 \pm 1 \text{ mM}$  and  $0.20 \pm 0.02 \text{ s}^{-1}$ , respectively.

Next, we confirmed gTG2-mediated peptide synthesis by LC-MS analysis to identify the reaction products. We assayed derivatives of three different amino acids previously shown to act as acyl-acceptor substrates of gTG2 [18] in which the negatively-charged carboxylate is neutralized under the form of a primary amide or a methyl ester. Chromatograms of the reaction time-course of Cbz-Gly-7HC with acceptors GlyNH<sub>2</sub> (Fig. 2A), AlaNH<sub>2</sub> (Fig. 2B), and LeuOMe

(Fig. 2C) unequivocally demonstrate the time-dependent increase in concentration of the corresponding dipeptide product. Since the Cbz-Gly-7HC donor substrate is an activated ester, it is highly reactive with nucleophilic amines such as the amino acid derivatives tested herein. This is illustrated by the fact that the LC-MS chromatograms also show significant amounts of dipeptide product being formed in the absence of gTG2 (2.4- to 4-fold catalyzed/uncatalyzed product ratio at the 20 min time-point). However, dipeptide products are formed more rapidly in the presence of enzyme, confirming the intrinsic peptide synthase activity of wild-type gTG2. This is particularly clear at the earlier time points, where the catalyzed/uncatalyzed product ratio is 4- to 8-fold after 10 min and greater than 10-fold for GlyNH<sub>2</sub> at the 2 min time point.

### 3.2 Donor substrate specificity of gTG2-catalyzed peptide synthesis

Having confirmed that gTG2 catalyzes the synthesis of Cbz-Gly-Xaa dipeptides, we next investigated whether alternate donor substrates could be utilized. Because the natural substrate of gTG2 is the  $\gamma$ -carboxamide group of an L-Gln residue, we first tested the amide analog of Cbz-Gly-7HC, Cbz-GlyNH<sub>2</sub>, using GlyNH<sub>2</sub> as the acceptor substrate. Cbz-GlyNH<sub>2</sub> does not react with gTG2 at concentrations up to 50 mM (Fig. S2), its solubility limit. It has previously been observed that aromatic ester acyl-donor substrates of gTG2 have a lower  $K_M$  value than the corresponding amide: Cbz-L-Gln-Gly has an apparent  $K_M$  of 4.1 mM in the hydrolysis reaction [38] whereas its aromatic ester analog *N*-carbobenzyloxy-L-glutamyl( $\gamma$ -*p*-nitrophenyl ester)glycine has an apparent  $K_M$  of 0.02 mM [12]. The two orders of magnitude lower  $K_M$  of the aromatic ester could be due to improved binding conferred by the *p*-nitrophenol aromatic leaving group. This improved binding could also occur in Cbz-Gly-7HC relative to Cbz-GlyNH<sub>2</sub> through

beneficial  $\pi$ -stacking interactions between the aromatic leaving group and the aromatic side chains of the tunnel-wall residues of gTG2 (Trp241 and Trp332) [23, 39].

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

### *3.3 Structural basis for donor substrate specificity of gTG2*

To elucidate the structural basis for the observed acyl-donor specificity of gTG2, we generated models of the acyl-enzyme intermediates formed during the gTG2-catalyzed hydrolysis of the Cbz-Gly-7HC and Cbz-D-Ala-7HC substrates. We did not generate an acyl-enzyme intermediate structure for hydrolysis of Cbz-L-Ala-7HC as this compound is not a gTG2 substrate. In the acyl-enzyme intermediates, the catalytic Cys277 residue of gTG2 is covalently bound to the Cbz-glycyl or Cbz-D-alanyl moiety through a thioester bond. In the acyl-enzyme

intermediate for gTG2-catalyzed hydrolysis of Cbz-Gly-7HC, the glycyl group fits, with no steric clashes, into a tunnel formed by residues Trp241, Gln276, Trp278, Trp332, and Phe334 (Fig. S3), while the Cbz phenyl ring is positioned outside the tunnel and lies in a cleft on the surface of the enzyme. Closer inspection of this acyl-enzyme intermediate suggests that the presence of a methyl side chain on the C<sub>α</sub> atom resulting in an L or D configuration would be detrimental to binding as it would clash with either residue Phe334 or Trp332, respectively (Fig. 3A). This is indeed what is observed in the acyl-enzyme intermediate structure of gTG2 with a Cbz-D-alanyl moiety (Fig. 3B). In this model structure, the side chain of Trp332 adopts an alternate conformation, presumably to alleviate unfavorable steric interactions with the methyl side chain of D-Ala.

Based on these observations, we hypothesized that the methyl side chain of alanine is detrimental to activity because it decreases productive binding for Cbz-D-Ala-7HC and abolishes binding for Cbz-L-Ala-7HC. To test these hypotheses, we generated noncovalent complexes of gTG2 bound to the hydrolysis products Cbz-Gly, Cbz-L-Ala, and Cbz-D-Ala from the acyl-enzyme structures, and used these complexes as input structures for molecular dynamics simulations. The goal of these simulations was to evaluate the binding modes of products in the gTG2 active site. Carboxylic acid products of hydrolysis were selected as ligands because gTG2 should be able to bind these compounds due to microscopic reversibility and because we could not unambiguously specify where the 7HC group would bind. To evaluate the efficiency with which these compounds are bound in the gTG2 active site, we measured the distance between the nucleophilic sulfur atom of the catalytic residue Cys277 and the electrophilic carbonyl carbon of the Cbz-Gly, Cbz-D-Ala, or Cbz-L-Ala products during the course of a 1.5-nanosecond MD simulation. As shown in Figure 4A, sulfur-carbon distances for Cbz-Gly and Cbz-D-Ala are

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

In light of our results, we propose that Cbz-D-Ala-7HC is a poor substrate and that Cbz-L-Ala-7HC is not a substrate of gTG2 because their methyl side chain clashes with tunnel wall residue Trp332 or Phe334, respectively. We postulate that the clash between the methyl side chain of D-Ala and Trp332 can be more easily accommodated in the active site because Trp332 is located on a loop formed by residues Asn318-Asn333. This loop is highly flexible, as illustrated by the fact that no electronic density is present for most of the residues comprising it in the human TG2 crystal structure [23]. The higher mobility of this loop would enable Trp332 to move out of the way from the methyl side chain of D-Ala, allowing retention of catalytic activity, albeit at a lower level. On the other hand, Phe334 is part of a  $\beta$ -sheet formed by residues Thr295-Phe301 and Phe334-Trp341, and its phenyl side chain is stacked against the backbone of residues Gln169 and Gly170. These interactions would make Phe334 more rigid, preventing it from moving away from the methyl side chain of L-Ala. These observations raise the possibility that the acyl-donor substrate scope of gTG2 may be expanded by mutating Phe334 and Trp332 in order to increase the space available for substrates containing alternate side chains.

#### 4. Discussion

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

Protease-catalyzed peptide synthesis is the topic of much current research [3]. Serine and cysteine proteases can catalyze peptide synthesis through a kinetically controlled process in which the protease (hydrolase) acts as a transamidase [44, 45]. This process requires a protease that can form a covalent acyl-enzyme intermediate, as is the case with gTG2. Competition between hydrolysis and aminolysis is always present during the degradation of this acyl-enzyme intermediate, resulting in lower yields for the synthesis of peptides, since proteases are not efficient at excluding water from their active site. Furthermore, proteases hydrolyze the peptide products, further lowering the overall yield of peptide synthesis.

Proteases used in kinetically controlled peptide synthesis have reported transamidation/hydrolysis ratios in the range of  $10^2$ - $10^4$  [46, 47] whereas gTG2 has a similar  $10^2$  increase in rate of transamidation relative to hydrolysis when the acceptor substrate is hydroxylamine [12]. Further, the catalytic efficiency of wild-type gTG2 for the synthesis of

various Cbz-Gly-L-Xaa dipeptides ranges from 12 to 141  $\text{M}^{-1}\text{s}^{-1}$  [18]. Papain, a cysteine protease that has been used in peptide synthesis, has catalytic efficiencies of 5 and 49  $\text{M}^{-1}\text{s}^{-1}$  for the synthesis of the Boc-Gly-L-Phe- $\text{N}_2\text{H}_2\text{Ph}$  dipeptide and the Boc-L-Tyr(Bzl)-Gly-Gly-L-Phe-L-Leu- $\text{N}_2\text{H}_2\text{Ph}$  pentapeptide, respectively [48]. These comparisons suggest that gTG2 could also be used as a catalyst for the synthesis of peptide bonds.

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

## 5. Conclusion

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

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

### **Acknowledgements**

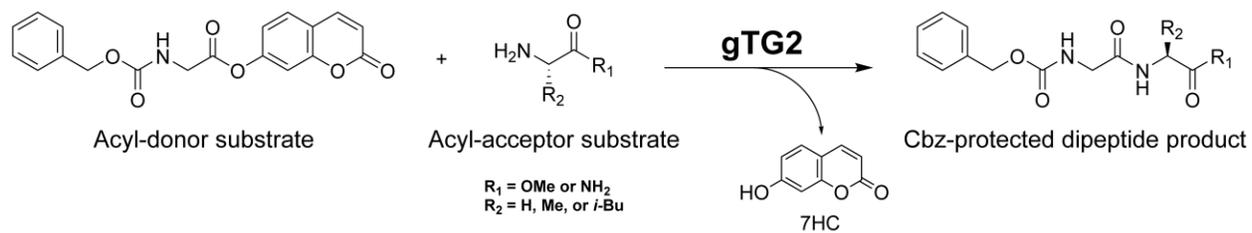
R. A. C., J. N. P and J. W. K. acknowledge grants from the Natural Sciences and Engineering Research Council of Canada (NSERC), and the Canada Foundation for Innovation. R. A. C. also acknowledges a grant from the Ontario Research Fund; J. N. P and J. W. K. acknowledge a grant from FRQNT. A. D. S. is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

## References

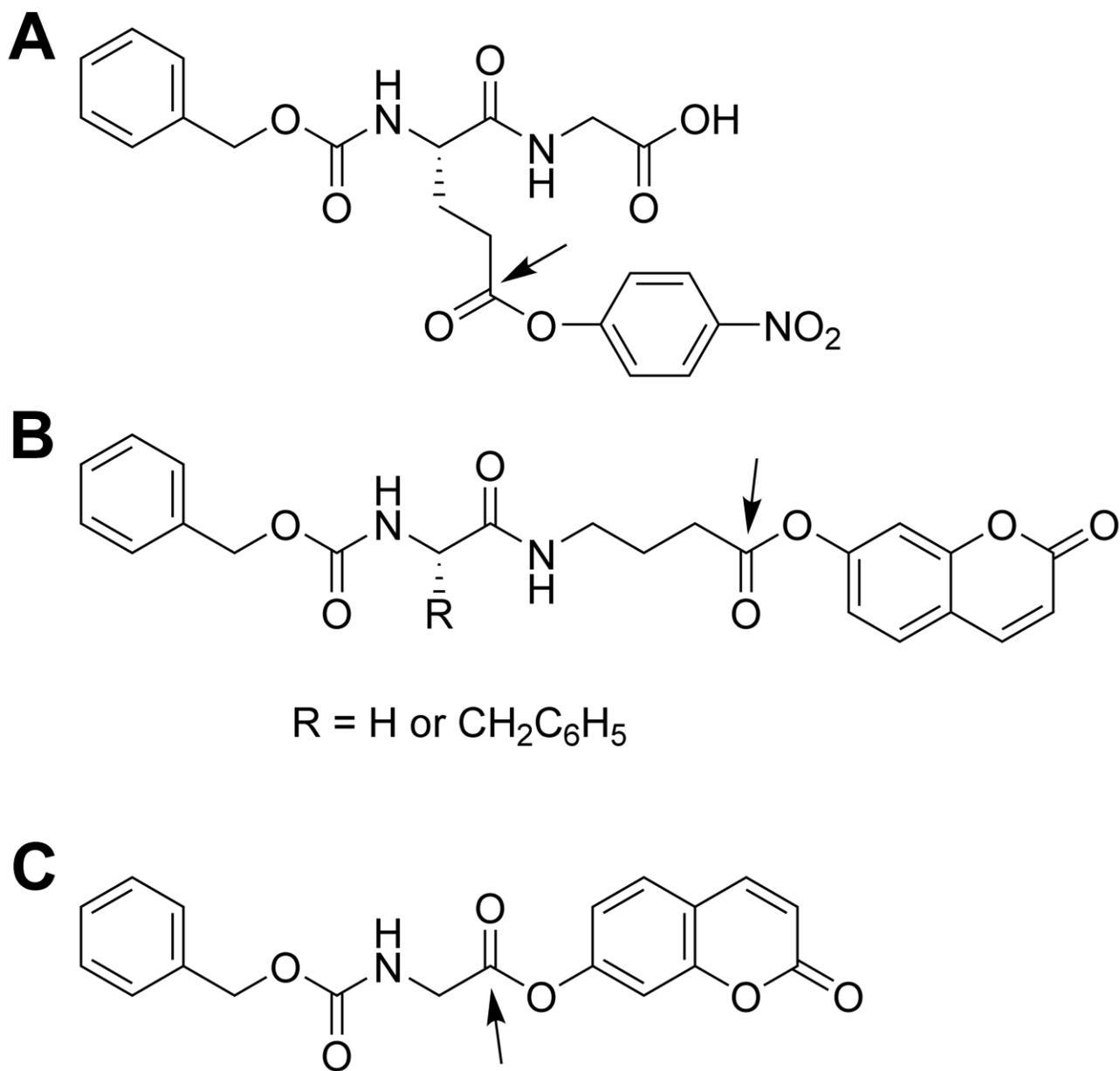
- [1] V.R. Pattabiraman, J.W. Bode, *Nature*, 480 (2011) 471-479.
- [2] D.J.C. Constable, P.J. Dunn, J.D. Hayler, G.R. Humphrey, J.L. Leazer, R.J. Linderman, K. Lorenz, J. Manley, B.A. Pearlman, A. Wells, A. Zaks, T.Y. Zhang, *Green Chem*, 9 (2007) 411-420.
- [3] K. Yazawa, K. Numata, *Molecules*, 19 (2014) 13755.
- [4] L. Fesus, M. Piacentini, *Trends in biochemical sciences*, 27 (2002) 534-539.
- [5] M. Griffin, R. Casadio, C.M. Bergamini, *Biochem J*, 368 (2002) 377-396.
- [6] L. Lorand, R.M. Graham, *Nature reviews. Molecular cell biology*, 4 (2003) 140-156.
- [7] J.W. Keillor, C.M. Clouthier, K.Y. Apperley, A. Akbar, A. Mulani, *Bioorganic chemistry*, 57 (2014) 186-197.
- [8] D. Aeschlimann, M. Paulsson, *Thrombosis and haemostasis*, 71 (1994) 402-415.
- [9] J.E. Folk, S.I. Chung, *Methods Enzymol*, 113 (1985) 358-375.
- [10] D.D. Clarke, M.J. Mycek, A. Neidle, H. Waelsch, *Archives of biochemistry and biophysics*, 79 (1959) 338-354.
- [11] P. de Macedo, C. Marrano, J.W. Keillor, *Analytical Biochemistry*, 285 (2000) 16-20.
- [12] A. Leblanc, C. Gravel, J. Labelle, J.W. Keillor, *Biochemistry*, 40 (2001) 8335-8342.
- [13] D. Halim, K. Caron, J.W. Keillor, *Bioorganic & Medicinal Chemistry Letters*, 17 (2007) 305-308.
- [14] S.M. Gillet, J.N. Pelletier, J.W. Keillor, *Anal Biochem*, 347 (2005) 221-226.
- [15] C. Klock, X. Jin, K. Choi, C. Khosla, P.B. Madrid, A. Spencer, B.C. Raimundo, P. Boardman, G. Lanza, J.H. Griffin, *Bioorg Med Chem Lett*, 21 (2011) 2692-2696.
- [16] K. Choi, M. Siegel, J.L. Piper, L. Yuan, E. Cho, P. Strnad, B. Omary, K.M. Rich, C. Khosla, *Chemistry & Biology*, 12 (2005) 469-475.
- [17] C. Pardin, S.M.F.G. Gillet, J.W. Keillor, *Bioorgan Med Chem*, 14 (2006) 8379-8385.
- [18] J.W. Keillor, R.A. Chica, N. Chabot, V. Vinci, C. Pardin, E. Fortin, S.M.F.G. Gillet, Y. Nakano, M.T. Kaartinen, J.N. Pelletier, W.D. Lubell, *Can J Chem*, 86 (2008) 271-276.
- [19] A.D. Pehere, A.D. Abell, *Tetrahedron Lett*, 52 (2011) 1493-1494.
- [20] S.M. Gillet, R.A. Chica, J.W. Keillor, J.N. Pelletier, *Protein Expr Purif*, 33 (2004) 256-264.
- [21] C. Pardin, I. Roy, R.A. Chica, E. Bonneil, P. Thibault, W.D. Lubell, J.N. Pelletier, J.W. Keillor, *Biochemistry*, 48 (2009) 3346-3353.
- [22] M.A. Larkin, G. Blackshields, N.P. Brown, R. Chenna, P.A. McGettigan, H. McWilliam, F. Valentin, I.M. Wallace, A. Wilm, R. Lopez, J.D. Thompson, T.J. Gibson, D.G. Higgins, *Bioinformatics*, 23 (2007) 2947-2948.
- [23] D.M. Pinkas, P. Strop, A.T. Brunger, C. Khosla, *PLoS Biol*, 5 (2007) e327.
- [24] N. Eswar, B. Webb, M.A. Marti-Renom, M.S. Madhusudhan, D. Eramian, M.Y. Shen, U. Pieper, A. Sali, *Current protocols in bioinformatics / editorial board, Andreas D. Baxeavanis ... [et al.]*, Chapter 5 (2006) Unit 5 6.
- [25] P. Labute, *Proteins-Structure Function and Bioinformatics*, 75 (2009) 187-205.
- [26] Chemical Computing Group Inc., in, *Chemical Computing Group Inc.*, 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2012.
- [27] J. Wang, P. Cieplak, P.A. Kollman, *Journal of Computational Chemistry*, 21 (2000) 1049-1074.
- [28] I.W. Davis, A. Leaver-Fay, V.B. Chen, J.N. Block, G.J. Kapral, X. Wang, L.W. Murray, W.B. Arendall, 3rd, J. Snoeyink, J.S. Richardson, D.C. Richardson, *Nucleic Acids Res*, 35 (2007) W375-383.
- [29] S.N. Murthy, S. Iismaa, G. Begg, D.M. Freymann, R.M. Graham, L. Lorand, *Proc Natl Acad Sci U S A*, 99 (2002) 2738-2742.
- [30] B.D. Allen, A. Nisthal, S.L. Mayo, *Proc Natl Acad Sci U S A*, 107 (2010) 19838-19843.
- [31] R.A. Chica, M.M. Moore, B.D. Allen, S.L. Mayo, *Proc Natl Acad Sci U S A*, 107 (2010) 20257-20262.

- [32] H.K. Privett, G. Kiss, T.M. Lee, R. Blomberg, R.A. Chica, L.M. Thomas, D. Hilvert, K.N. Houk, S.L. Mayo, *Proc Natl Acad Sci U S A*, 109 (2012) 3790-3795.
- [33] R.L. Dunbrack, F.E. Cohen, *Protein Science*, 6 (1997) 1661-1681.
- [34] S.L. Mayo, B.D. Olafson, W.A. Goddard, *J. Phys. Chem.*, 94 (1990) 8897-8909.
- [35] B.I. Dahiyat, S.L. Mayo, *Proc Natl Acad Sci U S A*, 94 (1997) 10172-10177.
- [36] P. Gerber, K. Müller, *Journal of computer-aided molecular design*, 9 (1995) 251-268.
- [37] J.C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R.D. Skeel, L. Kalé, K. Schulten, *Journal of Computational Chemistry*, 26 (2005) 1781-1802.
- [38] N. Day, J.W. Keillor, *Anal Biochem*, 274 (1999) 141-144.
- [39] R.A. Chica, P. Gagnon, J.W. Keillor, J.N. Pelletier, *Protein Sci*, 13 (2004) 979-991.
- [40] M. Gross, J.E. Folk, *J Biol Chem*, 248 (1973) 1301-1306.
- [41] M. Gross, J.E. Folk, *J Biol Chem*, 248 (1973) 6534-6542.
- [42] Z. Nemes, T. Csoz, G. Petrovski, L. Fesus, *Prog Exp Tumor Res*, 38 (2005) 19-36.
- [43] T. Kashiwagi, K. Yokoyama, K. Ishikawa, K. Ono, D. Ejima, H. Matsui, E. Suzuki, *Journal of Biological Chemistry*, 277 (2002) 44252-44260.
- [44] C. Lombard, J. Saulnier, J.M. Wallach, *Protein Peptide Lett*, 12 (2005) 621-629.
- [45] J. Mitsuhashi, T. Nakayama, A. Narai-Kanayama, *Enzyme and Microbial Technology*, 75-76 (2015) 10-17.
- [46] D. Kumar, T.C. Bhalla, *Applied microbiology and biotechnology*, 68 (2005) 726-736.
- [47] L. Riechmann, V. Kasche, *Biochimica Et Biophysica Acta*, 830 (1985) 164-172.
- [48] W. Kullmann, *Biochemical Journal*, 220 (1984) 405-416.
- [49] W. Kullmann, *Journal of Biological Chemistry*, 255 (1980) 8234-8238.

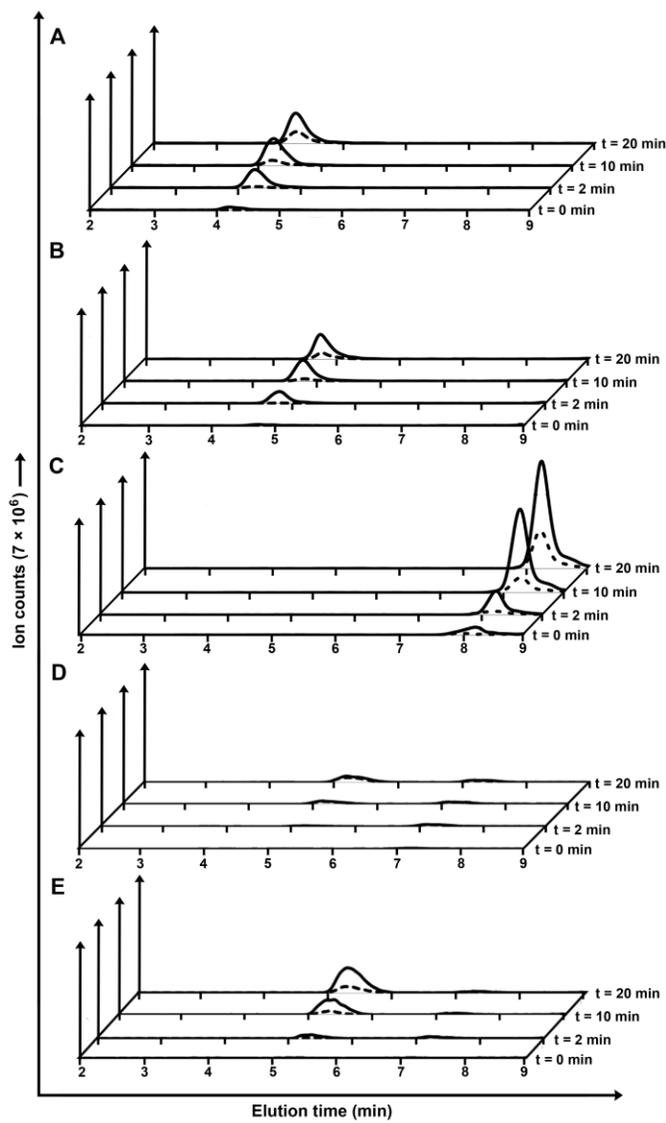
## Figure Captions



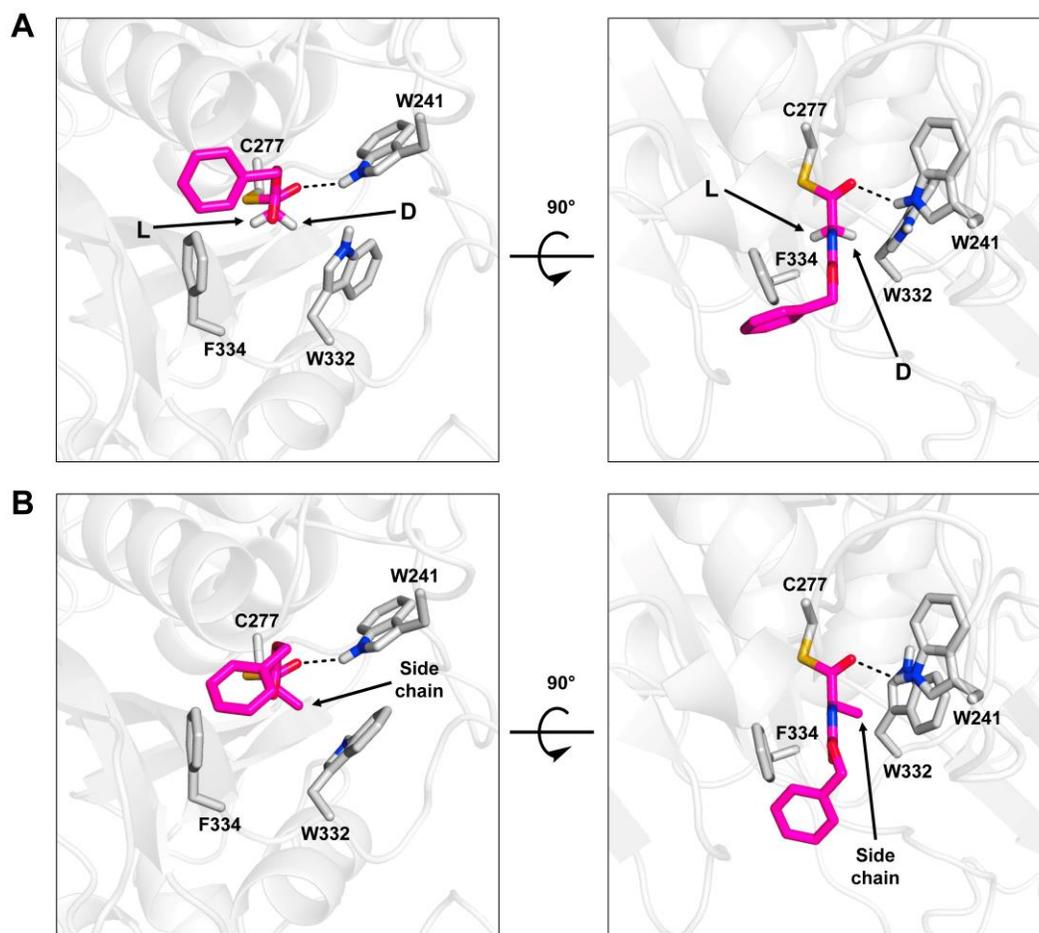
Scheme 1. Peptide synthesis reaction catalyzed by gTG2.



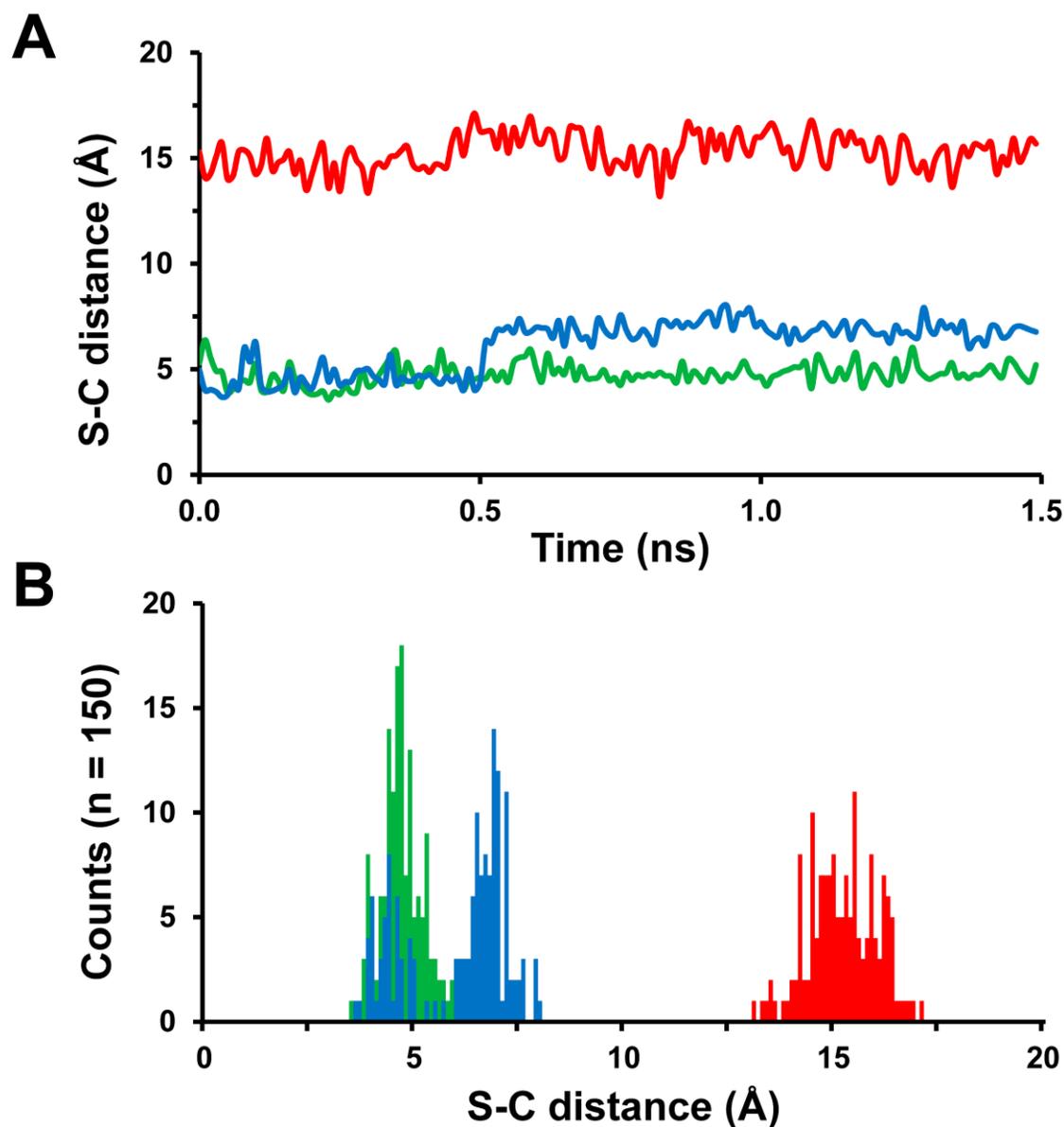
**Figure 1. Aromatic ester donor substrates of gTG2.** The reactive carbonyl group of these compounds is indicated by an arrow. (A) *N*-carbobenzyloxy-L-glutamyl( $\gamma$ -*p*-nitrophenyl ester)glycine; (B) Cbz-Gly-GABA-7HC and Cbz-Phe-GABA-7HC; (C) Cbz-Gly-7HC.



**Figure 2. LC-MS traces of gTG2-catalyzed peptide synthesis reaction mixtures.** gTG2-catalyzed synthesis of dipeptide products (solid lines) and uncatalyzed control reactions (dashed lines). Various combination of aromatic ester donors and amino acid derivative acceptors were tested: (A) Cbz-Gly-7HC + GlyNH<sub>2</sub>; (B) Cbz-Gly-7HC + AlaNH<sub>2</sub>; (C) Cbz-Gly-7HC + LeuOMe; (D) Cbz-L-Ala-7HC + GlyNH<sub>2</sub>; (E) Cbz-D-Ala-7HC + GlyNH<sub>2</sub>.



**Figure 3. gTG2 acyl-enzyme intermediate models.** The active site of gTG2 (white) with the catalytic Cys277 residue acylated by the Cbz-Gly (A) or Cbz-D-Ala (B) moiety (magenta) is shown. An H-bond between Trp241 and the carbonyl oxygen of the acyl group is indicated by a dashed line.  $C_{\alpha}$  hydrogens of Cbz-Gly are shown as sticks, and the resulting configuration of  $C_{\alpha}$  following replacement of each hydrogen by a methyl group side chain is indicated. The methyl side chain of Cbz-D-Ala is indicated by an arrow.



**Figure 4. Molecular dynamics simulations results.** (A) Distance between the nucleophilic sulfur atom of catalytic residue Cys277 and the electrophilic carbonyl carbon atom of the Cbz-Gly (green), Cbz-D-Ala (blue), or Cbz-L-Ala (red) products during the course of a 1.5-nanosecond MD simulation. (B) Distributions depict the number of MD snapshots with sulfur-carbon distances grouped in incremental bins of 0.1 Å. Each snapshot was taken at 10-picosecond intervals.

Tables

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

Substrate	$K_M$ ( $\mu\text{M}$ )	$k_{\text{cat}}$ ( $\text{s}^{-1}$ )	$k_{\text{cat}}/K_M$ ( $\times 10^4 \text{M}^{-1} \text{s}^{-1}$ )
Cbz-Gly-7HC	$15 \pm 2$	$0.128 \pm 0.007$	0.85
Cbz-L-Ala-7HC	N.D. <sup>b</sup>	N.D.	N.D.
Cbz-D-Ala-7HC	- <sup>c</sup>	-	0.018
Cbz-Gly-GABA-7HC <sup>a</sup>	$9 \pm 2$	$1.25 \pm 0.08$	14

<sup>a</sup> Data from [14].<sup>b</sup> No detectable activity.<sup>c</sup> Saturation could not be achieved within solubility limit of donor substrate.

Errors indicated are standard errors of best-fit parameters.