

# **Design and Evaluation of Novel Fluorescent Molecular Probes Targeting Cathepsin B**

---

A thesis presented to  
The Faculty of Graduate Studies  
of  
Lakehead University, Thunder Bay, Ontario  
by  
**Sepideh Dadgar**

In partial fulfillment of the requirements  
for the degree of  
Doctor of Philosophy in Biotechnology  
June 30<sup>th</sup>, 2020

© Sepideh Dadgar, 2020

## Abstract

Lysosomal cysteine proteinase cathepsin B (CTB) is a member of the cysteine protease family known to participate in intracellular degradation processes and protein turnover in the lysosomes of healthy cells. Cathepsin B plays a crucial role in tumor invasion and progression by controlling extracellular degradation and participating in a proteolytic cascade activation (Gondi and Rao 2013). Its role in tumor invasion and progression makes CTB a promising biomarker and target for antibody-directed prodrug therapy (Dheer, Nicolas et al. 2019). The development of novel CTB-specific molecular probes opens new possibilities for image-based diagnostic methods for different types of cancers (Podgorski and Sloane 2003, Tan, Peng et al. 2013). Since aberrant expression of this protein has been an indicator of cancer development, detecting CTB expression and activity might be beneficial for the early detection of cancer or revealing aggressive lesions (Gondi and Rao 2013). Developing probes capable of binding with CTB is challenging due to binding site homology to other members of the cysteine cathepsin family (Turk, Stoka et al. 2012). In this study, to identify unique residues in human CTB compare to other members of the cathepsin family, amino acid sequences of these proteins were exposed to. multiple sequence alignments. Cathepsin B in humans has three active site residues critical for catalysis: Cys108; His278, and Asp298 (Ruan, Hao et al. 2015) were confirmed with multiple sequence alignment as fully conserved residues. The initial step in the development of a detection assay for CTB is finding appropriate fluorescent small molecules for enzyme binding. In this study, two ligand candidates CID8795 and CID535684 were identified and successfully conjugated to the dye ATTO680 and were tested for binding affinity and specificity to CTB. For CID535684ATTO680, no binding interaction was observed in the fluorescence polarization (FP) assay. CID8795ATTO680 demonstrated increases in fluorescence polarization assays in the presence of CTB with the half-

maximal effective concentration ( $EC_{50}$ ) at  $3.27 \pm 1.27$  nM. A third probe, Benzyloxycarbonyl (Cbz)-Lys-Lys-*p*-Aminobenzyl alcohol (PABA)-2', 7'-dichloro-6'-methoxy-fluorescein (DCMF), was designed based on a known substrate scaffold for CTB. This novel substrate-based fluorescent probe was shown to be hydrolyzed by CTB having a specificity constant  $k_{cat}/K_M = 41.9 \pm 0.07$   $mM^{-1} \times s^{-1}$ . Finally, we investigated single nucleotide polymorphisms (SNPs) within the coding region of the *CTSB* gene within the general population (random data from 2,504 samples) included in the 1,000 Genomes project. The mapping of SNPs onto the 3D structure of cathepsin B indicates that the active site of CTB is fully conserved among humans – as no SNPs were identified within the binding pocket of CTB. According to these results, probes that bind to the enzyme's active site should be generally useful for detecting CTB in all populations studied in the 1,000 Genomes project.

**Keywords:** Cathepsin B, Fluorescent Probe, *Dichlorofluorescein* Reporter, Fluorescent Substrate, SNPs, and, 1,000 Genomes.

## **Acknowledgments**

I am sincerely grateful to my supervisor, Dr. Wely Floriano, for her guidance, endless encouragement, continuous support, and allowing me to conduct such an exciting project in my Ph.D. study. I would also like to extend my thanks to my amazing supervisor Dr. Christopher Phenix for spending time to lend his knowledge and experience while conducting my research. I am also very grateful to my committee members Dr. Laura Curiel and Dr. Campbell. Special thanks to Dr. Campbell for lending his lab space during my previous year of Ph.D. study and all his non-stop support and advice during my study. I would also like to thank Dr. Simon Lees and Sarah Niccoli for helping me prepare the cell lysate. I also wish to express my gratitude to Dr. Morshed Chowdhury and Mr. Daniel Tesolin, who led all efforts to synthesize the various probes reported in this thesis. Thank you to Brady Vigliarolo, who provided advice and support throughout my thesis project. Thank you to Kimberley Christopher for proofreading. Finally, I would like to acknowledge Brenda Magajna for her help and reassurance during my thesis.

## **Dedication**

To my lovely parents, who have supported me in all my endeavors.

To my lovely sister (Saedeh Dadgar) and her supportive husband (Mehdi Dashtban).

## Table of Contents

Abstract.....	2
Acknowledgments.....	4
Dedication.....	5
Table of Figures.....	9
List of Tables.....	16
Abbreviations.....	17
Chapter 1: Introduction.....	20
1.1. Background.....	20
1.2. Research Problem and Hypothesis.....	22
1.3. Research Strategy.....	25
1.4. Research Significance.....	28
Chapter 2: Literature Review.....	30
Chapter 3: Cathepsin B: Sequence, Structure, and Function.....	42
3.1. Introduction.....	42
3.2. Procedure and Materials.....	42
3.2.1. <i>CTSB</i> Gene.....	42
3.2.2. CTB Protein.....	43
3.2.3. Similarity and Differences between Cathepsin B and Different Members of Cathepsin Family.....	43
3.2.4. 3D Structure of CTB.....	44
3.2.5. Binding Interaction Analysis.....	45
3.3. Results.....	46
3.3.1. <i>CTSB</i> Gene.....	46
3.3.2. CTB Protein.....	46
3.3.3. Similarity and Differences between Cathepsin B and Other Members of the Cathepsin Family.....	48
3.3.4. Mapping Variable Residues onto the 3D Structure of CTB.....	56
3.3.5. Computationally Identified Binding Interactions.....	63
3.4. Summary and Conclusions.....	67
Chapter 4: Genetic Variation in Cathepsin B Gene ( <i>Homo sapiens</i> ).....	69
4.1. Introduction.....	69
4.2. Procedures and Materials.....	70
4.2.1. Variant Identification in the 1,000 Genomes Data.....	70
4.2.2. Identification of SNPs that Affect the Protein Sequence.....	71

4.2.3. Mapping SNPs in the Amino Acids Sequence and 3D Structure of CTB .....	72
4.3. Results.....	73
4.3.1. Variant Identification in the 1,000 Genomes Data.....	73
4.3.2. Identification of SNPs that Affect the Protein Sequence .....	73
4.3.3. Mapping SNPs in the Amino Acids Sequence and 3D Structure of CTB .....	78
4.4. Summary and Conclusions.....	81
Chapter 5: Synthesis and Characterization of Fluorescent Probes.....	82
5.1. Introduction.....	82
5.1.1. Selection of NIR Probe Candidate .....	82
5.1.2. Selection of Substrate-Based Fluorescent Probe.....	87
5.2. Procedures and Materials .....	89
5.2.1. Synthesis of the NIR Probe.....	89
5.2.2. Synthesis of the Fluorescent Substrate.....	90
5.2.3. Spectrophotometric Characterization.....	93
5.3. Results.....	94
5.3.1. Synthesis and Conjugation of Selected Hits .....	94
5.3.2. An Improved Substrate-Based Fluorescent Probe for CTB .....	96
5.3.3. Spectrophotometric Characterization.....	97
5.4. Summary and Conclusions.....	101
Chapter 6: Evaluation of Binding Interaction of CID8795ATTO680 and CID535684ATTO680 to CTB .....	102
6.1. Introduction.....	102
6.2. Procedures and Materials .....	102
6.2.1. Evaluating the Inhibition Activity of the Ligands on CTB.....	102
6.2.2. Evaluating the Inhibition Activity of the Two NIR Probes on CTB.....	104
6.2.3. Fluorescence Polarization .....	106
6.2.4. Fluorescence Polarization Protein Concentrations Curve.....	107
6.2.5. Probe Aggregation .....	108
6.2.6. Selectivity Assay for ATTO 680 Conjugates.....	109
6.2.7. Cell Lysate Assay.....	110
6.3. Result for Two Novels NIR Non-Reactive Probes .....	112
6.3.1. Inhibition Assay for Seven Ligands.....	112
6.3.2. Inhibition Assay for CID8795ATTO680 and CID535684ATTO680.....	115
6.3.3. Fluorescence Polarization Results.....	118
6.3.4. Fluorescence Polarization Protein Concentrations Curve.....	122

6.3.5. Probe Aggregation Results.....	125
6.3.6. Selectivity Confirmation of the CID8795ATTO680 towards CTB .....	134
6.3.7. Cell Lysate Result for CID8795ATTO680 .....	135
6.4. Summary and Conclusions.....	138
Chapter 7: Rate of Hydrolysis of Cbz-Lys-Lys-PABA-DCMF by CTB and CTL.....	142
7.1. Introduction.....	142
7.2. Procedures and Materials for Cbz-Lys-Lys-PABA-DCMF .....	142
7.2.1. Enzyme Kinetics .....	142
7.2.2. The Selectivity of the Novel Substrate towards CTB .....	143
7.3. Results for Cbz-Lys-Lys-PABA-DCMF.....	143
7.3.1. Enzyme Kinetics Assay for Novel Substrate .....	143
7.3.2. The Selectivity of Cbz-Lys-Lys-PABA-DCMF Towards CTB and CTL .....	146
7.4. Summary and Conclusions.....	147
Chapter 8: Cell Viability and Cell Imaging Assays for CID8795ATTO680 and Cbz-Lys-Lys-PABA-DCMF .....	149
8.1. Introduction.....	149
8.2. Procedures and Materials .....	149
8.2.1. Cell Viability.....	149
8.2.2. Cell Imaging.....	150
8.3. Result for Two Novels NIR Non-Reactive Probes .....	151
8.3.1. Cell Viability Results .....	151
8.3.2. Cell Imaging Results .....	153
8.4. Summary and Conclusions.....	155
Chapter 9: Summary and Conclusions.....	156
Chapter 10: Future Work .....	159
Appendix:.....	160
1. Reagents, Buffers, and Enzymes.....	160
2. Fluorescence Polarization and Anisotropy .....	161
3. Enzyme Kinetic.....	163
References.....	167



## Table of Figures

<b>Figure 1.</b> The chemical structure of Cbz-Arg-Arg-AMC. ....	24
<b>Figure 2.</b> The chemical structure of <i>CB-CNP</i> , <i>Cy5.5</i> (dye), and <i>BHQ-3</i> (quencher).....	26
<b>Figure 3.</b> Involvement of cysteine cathepsins in cancer progression. Cancer can progress in two steps: extracellular and intracellular proteolysis, which goes under E-cadherin inactivation, and Bid cleavage, respectively.....	31
<b>Figure 4.</b> The three different mechanisms of the cathepsin B involvement in the activation of the proteolytic cascade. Cathepsin B could involve in this process in different ways: degradation of the extracellular matrix; converting pro-MMP to MMP, and activation of the pre-TGF-B1 to TGF-B1. ....	32
<b>Figure 5.</b> Representation of signaling and protease cascade events induced by cathepsin B on cancer cells. ....	34
<b>Figure 6.</b> The mode of action of the fluorescent FRET-based probe. ....	39
<b>Figure 7.</b> Annotation of the protease-substrate interaction, where <b>P</b> positions refer to the positions of the substrate residues and <b>S</b> positions refer to subsites where they bind to the surface of proteases.....	45
<b>Figure 8.</b> Human cathepsin B amino acid sequence in FASTA format with the positions of three active site residues, Cys108, His278, and, Asn289, underlined in green. (NCBI 2018, Nov 13) .....	48
<b>Figure 9.</b> Section of multiple sequence alignment of human CTB and <i>IHUC</i> with other members of the cathepsins family (same organism), obtained from. CLUSTALX-2.1 software (Clustal: Multiple Sequence Alignment AUG,2012). The red boxes show the homology of three active site residues of CTB (Cys (108), His (278), and Asn (298)) with seven members of the cathepsin family (CTZ, CTL2, CTH, CTC, CTS, CTK, and CTL1). The horizontal red box shows the positions of the human CTB and <i>IHUC</i> structure in the alignment. ....	51
<b>Figure 10.</b> View of the parts of aligned multiple sequences of 12 members of the cathepsin family and <i>IHUC</i> structure by CLUSTALX-2.1 software (Clustal: Multiple Sequence Alignment AUG,2012), which indicates the fully, strongly and weakly conserved residues. The brown and red and yellow boxes are marked the fully, strongly and weakly conserved residues, respectively. ....	53
<b>Figure 11.</b> Multiple sequence alignment of 12 members of the cathepsin family and <i>IHUC</i> structure obtained using CLUSTALX-2.1 software (Clustal: Multiple Sequence Alignment AUG,2012). The vertical black boxes mark the unique residues in human CTB relative to other members of the cathepsin family, in terms of different biochemical characteristics. The horizontal black boxes show the positions of the human CTB and <i>IHUC</i> structure in the alignment.....	54
<b>Figure 12.</b> Multiple sequence alignment of 12 members of the cathepsin family and <i>IHUC</i> structure obtained with CLUSTALX-2.1 software (Clustal: Multiple Sequence Alignment AUG,2012).	

The location of two adjacent histidines at positions 189 and 190 in CTB compared to other members of the family is marked by a vertical red box. The horizontal red box shows the positions of the human CTB and *IHUC* structure in the alignment. .... 56

**Figure 13.** The chemical structure of ligand (EP0). .... 58

**Figure 14.** The 3D structure of CTB (PDB code *ICSB*) which is visualized by VMD software (VMD APR,2008), the red, light and dark blue balls represent six unique residues that were identified by multiple sequence alignment (Glu115, Ser118, Glu132, Ser254, Arg281, and, Asn301). All these residues are located above 11 Å distance of ligand (EP0 (orange stick)). The position of unique residue Glu324, which is shown in a white circle, is located within close distance from the ligand (5.97 Å). .... 59

**Figure 15.** The 3D structure of CTB (PDB code *ICSB*) with bound ligand (EP0 (yellow stick) after loading *ICSB* under the VMD main window (VMD APR,2008). The red balls represent the residues located within 6 Å of the ligand. The list of these residues is as follows: Gly102, Ser104, Cys105, Gly106, Trp109, Asn151, Gly152, Gly153, His189, His190, Gly200, Glu201, Gly202, Val255, Met275, Gly276, Gly277, Trp300, and the unique residue: Glu324. The active site residues of human CTB marked as green balls: Cys108, His278, and Asn298. .... 61

**Figure 16.** The 3D structure of cathepsin D (PDB code *ILYB*) which is visualized by VMD software (VMD APR,2008), with the position of two active site residues (Asp97 and Asp295 (green balls)) and their distance from two adjacent histidines at positions 120 and 121 (red balls). .... 62

**Figure 17.** The 3D structure of cathepsin C (PDB code *3PDF*) which is visualized by VMD software (VMD APR,2008), with the position of three active site residues (Cys258, His 405, and Asn427 (blue balls)) and their distance from two pairs of adjacent histidine at positions 366-367 and 389-390 (red balls). .... 63

**Figure 18.** Predicted binding interaction between CID8795ATTO680 and CTB (PDB code *IHUC*). Diagram created with MOE (Molecular Operating Environment MOE March,2020). using the computationally generated structure of CID8795ATTO680 bound to CTB. **P** positions refer to positions in the substrate responsible for CTB recognition. .... 65

**Figure 19.** Predicted binding interaction between Cbz-Arg-Arg-AMC and CTB visualized with MOE (Molecular Operating Environment MOE March,2020). The structure of Cbz-Arg-Arg-AMC docked to CTB (PDB code *IHUC*). **P** positions refer to positions in the substrate responsible for CTB recognition. .... 66

**Figure 20.** The predicted binding interaction between Cbz-Lys-Lys-PABA-DCMF and CTB. **P** positions refer to positions in the substrate responsible for CTB recognition. .... 67

**Figure 21.** Population-based allele frequency of two variants associated with amino acid substitutions Ser53Gly and Leu26Val in the CTB protein. .... 75

**Figure 22.** Population-based allele frequency of five variants (located on mature CTB) associated with amino acid substitutions Pro91Leu, Ser235Asn, Lys237Glu, Asn246Thr, and Asp317Asn in the CTB protein. .... 77

- Figure 23.** The protein sequence of human CTB with marking SNPs. The red box indicates the propeptide of CTB (position at 18-79 amino acid) which is cleaved during enzyme maturation. Mutations with MAF < 0.5 are highlighted in blue (Leu26Val and Ser53Gly). The rare variations with a MAF < 0.01 that are located on the mature CTB identified by pink, meanwhile, those rare mutations which are not present in the mature CTB highlighted by orange color. Two mutations occur at position Leu26, common highlighted as blue and rare underlined by orange color. The red highlighted amino acids indicate the predicted binding site of the protein for the selected probes (Cbz-Lys-Lys-PABA-DCMF and CID8795ATTO680). Finally, the start and end positions of the mature protein are highlighted as green color. .... 79
- Figure 24.** View of docked structure of CID8795ATTO680 (Daniel Tesolin, Private Communication) with mapping residues within 3.5 Å of the probe by VMD software (VMD APR,2008). The red ribbons show the *IHUC* carbon alpha trace. The probe is shown in blue. Yellow represents the residues within 3.5 Å of the CID8795ATTO680. The name and position of these amino acids are as follows: Gly153, His189, His190, Gly200, Glu201, Val255, Met275, Gly276, Gly277, His278, Trp300, and Glu324. White represents the three active site residues (Cys108, His278, and Asn298) of human CTB. Green shows the positions of rare mutations on this structure which are located above 20 Å from the binding pocket of CTB..... 80
- Figure 25.** The chemical structure of ATTO 680 NHS-ester. The red circle indicates the site of conjugation. .... 86
- Figure 26.** The chemical composition of 2', 7'-dichloro -fluorescein. The red circles indicate sites of conjugation. .... 89
- Figure 27.** The chemical structure of the CID8795ATTO680. .... 95
- Figure 28.** The chemical structure of CID535684ATTO680. .... 95
- Figure 29.** Fluorescence profile of CID8795ATTO680. A) Blank-discounted excitation and emission spectra of CID8795ATTO680 at 40 µM, pH 7, showing maximum excitation and emission at 675 and 705 nm respectively. B) Fluorescence intensity measurement (at 0-time point) of different concentrations of CID8795ATTO680 (two-fold serial dilution with a starting concentration at 0.195 to 200 µM), which is increased linearly with the increase of the conjugate concentration. Linear regression was performed using Prism 5 software (Prism 2018, November 16) with the R-squared ( $R^2$ ) at 0.99. FP results presented in (C) and (D) refer to the log concentrations of CID8795ATTO680 at 60 minutes and pH 7 (C) and 4 (D) at the 0-time point. All assays were carried out at 37° temperature..... 98
- Figure 30.** Fluorescence profile of CID535684ATTO680. A) Blank-discounted excitation and emission spectra of the CID535684ATTO680 (40 µM) at 675 and 690 nm respectively at pH 7. B) Fluorescence intensity measurement (at the 0-time point) of the different concentrations of CID535684ATTO680(a two-fold serial dilution with a starting concentration at 0.195 to 200 µM), which is increased linearly to the increase of the conjugate concentration, the linear regression fit has been applied by using Prism 5 software (Prism 2018, November 16) with the  $R^2$  at 0.99, non-linearity was observed at concentrations above 30 µM. FP results represented in (C) and (D) refer to the log concentrations of CID535684ATTO680 at pH 7 and 7.2 at the 0-time point, respectively. All assays were carried out at 37° temperature..... 99

<b>Figure 31.</b> The chemical structure of the Cbz-Lys-Lys-PABA-DCMF.....	100
<b>Figure 32.</b> Fluorescence profile (blank-discounted excitation and emission) of DCMF at 1 $\mu$ M, pH 5.5, and 37°C.....	100
<b>Figure 33.</b> The chemical structure of Acetyl-Arg-Arg-7-Amino-4-trifluoromethyl coumarin.....	103
<b>Figure 34.</b> The chemical structure of Phe-Phe-Fluoromethyl ketone.....	103
<b>Figure 35.</b> Inhibitor screening results in the presence of ligands in three concentrations according to the average of the last seven minutes of the progression curve for the AC-R-R-AFC substrate. F-F-FMK is a known CTB inhibitor and was used as the control. The dotted line marks the 70% activity threshold and maximum activity marked as 100%. Ligands were tested at 100 $\mu$ M (black), 10 $\mu$ M (gray), and 1 $\mu$ M (light gray) in duplicates. The experiment was performed at pH 7, and a temperature of 37°C. The concentration of the substrate, F-F-FMK, and CTB was fixed at 200 $\mu$ M, 10 $\mu$ M, and 0.5 $\mu$ M, respectively. Statistical differences in inhibition activity were determined by ANOVA with post-hoc Dunnett's multiple comparison test using Prism 5 software (Prism 2018, November 16); $p < 0.05$ was considered significant and presented by ***.....	113
<b>Figure 36.</b> Inhibitor screening results for ligands according to the initial velocity for the first few minutes of the progression curve for the AC-R-R-AFC substrate. F-F-FMK is a known CTB inhibitor and was used as a control. The dotted line marks the 70% activity threshold and maximum activity marked as 100%. Ligands were tested at 100 $\mu$ M (black), 10 $\mu$ M (gray), and 1 $\mu$ M (light gray) in duplicates. The experiment was performed at pH 7, and a temperature of 37°C. The concentration of the substrate and, F-F-FMK, and CTB were fixed at 200 $\mu$ M, 10 $\mu$ M, and 0.5 $\mu$ M, respectively. Statistical differences in inhibition activity were determined by ANOVA post-hoc Dunnett's multiple comparison test using Prism 5 software (Prism 2018, November 16); $p < 0.05$ was considered significant and presented by **.....	114
<b>Figure 37.</b> Inhibitor screening results for ATTO 680 conjugates; CID8795ATTO680 (a) and CID535684ATTO680 (b) at 25, 10, and 1 $\mu$ M according to the average of last seven minutes of the substrate (Ac-RR-AFC) progression curve. F-F-FMK is a known CTB inhibitor and was used as a control. The dotted line marks the 70% activity threshold and maximum activity marked as 100%. The experiment runs at 37°C and pH 7. The concentration of the substrate (Ac-RR-AFC) was 200 $\mu$ M, and F-F-FMK was 10 $\mu$ M. The CTB concentration was 0.5 $\mu$ M. Statistical differences in inhibition activity of samples (3 concentrations of NIR probe or F-F-FMK with a fixed concentration of substrate and CTB) compare to the maximum activity (only substrate and CTB) were determined by ANOVA post-hoc Dunnett's multiple comparison test, using Prism 5 software (Prism 2018, November 16); $p < 0.05$ was considered significant and presented by ***.....	116
<b>Figure 38.</b> Inhibitor screening results for ATTO 680 conjugates; CID8795ATTO680 (a) and CID535684ATTO680 (b) according to the initial velocity for the first few minutes of the Ac-RR-AFC activity toward CTB. F-F-FMK is a known CTB inhibitor and was used as a control. The dotted line marks the 70% activity threshold and maximum activity marked as 100%. The experiment runs at 37°C and pH 7. The concentration of the substrate was 200 $\mu$ M, and F-F-FMK was 10 $\mu$ M. The CTB concentration was 0.5 $\mu$ M. Statistical differences in inhibition activity of samples (3 concentrations of NIR probe or F-F-FMK with a fixed concentration of substrate)	

compare to the maximum activity (only substrate and CTB) were determined by ANOVA post-hoc Dunnett's multiple comparison test, using Prism 5 software (Prism 2018, November 16);  $p < 0.05$  was considered significant and presented by \*..... 117

**Figure 39.** Inhibitor screening results for ATTO 680 conjugates CID8795ATTO680 at 10, 1, and 0.1  $\mu\text{M}$ . a) Inhibitor screening results based on the initial velocity of the pNA formation at first ten minutes b) The average of the last seven minutes of the substrate (Z-Arg-Arg-pNA) progression curve (pNA formation). F-F-FMK is a known CTB inhibitor and was used as a control. The red dotted line marks the 70% activity threshold and maximum activity marked as 100%. The experiment runs at 37°C and pH 5.5. The concentration of the substrate (Z-Arg-Arg-pNA) was 200  $\mu\text{M}$ , and F-F-FMK was 10  $\mu\text{M}$ . The CTB concentration was 0.01  $\mu\text{M}$ . Statistical differences in inhibition activity of samples (3 concentrations of NIR probe or F-F-FMK with a fixed concentration of substrate) compare to the maximum activity (only substrate and CTB) were determined by ANOVA post-hoc Dunnett's multiple comparison test, using Prism 5 software (Prism 2018, November 16);  $p < 0.05$  was considered significant and presented by \*\*\*..... 118

**Figure 40.** Change in anisotropy in the presence and the absence of CTB (100, 10 and 1 nM) for CID8795ATTO680 (A) and CID535684ATTO680 (B) at 1, 0.1 and 0.01  $\mu\text{M}$ . Assays were performed in duplicate at pH 5.5 and read at 60 minutes which were incubated for 5 minutes at 37°C. Statistical differences in anisotropy were determined by ANOVA post-hoc Dunnett's multiple comparison test, using Prism 5 software (Prism 2018, November 16);  $p < 0.05$  was considered significant and presented by \*\*\* and \*\*..... 121

**Figure 41.** Anisotropy as a function of the log CTB concentrations started from 0.25 nM to 180.68 nM against CID8795ATTO680 at 0.02  $\mu\text{M}$  (a) and 0.05  $\mu\text{M}$  (b), and the log CTB concentrations started at 0.42 nM to 307.16 nM against CID8795ATTO680 at 0.1  $\mu\text{M}$  (c). The non-linear approximation curve fitting applied for measuring the  $\text{EC}_{50}$  of the CID8795ATTO680 by using Prism 5 software (Prism 2018, November 16). Assays were performed in triplicate at pH 5.5 and read at 60 minutes, which were incubated for 5 minutes at 37°C. The average of all results (CV < 10%) shows as points. The dotted-line in each dataset corresponding to the binding fit model. .... 123

**Figure 42.** Change in maximum fluorescence intensity at 665 nm in the presence and the absence of CTB (100, 10 and 1 nM) for CID8795ATTO680 (A) and CID535684ATTO680 (B) at 1, 0.1 and 0.01  $\mu\text{M}$ . Assays were performed in duplicate at pH 5.5 and read at 60 minutes which were incubated for 5 minutes at 37°C. The backgrounds correspond to free probes in acetate buffer which are represented as black columns. .... 126

**Figure 43.** Change in maximum fluorescence intensity at 665 nm as a function of the CTB concentrations started from 0.25 nM to 180.68 nM against CID8795ATTO680 at 0.02  $\mu\text{M}$  (a) and 0.05  $\mu\text{M}$  (b), and the CTB concentrations started at 0.42 nM to 307.16 nM against CID8795ATTO680 at 0.1  $\mu\text{M}$  (c). Assays were performed in triplicate at pH 5.5 and read at 60 minutes, which were incubated for 5 minutes at 37°C. The average of all results with CV < 10% shows as a graph. The backgrounds correspond to free probes in acetate buffer which are represented as black columns. Statistical differences in anisotropy were determined by ANOVA post-hoc Dunnett's multiple comparison test, using Prism 5 software (Prism 2018, November 16);  $p < 0.05$  was considered significant and presented by \*, \*\*, and \*\*\*..... 130

- Figure 44.** Fluorescence emission spectrums with a blue shift. CTB concentrations started from 0.25 nM to 180.68 nM against CID8795ATTO680 at 0.02  $\mu$ M (A) and 0.05  $\mu$ M (B), and the CTB concentrations started at 0.42 nM to 307.16 nM against CID8795ATTO680 at 0.1  $\mu$ M (C). Assays were performed in triplicate at pH 5.5 and read at 60 minutes, which were incubated for 5 minutes at 37°C. Nonlinear regression curve fit (Gaussian) used for fitting by using Prism 5 software (Prism 2018, November 16). All the  $R^2$  values were above 0.99. .... 132
- Figure 45.** Change in anisotropy of the free CID8795ATTO680 at 0.02  $\mu$ M, 0.05  $\mu$ M, and 0.1  $\mu$ M over time (0 and 60 minutes). Assays were performed in triplicate at pH 5.5 and read at 0 and 60 minutes, which were incubated for 5 minutes at 37°C. The average of all results with CV < 10% shows as a graph. Statistical differences in anisotropy at zero and 60 minutes were determined by ANOVA post-hoc Dunnett's multiple comparison test, using Prism 5 software (Prism 2018, November 16);  $p < 0.05$  was considered significant and presented by \*\*. .... 133
- Figure 46.** Absorbance spectrum of the CID8795ATTO680 at 40  $\mu$ M in CTB reaction buffer at pH 7.2 and 37°C. The peak at lower wavelength represents the H-band induced by probe aggregation. ... 134
- Figure 47.** Changing in anisotropy of CID8795ATTO680 at 0.02  $\mu$ M in the presence of CTB and BSA. A) Anisotropy of the NIR probe in the presence of BSA with two-fold serial dilution with a starting concentration at 0.25 nM to 180.45 nM in PBS buffer pH 7.4. B) Anisotropy of the NIR probe in the presence of CTB with a two-fold serial dilution with a starting concentration at 0.25 nM to 180.68 nM in acetate buffer pH 5.5 and incubated for 5 minutes at 37°C and read at 60 minutes. The result is based on the average of all triplicates (CV < 10%). Statistical differences in anisotropy were determined by ANOVA post-hoc Dunnett's multiple comparison test, using Prism 5 software (Prism 2018, November 16);  $p < 0.05$  was considered significant and presented by \*\*. .... 135
- Figure 48.** Average OD of the samples versus serial concentration of the BSA (2, 1, 0.5, 0.25, and 0.125 mg/ml) to quantify the total protein concentration of the lysate..... 136
- Figure 49.** Change in anisotropy in the presence and the absence of *MDA-MB-231* cancer cells lysate at 0.24 mg/ml (blue graph) and *H9C2* normal cells lysate at 0.0061 mg/ml (orange graph) for CID8795ATTO680 at 5  $\mu$ M at 240, 270, and 300 minutes. Assays were performed in triplicate at pH 5.5 and read for 300 minutes, which were incubated for 15 minutes before adding probe at 37°C. The result is based on the average of all triplicates (CV < 6%). Statistical differences in anisotropy were determined by ANOVA post-hoc Dunnett's multiple comparison test, using Prism 5 software (Prism 2018, November 16);  $p < 0.05$  was considered significant and presented by \*. .... 138
- Figure 50.** The reaction steps for hydrolysis of Cbz-Lys-Lys-PABA-DCMF, PABA immolation, and releasing DCMF. .... 144
- Figure 51.** Lineweaver-Burk (A) and Michaelis-Menten (B) plots applied to determine kinetic parameters for Cbz-Lys-Lys-PABA-DCMF by using Prism 5 software (Prism 2018, November 16). Experiments were performed in duplicate at 5 nM concentration of CTB, pH 5.5, and temperature of 37°C. The substrate concentration ranged from 0.019 mM to 5 mM. .... 145

- Figure 52.** Selectivity assay of the Z-Lys-Lys-PABA-DCMF at a fixed concentration (0.078 mM) versus CTB or CTL at 5 nM. Fluorescence intensity at excitation and emission wavelengths  $470 \pm 9$  nm and  $540 \pm 9$  nm applied to evaluate the hydrolysis of Z-Lys-Lys-PABA-DCMF by CTB and CTL. A graph showing the endpoint RFU corresponding to releasing DCMF with CTL and CTB. . 147
- Figure 53.** Effects of Cbz-Lys-Lys-PABA-DCMF (a) and CID8795ATTO680 (b) on cell viability. *MDA-MB-231* and *H9C2* cells were treated with candidate probes at 1, 10, and 40  $\mu$ M. The % viability of cells was quantified as a ratio of absorbance of the untreated and treated cells using MTT assay. Statistical differences in viability were determined by performing ANOVA post-hoc Dunnett's multiple comparison test, using Prism 5 software (Prism 2018, November 16);  $p < 0.05$  was considered significant and marked by \*. ..... 152
- Figure 54.** Fluorescence microscopy and overlayer "Bright filed" images of *MDA-MB-231* and *H9C2* cell lines treated with Cbz-Lys-Lys-PABA-DCMF (A (zoom at 200  $\mu$ m) and B (zoom at 100  $\mu$ m) and CID8795ATTO680 (C (zoom at 1000  $\mu$ m) and D (zoom at 1000  $\mu$ m)). Fluorescent microscopy images taken of cells treated for 4 h with (A (cancer cells) and B (normal cells)) with 40  $\mu$ M of Cbz-Lys-Lys-PABA-DCMF ( $\lambda_{\text{ex}} = 469$  nm and  $\lambda_{\text{em}} = 525$  nm) (C (cancer cells) and D (normal cells)) 10  $\mu$ M of CID8795ATTO680 ( $\lambda_{\text{ex}} = 628$  nm and  $\lambda_{\text{em}} = 685$  nm). To evaluate probe specificity cells were treated with 10  $\mu$ M CA-074Me for 12 hours before incubation with each probe for 4 h. In all images, the left side panel shows uninhibited cells versus inhibited in the right panel. .... 154
- Figure 55.** Linear regression fitting model for initial velocity (RFU/Time) calculation..... 164
- Figure 56.** Michaelis-Menten kinetics graph. The X-axis is the substrate concentration, and the reaction velocity is shown on the Y-axis.  $K_M$  and  $V_{\text{max}}$  could be defined with the following equation .. 165
- Figure 57.** The double-reciprocal plot of Michaelis-Menten equation, Lineweaver-Burk plots ( $1/V$  versus  $1/[S]$ ), where  $-1/K_M$  is X-intercept,  $1/V_{\text{max}}$  is Y-intercept, and  $K_M/V_{\text{max}}$  is the slop. .... 166

## List of Tables

<b>Table 1.</b> Selected information of 12 human proteins from the Cathepsin family obtained from the UniProt website (UniProtKB 2018, Sep 09). .....	49
<b>Table 2.</b> Information obtained from multiple sequence alignment (CLUSTALX-2.1) of the amino acids of 12 members of the cathepsins family as well as <i>IHUC</i> structure (Clustal: Multiple Sequence Alignment AUG,2012). The results show the fully, strongly (amino acid alter to another amino acid with similar biochemical characteristics) and weakly conserved (amino acid alter to another amino acid with different biochemical characteristics) variable residues of CTB and <i>IHUC</i> compared to the rest of the cathepsin family. ....	52
<b>Table 3.</b> Information obtained from multiple sequence alignment (CLUSTALX-2.1) of the amino acids of the cathepsins family (Clustal: Multiple Sequence Alignment AUG,2012)., which used to identify which cathepsin is the most closely related to CTB. ....	52
<b>Table 4.</b> Information obtained from the RCSB PDB website (RCSB PDB 2018, OCT 30) about five of the experimentally determined structures for human CTB, as well as two experimentally determined structures of human CTC and CTD. ....	57
<b>Table 5.</b> The position of SNPs that cause amino acid changes in CTB found in MAF between 0.01 to 0.5 for the entire population which has been studied in 1,000 Genomes databased These amino acid substitutions are located on the propeptide of CTB. ....	74
<b>Table 6.</b> The position of SNPs that cause amino acid changes in CTB found in MAF between 0.001 and 0.01 of the population. ....	76
<b>Table 7.</b> Lead probe compounds discovered computationally (Daniel Tesolin, Private Communication) (Molport 2018, November 16). ....	83
<b>Table 8.</b> The structure of the selected ligands for ATTO680 conjugation (Kamstra, Dadgar et al. 2014). ....	84
<b>Table 9.</b> Comparison of kinetic constants between Cbz-Lys-Lys-PABA-DCMF and other substrate-based probes. Nine (9) concentration of the fluorescent substrate was used to determine $K_M$ , $V_{max}$ , and $k_{cat}$ values under the Michaelis-Menten model. Assays were performed at pH 5.5 and 37°C, with an enzyme concentration of 5 nM. ....	146



## Abbreviations

a.u.: arbitrary units .....	98
ABPs: activity-based probes .....	22
Ac: Acetyl .....	37
Ac-Arg-Arg-AFC: Acetyl-Arg-Arg-7-Amino-4-trifluoromethyl coumarin .....	101
AIBS: aggregation-induced blue shift.....	107
AIE: aggregation-induced emission.....	107
AMC: 7-Amino-4-methylcoumarin .....	23
AOMK: acyloxymethyl ketone.....	35
BCA: bicinechonic acid.....	110
Boc: tert-butoxycarbonyl protecting group.....	95
BSA: Bovine Serum Albumin .....	108
CB-CNP: cathepsin B glycol chitosan nanoparticle.....	24
Cbz: Benzyloxycarbonyl.....	3
CD44: cell surface adhesion receptor .....	32
CTB: Cathepsin B.....	2
CTK: Cathepsin K.....	21
CTL: Cathepsin L .....	21
CV: coefficient of variation .....	106
$C\alpha$ : alpha carbon .....	57
DCMF: 2', 7'-dichloro-6'-Methoxy-Fluorescein.....	3
DMEM: Dulbecco's Modified Eagle Medium.....	109
EC <sub>50</sub> : half-maximal effective concentration.....	3
ECM: extracellular matrix .....	20

EMT: epithelial-to-mesenchymal transition .....	31
EP0: N-[(3R)-4-ethoxy-3-hydroxy-4-oxobutanoyl]-L-isoleucyl-L-proline.....	43
E-S: enzyme-substrate .....	141
F-F-FMK: Phe-Phe-Fluoromethyl ketone.....	101
FP: fluorescence polarization.....	2
<i>GB123: Carbo benzoxy-capped Phe-Lys(Cy5)-acyloxymethyl ketone.....</i>	<i>22</i>
GM: <i>Geldanamycin</i> .....	23
Hsp90: heat shock protein 90.....	23
IGSR: International Genome Sample Resource.....	27
kDa: kilodaltons .....	20
$K_i$ : Dissociation constant for the enzyme-inhibitor complex .....	34
MAF: Minor Allele frequency .....	70
MFAM: multiphoton fluorescence anisotropy microscopy .....	39
MOE: Molecular Operating Environment .....	44
<i>MTT: 3-[4,5-dimethylthiazol-2-yl]-2,5 diphenyl tetrazolium bromide .....</i>	<i>148</i>
NHS: <i>N</i> -hydroxysuccinimide .....	26
NIR: Near-infrared.....	22
ns: nanosecond .....	26
p11: annexin II tetramer.....	32
PABA: <i>p</i> -Aminobenzyl alcohol.....	3
PBS: phosphate-buffered saline .....	108
R <sup>2</sup> : R-squared .....	97
RER: Rough Endoplasmic Reticulum.....	45
SNPs: single nucleotide polymorphisms.....	3
TFA: Trifluoroacetic acid .....	95

TGF- $\beta$ : transforming growth factor-beta .....	32
TICT: twisted intermolecular charge transfer .....	107
TLR3: toll-like receptor 33 .....	33
tPA: Tissue plasminogen activator .....	32
uPA: urokinase-type plasminogen activator .....	32
uPAR: urokinase plasminogen activator receptor.....	32
VCF: Variant Call Format .....	69

## **Chapter 1: Introduction**

### **1.1. Background**

According to the Canadian Cancer Society, approximately 50% of all Canadian people are expected to develop cancer (Ahmed and Shahid 2012, Turk, Stoka et al. 2012). Cathepsins are proteases that are commonly found in the lysosomes of healthy cells. Cysteine cathepsins belong to the family of papain-like cysteine proteases. These enzymes have unique active-site properties and are found expressed in various healthy tissues. Cathepsins have known roles in extracellular space (Puri and Roche 2008, Wartmann, Mayerle et al. 2010), cytosol (Reiser, Oh et al. 2004, Sever, Altintas et al. 2007), and the nucleus (Goulet, Baruch et al. 2004, Alcalay, Sharma et al. 2008, Chen, Dong et al. 2017). Additionally, substantial evidence confirms the role of cathepsins in many physiologic and pathophysiologic cellular processes, including digestion and regulation of other proteases (Turk, Stoka et al. 2012). Classification of different cathepsins are based on the active site amino acids responsible for peptide hydrolysis: serine (cathepsins A and G); aspartic (cathepsins D and E), and lysosomal cysteine cathepsins consisting of eleven different proteases (cathepsins B, C, L, O, S, W, V, F, H, K, X, and Z), (Reiser, Adair et al. 2010). In general, cathepsins are capable of efficiently cleaving a variety of protein and peptide-based substrates while having high stability in acidic environments characteristic of the lysosome as well as endosomes. In addition to the normal function of this family, many studies performed using human tumor samples have implicated several cathepsins in cancer invasion and progression (Gocheva, Zeng et al. 2006). To date, cathepsins B, L, S, X are recognized to play pivotal roles in cancer progression. According to a recent report, cathepsin S is found in the serum of gastric cancer patients (Liu, Liu et al. 2016). Additionally, other reports confirmed the overexpression of

cathepsins L and X in gastrointestinal stromal tumors and gastritis and gastric cancer, respectively (Fox, Batchelder et al. 1995, Miyamoto, Iwadate et al. 2011).

A review of the scientific literature demonstrates that cathepsin B plays a critical role in cancer as it is involved in a variety of pathological processes (Gondi and Rao 2013). In this study, our focus is on CTB, since it has great importance to cancer therapy (Gondi and Rao 2013). In humans, CTB is encoded by the *CTSB* gene, which is located on the short arm of chromosome 8. CTB has comprised of two chains: a heavy chain (amino acid 129-333 with a molecular weight of 25-26 kilodaltons (kDa)) and a light chain (amino acid 80-126 with a molecular weight of 5 kDa) with a total molecular weight of 38 kDa; (Mach, Stuwe et al. 1992, Iacobuzio-Donahue, Shuja et al. 1997). CTB is expressed as an inactive zymogen that requires proteolytic processing to become catalytically active within the low pH environment of the lysosomes (Turk, Stoka et al. 2012, Aggarwal and Sloane 2014). This enzyme participates in intracellular degradation and protein turnover and possesses both endopeptidase and exopeptidase activity (Sobic, Mirkovic et al. 2013). Overexpression of CTB has been reported in human prostate, colorectal, glioma, melanoma, and breast cancers in humans (Gondi and Rao 2013). Experimental and clinical evidence has linked CTB to tumor invasion and metastasis (Podgorski and Sloane 2003, Li, Chen et al. 2011, Aggarwal and Sloane 2014). CTB has been proposed as a biomarker for various cancers (Terasawa, Hotani et al. 2015, Chen, Dong et al. 2017) and represents a potential target for diagnostic purposes and the monitoring of treatment outcomes. The enzyme is generally found in the lysosomes but is excreted into the extracellular matrix (ECM) during cancer progression (Szpaderska and Frankfater 2001). A variety of techniques have been developed for monitoring the expression of CTB, including DNA, RNA, or nucleic acid sensors (Nguyen, Lee et al. 2019). A variety of optical imaging methods have also been developed for studying CTB activity in cell and animal studies.

These methods rely on the design and synthesis of fluorogenic probes capable of detecting protease activity in complex biological mixtures (Lai, Chang et al. 2004, Dai and Kool 2011).

## **1.2. Research Problem and Hypothesis**

Due to the roles of CTB in cancer progression and invasion (Duffy 1992), this enzyme has been targeted for the design of specific inhibitors to prevent activity and, as a result, inhibit tumors from metastasizing. This approach faces two main challenges: first: other members of this family, especially cathepsin K (CTK) and cathepsin L (CTL) can compensate for the loss of CTB activity; second: the active site residues in CTB are common to most members of this family (Khouri, Plouffe et al. 1991). Although these are considerable challenges for inhibitor design, they are less challenging for the design of diagnostic probes, since the diagnostic probes can target the amino acids within close distance of the active site residue for binding interaction. Evaluation of CTB activities associated with tumors by fluorescent probe considered to be a diagnostic or therapeutic target for different types of cancer. (Alford, Ogawa et al. 2009, Ofori, Withana et al. 2015, Kramer, Renko et al. 2017). Molecular imaging probes do not need to bind directly to the active site of the enzyme, but they still could bind to distant pockets within the protein allowing for more flexibility in the design of probes specific to CTB. This flexibility is absent in the design of inhibitors and probes that depend on enzymatic activity for activation. Besides, similar to CTB, overexpression of CTK and CTL has also been associated with tumor progression and invasion (Podgorski, Linebaugh et al. 2009, Sui, Shi et al. 2016), and may serve as a biomarker for cancers; therefore, small organic molecules intended for use as molecular probes may not need the same degree of specificity as inhibitors of CTB activity. Since the role of CTB in cancer progression has been proven clinically (Podgorski and Sloane 2003, Gondi and Rao 2013, Aggarwal and Sloane 2014),

targeting the over-activity or overexpression of this protein with fluorescent small molecules may be used as a promising approach for cancer detection (Li, Chen et al. 2011, Gondi and Rao 2013).

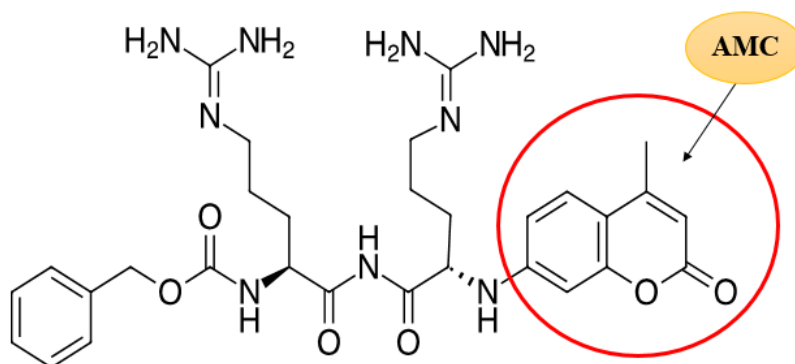
Currently, two classes of fluorescent molecular probes targeting proteases have been reported in the literature (Blum, Weimer et al. 2009): 1) activity-based probes (ABPs) that are covalent inhibitors and can be made in a quenched form (Kato, Boatright et al. 2005, Ofori, Withana et al. 2015), and an unquenched form (Blum, von Degenfeld et al. 2007), 2) probes that act as a substrate and become detectable or fluorescent upon enzymatic cleavage (Chowdhury, Moya et al. 2014, Garland, Yim et al. 2016, Wang, Li et al. 2016, Chen, Lee et al. 2017).

In addition to these probes, there is another series of probes (with no inhibitory activity) that are unquenched and do not need enzymatic activation for becoming fluorescent (Du, Moulick et al. 2007).

The two best examples of quenched ABPs before enzyme activation would be *MB-074* (Ofori, Withana et al. 2015) and *6QCNIR* (Ofori, Withana et al. 2015) reported by Dr. Matthew Bogyo. Matthew Bogyo designed and synthesized a series of elegant activity-based, near-infrared (NIR) fluorogenic probes (Kato, Boatright et al. 2005, Blum, von Degenfeld et al. 2007) for CTB that were used in cell and animal studies as well as to identify tumor margins during surgery (Kato, Boatright et al. 2005, Verdoes, Bender et al. 2013, Ofori, Withana et al. 2015). An example of unquenched ABPs would be *Carbo benzoxy-capped Phe-Lys(Cy5)-acyloxymethyl ketone (GB123)* designed by Blum group (Blum, von Degenfeld et al. 2007) which freely penetrates cells and has a bright signal accumulation in the living mice tumor. The ABPs (quenched and unquenched) irreversibly bind to the active site of the specific enzyme classes; therefore, despite their favorable outcome these probes still suffer from the number of limitations: 1) lack of selectivity when most of the family of the target enzyme share the common active site residues; 2) 1:1 stoichiometry

between probe and protein and lack of amplification by multiple processing events; 3) trapping in the lysosome.

An example of a substrate-based probe is Cbz-Lys-Lys-PABA-7-Amino-4-methylcoumarin (AMC) (Chowdhury, Moya et al. 2014), which was reported to be highly selective toward human CTB in live cells over other members of the cathepsin family (Chowdhury, Moya et al. 2014). Other similar probes include Cbz-Arg-Arg-AMC (Khouri, Plouffe et al. 1991), which is shown in **Figure 1**. A significant problem facing these probes is that they compete with the high fluorescent background of live cells due to autofluorescence ( $\lambda_{em} = 420\text{--}470\text{ nm}$ ) (Andersson, Baechi et al. 1998) and the released fluorophore AMC rapidly diffuses out of cancer cells.



**Figure 1.** The chemical structure of Cbz-Arg-Arg-AMC.

The fluorescently labeled *Geldanamycin* (*GM-cy3B*) would be an example of an unquenched probe that does not require protein activation to become fluorescent, reported by the Moulick group (Du, Moulick et al. 2007). They identified the high affinity and specificity of *GM-cy3B* toward heat shock protein 90 (Hsp90) with  $EC_{50}$  value at  $58.5 \pm 20\text{ nM}$  with fluorescence



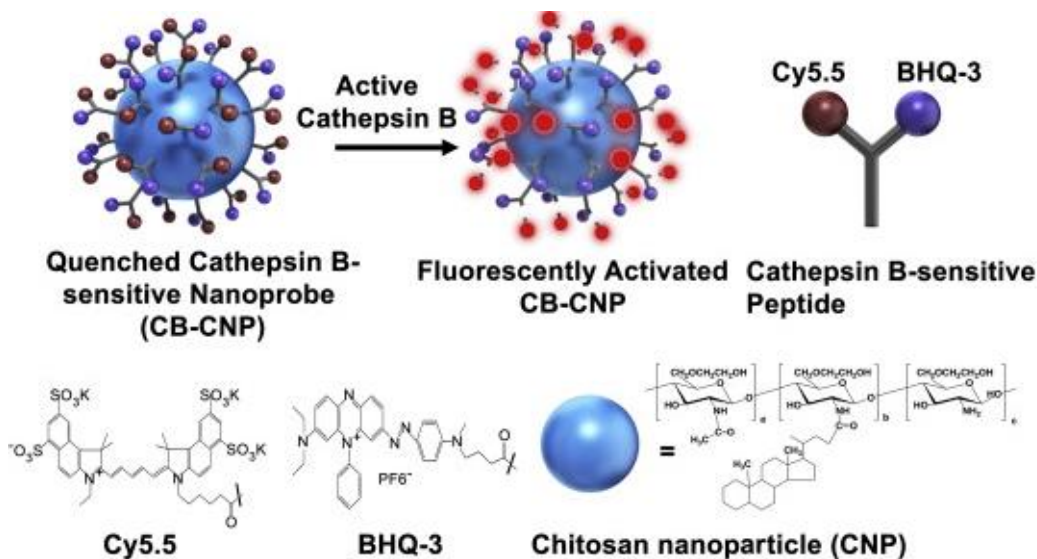
polarization assay in breast cancer cell line. There was no example of this type of probes toward CTB reported in the literature.

### 1.3. Research Strategy

To circumvent the obstacles mentioned above, as an alternative to ABPs, using substrate-based probes as reporters can be beneficial for signal amplification due to the processing of many substrates by an enzyme over time. The substrate-based probes require CTB catalytic activity for activation and the release of a fluorophore or quencher.

Using a fluorophore with a certain wavelength, preferably above autofluorescence of the live cells and ideally in the near-infrared (spectral range from 650 to 900 nm) part of the electromagnetic spectrum (Kisin-Finfer, Ferber et al. 2014), would be beneficial to distinguish between the signal releasing from the substrate hydrolysis and background. Near-infrared probes have an advantage over other conventional probes (spectral range from 380 to 650 nm) in cellular and tissue imaging because of the low absorption of the excitation radiation by endogenous biological molecules at higher frequencies. Moreover, at the NIR range, there is less autofluorescence compared to visible wavelengths – allowing for a higher signal-to-noise level (Zhang, Bloch et al. 2012). Besides that, using NIR probes enable scientists to achieve imaging in deeper tissues due to their longer excitation wavelengths (Chen, Lee et al. 2017). Recently, these probes have become favored due to their low interference and reduced damage to living cells. Past research has been dedicated to designing NIR probes and substrates for CTB (Habibollahi, Figueiredo et al. 2012, Kisin-Finfer, Ferber et al. 2014, Wang, Li et al. 2016, Chen, Lee et al. 2017). An example of the substrate-based probe with NIR spectrum is the cathepsin B *glycol chitosan nanoparticle (CB-CNP)* with chemical structures of *Cy5.5*, *BHQ-3*, containing NIR dye

(Cy5.5) with the  $\lambda_{\text{ex}} = 678 \text{ nm}$  and  $\lambda_{\text{em}} = 694 \text{ nm}$  and a quencher (*BHQ-3*) which was reported by Ryu group (Ryu, Na et al. 2014) (**Figure 2**).



**Figure 2.** The chemical structure of *CB-CNP*, *Cy5.5* (dye), and *BHQ-3* (quencher).

Adapted from:(Ryu, Na et al. 2014)

The next strategy requires selecting the fluorophore to cover a broad range of pH – ideally acidic pH (pH of the lysosome (5.5) in which CTB is localized) (Bohley and Seglen 1992). *Cyanine* dye would be a great example of covering the wide range of pH (6.0–8.0), although it subsequently loses its fluorescent characteristics in the less acidic environment (Briggs, Burns et al. 2000).

In addition to the substrate-based probe, designing the unquenched, non-reactive probes which efficiently and tightly bind to the protein with no inhibitory activity – thereby revealing

enzyme accumulation in tissues, would be another strategy. This strategy brings novelty to our study, as there are no examples of these types of probes for measuring CTB overexpression.

For resolving the limitations mentioned above and obtaining a direct measurement of CTB overexpression in the context of tumor metastasis and progression, two non-reactive NIR fluorescent probes and one substrate-based probe were synthesized and evaluated in this study.

For non-reactive probes, the initial step was selecting computationally identified ligands from previous work reported by Kamstra et al. (Kamstra and Floriano 2014) and converting them into fluorescent probes through conjugation to a dye having NIR emission in the range of 700 nm. All the ligands identified by the Kamstra group contain an amine group which can react with an amine-reactive fluorophore for conjugation. *ATTO-TEC* offers a variety of new generation fluorescent dyes for amino groups labeling with the spectral region from 350 nm to 750 nm (Buschmann, Weston et al. 2003). The most common amine-reactive group in commercial dyes among these derivatives is the dye contains *N*-hydroxysuccinimide (NHS)-esters group (Hughes, Rawle et al. 2014). ATTO680 NHS-esters with excitation and emission in the NIR spectral range ( $\lambda_{\text{ex}} = 675 \text{ nm}$  and  $\lambda_{\text{em}} = 707 \text{ nm}$ ) were selected for forming a chemically stable amide bond with the subjected ligands. ATTO680 NHS-esters is a hydrophilic fluorophore with a fluorescence lifetime of 3.04 nanosecond (ns) (Buschmann, Weston et al. 2003, Hughes, Rawle et al. 2014). The hydrophilic characteristic of this dye allows better utilization *in vitro* and *in vivo*. One of the most important characteristics of the dye which has a strong effect on fluorescence polarization is a lifetime. Fluorophores with lifetimes above 3 ns, are considered suitable for providing a distinct FP between free fluorophore and fluorophore bound with protein (molecular weight above 10 kDa) (French, Burton et al. 2000, Owicki 2000). To increase specificity towards CTB, the probes were selected based on predicted contacts to unique residues in the active site of CTB. These probes

were synthesized and subjected to inhibition and fluorescence polarization assays to confirm binding to CTB.

Besides, we reassessed the scaffold used to produce substrates previously developed with Dr. Phenix's group (Chowdhury, Moya et al. 2014). Here, a similar substrate-based probe was designed and synthesized but with the dye, DCMF (**Figure 26**) as a fluorophore, instead of AMC. This new probe would be expected to be quenched in the intact substrate but have fluorescence emission at higher wavelengths upon CTB cleavage (DCMF  $\lambda_{em} = 522$  nm). Having an improved fluorescent reporter should lead to lower background noise while also maintaining fluorescence at low pH (Mugherli, Burchak et al. 2006). In an activity assay toward two proteases proteins (CTL and CTB), we confirmed the binding affinity and selectivity of this probe toward CTB.

Finally, the International Genome Sample Resource (IGSR) (1,000 Genomes database) (IGSR: The International Genome Sample Resource 2015, June 30) was used to investigate single nucleotide polymorphisms present in the gene encoding CTB in human populations. This was done for two main reasons: first, to detect any mutations located on the *CTSB* gene, that may be involved in the development of cancers; second, to determine if there are SNPs in the population (1,000 Genomes project) that prevent binding of the developed probes to the protein, CTB.

#### **1.4. Research Significance**

In this study, the two novel NIR fluorescent non-reactive probes: CID8795ATTO680 (**Figure 27**) and CID535684ATTO680 (**Figure 28**) and a novel substrate-based fluorescent probe Cbz-Lys-Lys-PABA-DCMF (**Figure 31**) for CTB were synthesized, characterized, and evaluated in preliminary biochemical studies. This study attempts to overcome the limitations mentioned above, including overcoming the low specificity of existing probes towards CTB due to high active

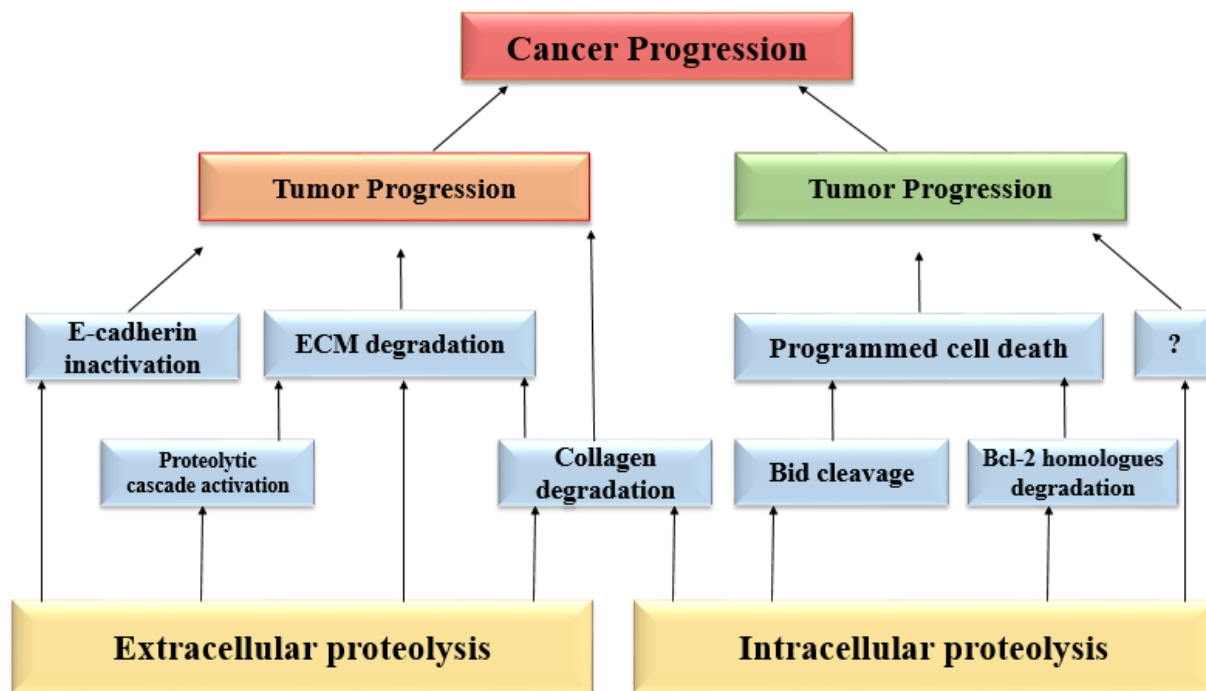
site homology with other members of the cysteine protease family. Also, the expansion of the number of CTB probes that have excitation and emission wavelengths beyond autofluorescence associated with live cells. To accomplish this, we describe the synthesis of a non-reactive probe (CID8795ATTO680 **Figure 27**) and its evaluation as a NIR probe for detecting CTB. Additionally, we describe the development of a new CTB substrate (Cbz-Lys-Lys-PABA-DCMF **Figure 31**). These probes would be an excellent low-cost method for CTB detection in cancer as well as potentially useful as an intraoperative tool to reveal tumor margins during surgical procedures.

## Chapter 2: Literature Review

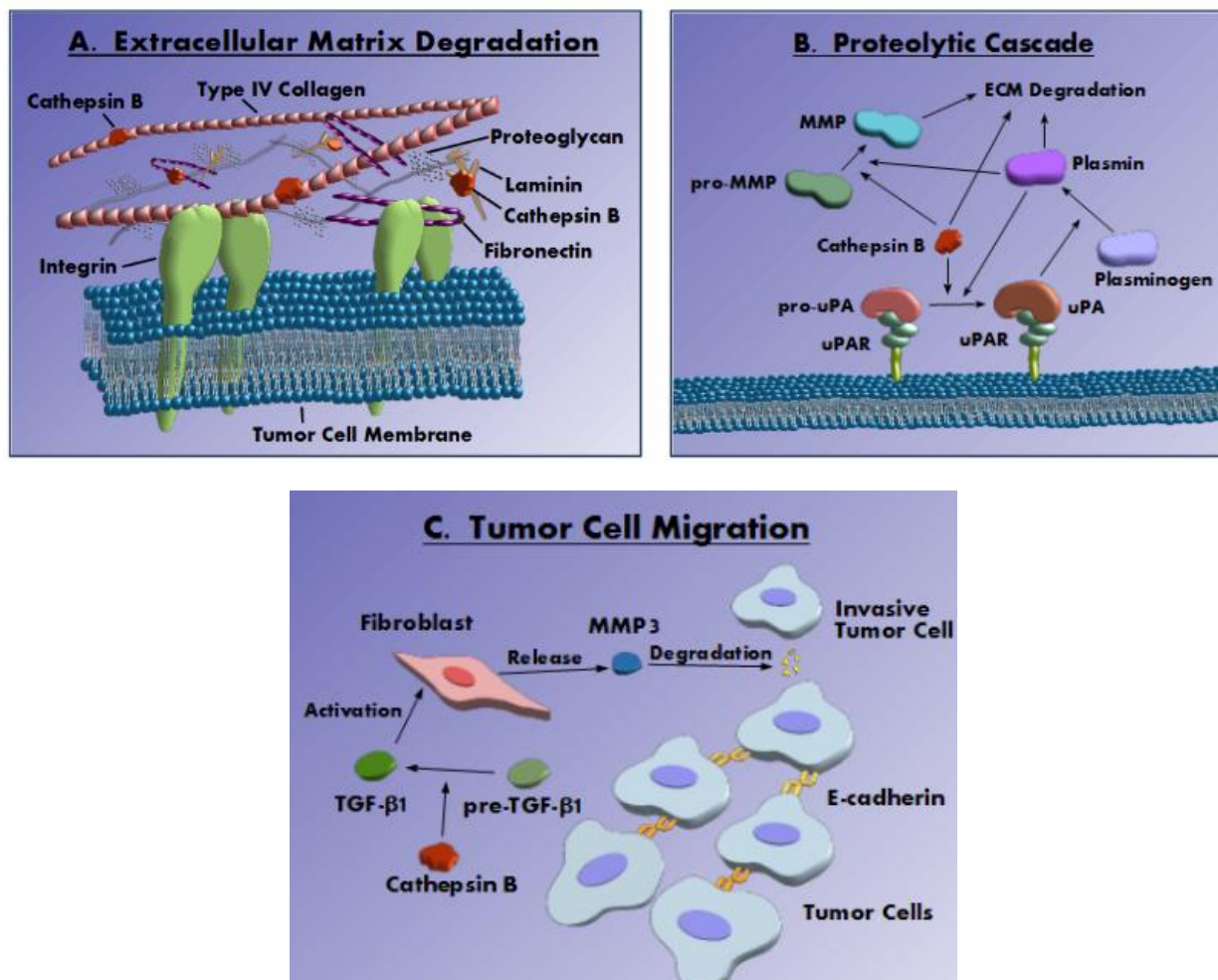
Currently, cancer is the most common cause of death in Canada leading to intense research into finding better diagnostic techniques and potential cures (Public Health Agency of Canada 2010, March 29). So far, more than 7000 reviews have confirmed the presence of proteases playing an important role in various types of cancers (Emmert-Buck, Roth et al. 1994), (Bohley and Seglen 1992), (Aggarwal and Sloane 2014). Cathepsins are a diverse group of proteases categorized into different families based on their active site structure – each of which employs a different catalytic nucleophile (e.g. aspartate, serine or cysteine) that efficiently hydrolyzes peptide bonds in the acidic environment of the lysosome (Bohley and Seglen 1992, Jedeszko and Sloane 2004, Patel, Homaei et al. 2018). The cysteine cathepsin family is responsible for protein turnover in many normal physiological processes but are linked to a variety of disease processes, including: Alzheimer's disease; pancreatitis; osteoarthritis; several ocular disorders, and different types of cancers (Lai, Chang et al. 2004, Im, Venkatakrishnan et al. 2005, Hook, Jacobsen et al. 2015, Oortveld, van Vlijmen-Willems et al. 2017, Vizovisek, Fonovic et al. 2019).

Among all members of the cysteine cathepsin family, cathepsin B has high expression in normal tissues, such as thymus; blood; corpus callosum; pituitary gland, and low expression in heart and uterus (UCSC Genome Browser 2013, Dec). The role of CTB in cancer-promoting processes includes several steps: the first step of the cancer progression requires the surrounding tissue to be cleared and the degradation of the extracellular matrix creating a pathway for tumor growth. This is followed by its involvement in the proteolytic cascade – a complex and redundant network of protease enzymes that are responsible for regulating another protease activity. For example, cysteine cathepsins help to activate matrix metalloproteinase and urokinase plasminogen activator resulting in further tissue invasion. The last step is E-cadherin cleavage at cadherin

junctions, which allows cancer cell migration (Kobayashi, Moniwa et al. 1993, Guo, Mathieu et al. 2002). These steps are summarized in **Figure 3** and **Figure 4**.



**Figure 3.** Involvement of cysteine cathepsins in cancer progression. Cancer can progress in two steps: extracellular and intracellular proteolysis, which goes under E-cadherin inactivation, and Bid cleavage, respectively.



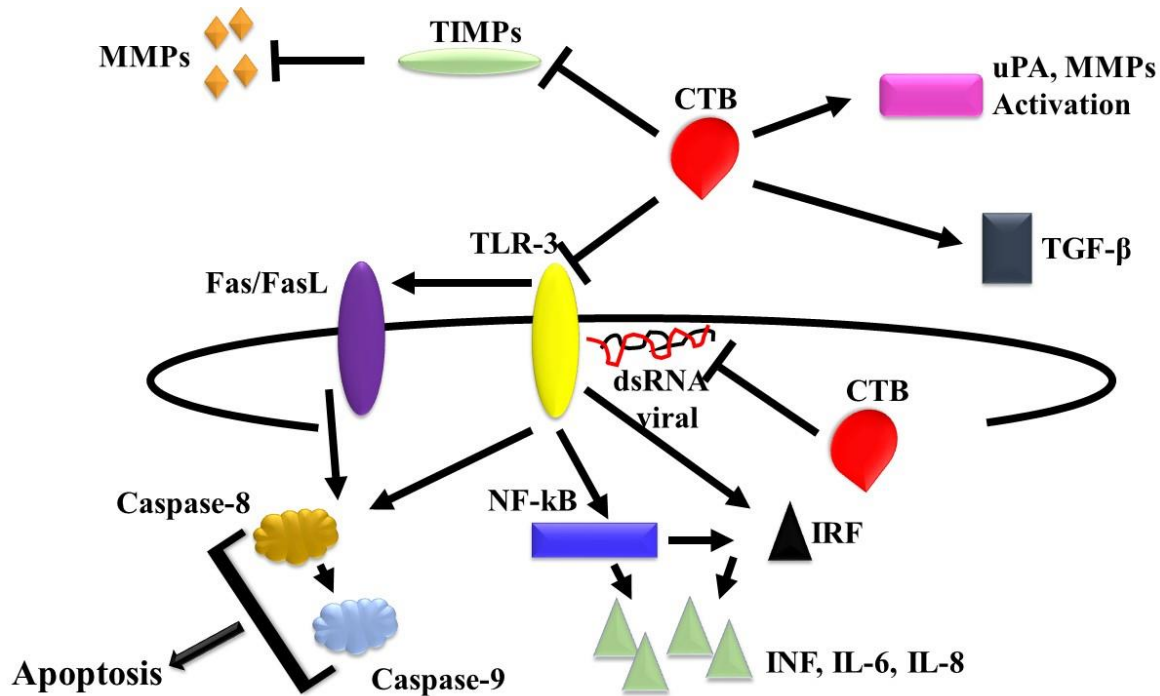
**Figure 4.** The three different mechanisms of the cathepsin B involvement in the activation of the proteolytic cascade. Cathepsin B could involve in this process in different ways: degradation of the extracellular matrix; converting pro-MMP to MMP, and activation of the pre-TGF-B1 to TGF-B1.

Adapted from: Lampe CM, Gondi CS (2014) cathepsin B Inhibitors for Targeted Cancer Therapy.

Initially, cancer cells undergo sequential steps, which include tissue architecture disruption by modulation of cell-matrix and cell-cell contacts. The epithelial-to-mesenchymal transition (EMT) controls the flexibility of cancer cells. This mechanism is also involved in wound healing and organism development under normal physiological conditions. During EMT, cells lose their polarity and connections and obtain pro-migratory characteristics to cross the extracellular matrix



(ECM) or the basement membrane. Indeed, EMT contributes to increased cancer cell motility, invasion, and progression (Thiery, Acloque et al. 2009, Micalizzi, Farabaugh et al. 2010, Kryczka, Papiewska-Pajak et al. 2019). Recent studies suggest a role of protease activity in areas of the cell which have direct contact with the ECM; therefore, precisely controlling cell invasion *in vivo* (Rottiers, Saltel et al. 2009). The role of cathepsins in both tumor progression and metastasis supported by numerous clinical reports – although the precise mechanisms are still not clear (Gocheva, Zeng et al. 2006, Bakst, Xiong et al. 2017). As an oncogenic role of cathepsin B, this protein interacts with cystatins (Pavlova and Bjork 2003) and annexin II tetramer (p11) (Ma, Finley et al. 2000). These interactions put CTB at key positions for the proteolytic activation of ECM; therefore, inducing ECM degradation. The active role of cathepsin B in initiating the proteolytic cascade has been confirmed with various studies, which include involving: urokinase-type plasminogen activator (uPA); plasminogen to plasmin, and activation of latent transforming growth factor-beta (TGF- $\beta$ ) (Somanna, Mundodi et al. 2002). Cathepsin B appears to increase the activity of the matrix metalloproteinase (MMPs) by destroying their inhibitors, thereby enabling ECM degradation (Kostoulas, Lang et al. 1999). The *CTSB* gene activation in cancer cells is mediated by extracellular signals. These signals indeed are mediated through the urokinase plasminogen activator receptor (uPAR)-uPA-integrin and the cell surface adhesion receptor (CD44)-MMP, resulting in overexpression of CTB mRNA and processed to pro-cathepsin B. The pro-cathepsin B activation goes through multiple mechanisms such as tissue plasminogen activator (tPA) and uPA). Of note, the essential role of CTB for organizing protease cascade and addressing it towards invasive fronts of metastatic cells is considerable (**Figure 5**) (Murphy, Ward et al. 1992).



**Figure 5.** Representation of signaling and protease cascade events induced by cathepsin B on cancer cells.

Furthermore, the function of CTB for activation of trypsinogen-1 in pancreatitis was confirmed by Lindkvis' group (Lindkvist, Fajardo et al. 2006); pancreatic inflammation induces the risk of pancreatic cancer (Whitcomb 2004). In 2012, Garcia-Cattaneo reported the signaling and cleavage of toll-like receptor 3 (TLR3) dependency on cathepsin B (Garcia-Cattaneo, Gobert et al. 2012). The TLR3 activation is also known to enable apoptosis in tumor cells through caspases 8 and 9 activations (Sun, Zhang et al. 2011). Cathepsin B directs tumor cells to survive by inactivation of the TLR3-mediated apoptotic pathway.

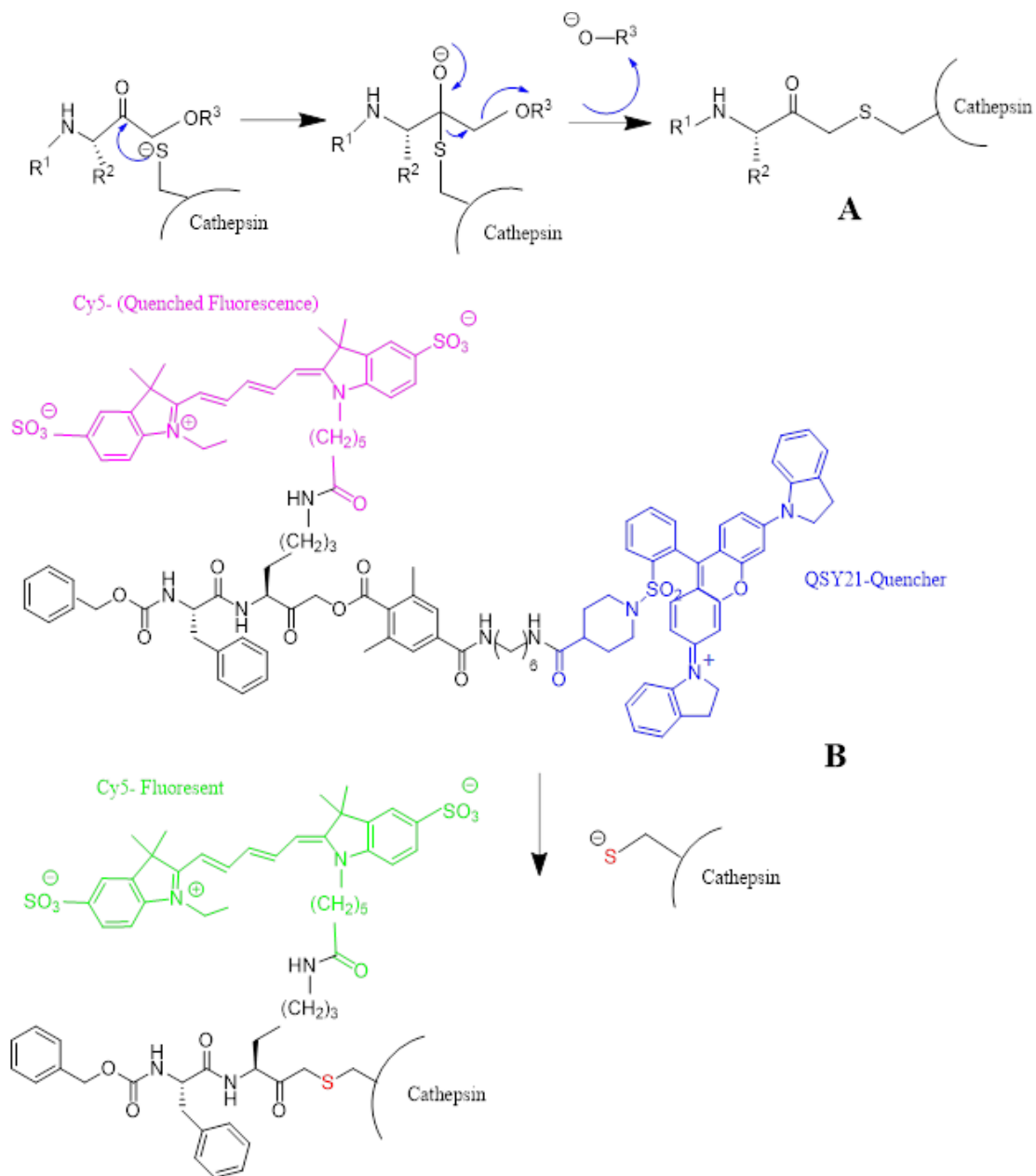
Cathepsin B activity has a dual nature that could serve as a pro-oncogenic molecule (involve in ECM degrading, angiogenesis, and metastatic inducing) and anti-apoptotic molecule (involve in autophagy and immune response). The low pH dependency of CTB activity is significant due to the observation of the same environment around the invasive front of the tumor cells (more

acidic than normal cells), (Tannock and Rotin 1989). This confirms the hypothesis that the tumor microenvironment (low pH) promotes CTB activity. Cathepsin B is localized in the lysosomes (Bohley and Seglen 1992); it was also reported to be localized in the mitochondria (Frohlich, Schaumburg-Lever et al. 1993) where cell death is initiated (Muntener, Zwicky et al. 2004).

Studies have shown that CTB hydrolytic activity can be targeted to release drugs for anti-cancer therapy (Wei, Gunzner-Toste et al. 2018). Cathepsin B has also been reported to be a potential diagnostic target for different types of cancers (Campo, Munoz et al. 1994, Ruan, Hao et al. 2015). Researchers have attempted to design specific inhibitors of CTB activity to impede tumor metastasis. For example, the Fox group in 1992 reported a synthetic inhibitor (*PCBI*) for cathepsin B with yielding the  $K_i$  at 0.4 nM (Fox, Demiguel et al. 1992). Also, in 2013, Izidor Sosic and his team described a reversible inhibitor (*nitroxoline (5-nitro-8-hydroxyquinoline)*) to inhibit endopeptidases activity of the CTB (Sosic, Mirkovic et al. 2013). However, this approach has many limitations, primarily due to the overlapping activity and redundancy of cysteine cathepsin activity in tumor progression (Aggarwal and Sloane 2014). For example, any therapeutic benefits of CTB inhibition are blunted by the compensation of proteolytic activity of cathepsin K and L leading to tumor metastasis. This overlapping proteolytic activity results from high homology of the residue of the active site between most of the members of this family with CTB thus making selective probes and inhibitors a major challenge (Gopinathan, Denicola et al. 2012, Loser and Pietzsch 2015). Designing a probe selective for human CTB localization and detection *in vitro* and *in vivo* would be useful for investigating the role of this protease in cancer (Wang, Li et al. 2016). There have been numerous reports that describe the design and synthesis of activity-based and substrate-based molecular imaging probes studying CTB activity in cells and animals (Blum, von Degenfeld et al. 2007, Kisin-Finifer, Ferber et al. 2014, Lee, Kim et al. 2014, Mertens, Schmitz et

al. 2014, Tian, Li et al. 2014, Wang, Li et al. 2016), and as well, tools for surgical resection to identify tumor margins (Ofori, Withana et al. 2015).

The first series of probes for CTB localization and detection would be activity-based probes, which are suicide inhibitors, allow dynamic monitoring of protease activity. These activity-dependent small molecules are processed by the enzyme into a reactive intermediate and covalently bonds a reporter group to the target protease; therefore, it provides the activity level of proteases in cell or tissue (Verdoes, Oresic Bender et al. 2013). The activity-based probes covalently bind to the target enzymes while simultaneously releasing a quenching group, thereby generating enhanced fluorescence and allowing localization of the target. The first example of ABPs was reported by Bogoy in 2005, the Bogoy group, via exploring irreversible inhibitors, developed activity-based fluorescent probes for CTB; however, these probes also reacted efficiently with CTL (Blum, Mullins et al. 2005, Blum, von Degenfeld et al. 2005). Bogoy and his team in 2007, also reported the fluorogenic mechanism-based inhibitors using *acyloxymethyl ketone* (*AOMK*) as a reactive group. The reaction mechanism of the *AOMK* group with Cysteine active site illustrates in **Scheme 1**. This series of probes were quenched before being recognized and hydrolyzed by the target protease (Blum, von Degenfeld et al. 2007).

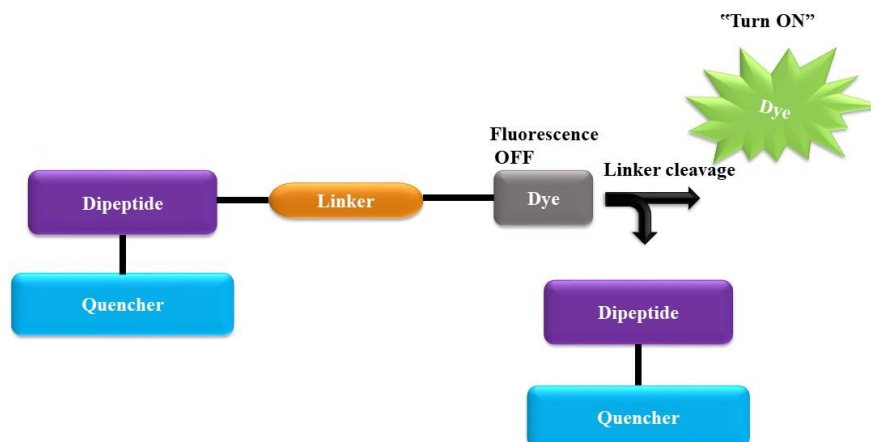


**Scheme 1:** Covalent labeling of cathepsin active sites by employing *AOMK* inhibitors. (A) The reaction mechanism of the *AOMK* inhibitors; (B) Quenched activity-based probes (Blum, von Degenfeld et al. 2007). The fluorescence of probes is low. Prior to enzymatic labeling since the fluorophore (Cy5), and internal quencher (QSY21) are held in close vicinity to each other.

Although these probes are excellent for profiling CTB and CTL in human models of cancer with high enzyme expression, this strategy suffers from a 1:1 stoichiometry between probe and enzyme. In cases where CTB expression is low, the fluorescent signal may be below detection limits and some derivatives suffer from a small fluorescence turn-on ratio.

Alternately, substrate-based probes for CTB may have enhanced sensitivity *in vitro* and *in vivo* compared to the activity-based probes since one enzyme can continuously activate multiple substrates leading to the accumulation of the fluorescent reporter group (Ryu, Na et al. 2014). Substrate-based probes for detecting protease activity are attractive since they can be quenched in the intact peptide; therefore, they have low background signals until hydrolysis and activation by the target enzyme.

The simplest examples of a fluorogenic substrate-based probe is a short peptide sequence attached to a quenched dye, such as Ac-Arg-Arg-7-Amino-6,8-difluorocoumarin (AFC) (commercial substrate for CTB) or Z-Phe-Arg-AMC (commercial substrate for CTB and CTL) (Los, Walczak et al. 2000). The two dyes, 7-Amino-4-methylcoumarin (AMC) or 7-Amino-6,8-difluorocoumarin (AFC) have been conjugated to a tetra-peptide cleavage site generating probes for caspase (Thornberry, Chapman et al. 2000). The general design and mode of action of these series of substrates are shown in **Figure 6**.



**Figure 6.** The mode of action of the fluorescent FRET-based probe.

During 2014, Ryu and his researchers reported a highly sensitive substrate-based probe for CTB consisting of a self-quenching CTB-cleavable peptide and chitosan nanoparticles (Ryu, Na et al. 2014). Fujii and his team developed a substrate-based fluorescent small-molecule probe that becomes highly fluorescent upon CTB activation (Fujii, Kamiya et al. 2014). These probes showed low selectivity toward CTB suffering from cross-reactivity with CTL. Recently, Dr. Phenix group developed the Cbz-Lys-Lys-PABA-AMC fluorogenic probe with high specificity to CTB. Although this probe is rapidly and selectively hydrolyzed by CTB, AMC readily diffuses out of the lysosomes thus rapidly losing fluorescent signals and requires excitation wavelengths that induce intracellular autofluorescence (Chowdhury, Moya et al. 2014). Others built upon this strategy by designing a substrate that employed the Cbz-Lys-Lys-PABA scaffold but used a lysosomotropic dye that remained trapped in the lysosomes after CTB hydrolysis (Wang, Li et al. 2016).

In addition to ABPs and substrate-based probes, there is another group of the small fluorescent molecule which are unquenched and do not need proteins for activation (non-reactive probes). Non-reactive probes refer to small-molecule fluorescent probes (unquenched) that bind to a protein and can potentially be measured by fluorescence polarization assays *in vitro*. Performing fluorescence polarization *in vivo* is challenging due to the lack of cellular spatial resolution, in addition to the potential to make temporal measurements (Piston 2010, Vinegoni, Dubach et al. 2016). Despite that, Dubach and co-workers in 2014, reported a ) live-cell imaging technique to locate and profile fluorescent small molecules, and target interaction in real-time at subcellular resolution *in vitro* and *in vivo* (Dubach, Vinegoni et al. 2014). Fluorescence polarization assays have been applied to study protein-protein, protein-DNA, and protein-ligand molecular interactions analysis (Jameson and Ross 2010). In 2007, Yuhong Du group reported the non-reactive probe (*GM-cy3B*) for Hsp90 which is an important target for cancer. They successfully measured the binding interaction of this protein toward their probe by fluorescence polarization *in vitro* (Du, Moulick et al. 2007). However, there is no example of non-reactive probes for CTB.

Overall, the limitations that we need to address in this study are as follow: 1) low selectivity of the probe toward CTB, due to active site homology of this protein with other members of cathepsin family, 2) the inhibitory effects of the probe toward CTB, due to compensatory feature of other members of this family for the loss of CTB activity, 3) high fluorescent background of live cells due to intracellular autofluorescence ( $\lambda_{em} = 420\text{--}470\text{ nm}$ ), 4) diffusion of the reporter (fluorescent dye) out of cancer cells due to low pH of the tumor microenvironment and CTB localization (lysosomes).



To address these shortcomings, we synthesized and evaluated two novel probes for assessing overexpression of CTB; a non-reactive probe with the desired feature of non-inhibitory activity toward CTB (CID8795ATTO680 **Figure 27**) with emission and excitation wavelengths in the NIR range, and a substrate-based probe (Cbz-Lys-Lys-PABA-DCMF **Figure 31**) with high specificity and selectivity toward CTB with the fluorophore reporter (DCMF) with excitation (480 nm) and emission (522 nm) above autofluorescence of the live cells as well as maintaining fluorescence at low pH.

## **Chapter 3: Cathepsin B: Sequence, Structure, and Function**

### **3.1. Introduction**

In this chapter, we compared the nucleotide sequence of human CTB with other members of the cathepsin family to identify the unique residues in this protein and map them on the 3D structure of human CTB. Besides, the predicted binding interaction of two subjected probes (CID8795ATTO680 and Cbz-Lys-Lys-PABA-DCMF) with human CTB is discussed. A variety of bioinformatics applications such as molecular docking and rational probe and drug design require a good understanding of functional sites within the target protein. The first strategy we employ is to compare the amino acid sequence of human CTB to other members of the cathepsin family to discover evolutionarily conserved residues that have functional activity. To predict the selectivity of novel probes, the active site residues of human CTB were compared to other members of the cathepsin family (in the same organism). Additionally, the 3D structure and binding pocket of CTB have been studied to assess the specificity of the probes. Finally, the two probes, CID8795ATTO680 and Cbz-Lys-Lys-PABA-DCMF, (docked structure with CTB) also screened to identify the binding interaction toward CTB (Daniel Tesolin, Private Communication)

### **3.2. Procedure and Materials**

#### **3.2.1. *CTSB* Gene**

Information about the *CTSB* gene (*Homo sapiens*) (ID: 1508) was found by searching through the NCBI gene bank (NCBI 2019, Aug).

### 3.2.2. CTB Protein

General information about proteins such as the length of the amino acid sequence, number of chains, and name and position of the active site residues can be obtained from the UniProt website (UniProtKB 2018, Sep 09). The amino acid sequence of human CTB was retrieved from the UniProt website (UniProtKB 2018, Sep 09) and saved as FASTA format (**Figure 8**). This sequence was used for multiple sequence alignment with CLUSTALX-2.1 software (Clustal: Multiple Sequence Alignment AUG,2012).

### 3.2.3. Similarity and Differences between Cathepsin B and Different Members of Cathepsin Family

The amino acid sequences of human CTB and eleven members of the human cathepsin family were obtained from the UniProt website (UniProtKB 2018, Sep 09). We compared the amino acids sequence of human CTB to other members of the human cathepsin family to identify any dissimilarities in the active sites that could guide the design of probes with high specificity toward CTB (UniProtKB - P07858 (CATB\_HUMAN) 2018, May 25, UCSC Genes 2018, NOV 24, NCBI 2018, OCT 08) relative to other members of the cathepsin family. The CLUSTALX-2.1 software was used to align the amino acid sequence of human CTB (UniProtKB - P07858 (CATB\_HUMAN)), the amino acid sequence corresponding to the human CTB crystal structure *IHUC* (amino acid sequence of mature enzyme), and eleven members of human cathepsin family including; cathepsins Z, L2, H, E, G, C, S, K, L1, L3, and, D. The reason for using CLUSTALX2.1 software was to include the *IHUC* amino acid sequence to limit the variable residues to those are within mature CTB.

The parameters used for running CLUSTALX-2.1 were as follows: substitution matrix: BLOSUM 30, gap opening penalty: 10, and gap extension penalty: 0.10. In the output of this program, the positions shown with "\*" indicate the fully conserved residue and "-" indicate a gap, ":" shows strongly conserved residue and "." means the residue is weakly conserved. In general, conserved residues refer to a nucleotide or amino acid which is identical across or within the species. Fully conserved residue refers to a single residue that is conserved in all sequences compared. Strongly conserved residue indicates conservation between groups of amino acids which has strongly similar properties, in contrast, the weakly conserved residue indicates conservation between groups of amino acids which has poorly similar properties (Chenna, Sugawara et al. 2003).

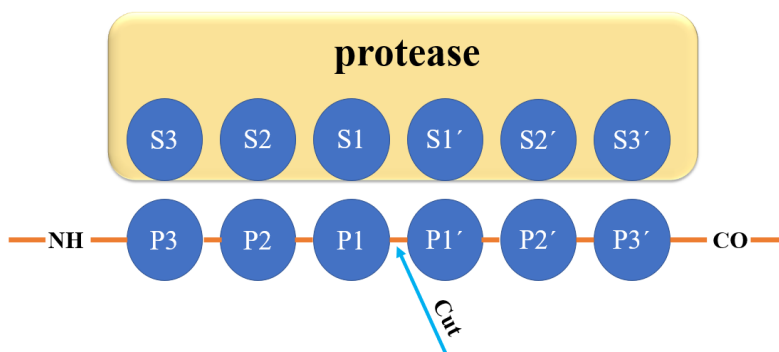
#### **3.2.4. 3D Structure of CTB**

Experimentally determined 3D structure of human CTB were identified at the RCSB website (RCSB PDB 2018, OCT 30) The list of the PDB files for CTB are as follows: *2IPP*; *1PBH*; *2PBH*; *3PBH*; *3AI8*; *1HUC*; *3CBJ*; *3CBK*; *3K9M*; *1GMY*, and *1CSB* (RCSB PDB 2018, OCT 30). The *1HUC* structure is a complete and non-complex structure of human CTB which was used for molecular docking of CID8795ATTO680 and Cbz-Lys-Lys-PABA-DCMF by a collaborator Daniel Tesolin. The *1CSB* is an inhibited complex structure of CTB in which the presence of ligand *N-[(3R)-4-ethoxy-3-hydroxy-4-oxobutanoyl]-L-isoleucyl-L-proline* (EP0) makes this structure a good target for analyzing the inhibitory interaction within the binding pocket of the protein. The 3D structure of human CTB (*1CSB*) (RCSB PDB 2018, OCT 30) was used for visualization of the active sites and variable residues within the binding site with VMD software (VMD APR,2008). In addition to the 3D structure of CTB, the 3D structures of human cathepsin D and C were used to compare the results obtained from multiple sequence alignment. The PDB codes *ILYB* and

3PDF which were obtained under the RCSB website (RCSB PDB 2018, OCT 30) used for visualizing the two adjacent histidines in human cathepsin D and C, respectively.

### 3.2.5. Binding Interaction Analysis

In this study, two probes (CID8795ATTO680 and Cbz-Lys-Lys-PABA-DCMF) were evaluated for binding interactions with CTB. For CID8795ATTO680-CTB visualization of the docked structure was done with the Molecular Operating Environment (MOE) software (Molecular Operating Environment MOE March,2020). This structure was kindly provided by Daniel Tesolin (Private Communication). To predict the binding interaction of Cbz-Lys-Lys-PABA-DCMF-CTB, the docked structure of Cbz-Arg-Arg-AMC-CTB (Daniel Tesolin, Private Communication) was used as a surrogate. The standard view orientation of the binding site and target residues for the description of the protease-substrate interaction is shown in **Figure 7** (Loser and Pietzsch 2015).



**Figure 7.** Annotation of the protease-substrate interaction, where **P** positions refer to the positions of the substrate residues and **S** positions refer to subsites where they bind to the surface of proteases.

### 3.3. Results

#### 3.3.1. *CTSB* Gene

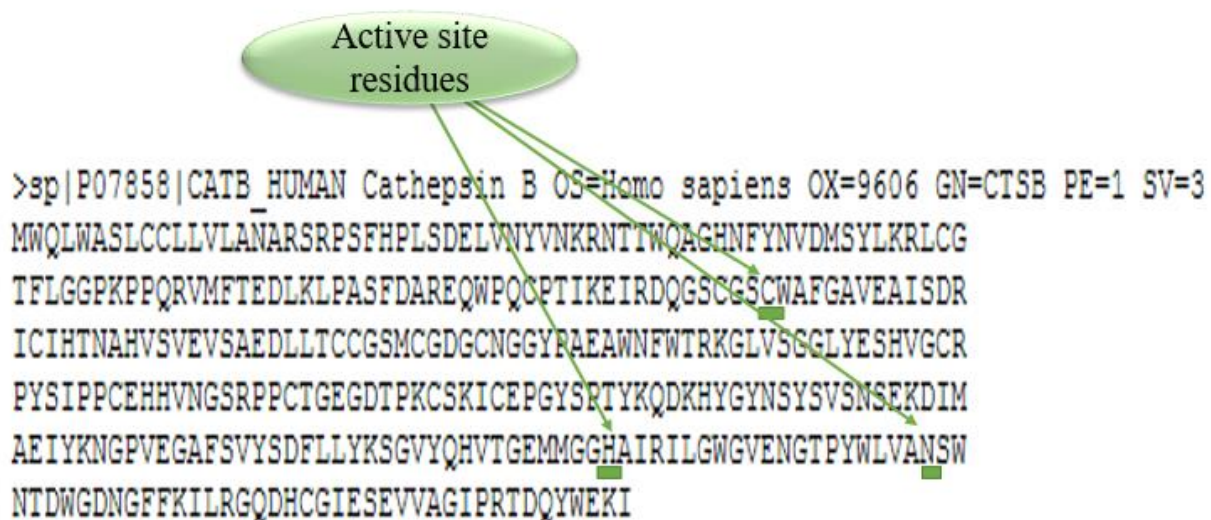
The *CTSB* (*Homo sapiens*) gene is a protein-coding sequence containing 15 exons with 9 of them are coding exons. The length of the *CTSB* gene is 25613 base pairs. The *CTSB* gene is located on the 8p22 chromosome in humans (Oliveira, Ferro et al.). The length of the transcript is 1017 bps which are translated to 339 amino acids (Atlas of Genetics and Cytogenetics in Oncology and Haematology 2008, Aug). The protein encoded by the *CTSB* gene belongs to the superfamily of papain-like cysteine proteases.

#### 3.3.2. CTB Protein

The protein encoded by the *CTSB* gene (*Homo sapiens*) contains one signaling peptide sequence located at positions 1-17 and one propeptide in the position of 18-79, both of which are cleaved during enzyme maturation and activation. CTB is comprised of a light chain containing 47 amino acids from sequence 80 to 126 and a heavy chain located at positions 129-333 (UniProtKB - P07858 (CATB\_HUMAN) 2018, May 25). CTB is synthesized in the rough endoplasmic reticulum (RER) as a pre-proenzyme (the inactive form of the enzyme which needs modification for activation) with 339 amino acids. It also contains 17 amino acids, known as a signal peptide, that directs the protein to the lumen of RER. Pro-cathepsin B is formed by removal of the signaling peptide and is transferred into the Golgi apparatus through the RER where it is glycosylated at two asparagine residues. The N-linked glycans are mannose-containing oligosaccharides with phosphorylated mannose residues that bind to the mannose-6-phosphate receptors in the trans-Golgi leading to the transportation of the enzyme to the lysosomes via transport vesicles. In the acidic environment of the lysosomes, pro-CTB undergoes autocatalytic activation leading to the formation of the active and mature CTB. The final proteolytic event

occurs through cleavage between residues 47 and 50 to yield two chains (Turk, Kos et al. 2004, Aggarwal and Sloane 2014). Finally, in the acidic environment of the lysosome, the active CTB is formed (Rowan, Mason et al. 1992). This enzyme has both endopeptidase and exopeptidase activity (Hasnain, HIRAMA et al. 1993). In terms of endopeptidase activity, CTB can cleave peptide bonds in the interior of a protein sequence. The substrate-binding site of the enzyme is located between the two lobes where cysteine (Cys108) on the left section of the lobe acts as a catalytic nucleophile and is responsible for peptide bond cleavage. Cys108 is activated by His278 – which is located on the right lobe (by convention).

Other active residues of the enzyme are Asn289 (Ruan, Hao et al. 2015), which provide the site where the substrates bind to the enzyme and the hydrolysis reaction takes place. Additionally, CTB can function as peptidyl dipeptidase which is responsible for dipeptides cleavage from the C-termini of the proteins. A unique feature of CTB is an occluding loop that allows this enzyme to hydrolyze dipeptides at the C-terminus acting as exopeptidase (Cygler, Sivaraman et al. 1996, Mort and Buttle 1997). The amino acid sequence in FASTA format of the CTB protein is shown in **Figure 8**.



**Figure 8.** Human cathepsin B amino acid sequence in FASTA format with the positions of three active site residues, Cys108, His278, and, Asn289, underlined in green. (NCBI 2018, Nov 13)

### 3.3.3. Similarity and Differences between Cathepsin B and Other Members of the Cathepsin Family

All general information about the human CTB and other members of the human cathepsin family was obtained from the UniProt website (UniProtKB 2018, Sep 09) (**Table 1**). This information reveals the active site homology of CTB with half of the cathepsin family. All the residue numbers in this section are based on human CTB (full sequence 339 amino acid).



**Table 1.** Selected information of 12 human proteins from the Cathepsin family obtained from the UniProt website (UniProtKB 2018, Sep 09).

Family member	Accession number	Gene	Protein	Length aa	Active sites	Function
CTB	P07858	<i>CTSB</i>	Cathepsin B	339	108(C), 278(H), 298(N)	Cysteine protease which is participates in intracellular degradation and turnover of proteins. Has also been implicated in tumor invasion and metastasis
CTL2	O60911	<i>CTSV</i>	Cathepsin L2	334	138(C), 277(H), 301(N)	Cysteine protease. May have an important role in corneal physiology
CTH	P09668	<i>CTSH</i>	Pro-cathepsin H	335	141(C), 281(H), 301(N)	Important for the overall degradation of proteins in lysosomes.
CTE	P14091	<i>CTSE</i>	Cathepsin E	401	96(D), 286(D)	May have a role in immune function. Probably involved in the processing of antigenic peptides during MHC class II-mediated antigen presentation. May play a role in activation-induced lymphocyte depletion in the thymus, and neuronal degeneration and glial cell activation in the brain.
CTG	P08311	<i>CTSG</i>	Cathepsin G	255	64(H), 108(D), 201(S)	Serine protease with trypsin- and chymotrypsin-like specificity. Cleaves complement C3. Has antibacterial activity against the Gram-negative Bacterium <i>P. aeruginosa</i> , antibacterial activity is inhibited by LPS from <i>P. aeruginosa</i> .
CTC	P53634	<i>CTSC</i>	Dipeptidyl peptidase 1	463	258(C), 405(H), 427(N). Binding site 302(F), 304(Y), 347(Y)	Cysteine protease. Has dipeptidyl peptidase activity. Active against a broad range of dipeptide substrates composed of both polar and hydrophobic amino acids. Proline cannot occupy the P1 position, and arginine cannot occupy the P2 position of the substrate. It can act as both an exopeptidase and an endopeptidase. Activates serine proteases such as elastase, cathepsin G and granzymes A and B. Can also activate neuraminidase and factor XIII.

Family member	Accession number	Gene	Protein	Length aa	Active sites	Function
CTS	P25774	<i>CTSS</i>	Cathepsin S	331	139(C), 278(H), 298(N)	Cysteine protease. Key protease responsible for the removal of the invariant chain from MHC class II molecules. The bond-specificity of this proteinase is in part similar to the specifics of cathepsin L and cathepsin N.
CTK	P43235	<i>CTSK</i>	Cathepsin K	329	139(C), 276(H), 296(N)	Closely involved in osteoclastic bone resorption and may participate partially in the disorder of bone remodeling. Displays potent end protease activity against fibrinogen at acid pH. May play an important role in extracellular matrix degradation.
CTL1	P07711	<i>CTSL</i>	Cathepsin L1	333	138(C), 276 (H), 300(N)	Essential for the overall degradation of proteins in lysosomes.
CTL3	Q5NE16	<i>CTSL3P</i>	Putative inactive cathepsin L-like protein CTSL3P	218		It has a collagenolytic activity and participates in Collagen degradation.
CTD	P07339	<i>CTSD</i>	Cathepsin D	421	97(D), 295(D)	Aspartic acid protease and is active in intracellular protein breakdown. Involved in the pathogenesis of several diseases such as breast cancer and possibly Alzheimer's disease.

To find a selective probe for human cathepsin B, it is necessary to identify unique residues within a binding pocket of this protein compared to other members of the cathepsin family. Therefore, PDB structure (*IHUC*) was added to the multiple sequence alignment to limit the variable residues to those that are within mature CTB. The active site homology of human CTB and *IHUC* (Cys108; His278, and Asn298) with seven members (cathepsin Z (CTZ); cathepsin L2 (CTL2); cathepsin H (CTH); cathepsin C (CTC); cathepsin S (CTS); cathepsin K (CTK), and cathepsin L1 (CTL1)) of cathepsin family which was obtained from multiple sequence alignment

with CLUSTALX-2.1 software (Clustal: Multiple Sequence Alignment AUG,2012), is shown in **Figure 9**. Given the active site similarity between human CTB and other members of the cysteine cathepsin family, designing a probe specific to CTB is difficult.



**Figure 9.** Section of multiple sequence alignment of human CTB and *1HUC* with other members of the cathepsins family (same organism), obtained from CLUSTALX-2.1 software (Clustal: Multiple Sequence Alignment AUG,2012). The red boxes show the homology of three active site residues of CTB (Cys (108), His (278), and Asn (298)) with seven members of the cathepsin family (CTZ, CTL2, CTH, CTC, CTS, CTK, and CTL1). The horizontal red box shows the positions of the human CTB and *1HUC* structure in the alignment.

To reveal fully, strongly, weakly conserved residues and, the variable amino acid of the human CTB relative to other members of the cathepsin family, the multiple sequence alignment obtained with CLUSTALX-2.1 software (Clustal: Multiple Sequence Alignment AUG,2012) was analyzed further. According to multiple sequence alignment results obtained with CLUSTALX-2.1 software (Clustal: Multiple Sequence Alignment AUG,2012) (**Table 2**), there were two fully

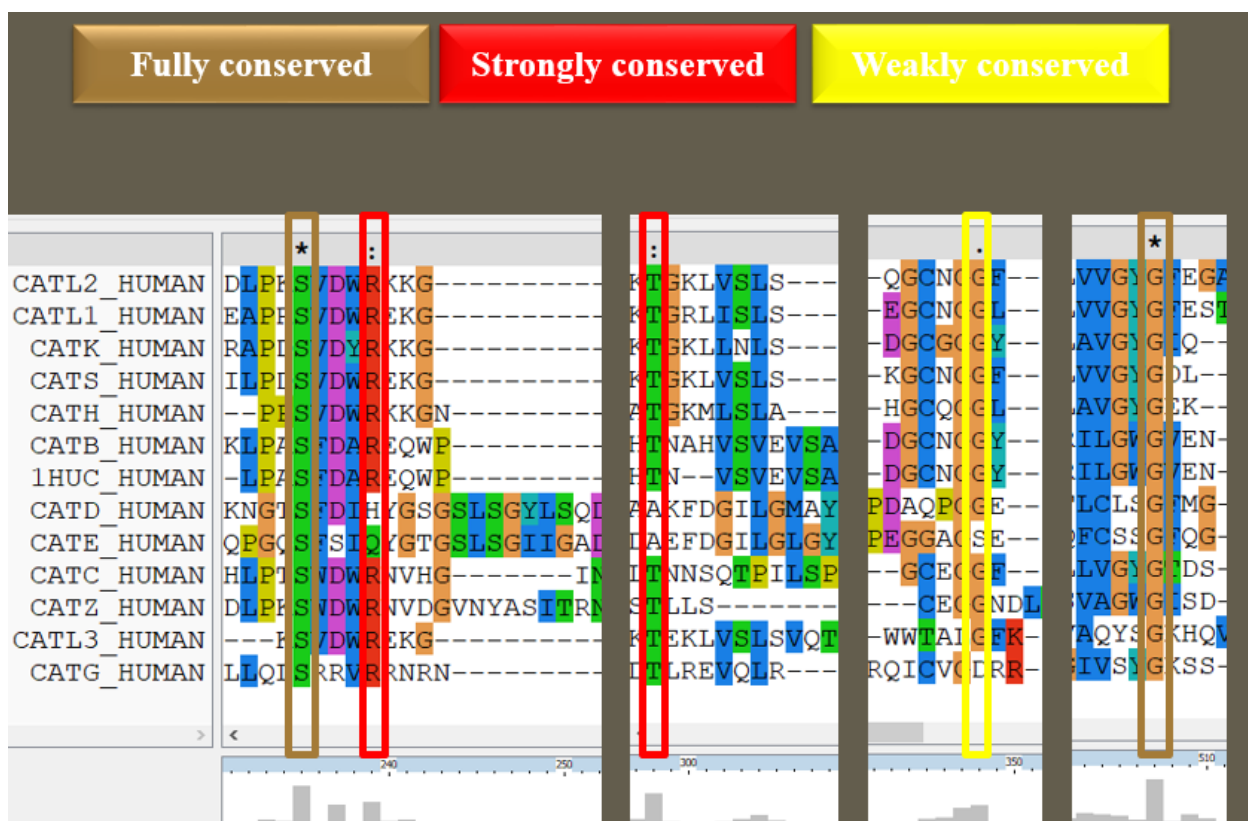
conserved residues (serine (83) and glycine(286), two strongly conserved residues (arginine (87) and threonine (125), and one weakly conserved residue (glycine (153) in human CTB compare to other members of cathepsin family. The variable residues which were obtained here would have led us to identify the unique residues in human CTB compare to others member of the cathepsin family. The position of these fully, strongly, and weakly conserved residues are shown in (**Figure 10**). As shown in **Table 3**, CTB shows the highest percentage of identity scores with Cathepsin C and the lowest percentage of identity scores with Cathepsin G. The identity percentage was calculated between each member relative to human CTB.

**Table 2.** Information obtained from multiple sequence alignment (CLUSTALX-2.1) of the amino acids of 12 members of the cathepsins family as well as *IHUC* structure (Clustal: Multiple Sequence Alignment AUG,2012). The results show the fully, strongly (amino acid alter to another amino acid with similar biochemical characteristics) and weakly conserved (amino acid alter to another amino acid with different biochemical characteristics) variable residues of CTB and *IHUC* compared to the rest of the cathepsin family.

<b>Variable position</b>	80-82, 84-86, 88-124, 126-152, 154-285, 287-333.
<b>fully conserved</b>	83 (S), 286 (G)
<b>strongly conserved</b>	87 (R), 125 (T).
<b>weakly conserved</b>	153(G).

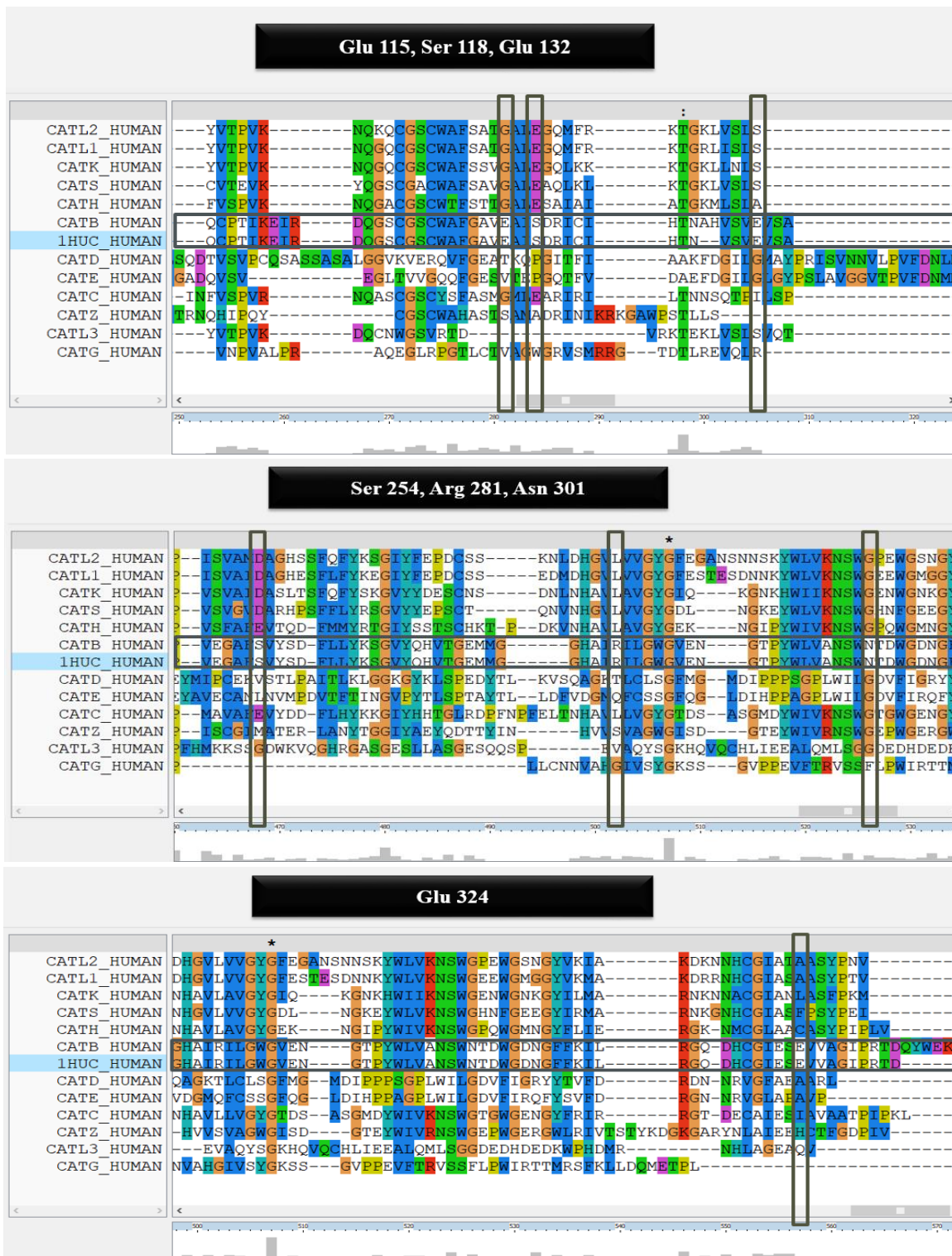
**Table 3.** Information obtained from multiple sequence alignment (CLUSTALX-2.1) of the amino acids of the cathepsins family (Clustal: Multiple Sequence Alignment AUG,2012)., which used to identify which cathepsin is the most closely related to CTB.

Cathepsin family members	Identity% scores
(CATB: CATZ)	17
(CATB: CATL2)	17
(CATB: CATH)	23
(CATB: CATE)	6
(CATB: CATG)	3
(CATB: CATC)	25
(CATB: CATS)	24
(CATB: CATK)	19
(CATB: CATL1)	20
(CATB: CATL3)	6
(CATB: CATD)	5



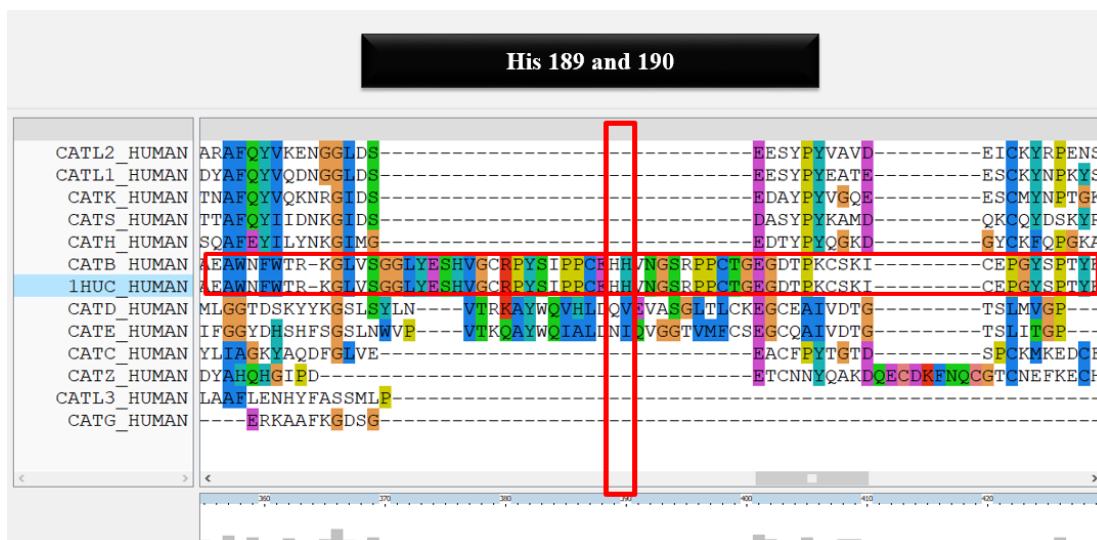
**Figure 10.** View of the parts of aligned multiple sequences of 12 members of the cathepsin family and *IHUC* structure by CLUSTALX-2.1 software (Clustal: Multiple Sequence Alignment AUG,2012), which indicates the fully, strongly and weakly conserved residues. The brown and red and yellow boxes are marked the fully, strongly and weakly conserved residues, respectively.

From variable residues which are shown in (Table 2), those were selected which have different side chains that result in chemical properties diversity compared to the rest of the cathepsin family. According to the results which are shown in, Figure 11, we found seven unique residues (Glu115, Ser118, Glu132, Ser254, Arg281, Asn301, and Glu324) in human CTB compared to other members of cathepsin family. For example, glutamic acid at position 115 of the human cathepsin B is an acidic amino acid which can form an ionic bond, while most of the family mostly contains hydrophobic amino acid in that position.



**Figure 11.** Multiple sequence alignment of 12 members of the cathepsin family and *IHUC* structure obtained using CLUSTALX-2.1 software (Clustal: Multiple Sequence Alignment AUG,2012). The vertical black boxes mark the unique residues in human CTB relative to other members of the cathepsin family, in terms of different biochemical characteristics. The horizontal black boxes show the positions of the human CTB and *IHUC* structure in the alignment.

In addition to those variable residues mentioned above, we also checked the residues in the occluding loop of cathepsin B which is a unique feature of this protein. The presence of two adjacent histidines (basic amino acid) at position 189 and 190 which are located near the catalytic binding site of the CTB, could be a good target for probe binding. According to multiple sequence alignment with CLUSTALX-2.1, the rest of the family did not have any basic amino acids in those locations (**Figure 12**). Additionally, none of the cathepsin members in their entire amino acid sequence have two adjacent histidines except cathepsins H, S, C, and D. Cathepsin H has two adjacent histidine residues at positions 61 and 62 and cathepsin S at 26 and 27; however, both residues are located within the propeptide segment which is cleaved during enzyme maturation. Cathepsin C has two pairs adjacent histidine residues (366-367 and 389-390), while cathepsin D has two at positions 120 and 121, both of which are present in the mature enzymes. Further investigation is required to estimate the distance of these histidines from active sites of these proteins in a 3D structure.



**Figure 12.** Multiple sequence alignment of 12 members of the cathepsin family and *IHUC* structure obtained with CLUSTALX-2.1 software (Clustal: Multiple Sequence Alignment AUG,2012). The location of two adjacent histidines at positions 189 and 190 in CTB compared to other members of the family is marked by a vertical red box. The horizontal red box shows the positions of the human CTB and *IHUC* structure in the alignment.

Variable residues that are unique in human CTB can be targeted as anchor amino acids for binding of designed probes if they are within a close distance (0.5 to 7 Å) (Angles, Arenas-Salinas et al. 2020) of the binding pocket of CTB; therefore, the next step would be mapping these residues on the 3D structure of CTB to assess their position within the binding pocket of CTB.

### 3.3.4. Mapping Variable Residues onto the 3D Structure of CTB

General information about five experimentally determined structures of CTB, as well as two experimentally determined structures of CTC and CTD, is shown in **Table 4**, which were obtained from the RCSB website (RCSB PDB 2018, OCT 30).

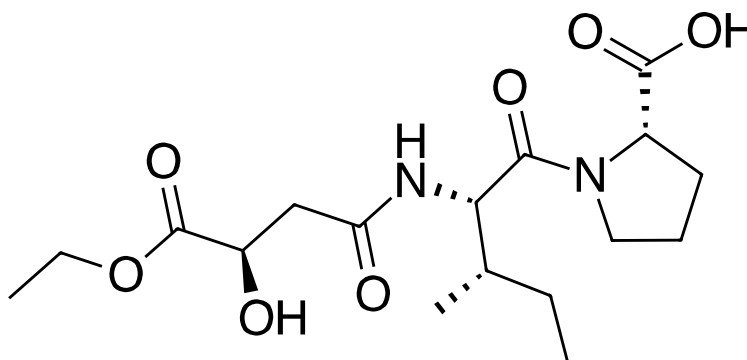


**Table 4.** Information obtained from the RCSB PDB website (RCSB PDB 2018, OCT 30) about five of the experimentally determined structures for human CTB, as well as two experimentally determined structures of human CTC and CTD.

<b>PDB code</b>	<b>Description</b>	<b>Protein</b>	<b>Experimental method</b>	<b>Resolution [Å]</b>	<b>Chain name and position</b>	<b>Ligand ID</b>
<i>1PBH</i>	Crystal structures of human pro-cathepsin B at 3.2 and 3.3 Å resolution reveals the propeptide folds on the surface, shielding the active site of the enzyme from exposure to solvent. (Turk, Podobnik et al. 1996)	CTB	X-RAY DIFFRACTION	3.20	A 18-333	None
<i>3PBH</i>	The crystal structure of the wild-type human pro-cathepsin B at 2.5 Å resolution reveals the native active site of a papain-like cysteine protease zymogen. (Podobnik, Kuhelj et al. 1997)	CTB	X-RAY DIFFRACTION	2.50	A 18-333	None
<i>1CSB</i>	Crystal structure of cathepsin B inhibited with CA030 at 2.0- Å resolution: This structure provides the basis for the design of specific epoxysuccinyl inhibitors. (Turk, Podobnik et al. 1995)	CTB	X-RAY DIFFRACTION	2.00	A/D 80-126 B/E 129-333	EPO
<i>3CBJ</i>	Displacement of the occluding loop by the parasite protein “chagasin” (act as a cysteine exopeptidase inhibitor) results from inefficient inhibition of human cathepsin B. Chagasin structure in complex with cathepsin B illustrating that the occluding loop is displaced to enable chagasin bind with its three loops and span the entire active site cleft. (Redzynia, Ljunggren et al. 2008)	CTB	X-RAY DIFFRACTION	1.80	A 74-339	PO4
<i>1HUC</i>	The refined structure of human liver cathepsin B shows a structural basis. (Musil, Zucic et al. 1991)	CTB	X-RAY	2.10	A/C 80-126 B/D 129-333	None

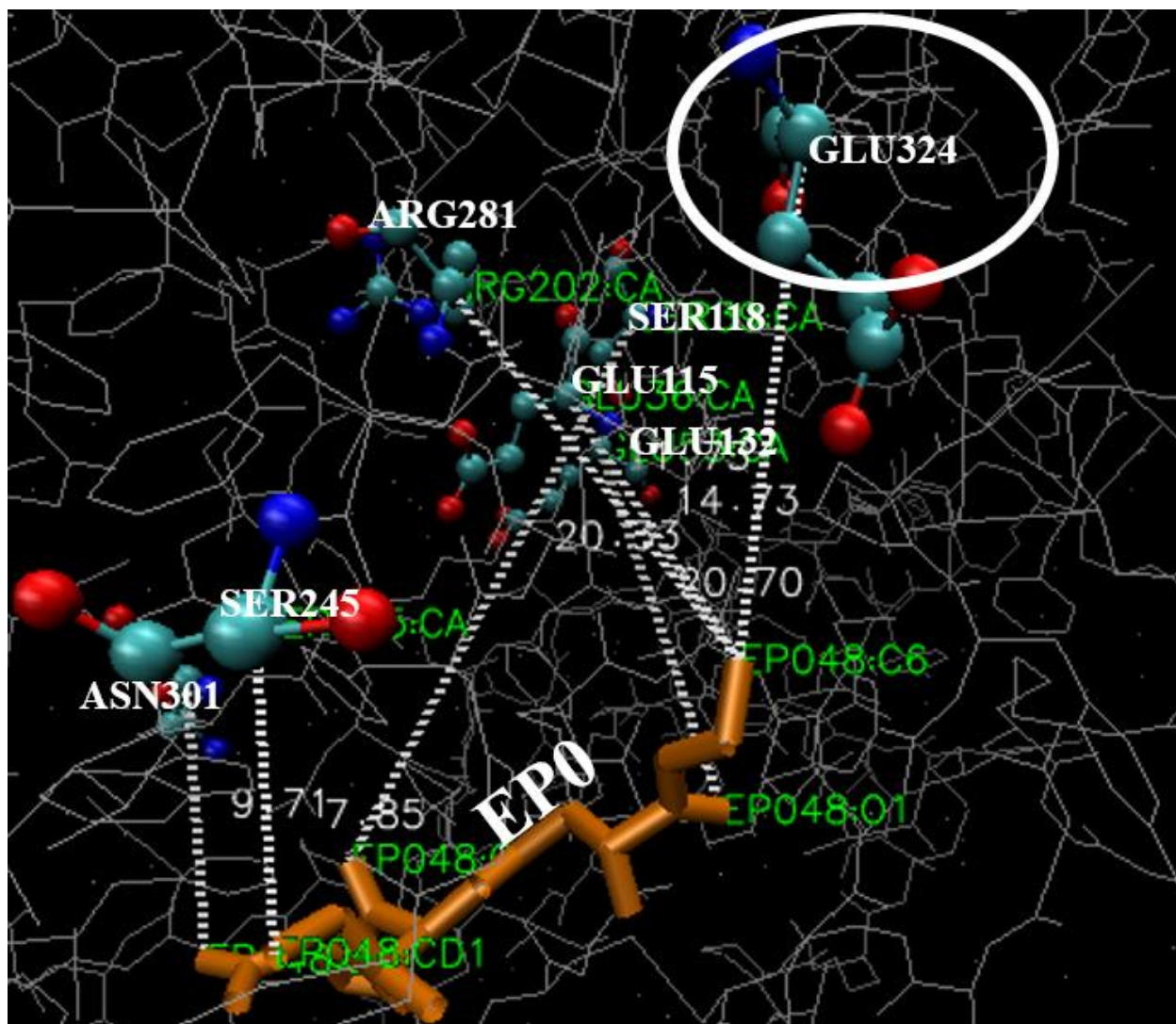
PDB code	Description	Protein	Experimental method	Resolution [Å]	Chain name and position	Ligand ID
3PDF	Crystal structure of human cathepsin C inhibited with Cyanamide-Based Inhibitors (Laine, Palovich et al. 2011).	CTC	X-RAY DIFFRACTION	1.85	A 441	LXV and NAG
1LYB	Inhibited and native crystal structure of human cathepsin D (Baldwin, Bhat et al. 1993).	CTD	X-RAY DIFFRACTION	2.50	A/C 97 B/D 241	NAG, BMA, and MAN

In *ICSB* structure, the ligand (EP0) with the molecular weight of 372.41 g/mol and the molecular formula (C<sub>17</sub>H<sub>28</sub>N<sub>2</sub>O<sub>7</sub>) (**Figure 13**) was visualized using VMD software (VMD APR,2008) to study the distance between the active site and unique residues of the human CTB and ligand.



**Figure 13.** The chemical structure of ligand (EP0).

To map the seven unique residues on the 3D structure of CTB and to find their distance from the ligand (EP0) within the binding pocket of this protein, the molecular graphics program VMD (VMD APR,2008) was used to visualize the *ICSB* PDB file (**Figure 14**). Distances were measured between the *alpha* carbon (*C $\alpha$* ) of all residues and atom C6, O1, and CD1 of the bound ligand (EP0).

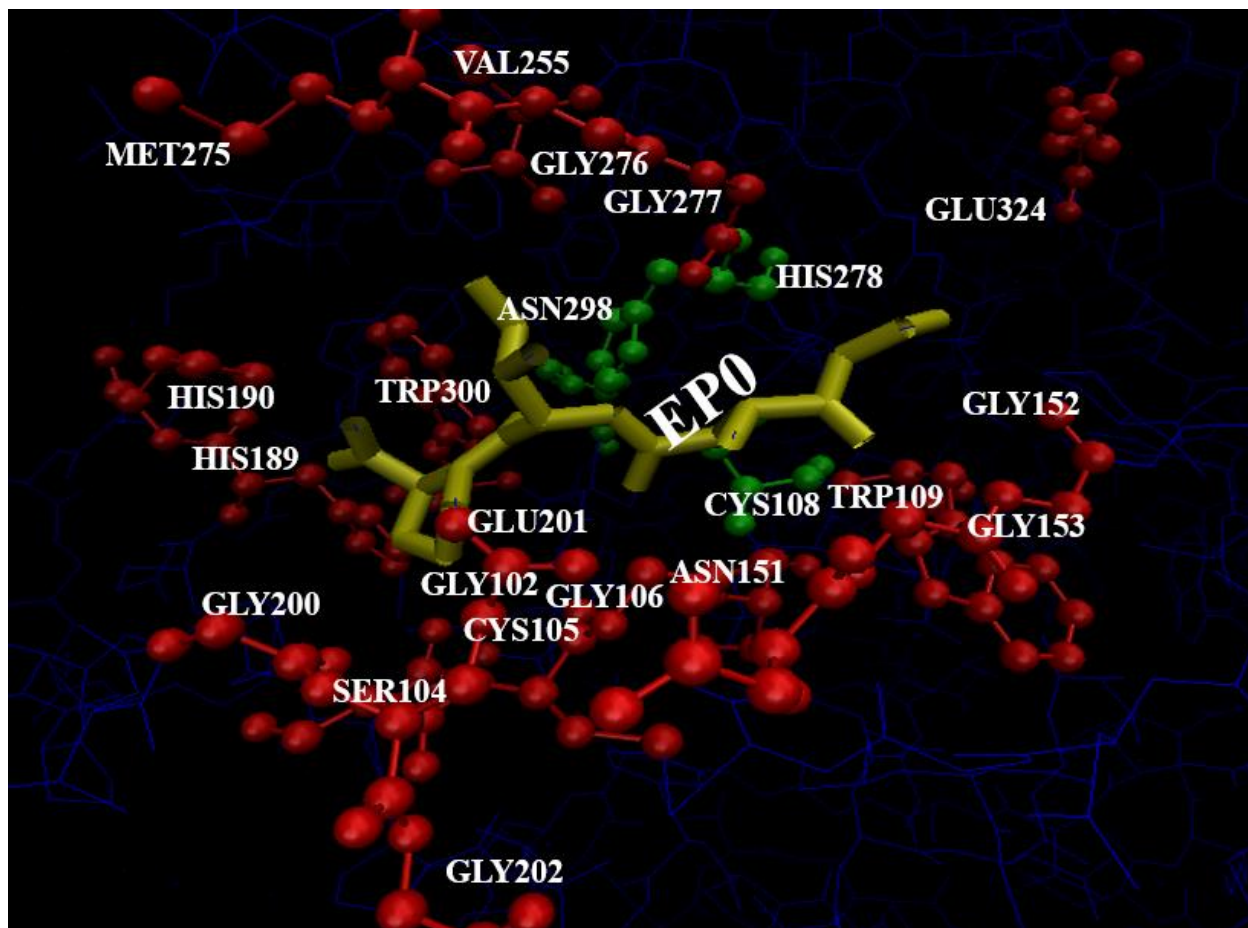


**Figure 14.** The 3D structure of CTB (PDB code *1CSB*) which is visualized by VMD software (VMD APR,2008), the red, light and dark blue balls represent six unique residues that were identified by multiple sequence alignment (Glu115, Ser118, Glu132, Ser254, Arg281, and, Asn301). All these residues are located above 11 Å distance of ligand (EP0 (orange stick)). The position of unique residue Glu324, which is shown in a white circle, is located within close distance from the ligand (5.97 Å).

According to **Figure 14**, all the unique residues for CTB which was identified for CTB through multiple sequence alignment with CLUSTALX-2 software are located in the far distance of the binding pocket of CTB (above 11 Å), except Glu324 which is located within close distance of the ligand EP0 (< 6 Å). Targeting Glu324, which is located within the close distance of the binding pocket of CTB as an anchor may rise the selectivity of ligands toward CTB over other

members of the cathepsin family, as the most of the cathepsin members have hydrophobic residue in that position.

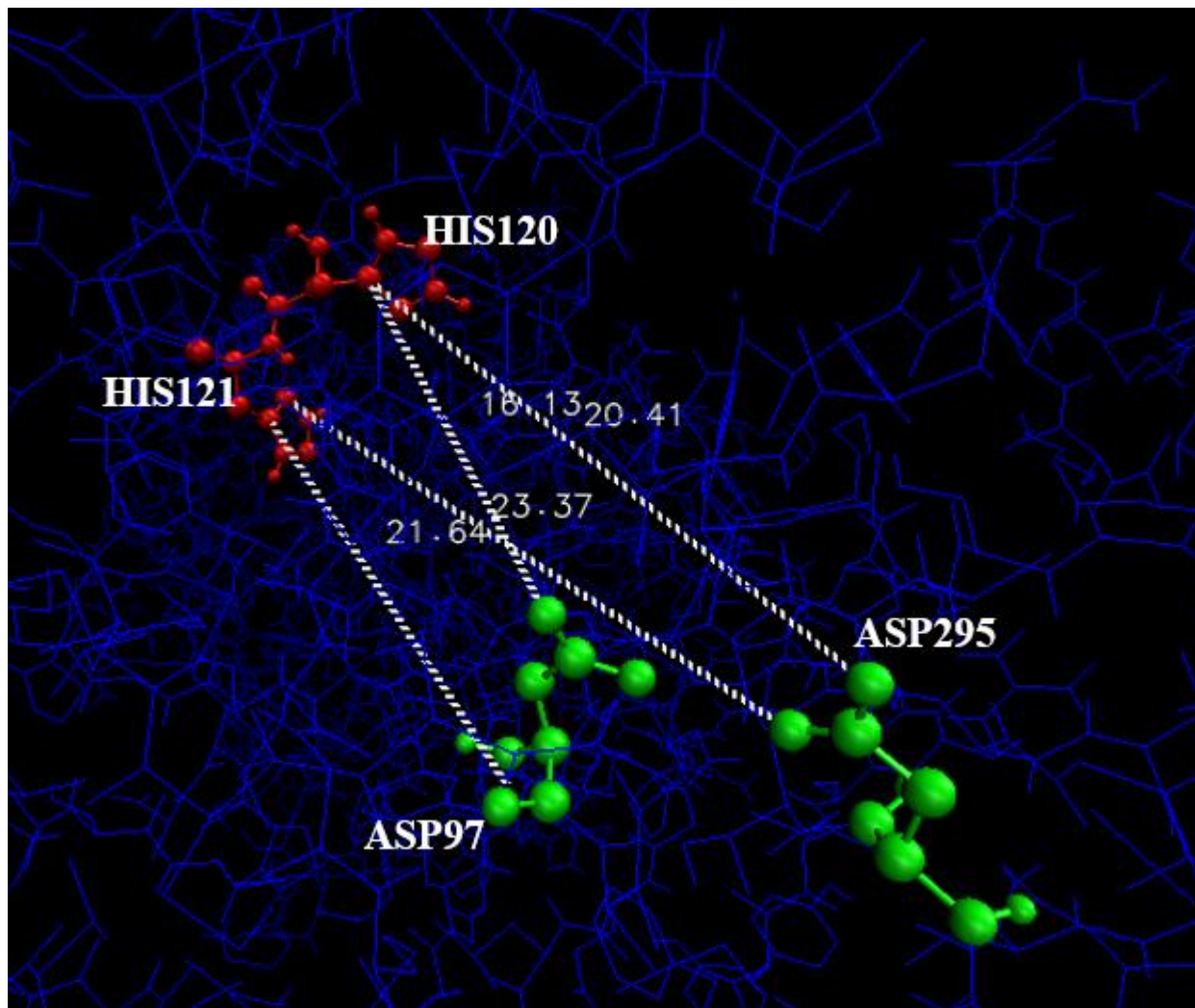
All the residues within 6 Å of the ligand were mapped to identify their names and positions. The identification of these residues would be useful since they could be targeted as an anchor for probes. The names and positions of all residues within 6 Å of the ligand (EP0) in the *ICSB* structure was obtained from the VMD program (VMD APR,2008) are as follows: Gly102, Ser104, Cys105, Gly106, Cys108, Trp109, Asn151, Gly152, Gly153, His189, His190, Gly200, Glu201, Gly202, Val255, Met275, Gly276, Gly277, His278, Asn298, Trp300, and the unique residue: Glu324 (**Figure 15**).



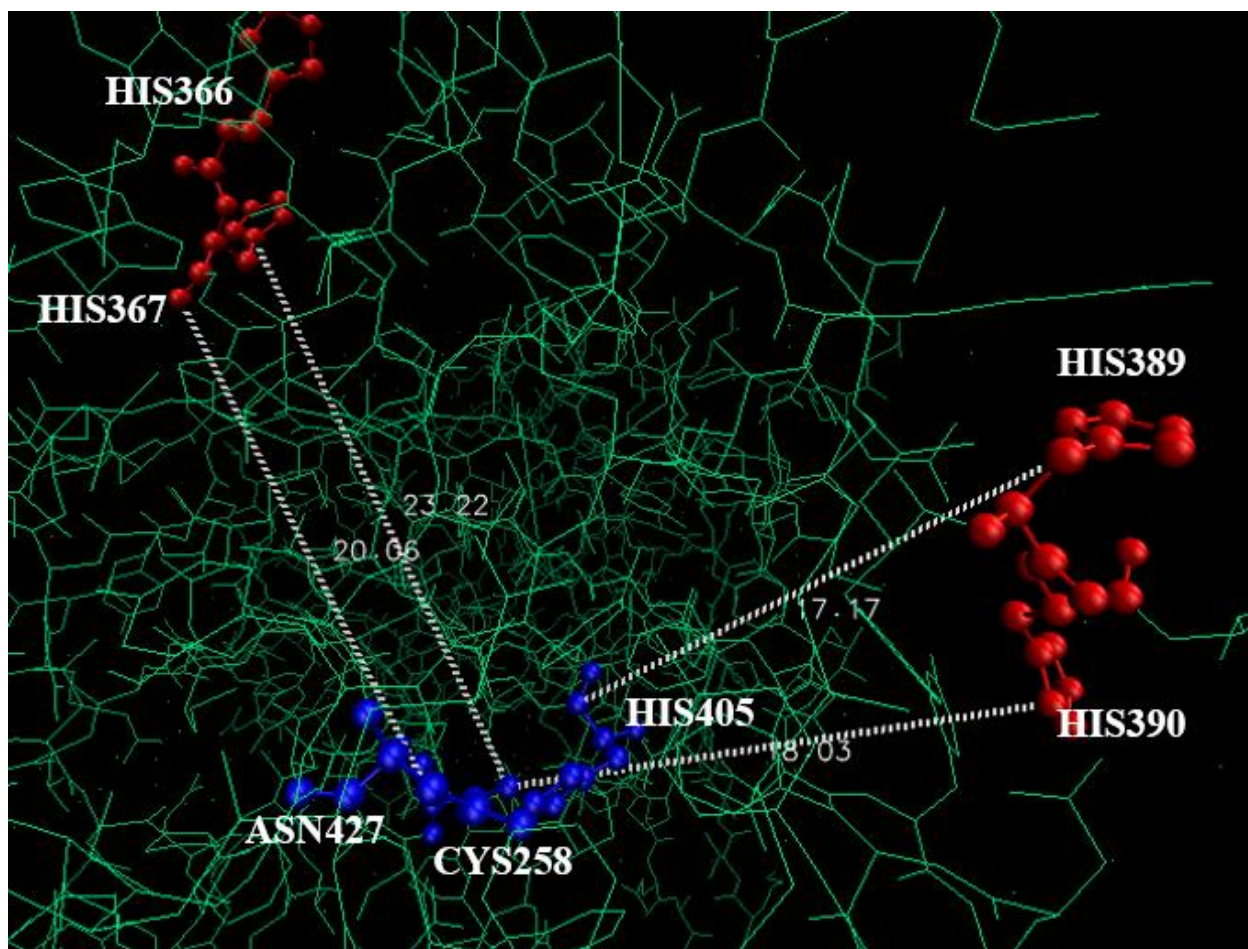
**Figure 15.** The 3D structure of CTB (PDB code *1CSB*) with bound ligand (EPO (yellow stick) after loading *1CSB* under the VMD main window (VMD APR,2008). The red balls represent the residues located within 6 Å of the ligand. The list of these residues is as follows: Gly102, Ser104, Cys105, Gly106, Trp109, Asn151, Gly152, Gly153, His189, His190, Gly200, Glu201, Gly202, Val255, Met275, Gly276, Gly277, Trp300, and the unique residue: Glu324. The active site residues of human CTB marked as green balls: Cys108, His278, and Asn298.

According to **Figure 15**, the two adjacent histidines (189 and 190) are within close distance from the active site of human CTB (less than 6 Å); however, the location of the two adjacent histidines (120 and 121) in human CTD (PDB code *1LYB*) is more than 16 Å from the active site of this protein (**Figure 16**). Additionally, mapping two pairs of adjacent histidine (366-367 and

389-390) on the 3D structure of human CTC (PDB code *3PDF*) reveals their distance from the active site of this protein which is  $> 17 \text{ \AA}$  (**Figure 17**). Distances were measured between the  $C\alpha$  of all residues.



**Figure 16.** The 3D structure of cathepsin D (PDB code *1LYB*) which is visualized by VMD software (VMD APR,2008), with the position of two active site residues (Asp97 and Asp295 (green balls)) and their distance from two adjacent histidines at positions 120 and 121 (red balls).



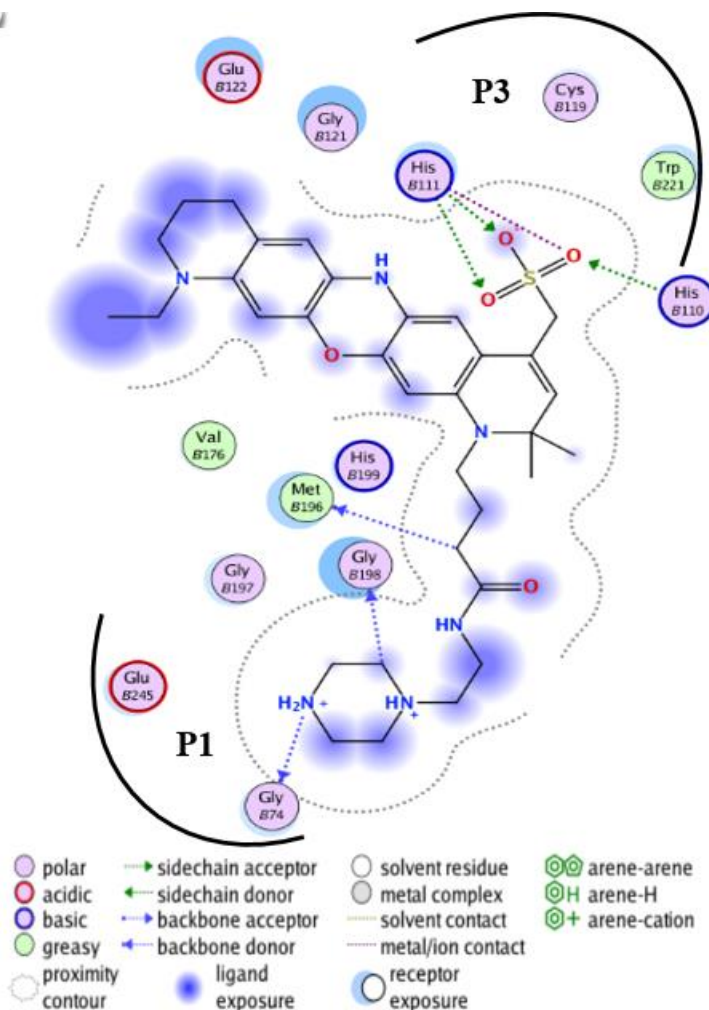
**Figure 17.** The 3D structure of cathepsin C (PDB code *3PDF*) which is visualized by VMD software (VMD APR,2008), with the position of three active site residues (Cys258, His 405, and Asn427 (blue balls)) and their distance from two pairs of adjacent histidine at positions 366-367 and 389-390 (red balls).

### 3.3.5. Computationally Identified Binding Interactions

The software MOE was used to predict the binding interaction of the target probe CID8795ATTO680 with human CTB (PDB code *1HUC*). All the residues numbers represented in this section are based on mature human CTB. As shown in **Figure 18**, the negatively charged sulfonate group of CID8795ATTO680 at the position **P3** ionically interacts with two histidine residues (110 and 111) in the occluding loop of CTB (within the binding pocket of cathepsin B)

though its ATTO680 side in an acidic environment. This probe can participate in hydrogen bond through its ligand side with two glycines at positions 74 and 198 at the position **P1**. As shown in **Figure 18**, we do not expect any inhibitory activity of this probe against CTB due to not targeting any active site residues of this protein, although probe may interfere with substrate binding since it targets almost the same residues as Cbz-Arg-Arg-AMC (**Figure 19**) at both positions **P1** and **P3**; therefore, this probe was subjected for inhibition assay with two commercial substrates (**Chapter 6**). From the amino acid sequence alignment results, in addition to CTB, cathepsins C and D also have two adjacent histidine residues in their mature protein. However, mapping these residues on the 3D structure of human CTD and CTC reveals they located in the far distance from the active site of these proteins (**Figure 16** and **Figure 17**); therefore, CID8795ATTO680 would not likely bind to these proteases with high affinity.

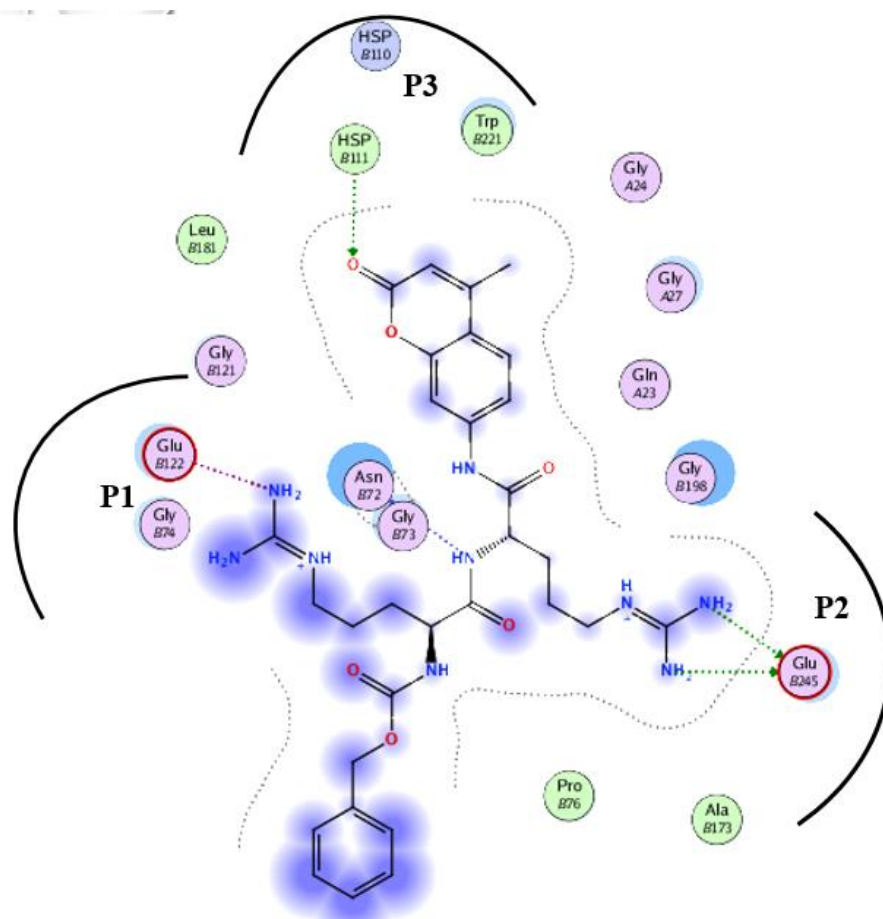




**Figure 18.** Predicted binding interaction between CID8795ATTO680 and CTB (PDB code *1HUC*). Diagram created with MOE (Molecular Operating Environment MOE March,2020). using the computationally generated structure of CID8795ATTO680 bound to CTB. **P** positions refer to positions in the substrate responsible for CTB recognition.

To determine binding interaction between Cbz-Lys-Lys-PABA-DCMF and CTB, the docked structure (Daniel Tesolin, Private Communication) of the Cbz-Arg-Arg-AMC-CTB used as a surrogate. As shown in **Figure 19**, this substrate may bind to CTB at three main positions within **P1**, **P2**, and **P3** pockets. At **P1** position most cathepsins will accept a positively charged amino acid side chain such as lysine or arginine (Chowdhury, Moya et al. 2014), due to the presence of glutamic acid (122) within **S1** pocket. However, at the **P2** position, only CTB will accept a positively charged group, because of the presence of glutamic acid 245 within the **S2**

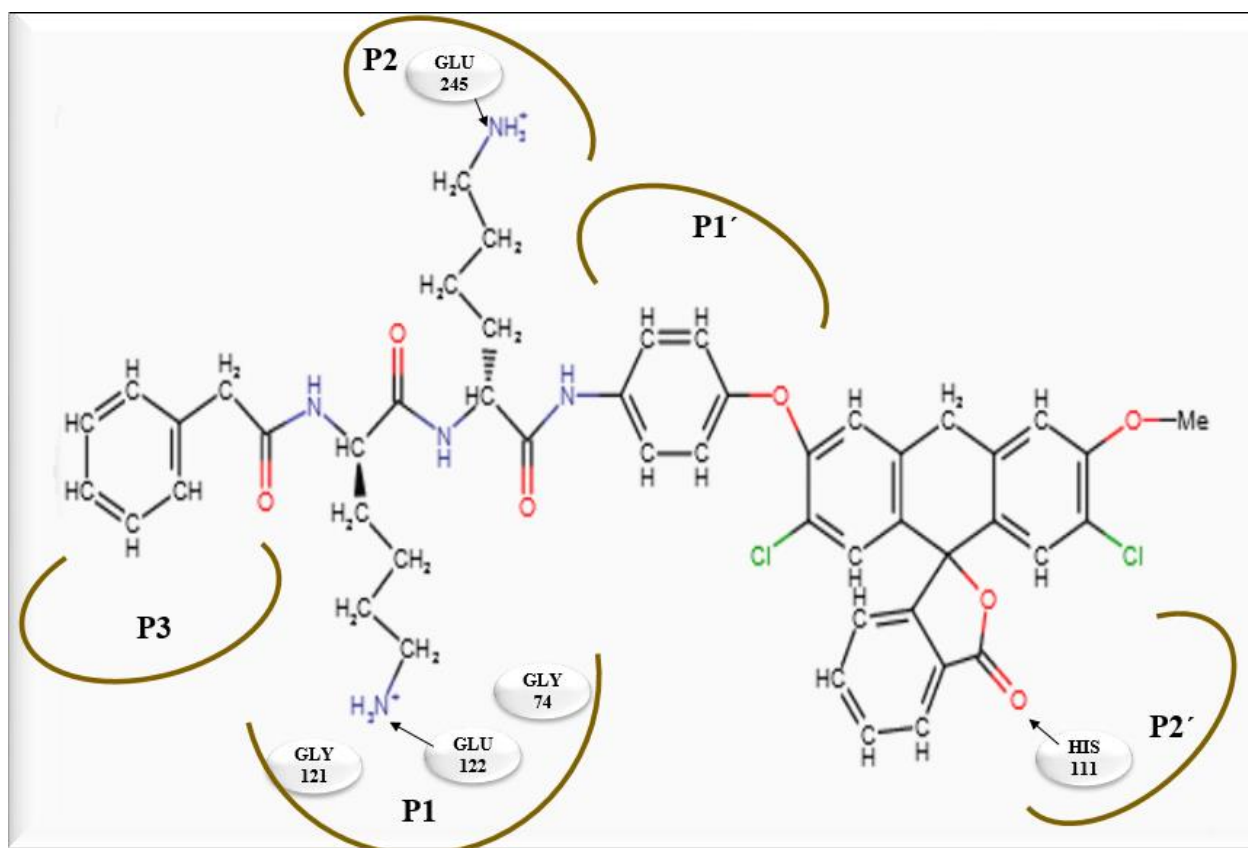
pocket. As mentioned in section 2.3.3 glutamic acid 245 (324 based on the CTB protein sequence) is unique to CTB as other members of the cathepsin family prefer amino acids with a hydrophobic side chain at this position. At **P3** position, the ester group of the AMC dye seems to interact with histidine (111) through hydrogen-bonding with ion-dipole interactions.



**Figure 19.** Predicted binding interaction between Cbz-Arg-Arg-AMC and CTB visualized with MOE (Molecular Operating Environment MOE March,2020). The structure of Cbz-Arg-Arg-AMC docked to CTB (PDB code *1HUC*). **P** positions refer to positions in the substrate responsible for CTB recognition.

From the structure of Cbz-Arg-Arg-AMC docked to CTB, we were able to identify the possible binding interactions of our substrate-based probe. As shown in **Figure 20**, CTB prefers

positively charged residues at **P1** and **P2**; however, other cathepsins prefer a large hydrophobic or aromatic amino acids at the **P2** position. The presence of the PABA linker extends the substrate to potentially introduce interactions between the aromatic group at **P3** with the **S3** site and DCMF at **P2'** with the **S2'** respectively.



**Figure 20.** The predicted binding interaction between Cbz-Lys-Lys-PABA-DCMF and CTB. **P** positions refer to positions in the substrate responsible for CTB recognition.

### 3.4. Summary and Conclusions

To develop a novel probe selective to cathepsin B, we studied this enzyme from nucleotide sequence to protein structure. We observed that cathepsin B shares common active site residues with more than half of the cysteine cathepsin family. The residues Glu115, Ser118, Glu132,

Ser254, Arg281, Asn301, and Glu324 in cathepsin B were found to be unique compared to other members of this family due to the presence of different side chain in their structures which lead to diversity in their chemical properties. Among all these unique residues only Glu324 are located within close distance of the catalytic site of human CTB (less than 6 Å). Glutamic acid is an amino acid that can form ionic bonds through the carboxylated group in its side chain with a positively charged species in acidic pH (~5.5). This residue can also act as a hydrogen bond acceptor and donor in its protonated form. From the analysis of docked structures and potential binding interaction, we expect this unique residue to increase the selectivity of Cbz-Lys-Lys-PABA-DCMF toward CTB since most of the cathepsin family member prefer hydrophobic residue in that position.

Two unique adjacent histidine residues which are located within 3.5 Å of the CID8795ATTO680, and targeted by ATTO680 of the probe, predict the probe selectivity toward CTB. Even though both cathepsin D and C have two adjacent histidine residues in their mature enzymes, the location of these histidine residues is more than ~16 Å of the active sites of both enzymes and is unlikely the NIR probe binds to these proteases with high affinity.

## Chapter 4: Genetic Variation in Cathepsin B Gene (*Homo sapiens*)

### 4.1. Introduction

The 1,000 Genomes project (The International Genome Sample Resource 2008) was launched in 2008 to create an accurate catalog of human genetic variation. Galaxy (Galaxy 2019) is a web-based platform and an open-source tool for computational biological research and analyzing genomic data with a simplified interface for bioinformatics tools (Cock, Gruning et al. 2013). In this project, the Lakehead Galaxy (Galaxy 2019) server has been used to access all the tools for analyzing, filtering, and converting the genomic data of the 1,000 Genomes database.

Here, the first objective of using the 1,000 Genomes database is to identify SNPs, which are located on the *CTSB* gene (*Homo sapiens*), and identify which of these SNPs are missense variants located on or near the active site of the protein. Missense variants refer to the genetic alteration of the nucleotide sequence that alters the genetic code and leads to the incorporation of a different amino acid in the protein. These missense mutations may affect the binding of the probe to CTB in the individuals that carry them. The second propose of this project is to discover the common genetic diversity with frequencies of 50% or less in 1,000 Genomes populations. The final objective was identifying any mutations in this gene that may be associated with cancer. For investigating single nucleotide polymorphisms (SNPs) in the human gene encoding CTB, phase 3 data from the 1,000 Genomes project have been used. This assembly used GRCh37/hg19 as a genome reference. However, more recently, GRCh38/hg38 has been used as a human reference genome (Devuyst 2015). This project allows access to the genome sequences of a large number of people, 2,504 individuals from 26 populations. Five populations were used in genetic analysis for

*CTSB* gene are as follows: African (AFR); American (AMR); East Asian (EAS); European (EUR), and South Asian (SAS) with 1018, 535, 617, 669 and 661 individuals respectively.

## **4.2. Procedures and Materials**

### **4.2.1. Variant Identification in the 1,000 Genomes Data**

The aim of using the 1,000 Genomes project (The International Genome Sample Resource 2008) is to investigate single nucleotide polymorphisms in a human *CTSB* gene in different populations; therefore, we need to inspect the frequency of each allele in the population. Allele frequency refers to the measure of the relative frequency of an allele in the population. In general, it is defined as a percentage or proportion, and it replicates the genetic diversity of a population. An allele frequency can be calculated as follows: the number of times that the target allele is recorded in a population divided by the total number of copies of the gene at that specific position in the population. To obtain allele frequency, the first step is to obtain genomic information as a Variant Call Format (VCF) and upload it to Galaxy (Galaxy 2019).

Variant Call Format is the specified model of a text file (encoded format of single nucleotide variants) in bioinformatics to store marker and genotype data. Generally, VCF files contain three parts: meta-information lines; one header line, and data lines contain marker and genotype data (Danecek, Auton et al. 2011). In VCF files, the users can select the specific genomic positions depends on their needs. The genomic position code refers to the chromosome, the position number of a first and last nucleotide in the sequence. In this study, 8.11697472-11728207 indicates the location of the *CTSB* gene within the reference human genome GRCh37/hg19. The SNPs location is identified through its position within the reference genome.

Genomic data from the 1,000 Genomes project corresponding to the *CTSB* gene (*Homo sapiens*) was downloaded from the NCBI website as a VCF file with no filter (NCBI 2013, May). Afterward, the VCF file was uploaded to Lakehead's Galaxy Server (Galaxy 2019). This VCF file contains all the allele frequencies of the *CTSB* gene. The next step was to filter SNPs to those located in the coding region of the *CTSB* gene and calculate genetic variation parameters. In general, rare mutations refer to variants that present in less than 1% of the population, whereas common mutations occur in more than 5% of the population (Gibson 2012). In Galaxy by using "*VCFtools Filter*" (Danecek, Auton et al. 2011) the SNPs were filtered in two following steps: i) minimum Minor Allele frequency (MAF) 0.001 and maximum MAF 0.01 for detecting rare polymorphisms which are presented in 1% or less of the population; and, ii) minimum MAF 0.01 and maximum MAF 0.5 for studying common variants which are presented between 50% and 1% of the population. The "*VCFtools Filter*" is useful for summarizing data, performing calculations on data, filtering, and converting data into other file formats (Danecek, Auton et al. 2011). "*VCFtools Filter*" (Danecek, Auton et al. 2011) result represents in two output datasets: a "*Log*" folder which contains the information about the job itself and the number of polymorphic sites passing the filter and "*VCFdataset*" with only polymorphisms passing the MAF filter that the user selected.

#### **4.2.2. Identification of SNPs that Affect the Protein Sequence**

Generally, SNPs represent diversity in a DNA sequence, which can generate biological variations between people. Sometimes they have a direct role in the disease when they occur within a gene or in a regulatory region and have an impact on the phenotype (mutations). In general, though, they are mostly benign (Haraksingh and Snyder 2013). To identify any effect of SNPs on the sequence and the function of the protein, missense variants were investigated. The "*SnpEff*"

tool (Cingolani, Platts et al. 2012) implemented in Galaxy has been used to predict the effects of SNPs on the encoded protein and evaluate how many variants are classified as SNPs: deletion or insertion, missense, and nonsense. "*SnpEff*" is a tool for annotating and predicting the effect of the variant on genes (Cingolani, Platts et al. 2012). The parameters were used for running "*SnpEff*" are as follow; "Input and output formats" set as "VCF", the "Genome source" set as "Named on demand", the "GRCh37/hg19" selected under " *SnpEff* Genome Version Name", the "Upstream/Downstream length" set as "No upstream/downstream intervals (0 bases)", for "Annotation options", "Only use canonical transcripts" and "Use Classic Effect names and amino acid variant" were selected and, the "Text to prepend to chromosome name" set to "chr". "*SnpEff*" (Cingolani, Platts et al. 2012) generates two output datasets: one is the "*Stats*" file containing statistics information and the next one is "*VCFdataset*" including predicted effects for each SNP on the encoded CTB. Here, we needed to proceed with the "*SnpEff*" analysis twice: once with the VCF file filtered for  $MAF < 50\%$ , and once with the VCF file filtered for  $MAF < 1\%$ .

#### **4.2.3. Mapping SNPs in the Amino Acids Sequence and 3D Structure of CTB**

The amino acid sequence of the human CTB downloaded through the UniProt website (UniProtKB - P07858 (CATB\_HUMAN)), 2018, May 25 #2614). Based on the "*SnpEff*" (Cingolani, Platts et al. 2012) output, all the common and rare mutations which were presented in (1% to 50%) and (1% or less) of the population, respectively, mapped on the human CTB amino acid sequence. Furthermore, the docked structure of CID8795ATTO680 (Daniel Tesolin, Private Communication) used for visualizing the location of these common and rare mutations on CTB by performing VMD software (VMD APR,2008).



## 4.3. Results

### 4.3.1. Variant Identification in the 1,000 Genomes Data

According to "*VCFtools Filter*" (Danecek, Auton et al. 2011) results, 112 out of 1862 and 370 out of 1598 variant sites in the *CTSB* gene were identified with MAF between (0.01 and 0.5) and (0.01 and 0.001), respectively. The variant filtering result with MAF between 0.01 and 0.5 represents the common mutations that occur between 1% to 50% of the population. The variant filtering result with MAF between 0.001 and 0.01 represents the rare mutations that are carried by 1% or less of the population.

### 4.3.2. Identification of SNPs that Affect the Protein Sequence

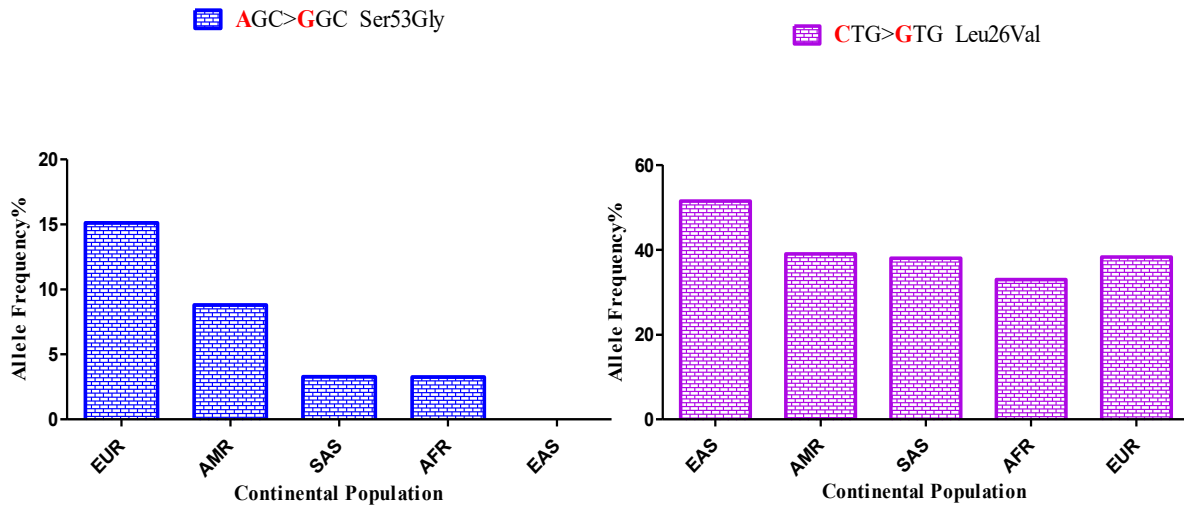
The "*SnpEff*" (Cingolani, Platts et al. 2012) predicts what effect each SNP is likely to have on the encoded CTB. In this study, the "*SnpEff*" analysis was carried out for two sets of analysis: one representing the effects of common SNPs on the coded CTB, and one for rare mutations which are present in 1% or less of the population.

The "*SnpEff*" (Cingolani, Platts et al. 2012) results for common SNPs are as follows; 62.7% were missense SNPs and 37.3% silent variants out of total 100 SNPs. Only two variants cause amino acid substitutions: Ser53Gly, and Leu26Val, which are located on the propeptide of CTB. The propeptide of CTB (position at 18-79 amino acid) is cleaved during maturation and activation. Once cleaved, a propeptide generally has no independent biological function. In terms of association with diseases, out of the two common CTB mutations (Leu26Val and Ser53Gly) (**Table 5**), only Leu26Val reported being associate with a significant number of tropical chronic pancreatitis patients (Mahurkar, Idris et al. 2006, HGMD 2020, June).

**Table 5.** The position of SNPs that cause amino acid changes in CTB found in MAF between 0.01 to 0.5 for the entire population which has been studied in 1,000 Genomes databased These amino acid substitutions are located on the propeptide of CTB.

Genome Position	Codon affected	Amino acid substitution	Allele Frequency
Chr8: 11710174(GRCh37.74) rs1803250	Agc/Ggc	p. Ser53Gly	0.057
chr8: 11710888 (GRCh37.74) rs12338	Ctg/Gtg	p. Leu26Val	0.395

The distribution of these SNPs in each continental population is shown in **Figure 21**. Based on the results, the highest percentage of allele frequency of (Leu26Val) belongs to the East Asian population (50%). Americans fit in the second level of allele frequency (39%), and African with 33% in the last group. The Ser53Gly mutation is also observed in all continental populations except East Asian. For Ser53Gly mutation, the European population shows the highest allele frequency at 15%, which is almost double of American, while Asians and Africans have the lowest frequencies (below 4%) for this SNP. The frequency of Ser53Gly mutation through the population is less significant compared to Leu26Val.

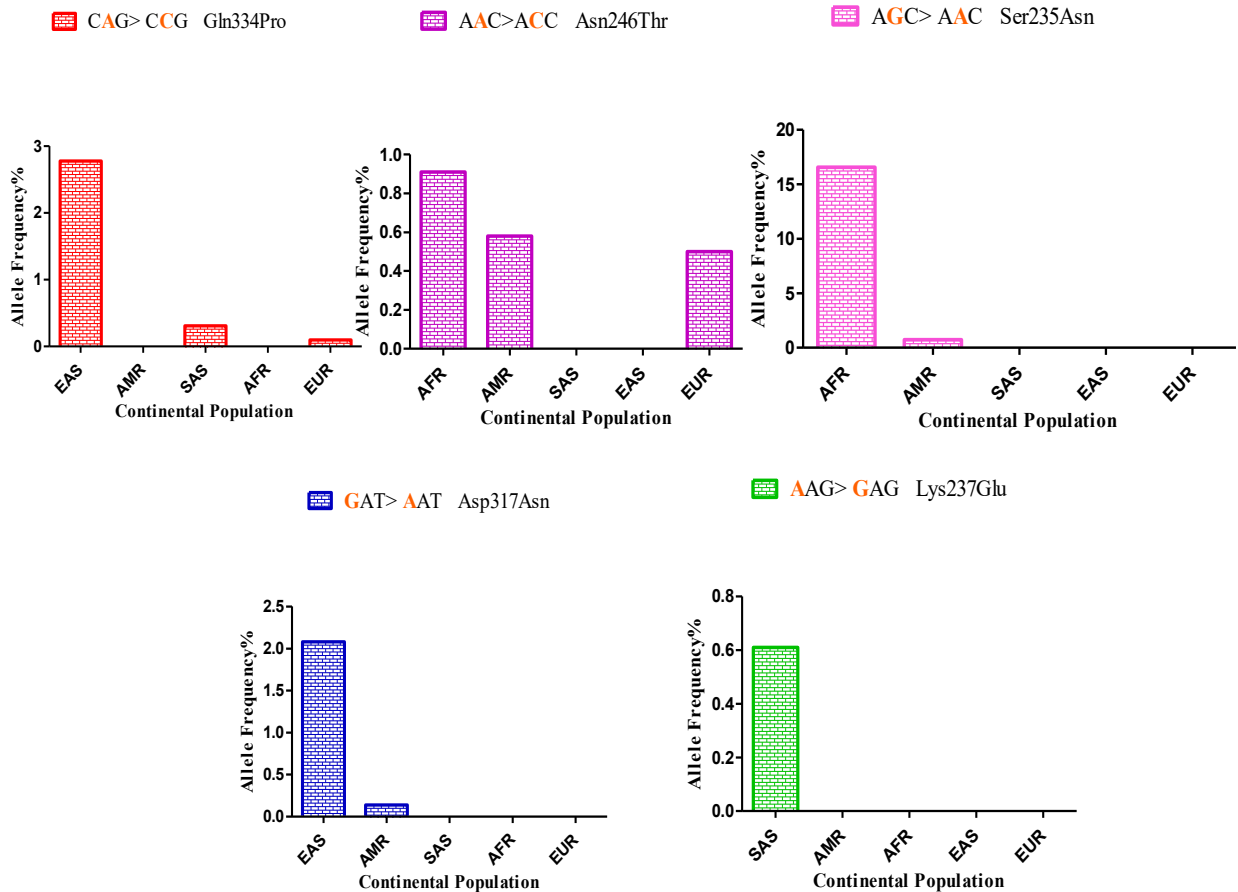


**Figure 21.** Population-based allele frequency of two variants associated with amino acid substitutions Ser53Gly and Leu26Val in the CTB protein.

The "*Snpeff*" (Cingolani, Platts et al. 2012) result based on rare mutations (present in 1% or less of the population) shows that the total number of SNPs effects by functional class was 39.2% missense and 60.7% silent variants out of 127 SNPs. Here, eight SNPs caused amino acid changes (Gln334Pro, Asp317Asn, Asn246Thr, Lys237Glu, Ser235Asn, Pro91Leu, Thr75Ala, and Leu26Pro) of which five are located within the coding region of the mature human cathepsin B (Pro91Leu, Ser235Asn, Lys237Glu, Asn246Thr, and Asp317Asn). The exact position and substitutions of all eight amino acids are listed in **Table 6**. The distribution of these SNPs in different populations is presented in **Figure 22**.

**Table 6.** The position of SNPs that cause amino acid changes in CTB found in MAF between 0.001 and 0.01 of the population.

<b>Genome Position</b>	<b>Codon affected</b>	<b>Amino acid substitution</b>	<b>Allele Frequency</b>
Chr8:11702653 (GRCh37.74) rs117613666	cAg/cCg	p. Gln334Pro	0.006
Chr8:11702705 (GRCh37.74) rs79487342	Gat/Aat	p. Asp317Asn	0.004
Chr8:11704617 (GRCh37.74) rs114308907	aAc/aCc	p. Asn246Thr	0.004
Chr8:11704645 (GRCh37.74) rs138489258	Aag/Gag	p. Lys237Glu	0.001
Chr8:11704650 (GRCh37.74) rs17573	aGc/aAc	p. Ser235Asn	0.044
Chr8:11708430 (GRCh37.74) rs11548596	cCa/cTa	p. Pro91Leu	0.003
Chr8:11708479 (GRCh37.74) rs74996838	Acc/Gcc	P. Thr75Ala	0.005
Chr8:11710887 (GRCh37.74) rs28605689	cTg/cCg	p. Leu26Pro	0.002



**Figure 22.** Population-based allele frequency of five variants (located on mature CTB) associated with amino acid substitutions Pro91Leu, Ser235Asn, Lys237Glu, Asn246Thr, and Asp317Asn in the CTB protein.

The highest percentage of the missense variant (Gln334Pro) and (Asp317Asn) belongs to East Asian 2.78% and 2.08%, respectively. For the missense variant (Asn246Thr), African people have the highest level of distribution (0.91%). East and South Asian populations illustrate 0% frequency for (Asn246Thr) mutation, which means they have no mutation in this location. African represents the highest percentage of allele frequency (16.57%) of the variant (Ser235Asn) meanwhile East and South Asian and European have no occurrence for this SNP (0%). Finally, South Asian has the most significant distribution (0.61%) for the variants (Lys237Glu) while other

populations have no change for this amino acid. The frequency of all the rare mutations mentioned here is not significant since 1% or less of all the population which have been studied here carry these variants.

#### **4.3.3. Mapping SNPs in the Amino Acids Sequence and 3D Structure of CTB**

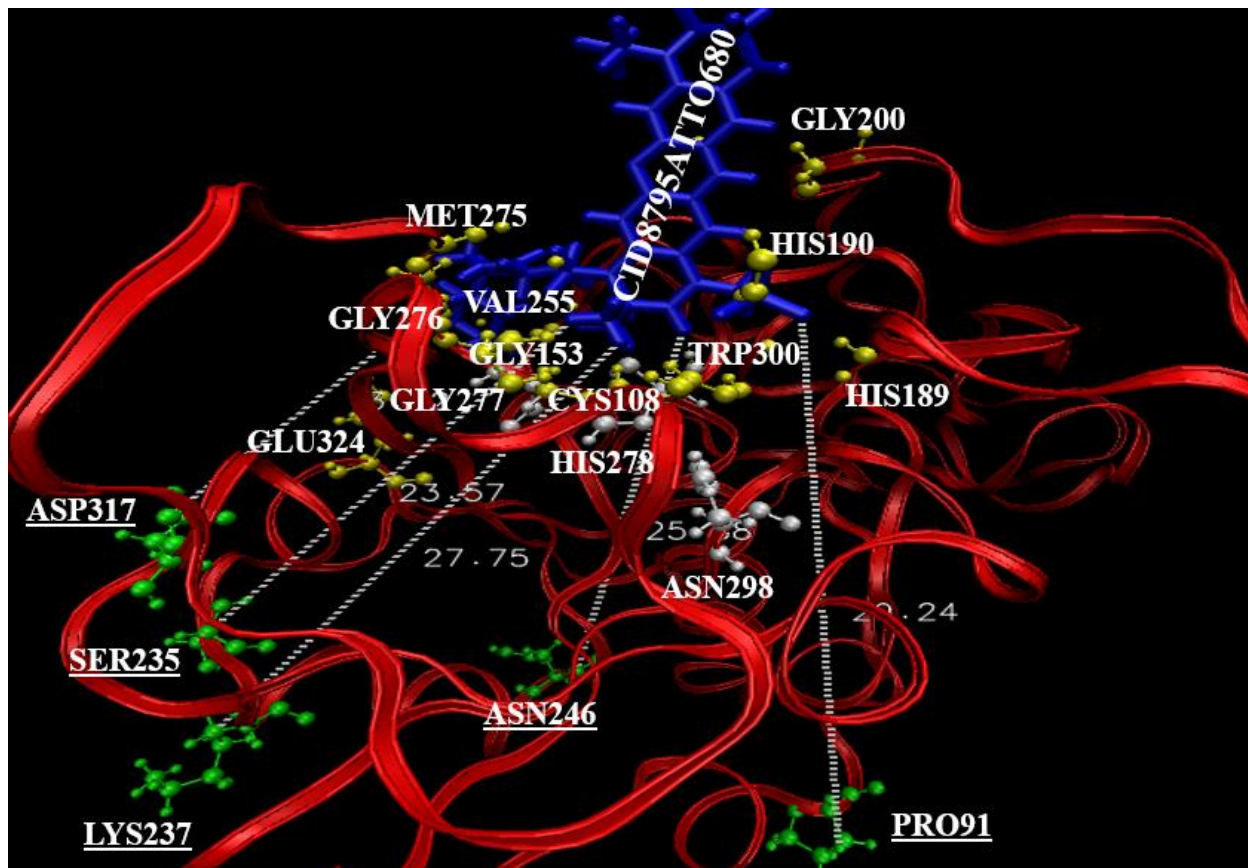
The position of each common (MAF between 0.5 and 0.01) and rare (MAF between 0.01 and 0.001) SNPs are marked in the amino acid sequence of human CTB in **Figure 23**. The two common variants are located on the propeptide portion of CTB. The mutations that may affect enzyme function are found in the mature and active CTB are Pro91Leu, Ser235Asn, Lys237Glu, Asn246Thr, and Asp317Asn. These mutations, however, are rare and are only present in 1% or less of the population.

10	20	30	40	50
MWQLWASLCC	LLVLANARSR	PSFH <u>LS</u> DEL	VNYVNRNTT	WQAGHNFYNV
60	70	80	90	100
<u>DMS</u> YLKRLCG	TFLGGPKPPQ	RVMF <u>T</u> EDLK <u>L</u>	PASFDAREQW	<u>P</u> QCPTIKEIR
110	120	130	140	150
DQGSCGSCWA	FGAVEAISDR	ICIHTNAHVS	VEVSAEDLLT	CCGSMCGDGC
160	170	180	190	200
<b>NGG</b> YPAEAWN	FWTRKGLVSG	GLYESHVGC	PYSIPPCE <b>HH</b>	VNGSRPPCT <b>G</b>
210	220	230	240	250
<b>E</b> GDTPKCSKI	CEPGYSPTYK	QDKHYGYNSY	SVSN <b>SEK</b> DIM	AEIYK <b>NP</b> VVE
260	270	280	290	300
GAFS <b>V</b> YSDFL	LYKSGVYQHV	TGEM <b>MGG</b> HAI	RILGWGVENG	TPYWLVA <b>NSW</b>
310	320	330		
NTDWGDNGFF	KILRGQ <b>D</b> HCG	IESE <b>V</b> VAGIP	RT <b>DQ</b> YWEKI	

**Figure 23.** The protein sequence of human CTB with marking SNPs. The red box indicates the propeptide of CTB (position at 18-79 amino acid) which is cleaved during enzyme maturation. Mutations with MAF < 0.5 are highlighted in blue (Leu26Val and Ser53Gly). The rare variations with a MAF < 0.01 that are located on the mature CTB identified by pink, meanwhile, those rare mutations which are not present in the mature CTB highlighted by orange color. Two mutations occur at position Leu26, common highlighted as blue and rare underlined by orange color. The red highlighted amino acids indicate the predicted binding site of the protein for the selected probes (Cbz-Lys-Lys-PABA-DCMF and CID8795ATTO680). Finally, the start and end positions of the mature protein are highlighted as green color.

The amino acid substitutions were also mapped onto the three-dimensional structure of CTB (**Figure 24**). In this study, the fluorogenic substrate-based probes Cbz-Lys-Lys-PABA-AMC and Cbz-Lys-Lys-PABA-DCMF, and the non-reactive probe CID8795ATTO680 (**Figure 24**) were hypothesized to bind to CTB by targeting amino acids: Asn151, Gly152, Gly153, His189, His190, Gly200, Glu201, Val255, Met275, Gly276, Gly277, His278, Trp300, and Glu324 (see **chapter 3**). Therefore, we investigate the mutations in the predicted anchor amino acids in CTB for these

probes. Distances were measured between the  $C\alpha$  of all residues and atom Z of the bound ligand (CID8795ATTO680).



**Figure 24.** View of docked structure of CID8795ATTO680 (Daniel Tesolin, Private Communication) with mapping residues within 3.5 Å of the probe by VMD software (VMD APR,2008). The red ribbons show the *IHUC* carbon alpha trace. The probe is shown in blue. Yellow represents the residues within 3.5 Å of the CID8795ATTO680. The name and position of these amino acids are as follows: Gly153, His189, His190, Gly200, Glu201, Val255, Met275, Gly276, Gly277, His278, Trp300, and Glu324. White represents the three active site residues (Cys108, His278, and Asn298) of human CTB. Green shows the positions of rare mutations on this structure which are located above 20 Å from the binding pocket of CTB.

Mapping rare amino acid substitutions on the three-dimensional structure of human CTB as visualized with VMD (VMD APR,2008) shows that these amino acid substitutions may not affect probe/CTB interactions since their distance to the bound probe is above 20 Å.



#### 4.4. Summary and Conclusions

In this section, by using the 1,000 Genomes database and software implemented at Lakehead's Galaxy server, we were able to identify SNPs located in the coding region of the *CTSB* gene. SNPs were classified into two groups: rare (MAF 0.01 to 0.001) and common (MAF 0.5 to 0.01) mutations. Here, our focus was on binding site residues of CTB for our probes. As described above, the analysis of genomic data did not identify the presence of mutations at these residues. All the SNPs, identified in this study, were mapped onto the CTB amino acid sequence and 3D structure of this protein to identify their location on this protein. The two common mutations which were found in 1% to 50% of the population are present on the proenzyme but, not the mature enzyme. These common SNPs cause the amino acid substitutions Ser53Gly (~ 6%) and Leu26Val (~ 39%) within the entire 1,000 Genomes database populations. *CTSB* polymorphism (Leu26Val) has been associated with pancreatic cancer (Mahurkar, Idris et al. 2006, HGMD 2020, June). Additionally, we found eight rare mutations in 1% or less of the population, five of which were located on the mature cathepsin B (Pro91Leu, Ser235Asn, Lys237Glu, Asn246Thr, and Asp317Asn). These rare variants were not located within the binding site of the protein and there were no diseases or conditions that have been reported as associated with them.

In summary, there is no expected effect of these SNPs on the CTB/probe interactions; therefore, these mapping results suggest no differences in probe recognition is expected within the populations studied in the 1,000 Genomes database.

## Chapter 5: Synthesis and Characterization of Fluorescent Probes

### 5.1. Introduction

Fluorescent probes have become indispensable tools for chemists and biologists to study the biological activities of proteins and enzymes in living cells and animals (Chen, Pradhan et al. 2012). They provide visual information with high spatial resolution, and probes with NIR emission are preferable for *in vivo* imaging because of the low fluorescent interference due to the autofluorescence of cells and tissues due to natural compounds. Compared to conventional fluorophores that are excited at low wavelengths, NIR fluorescent probes can provide imaging data in deeper tissues due to their longer excitation wavelengths. Indeed, the NIR excitation and emission range results in less damage to biological samples than the shorter wavelengths (Celli, Spring et al. 2010, Guo, Park et al. 2014).

#### 5.1.1. Selection of NIR Probe Candidate

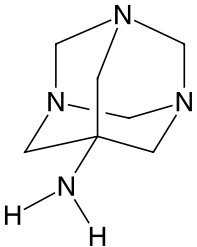
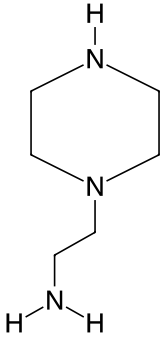
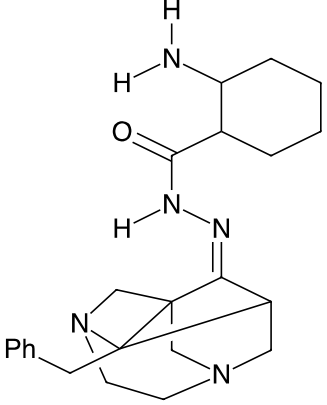
For detecting overexpression of CTB in aggressive and invasive cancers, designing fluorogenic probes that have high affinity and selectivity is required. To accomplish this, probe molecules can be inherently fluorescent or comprised of a potent targeting moiety labeled with a fluorescent dye that binds to the enzyme. A total of 14,862 compounds were constructed in Kamstra and et al., and screened against CTB intending to identify non-reactive probes with near-infrared (NIR) emission (Kamstra, Dadgar et al. 2014). From this library, seven ligands were selected for conjugation. General information about these seven potential ligands is shown in **Table 7** (Molport 2018, November 16), (Kamstra, Dadgar et al. 2014).

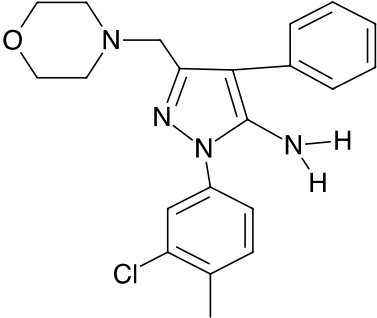
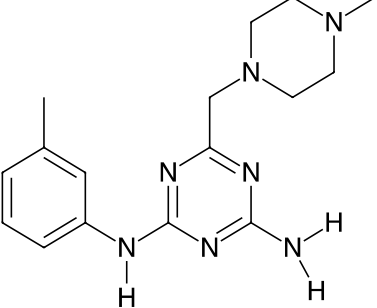
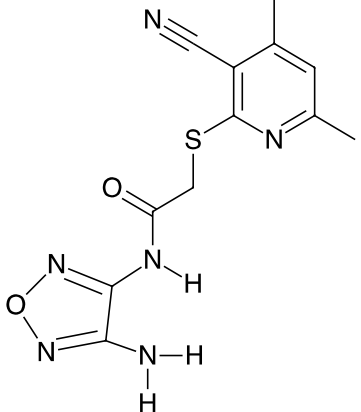
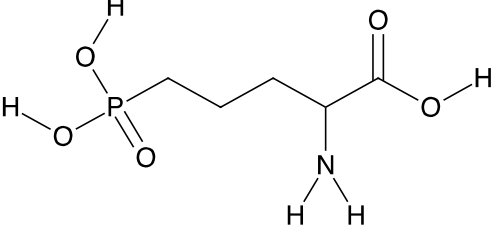
**Table 7.** Lead probe compounds discovered computationally (Daniel Tesolin, Private Communication) (Molport 2018, November 16).

<b>PubChem CID</b>	<b>Compound name (IUPAC)</b>	<b>MW (g/mol)</b>	<b>MF</b>	<b>Vendor (Product Code)</b>
<b>CID2434131</b>	<i>2-(3-chloro-4-methylphenyl)-5-(morpholin-4-ylmethyl)-4-phenylpyrazol-3-amine</i>	382.88	C <sub>21</sub> H <sub>23</sub> ClN <sub>4</sub> O	Vista-M Laboratory, Ltd.
<b>CID5939530</b>	<i>(2-aminophenyl)-N- {[1-benzyl-3,6-diazatricyclo [4.3.1.1 &lt; 3,8 &gt;] undec-9-ylidene]az amethyl} carboxamide</i>	389.49	C <sub>23</sub> H <sub>27</sub> N <sub>5</sub> O	Tim Tec (ST50912625)
<b>CID667134</b>	<i>2-N-(3-methylphenyl)-6-[(4-methylpiperazin-1-yl)methyl]-1,3,5-triazine-2,4-diamine</i>	313.40	C <sub>16</sub> H <sub>23</sub> N <sub>7</sub>	Vista-M Laboratory, Ltd.
<b>CID1256741</b>	<i>2-(4-chlorophenoxy)-2-methyl-N-(2-pyridin-4-yl-1,3-benzoxazol-5-yl)propanamide</i>	407.84	C <sub>22</sub> H <sub>18</sub> ClN <sub>3</sub> O	Vista-M Laboratory, L (STK236667)
<b>CID2999504</b>	<i>N-(5-cyclopropyl-1,3,4-thiadiazol-2-yl)-2-(6-methoxy-1-benzofuran-3-yl)acetamide</i>	329.37	C <sub>16</sub> H <sub>15</sub> N <sub>3</sub> O <sub>3</sub> S	Enamine (T5257017)
<b>CID8795</b>	<i>2-piperazine-1-ylethanamine</i>	129.20	C <sub>6</sub> H <sub>15</sub> N <sub>3</sub>	Ark Pharm, Inc (AK-59134)
<b>CID535684</b>	<i>1,3,5-triazatricyclo [3.3.1.1~3,7~] decan-7-amine</i>	154.21	C <sub>7</sub> H <sub>14</sub> N <sub>4</sub>	Enamine (EN300-68939)
<b>ATTO 680 NHS-ester</b>	<i>1-{4-[(2,5-dioxopyrrolidin-1-yl)oxy]-4-oxobutyl}-11-ethyl-2,2-dimethyl-4-(sulfonatomethyl)-2,8,9,10-tetrahydro-1H-13-oxa-1,6,11-triazapentacen-11-ium</i>	622.68	C <sub>31</sub> H <sub>34</sub> N <sub>4</sub> O <sub>8</sub> S	ATTO-TEC GmbH (AD 680-3)

Ligands in this library are novel and designed for straightforward synthesis. Each compound is commercially available, and, in most cases, the ligands bear an amino-group (**Table 8**) that can be converted into a fluorescent NIR probe by conjugation to commercially available ATTO680 dye with its *N*-hydroxysuccinimide (NHS)-ester group (amino-reactive) forming a stable amide linkage.

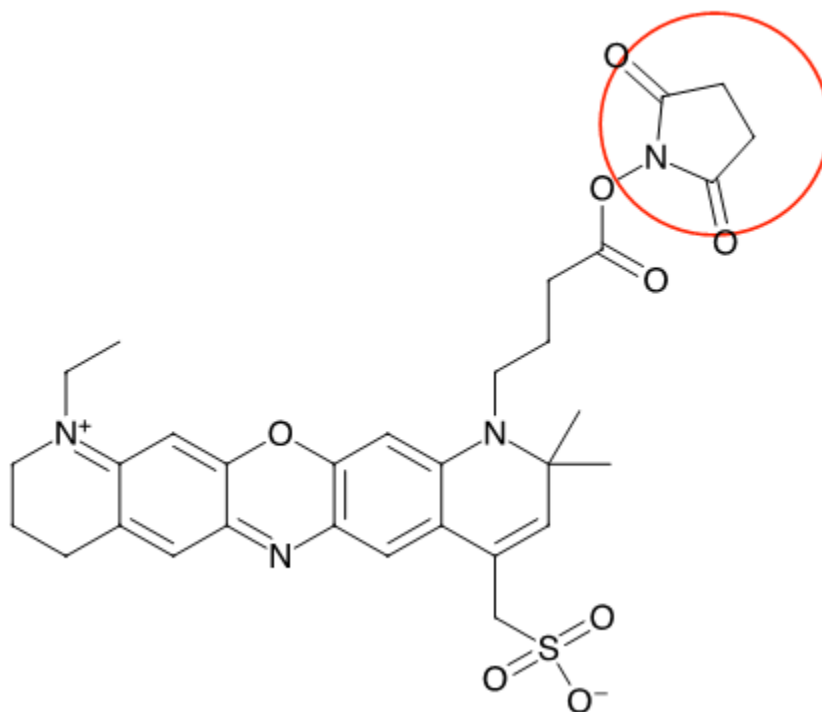
**Table 8.** The structure of the selected ligands for ATTO680 conjugation (Kamstra, Dadgar et al. 2014).

Base-Compound CID	Structure of base-compound
CID535684	
CID8795	
CID5939530	

Base-Compound CID	Structure of base-compound
CID2434131	
CID667134	
CID4917705	
CID1216	

NHS-esters are preferable for the labeling process for four main reasons: 1) NHS-esters have high specificity towards the primary amines group; 2) the resulting amide is a very stable linkage

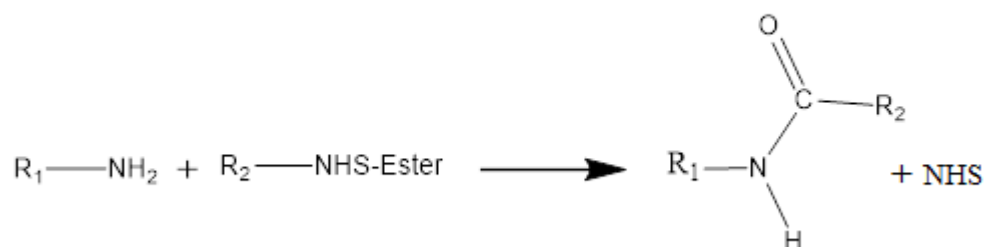
(Sameiro and Goncalves 2009), 3) ATTO680 is a hydrophilic dye which allows better functionality *in vitro* and *in vivo* (Buschmann, Weston et al. 2003), 4) it has a fluorescence lifetime at 3.04 ns which indicates better differentiation in FP assay between its free form and bound form (Hughes, Rawle et al. 2014). The chemical structures of the ATTO680 fluorescent dyes used in this study are displayed in **Figure 25**. In this figure, the reactive group targeted for conjugate is indicated by red circles.



**Figure 25.** The chemical structure of ATTO 680 NHS-ester. The red circle indicates the site of conjugation.

According to the reaction in **Scheme 2**, R<sub>1</sub>-NH<sub>2</sub> is the ligand, and R<sub>2</sub>-NHS-ester is the commercially available ATTO 680 dye. In this study, the synthesis of the fluorescent conjugates

was performed in collaboration with Dr. Morshed Chowdhury from Dr. Phenix's group. Seven candidates were identified as potential ligands of CTB and were used as starting material for conjugation chemistry employing an ATTO 680 NHS-ester dye. As described above, all seven candidates had a primary amino group that should attack the ATTO 680 NHS-ester forming an amide as the conjugation functional group. Of note, some of the ligands shown above have a primary amine that is attached to an aromatic ring (e.g. CID2434131, CID4917705, CID667134). Therefore, the lone pair of electrons on nitrogen is delocalized into the  $\pi$  system thus reducing the reactivity of the amine making conjugation with the ATTO680 more difficult compared to the ligands bearing a primary amine.



**Scheme 2.** Chemical reaction scheme of ligands (**Table 8**) with an amino group and ATTO 680 NHS-ester dye. Where R1-NH<sub>2</sub> is the ligand attached to the amine group, and R<sub>2</sub>-NHS-ester is ATTO 680 dye bound to the *N*-hydroxysuccinimide group.

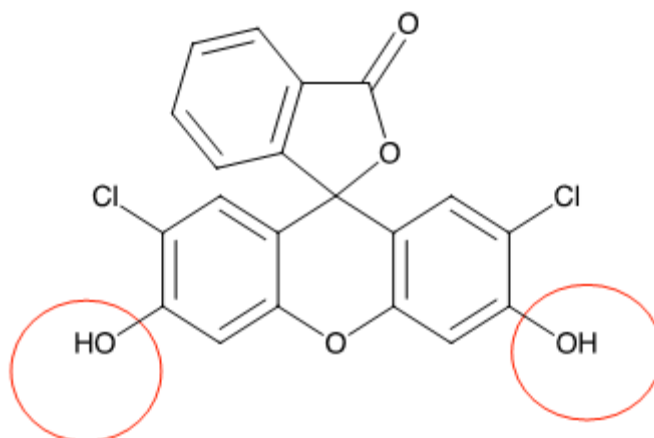
### 5.1.2. Selection of Substrate-Based Fluorescent Probe

Previously, Dr. Phenix's group designed and synthesized prodrug inspired substrate-based fluorogenic probes selective to CTB (Chowdhury, Moya et al. 2014). Each substrate consisted of an *N*-carbobenzyloxy (Cbz) protected dipeptide (Cbz-phenylalanyl-lysine and Cbz-lysine-lysine) conjugated to the prodrug linker *p*-Aminobenzyl alcohol (PABA) and 7-Amino-4-methylcoumarin

(AMC) as a fluorescent reporter group. They found that treating the fluorogenic peptides Cbz-Phe-Lys-PABA-AMC ( $k_{\text{cat}}/K_{\text{M}}$  of  $6.1 \pm 7 \text{ mM}^{-1} \text{ s}^{-1}$ ) and Cbz-Lys-Lys-PABA-AMC ( $231 \pm 70 \text{ mM}^{-1} \text{ s}^{-1}$ ) with CTB led to the efficient release of PABA and spontaneous immolation of the linker to liberate AMC as the fluorescent reporter. Importantly, the intact peptides had little to no fluorescence, and the enzymatic reactions could be monitored by measuring the fluorescence intensity of the released AMC ( $\lambda_{\text{ex}} = 380 \text{ nm}$ ;  $\lambda_{\text{em}} = 460 \text{ nm}$ ). A series of detailed biological experiments showed that Cbz-Lys-Lys-PABA-AMC was selective to CTB in cancer cell lysates and its hydrolysis could be detected in living cells monitored by confocal microscopy. In contrast, Cbz-Phe-Lys-PABA-AMC displays low specificity and high toxicity in live-cell fluorescence microscopy studies (Chowdhury, Moya et al. 2014). An additional limitation was that the substrate-based probes employed AMC ( $\lambda_{\text{ex}} = 380 \text{ nm}$ ;  $\lambda_{\text{em}} = 460 \text{ nm}$ ) as a fluorescent reporter group. At these excitation and emission wavelengths, live cells have high endogenous fluorescence due to natural fluorescent species found in cells (e.g., auto-fluorescence). More sensitive fluorescence imaging studies can be accomplished in live cells by using fluorophores with longer wavelengths or by incorporating a radioactive reporter. To address these shortcomings, a new substrate-based probe was designed using a fluorescent reporter group with longer excitation and emission wavelengths as well as a dye that remains fluorescent over a broader pH range. Molecular docking was used (results not shown here) to select fluorophores that are expected to maintain or improve binding affinity to CTB. To improve the fluorescent profile of the new probe at low pH, one of the fluorescein derivatives (*2', 7'-dichloro-6'-methoxy-fluorescein* (DCMF with  $\lambda_{\text{ex}} = 480 \text{ nm}$  and  $\lambda_{\text{em}} = 522 \text{ nm}$ )) has been selected. DCMF has been used previously for reporting on enzymatic activity in the lysosome of cells (Mugherli, Burchak et al. 2006, Harlan, Lusk et al. 2016). Also, its maximum emission wavelength is well above the cellular background. The



chemical structures of the DCMF fluorescent dyes used in this study are displayed in **Figure 26**. In this figure, the reactive group targeted for conjugation is indicated by red circles.



**Figure 26.** The chemical composition of 2', 7'-dichloro -fluorescein. The red circles indicate sites of conjugation.

## 5.2. Procedures and Materials

### 5.2.1. Synthesis of the NIR Probe

Conjugation is the process by which the dye and a compound of interest react to form a single chemical entity. The chemical conditions used for each conjugation process depends on the functional groups present on the reactant and the dye structure (Feixas, Matito et al. 2011).

To synthesize the dye-labeled conjugates, the dye was purchased with an activated ester utilizing *N*-hydroxysuccinimide as a leaving group. Nucleophilic attack by the primary amine of the binding ligand would generate the fluorescent probe (as an amide) required for testing. For the conjugation reactions to be successful, dry DMF was used since water may hydrolyze the *N*-

hydroxysuccinimide ester leading to poor yields. Triethylamine was added as a base to neutralize the protons produced during the conjugation reaction. The reaction mixture was stirred at 0 °C for 30 min and then at room temperature overnight. After overnight incubation, water was added to the reaction, and the mixture was extracted with ethyl acetate. The aqueous layer was separated and freeze-dried to get the title compound as a dark blue solid. Since CID8795ATTO680 and CID535684ATTO680 were obtained in low mg amounts, NMR was not done for these probes. For these products after running TLC, column chromatography was used for purification of the products. After purification, all spots on the TLC were subjected to TLC-mass spectrometry using the Advion Expression CMS, which confirmed the molecular weight of CID8795ATTO680 at 636.31 g/mol and CID535684ATTO680 at 661.30 g/mol as expected. Unexpectedly, none of the synthesized probes were soluble in the water; therefore, DMSO used for making the stock solution for CID8795ATTO680 and methanol used as a solvent for CID535684ATTO680.

### 5.2.2. Synthesis of the Fluorescent Substrate

#### *2', 7'-Dichlorofluorescein (1)*

*2', 7'-Dichlorofluorescein* was prepared following literature procedures (Mugherli, Burchak et al. 2006).

Characterization of the dye matches previously published NMR and mass spectra data. This section is done by Dr. Morshed Chowdhury from Dr. Phenix's group.

#### *2', 7'-Dichloro-6'-methoxy-fluorescein (2)*

MeI (1.33 g, 9.37 mmol) was added to the mixture of **1** (1.5 g, 3.74 mmol) and K<sub>2</sub>CO<sub>3</sub> (1.3 g, 9.37 mmol) in 15 ml of DMF at room temperature. After stirring for 24 hours, the reaction

mixture was quenched with H<sub>2</sub>O and solvents removed under reduced pressure. To this residue was added 30 mL of an aqueous solution of 10% NaOH and 60 mL of acetone. This mixture was subsequently refluxed for 1 hour at 60°C. The solution was cooled to room temperature, then diluted with 100 mL of EtOAc. Dilute HCl was added to acidify the mixture to approximately pH = 3 and then extracted the organic layer three times with 100 mL of EtOAc. The organic layers were combined and washed with brine, dried with sodium sulfate, and filtered. The solvent removed under vacuum, and the resulting solid was purified using silica gel chromatography (2:1 EtOAc: Hexanes with 0.1% AcOH) to get **2** (853 mg, 55%) as an orange solid. <sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>) δ 8.07 (d, *J* = 7.4 Hz, 1H), 7.70 (m, 2H), 7.17 (d, *J* = 7.5 Hz, 1H), 6.92 (s, 1H), 6.81 (d, *J* = 4.6 Hz, 1H), 6.77-6.68 (m, 2H), 3.95 (s, 3H). <sup>13</sup>C NMR (126 MHz, CDCl<sub>3</sub>) δ 169.08, 156.66, 152.09, 150.97, 150.77, 135.55, 130.34, 128.70, 128.07, 126.34, 125.47, 123.86, 118.27, 111.23, 108.81, 104.13, 103.56, 100.39, 56.48. LRMS (ESI): *m/z* Calcd for [C<sub>21</sub>H<sub>12</sub>C<sub>12</sub>O<sub>5</sub> + Na]<sup>+</sup>: 436.99; found: 437.00.

***Cbz-Lys-N-ε-BOC-Lys-N-ε-BOC-PAB-2', 7'-Dichloro-6'-methoxy-fluorescein (4)***

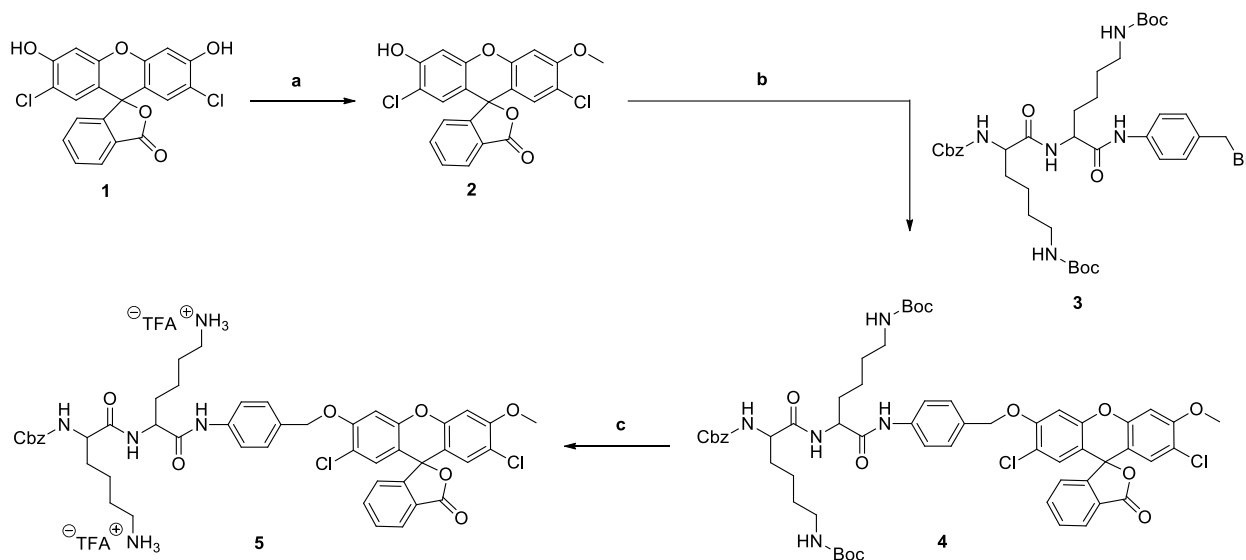
Compound **2** (257 mg, 0.168 mmol) was added to a mixture of **3** (400 mg, 0.515 mmol) and Ag<sub>2</sub>O (143 mg, 0.618 mmol) in 10 mL of benzene. The mixture was refluxed in the dark for 24 hours. After quenching the reaction with 50 mL of water the mixture was passed through filter paper into a separatory funnel. The filter paper was washed with EtOAc and added to the same separatory funnel. 100 mL EtOAc was added, and the aqueous phase was extracted three additional times with 50 mL of EtOAc. The organic layers were combined then washed with brine (saturated NaCl), dried with Na<sub>2</sub>SO<sub>4</sub>, filtered, and concentrated under vacuum. Silica gel chromatography (2:1 EtOAc: Hexanes → 5:1 EtOAc Hexanes) yielded compound **4** (130 mg, 23%) as a light orange

solid. **4**:  $^1\text{H}$  NMR (500 MHz,  $\text{CDCl}_3$ )  $\delta$  8.04 (d,  $J = 7.4$  Hz, 1H), 7.78-7.69 (m, 2H), 7.63 (d,  $J = 8.6$  Hz, 2H), 7.39 (d,  $J = 8.6$  Hz, 2H), 7.35-7.23 (m, 5H), 7.17 (d,  $J = 7.4$  Hz, 1H), 6.95 (s, 1H), 6.87 (s, 1H), 6.71 (s, 1H), 6.67 (s, 1H), 5.11 (s, 2H), 5.09 (s, 2H), 4.52-4.46 (m, 1H), 4.18-4.13 (m, 1H), 3.86 (s, 3H), 3.07-2.96 (m, 4H), 1.94-1.60 (m, 4H), 1.54-1.37 (m, 26H). LRMS (ESI):  $m/z$  Calcd for  $[\text{C}_{58}\text{H}_{65}\text{Cl}_2\text{N}_5\text{O}_{13} + \text{Na}]^+$ : 1132.38; found: 1132.39.

***Cbz-Lys-Lys-PAB-2', 7'-Dichloro-6'-methoxy-fluorescein 2TFA. (-5-)***

Compound **4** (100 mg, 0.09 mmol) was added to a dry flask and cooled to  $0^\circ\text{C}$  using an ice bath. In a separate container, a 1:1 TFA:  $\text{CH}_2\text{Cl}_2$  (anhydrous) stock was made and cooled to  $0^\circ\text{C}$ . Using a syringe, 0.5 mL of 1:1 TFA:  $\text{CH}_2\text{Cl}_2$  stock was added to the flask dropwise while stirring and maintaining a temperature of  $0^\circ\text{C}$  for 10 minutes. After this time, 5 mL of diethyl ether was added to precipitate the dipeptide. The mixture was transferred to a 15 ml Eppendorf tube and centrifuged at 10000 g for 1 minute at  $4^\circ\text{C}$ . The diethyl ether was subsequently discarded while keeping the precipitate intact. The precipitate was re-suspended in 5 mL of diethyl ether, and the centrifugation steps, including washing with ether, was repeated two more times. The final residue was then purified on a Waters Sep-Pak® tC18 plus cartridge using 2% ACN:  $\text{H}_2\text{O}$  as the eluent. The fractions containing the desired compound were then frozen and freeze-dried to yield compound **5** (11 mg, 11%) obtained as a faint orange solid. The NMR result for the new substrate is as follows: (2:1 EtOAc: Hexanes with 0.1% AcOH) to get **2** (853 mg, 55%) as an orange solid. **5**:  $^1\text{H}$  NMR (500 MHz,  $\text{CDCl}_3$ )  $\delta$  8.04 (d,  $J = 7.4$  Hz, 1H); 7.92-7.54 (m, 4H); 7.44 (d,  $J = 7.6$  Hz, 1H); 7.41-7.27 (m, 6H), 7.21 (d,  $J = 7.6$  Hz, 1H); 6.97-6.92 (m, 1H); 6.91-6.85 (m, 1H); 6.78-6.73 (m, 1H); 6.73-6.65 (m, 1H); 5.15 (s, 2H), 5.12 (s, 2H); 4.58-4.50 (m, 1H); 4.21-4.11 (m, 1H); 3.86 (s, 3H); 2.99-2.84 (m, 4H), and 2.01-1.44 (m, 12H). LRMS (ESI):  $m/z$  Calcd for  $[\text{C}_{48}\text{H}_{49}\text{Cl}_2\text{N}_5\text{O}_9 + \text{Na}]^+$ : 932.28, found: 932.29.

As described above, Cbz-Lys-Lys-PABA-DCMF was synthesized in 5 chemical steps which are shown below in **Scheme 3**.



**Scheme 3:** Fluorophore synthesis and peptide coupling. (a) i) MeI, K<sub>2</sub>CO<sub>3</sub>, DMF, rt; ii) 10% NaOH<sub>(aq)</sub>, acetone, reflux; iii) 0.1 M HCl<sub>(aq)</sub>, 55%; (b) **3**, Ag<sub>2</sub>O, benzene, reflux, 23%; (c) TFA, CH<sub>2</sub>Cl<sub>2</sub> at 0°C, 11%.

### 5.2.3. Spectrophotometric Characterization

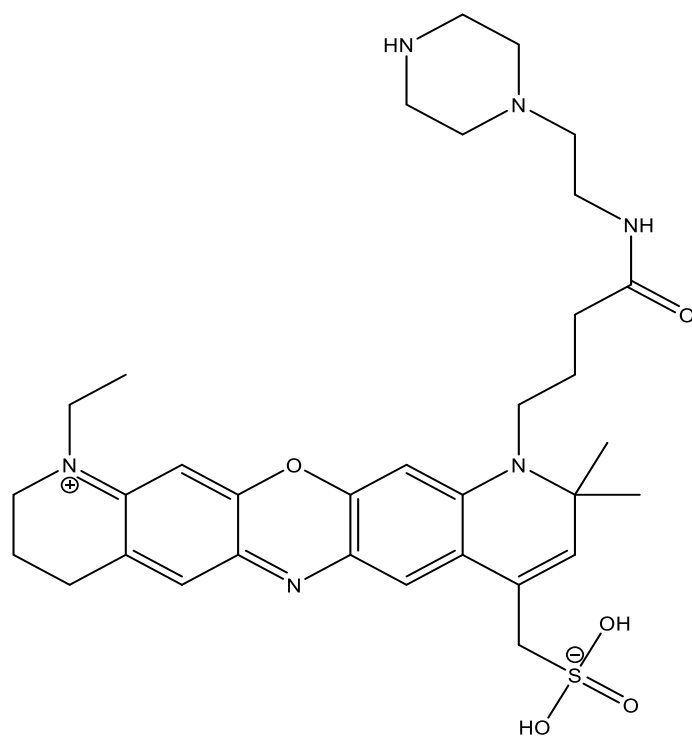
The emission and excitation of all the ligands at 100 μM, the NIR conjugates at 40 μM, and dye (DCMF) at 1 μM were initially measured prior to any binding experiments with the enzymes. Maximum absorbance, emission, and excitation spectra were determined for all fluorescent compounds at 37°C using a Synergy 4™ plate reader (Biotek Inc). The absorbance spectra were taken from 300 nm to 750 nm in 5 nm increments. For the emission spectra, the excitation wavelength maxima were chosen for each compound monitoring emission at 400 nm to 800 nm in 5 nm increments. Four buffers have been used in these experiments; CTB reaction buffer, Tris at 50 mM, MES at 50 mM, and acetate buffer at 50 mM. For ligands, the commercial CTB reaction

buffer at pH 7.2, for CID8795ATTO680 and CID535684ATTO680 as well as CTB reaction buffer, Tris (pH 7) and MES (pH 4) were used. Finally, for DCMF acetate buffer at pH 5.5 was used. For CID8795ATTO680 and CID535684ATTO680 a two-fold serial dilution with a starting concentration at 0.195 to 200  $\mu\text{M}$  was used for measuring anisotropy and fluorescence intensity assays at zero-time point, at pH 7.2 and 37°C.

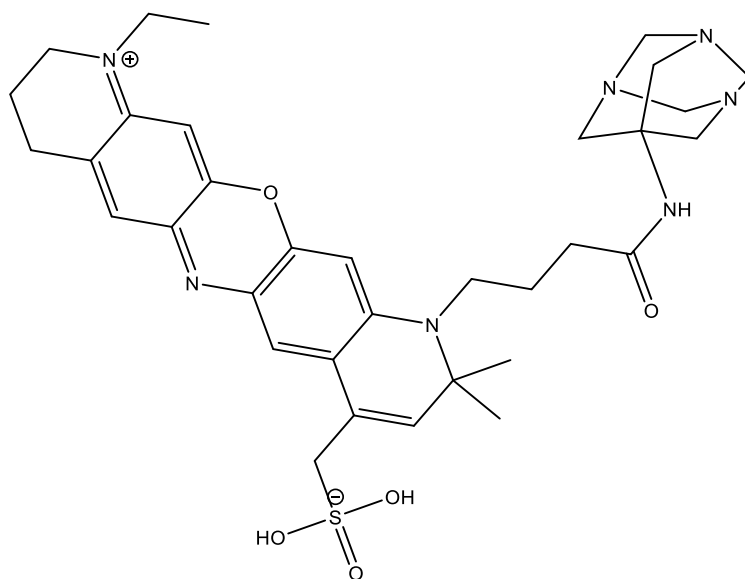
## 5.3. Results

### 5.3.1. Synthesis and Conjugation of Selected Hits

Seven ligands that were selected from a computational screening (Kamstra, Dadgar et al. 2014) were subjected to conjugation chemistry with ATTO 680 NHS-ester (CID2434131, CID5939530, CID667134, CID1256741, CID2999504, CID8795, and CID535684). As expected, CID8795 and CID535684 reacted smoothly with the ATTO dye resulting in novel compounds named CID8795ATTO680 (**Figure 27**) and CID535684ATTO680 (**Figure 28**), respectively. Unfortunately, the fluorescent conjugates were not produced by the conjugation chemistry for CID2434131, CID5939530, CID667134, CID1256741, and CID2999504. This is likely due to the low reactivity of the nitrogen atom in some of these ligands due to the delocalization of the lone pair of electrons into aromatic rings (e.g. amides).



**Figure 27.** The chemical structure of the CID8795ATTO680.



**Figure 28.** The chemical structure of CID535684ATTO680.

### 5.3.2. An Improved Substrate-Based Fluorescent Probe for CTB

In this work, we employed the same scaffold previously reported (Cbz-Lys-Lys-PABA) (Chowdhury, Moya et al. 2014) but with DCMF as a fluorescent reporter.

The 2',7'- *Dichlorofluorescein* (**1**) was methylated in DMF to form a methyl ester and methyl ether to produce compound **1** (compound was not isolated). The ester of compound **1** was then hydrolyzed using 10% NaOH to reform the lactone ring resulting in compound **2** in low yield. This fluorophore was then coupled with *Cbz-Lys-N-ε-BOC-Lys-N-ε-BOC-PAB-Br* (**3**), using Ag<sub>2</sub>O in a non-polar solvent to obtain peptide **4**. Compound **3** was already prepared and available in the Phenix lab. *Boc* (Tert-butoxy carbonyl protecting group) deprotection with Trifluoroacetic acid (TFA) yielded the final compound **5** (**Scheme 3**) ( (UniProtKB - P07858 (*CATB\_HUMAN*)) with an 11% yield.



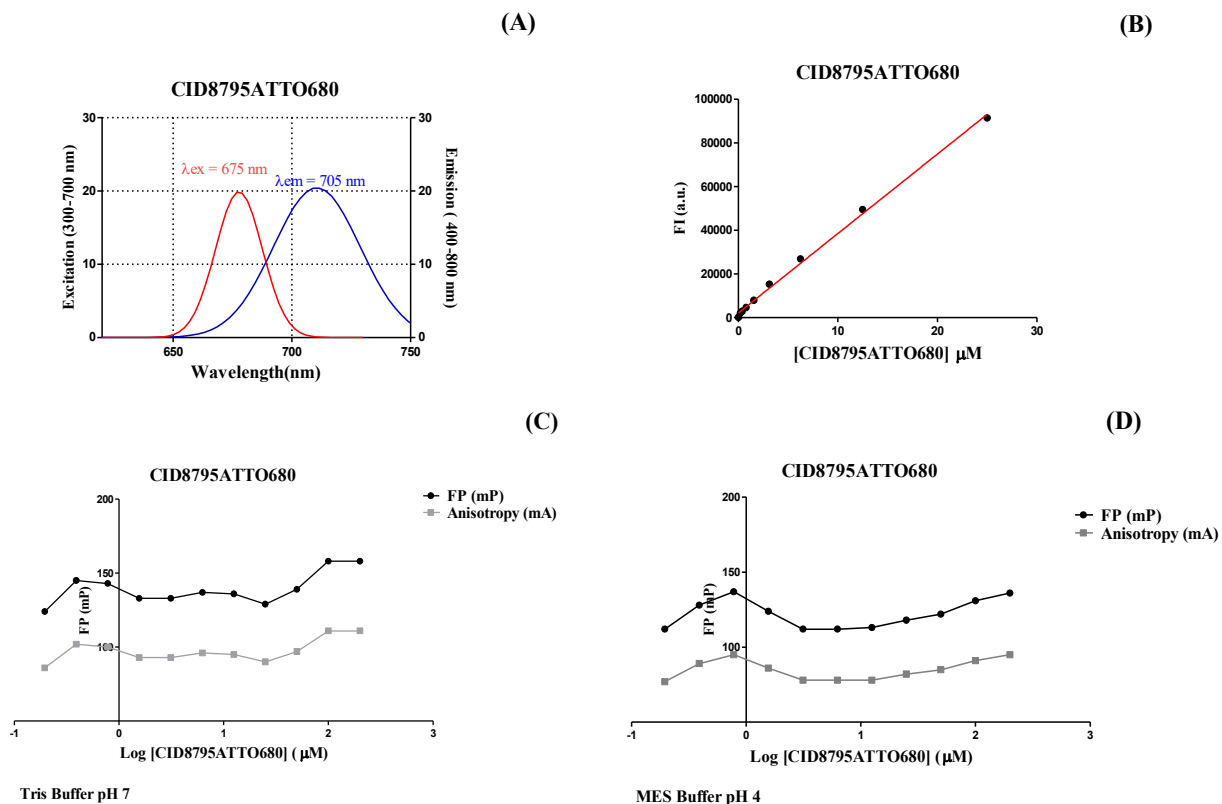
### 5.3.3. Spectrophotometric Characterization

To determine if the ligands identified by Tesolin, had inherent fluorescence, excitation, and emission spectra were measured for all ligands; CID2434131, CID5939530, CID667134, CID1256741, CID2999504, CID8795, and CID535684. None of these ligands showed any inherent fluorescence in their spectrophotometric characterization.

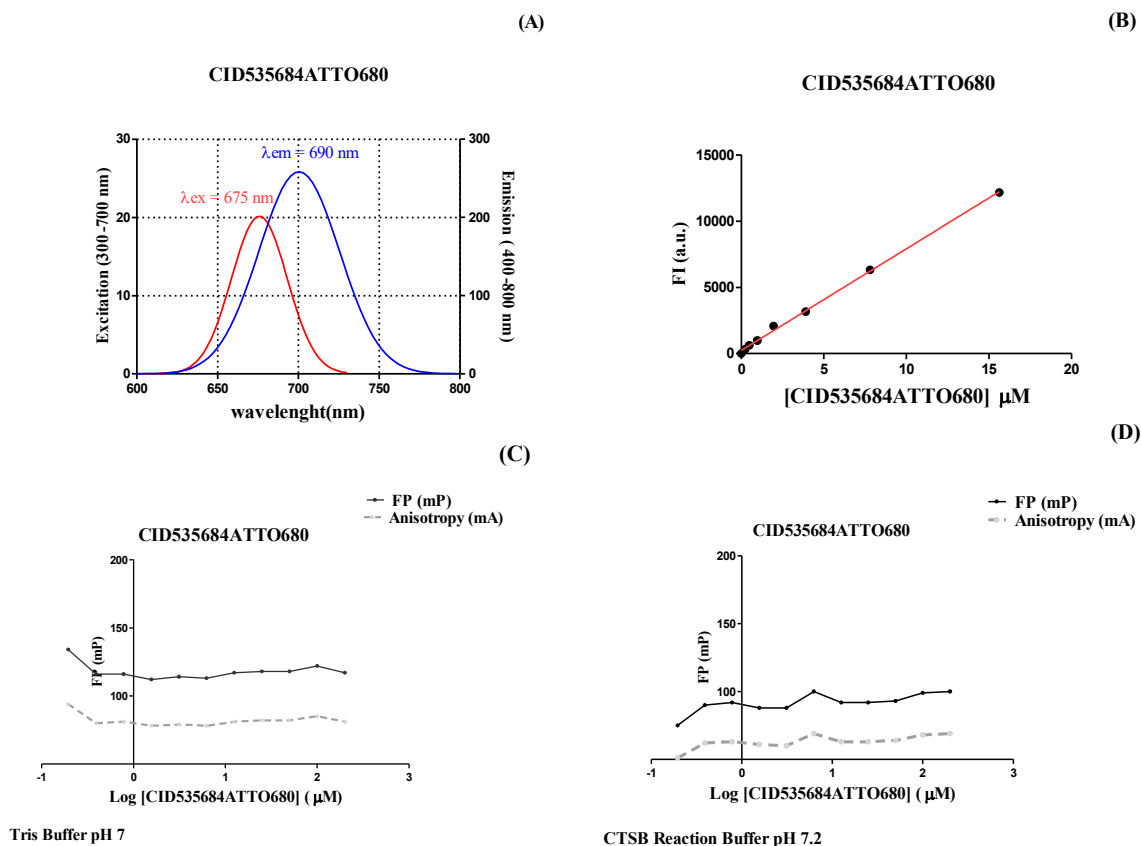
The conjugate CID8795ATTO680 (**Figure 29**) had an excitation maximum at 675 nm and an emission maximum at 705 nm (**Figure 29 A**) with emission increasing linearly with concentration from 0.195  $\mu\text{M}$  to 200  $\mu\text{M}$  (**Figure 29 B**). Conjugate CID535684ATTO680 (**Figure 30**) had a maximum excitation at 675 nm and emission at 690 nm (**Figure 30 A**); the fluorescence intensity rises linearly with the concentration from 0.195  $\mu\text{M}$  until 30  $\mu\text{M}$  (**Figure 30 B**). However, linearity was not maintained above 30  $\mu\text{M}$  of the conjugate CID535684ATTO680. it could be due to aggregation of the conjugate or reabsorption of emitted photons by a ground-state molecule that caused quenching.

Fluorescence polarization and anisotropy of both NIR probes were measured for a two-fold serial dilution with a starting concentration at 0.195  $\mu\text{M}$  and end concentration at 200  $\mu\text{M}$ . The objective was to identify a workable concentration range of the probes for further experiments. Two buffers, Tris pH 7 (**Figure 29 C**) and MES pH 4 (**Figure 29 D**) were selected for CID8795ATTO680 to observe the behavior of the probe at different pHs. For CID535684ATTO680; Tris pH 7 (**Figure 30 C**) and CTB reaction buffer pH 7.2 (**Figure 30 D**) were selected based on preliminary experiments which confirmed the better change in anisotropy for this probe in this range of pHs. Fluorescence polarization results of the probes did not vary significantly over the concentrations range tested at zero-time point; therefore, we chose 1, 10, and

25  $\mu\text{M}$  for further assays reported in **chapter 6**. Further, no change in polarization was observed at different pH tested (pH 4 and pH 7).



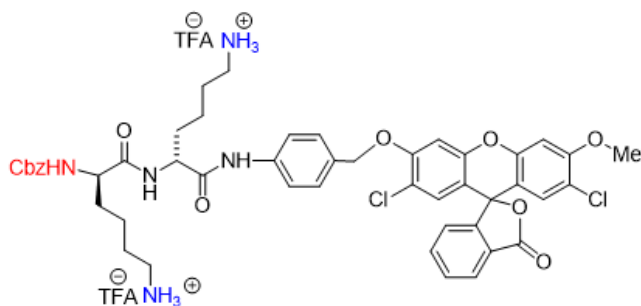
**Figure 29.** Fluorescence profile of CID8795ATTO680. A) Blank-discounted excitation and emission spectra of CID8795ATTO680 at 40  $\mu\text{M}$ , pH 7, showing maximum excitation and emission at 675 and 705 nm respectively. B) Fluorescence intensity measurement (at 0-time point) of different concentrations of CID8795ATTO680 (two-fold serial dilution with a starting concentration at 0.195 to 200  $\mu\text{M}$ ), which is increased linearly with the increase of the conjugate concentration. Linear regression was performed using Prism 5 software (Prism 2018, November 16) with the R-squared ( $R^2$ ) at 0.99. FP results presented in (C) and (D) refer to the log concentrations of CID8795ATTO680 at 60 minutes and pH 7 (C) and 4 (D) at the 0-time point. All assays were carried out at 37° temperature.



**Figure 30.** Fluorescence profile of CID535684ATTO680. A) Blank-discounted excitation and emission spectra of the CID535684ATTO680 (40  $\mu\text{M}$ ) at 675 and 690 nm respectively at pH 7. B) Fluorescence intensity measurement (at the 0-time point) of the different concentrations of CID535684ATTO680 (a two-fold serial dilution with a starting concentration at 0.195 to 200  $\mu\text{M}$ ), which is increased linearly to the increase of the conjugate concentration, the linear regression fit has been applied by using Prism 5 software (Prism 2018, November 16) with the  $R^2$  at 0.99, non-linearity was observed at concentrations above 30  $\mu\text{M}$ . FP results represented in (C) and (D) refer to the log concentrations of CID535684ATTO680 at pH 7 and 7.2 at the 0-time point, respectively. All assays were carried out at 37° temperature.

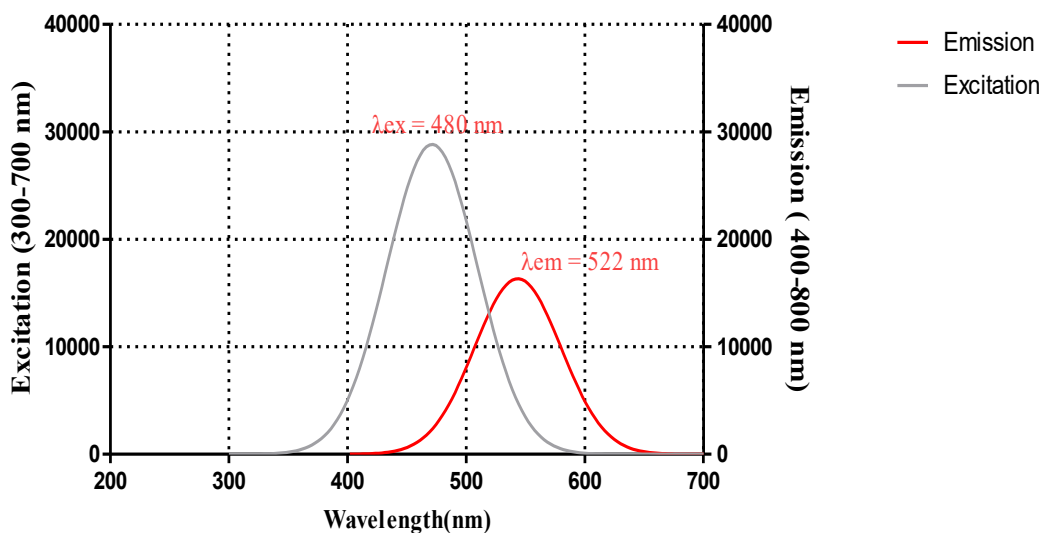
According to the excitation and emission spectra of both NIR conjugates, the relative intensity (arbitrary units (a.u.)) values were very low ranging (zero to 30 a.u.) for CID8795ATTO680 and (zero to 300 a.u.) for CID535684ATTO680 at 40  $\mu\text{M}$ , which could be a sign of self-quenching.

The chemical structure of Cbz-Lys-Lys-PABA-DCMF is presented in **Figure 31**. The spectrophotometric characterization of DCMF indicates the product of the enzymatic cleavage has maximum fluorescence excitation ( $\lambda_{ex} = 480$  nm) and emission ( $\lambda_{em} = 522$  nm) (**Figure 32**).



**Figure 31.** The chemical structure of the Cbz-Lys-Lys-PABA-DCMF.

2', 7'-Dichloro-6'-methoxy-fluorescein at  $1\mu\text{M}$



**Figure 32.** Fluorescence profile (blank-discounted excitation and emission) of DCMF at  $1\mu\text{M}$ , pH 5.5, and  $37^\circ\text{C}$ .

## 5.4. Summary and Conclusions

Three fluorescent probes (CID8795ATTO680, CID535684ATTO680, and Z-Lys-Lys-PABA-DCMF) were successfully synthesized. For the non-reactive probes, seven potential CTB ligands were subjected to conjugation chemistry with ATTO 680 NHS-ester. Two of them (CID8795ATTO680 and CID535684ATTO680) were successfully prepared and used in further binding experiments. Both non-reactive probes were fluorescent after conjugation with the ATTO dye, thus not requiring CTB for activation. A novel substrate-based probe was prepared using the same Z-Lys-Lys dipeptide scaffold as Dr. Phenix's group previously reported, but with DCMF as a fluorophore.

The spectrophotometry results of CID8795ATTO680 and CID535684ATTO680 indicate the excitation and emission wavelengths of 675 nm and 705 nm, 675 nm, and 690 nm, respectively, although the RFU values were too low and could be a proof of self-quenching for both probes. The excitation and emission wavelengths of DCMF is 480 nm and 522 nm, respectively. As expected, the spectrophotometric characterization of Z-Lys-Lys-PABA-DCMF indicated no fluorescence before enzyme cleavage.

## **Chapter 6: Evaluation of Binding Interaction of CID8795ATTO680 and CID535684ATTO680 to CTB**

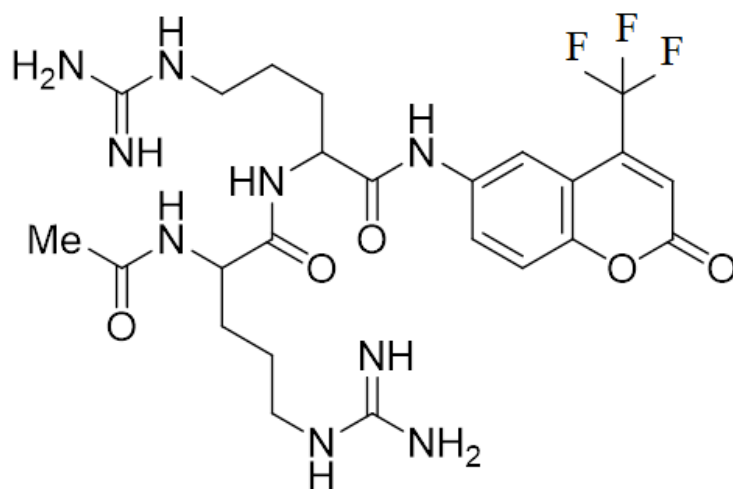
### **6.1. Introduction**

Recently, there has been tremendous progress in the design of NIR probes for imaging enzyme expression and activity in complex biological samples including cells and animals. We synthesized two novel NIR fluorescent probes for CTB: CID8795ATTO680 and CID535684ATTO680, which are fluorescent and intended to bind reversibly to CTB. In this chapter, we will evaluate the binding interactions of these non-reactive probes to CTB using competitive inhibition, fluorescence polarization, and fluorescence intensity measurements (**Appendix section 2**). In this study, anisotropy measured due to observing better prediction of the binding interaction between ATTO 680 conjugates to CTB by taking into account the total intensity of the samples.

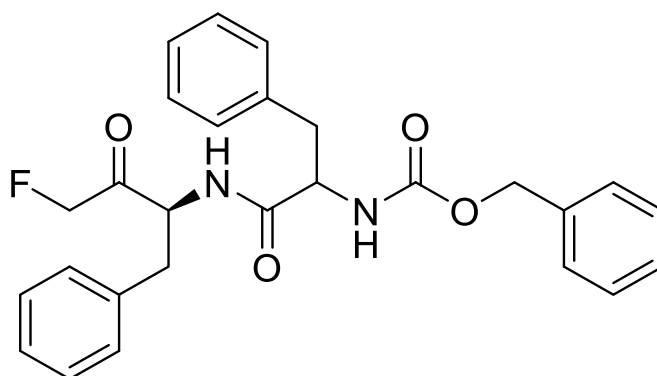
### **6.2. Procedures and Materials**

#### **6.2.1. Evaluating the Inhibition Activity of the Ligands on CTB**

Experiments were performed to determine if the ligands had inhibitory activity towards CTB using an inhibitor screening kit (Sigma Aldrich, Catalog number: MAK200). Each kit contained CTB reaction buffer at pH 7.2, CTB reagent, human recombinant cathepsin B, CTB substrate, Acetyl-Arg-Arg-7-Amino-4-trifluoromethyl coumarin (Ac-Arg-Arg-AFC) (**Figure 33**), and a known CTB irreversible inhibitor, Phe-Phe-Fluoromethyl ketone (F-F-FMK) (**Figure 34**).



**Figure 33.** The chemical structure of Acetyl-Arg-Arg-7-Amino-4-trifluoromethyl coumarin.



**Figure 34.** The chemical structure of Phe-Phe-Fluoromethyl ketone.

The assay concentrations of CTB (0.5  $\mu\text{M}$ ), peptide substrate (200  $\mu\text{M}$ ), and the control inhibitor F-F-FMK (10  $\mu\text{M}$ ) were used as recommended by the manufacturer. Ligands were tested for CTB inhibition at final ligand concentrations of 100, 10, and 1  $\mu\text{M}$ . Any inhibitory effect attributed to solvent (DMSO or ethanol) was accounted for by separate control reactions containing solvent instead of the test compound – with and without CTB. Reactions were carried out in the CTB reaction buffer at pH 7.2. The final reaction volumes were 100  $\mu\text{L}$ , replicates twice

and each assay was repeated twice. Fluorescence was measured using a Synergy 4™ microplate reader in kinetic mode for 60 minutes in 1-minute intervals, with excitation and emission wavelengths at 400 nm and 505 nm, respectively. All reactions proceeded at body temperature (37°C). The slope of the linear portion of the reaction progression curve (time versus fluorescence intensity) was used to calculate initial velocity (RFU/min) with and without the presence of the ligand at 100, 10, and 1 μM, as well as, the average of the last seven minutes of the reaction progression curves for evaluating the inhibition activity of the ligands on the total amount of AFC formation. All samples were blank (100 μl of the CTB reaction buffer) discounted. The AFC dye was not available to construct a standard curve. The statistical difference between samples (CTB and Ac-Arg-Arg-AFC with different concentrations of ligands or inhibitor) and baseline (CTB and Ac-Arg-Arg-AFC) was estimated by ANOVA with post-hoc Dunnett's multiple comparison test using Prism 5 software (Prism 2018, November 16). A  $p < 0.05$  was considered significant.

### **6.2.2. Evaluating the Inhibition Activity of the Two NIR Probes on CTB**

To evaluate the inhibitory activity of two new probes (CID8795ATTO680 and CID535684ATTO680) on CTB, the same inhibition procedure as for the ligands was repeated. All reagents used in these assays were at the same concentrations as those used for ligands inhibition assay (6.2.1 section). The slope of the linear portion of the progression curve was used to calculate initial velocity (RFU/min) in the absence and the presence of probe at 25, 10, and 1 μM. The average of the last seven minutes of the progression curve was also calculated and used to assess the effect of the non-reactive probes on the total amount of AFC formation. All the samples were blank (100 μl of the CTB reaction buffer) discounted. The statistical difference between samples (CTB, Ac-Arg-Arg-AFC, and different concentrations of non-reactive probes or inhibitor) and baseline (CTB and Ac-Arg-Arg-AFC) were measured by ANOVA with post-hoc Dunnett's



multiple comparison test using Prism 5 software (Prism 2018, November 16). A  $p < 0.05$  was considered significant.

To confirm inhibitory activity, a commercially available colorimetric CTB substrate, Z-Arg-Arg-pNA, was used in a similar inhibition screening assay to test CID8795ATTO680. A colorimetric substrate was chosen to minimize any potential interference from a fluorescent substrate with CID8795ATTO680. The second purpose of selecting this substrate was to test the probe at a lower pH. The Ac-RR-AFC substrate is more active at pH 7.2 than at pH 5.5 according to the manufacturer's instructions. However, a CTB probe should ideally work at the lower pH of the lysosome, for which the colorimetric substrate Z-Arg-Arg-pNA is adequate. Therefore, the colorimetric assay was performed at pH 5.5 in the acetated buffer corresponding to the pH of the biological system.

Absorbance at 410 nm was read on a Synergy 4™ plate reader (Biotek Inc) in kinetic mode every 1 minute for 60 minutes at 37°C. Seven concentrations of the serial dilution of pNA, 10-fold (1 to  $1 \times 10^{-6}$  mM) in assay buffer were conducted to construct a standard curve. The slope of the standard curve was used to evaluate the formation of pNA from the hydrolysis of Z-Arg-Arg-pNA under the presence of CTB. The colorimetric substrate was assayed at 200  $\mu$ M against 0.01  $\mu$ M CTB, in the absence and the presence of the CID8795ATTO680 at 3 different concentrations (10, 1, and 0.1  $\mu$ M). A known CTB irreversible inhibitor (F-F-FMK) was used as control at a fixed concentration of 10  $\mu$ M. All assays were conducted in duplicate and repeated once. All the samples were blank (100  $\mu$ l of the acetate buffer) discounted. The slope of the linear portion of the progression curve was used to calculate initial velocity ( $\mu$ M/min) with and without CID8795ATTO680, as well as the average of the last seven minutes of the progression curve for evaluating inhibition activity of the non-reactive probe on the total amount of pNA formation. The

statistical difference between samples (CTB, Z-Arg-Arg-pNA, and different concentrations of non-reactive probes or inhibitor) and baseline (CTB and Z-Arg-Arg-pNA) were measured by ANOVA post-hoc Dunnett's multiple comparison test, using Prism 5 software (Prism 2018, November 16). A  $p < 0.05$  was considered significant.

### 6.2.3. Fluorescence Polarization

In this study, anisotropy was used to verify the binding of the ATTO 680 conjugates to CTB. The objective of this experiment was to investigate the difference of the polarized light between ATTO 680 conjugates and when it is bound to CTB (**Appendix section 2**). The Gen5 program is as follows: the temperature at 37°C, filter sets as excitation at 620/40 nm, emission at 680/30 nm, and sensitivity was selected at 80. In all cases, black 96 well microplates were used (Costar catalog number 3915). Overall, anisotropy was measured at 0 and 60 minutes. After several preliminary assays (mostly used CTB human (catalog number: MAK200C) with the CTB reaction buffer (catalog number: MAK200A) pH 7.2 at 37°C)) for both NIR probes and considering the effect of all the reagents which have been used in these assays, the following experiments have been performed for both probes.

Subsequence to optimization results for anisotropy assays that have been done at pH 7.2, the binding confirmation assay for CID8795ATTO680, and CID535684ATTO680 was done with cathepsin B (Novoprotein, C398) in acetate at pH 5.5, corresponding to the pH of the biological system. Here, three CTB concentrations 100, 10, and 1 nM were used versus CID8795ATTO680 and CID535684ATTO680 at 1, 0.1 and, 0.01  $\mu$ M and anisotropy were read at zero and 60 minutes. The anisotropy calculated in Gen5 using **Equation 5 (Appendix section 2)**, where the G factor was set as 0.87 – according to Gen5 manufacture. The parallel and perpendicular intensity readings

have been corrected by blank subtraction (100  $\mu$ L acetate buffer at pH 5.5) to minimize the background contributions. These assays were performed in duplicate.

From the results of this section, we plotted the bar graphs to compare samples (probe/CTB) to the background (probe no protein) by using Prism 5 software (Prism 2018, November 16). To show preliminary data that suggest binding, bar graphs were utilized. If the preliminary results confirm binding, the next step is to build a protein concentration curve. The statistical difference between samples (CTB and non-reactive probes) and background (non-reactive probes no protein) was measured by ANOVA post-hoc Dunnett's multiple comparison test, using Prism 5 software (Prism 2018, November 16). A  $p < 0.05$  was considered significant.

#### **6.2.4. Fluorescence Polarization Protein Concentrations Curve**

A three-fold serial dilution with a starting CTB concentration at 0.25 to 180.68 nM was tested against 0.02  $\mu$ M and 0.05  $\mu$ M concentrations of CID8795ATTO680 in a fluorescence polarization assay. Assays were performed in triplicate at pH 5.5 and read at zero and 60 minutes. All the procedures are the same as mentioned in section **6.2.3**. In the following fluorescence polarization experiments, we used higher concentrations of the protein (0.42, 1.26, 3.79, 11.38, 34.13, 102.39, and 307.16 nM) versus a higher concentration of CID8795ATTO680 at 0.1  $\mu$ M. The reason for increasing the CTB concentrations is having more probe-protein complex in the solution; therefore, enhancing polarization. For all assays, the coefficient of variation (CV) (**equation 1**) was used to check the quality of replicates and eliminate replicates as needed to keep CV lower than 10%. Generally, %CV below 5 is recommended in medicine and pharmaceutical research. However in life sciences, such as biology, biochemistry, and biotechnology, up to 20% may be acceptable (Couto, Peternelli et al. 2013).

$$CV = \text{Standard Deviation}/\text{Mean} * 100$$

**Equation 1**

From these results, we would plot FP as a function of the protein concentrations at each fixed probe concentration. Generally, this plot has a sigmoidal shape where it begins at the polarized signal from the free ligand then reaches the plateau which is responsible for the maximum polarization corresponds to the complete binding of all fluorescent probes (Moerke 2009). In general, for building FP protein concentrations curves, the concentration of probe preferably should be low, as long as it reaches the detection limit of the instrument. In contrast, the protein concentration should be high enough to rise to the plateau of the curve while considering protein solubility limitation (Moerke 2009). The log CTB concentration versus fixed probe concentration was used for calculating EC<sub>50</sub> by using a non-linear approximation to fit the curve. An R-squared value above 0.99 considers a good fitting model. (Hamilton, Ghert et al. 2015).

### **6.2.5. Probe Aggregation**

Generally, aggregation-induced blue shift (AIBS) and aggregation-induced emission (AIE) are two common phenomena to prove of fluorescent probe aggregation. AIBS and AIE will occur when the rotation ability of the fluorescent probe restrains by a variety of factors such as; high viscosity of the buffer, different polarity preference of the fluorophore, temperature, and more. In general, fluorescence emission increase when intramolecular mobility restricted while shifting of spectra induced by changing in fluorescent probe environment or restriction of the twisted intermolecular charge transfer (TICT). Soon after the formation of aggregates, the solvent moves out of the aggregates; therefore, the fluorescent probe is facing different environments. Suppression of TICT state induced by aggregation leads to enhanced fluorescence and blue shift (Zhu, Yang et al. 2008). In this study, in the same fluorescence polarization experiments, change or shift in maximum fluorescence emission was evaluated.

Besides, aggregates of the conjugated fluorescent substance divided to two categories of electronic interactions: 1) H-aggregates or H-bond refers to side-by-side interaction manner of the chromophores with its closest-neighbor which cause displacement of the spectrum at higher energy and lower wavelength, 2) J-aggregates or J-bond occurs by head-to-tail interaction of the chromophores with its nearest-neighbor which cause displacement of the spectrum at lower energy and higher wavelength (Spano and Silva 2014). In this study, the absorbance of the CID8795ATTO680 at 40  $\mu\text{M}$  with the CTB reaction buffer (catalog number: MAK200A) at pH 7.2 and 37°C was measured. The Gen5 program is as follows: the temperature at 37°C, the wavelengths set at 300 to 700 nm. Clear 96-well microplates were used (Costar catalog number 3370).

#### **6.2.6. Selectivity Assay for ATTO 680 Conjugates**

To confirm the selective binding of the CID8795ATTO680 toward cathepsin B against other proteins, an FP assay was performed. Bovine Serum Albumin (BSA) was subjected to FP assay using the NIR probe CID8795ATTO680 at 0.02  $\mu\text{M}$ . BSA was chosen since it is pure, inexpensive, colorless, stable, and can interact with various small molecules easily (Phan, Bartelt-Hunt et al. 2015). All assays were performed in triplicate, in acetate buffer (pH 5.5) for CTB and phosphate-buffered saline (PBS) (pH 7.4) for BSA. Anisotropy was measured at 37°C after 60 minutes based on the results of assays optimization. A two-fold serial dilution of BSA with a starting concentration at 0.25 nM to 180.45 nM and a two-fold serial dilution with a starting concentration at 0.25 nM to 180.68 nM for the CTB was used against a fixed concentration of CID8795ATTO680 at 0.02  $\mu\text{M}$ .

From the result of this section, we plotted bar graphs to compare the difference in anisotropy of the probe in the presence of CTB and BSA by using Prism 5 software (Prism 2018, November

16). The statistical difference between samples (CTB and non-reactive probes or BSA and non-reactive probes) and background (non-reactive probes no protein) was measured by ANOVA post-hoc Dunnett's multiple comparison test using Prism 5 software (Prism 2018, November 16). A  $p < 0.05$  was considered significant. All the data in this section was normalized by dividing the anisotropy value of the samples by the background and expressed as a percentage to have an equal baseline for comparison.

### **6.2.7. Cell Lysate Assay**

The cell cultures were prepared by Sarah Niccoli, a member of Dr. Simon Lees' team. *MDA-MB-231* (breast cancer cells) and *H9C2* rat Cardio-myocytes (normal tissue cells) were cultured in Hyclone Dulbecco's Modified Eagle Medium DMEM/high-glucose media containing either 10% fetal clone III, 1% penicillin/streptomycin, 0.1% gentamicin and 1% sodium pyruvate, or 10% fetal bovine serum and 5% antibiotics-antimycotics. Cells were grown in T75 flasks at 37°C, 5% CO<sub>2</sub>.

Adhered cells were washed once with PBS, trypsinized, then centrifuged to obtain a cell pellet. The pellet was then washed twice with PBS and re-suspended in 1 ml acetate buffer (with or without triton 1%). The solution was lysed using tissuelyser beads at 20 Hz for 2 sets of two minutes. The resulting solution was centrifuged (4°C, 10 min, 16000\*g), and the supernatant moved to new tubes. Fifty µl was set aside for the protein concentration assay while the remaining lysate was used for the activity assay. A Pierce BCA protein assay (cat. # 23227) was performed as per manufacturer's instructions to measure total protein concentration.

In the case of probe CID8795ATTO680, we wanted to compare the difference between anisotropy of the normal and cancer cells. We expected to see increased anisotropy in cancer cells due to the overexpression of CTB compared to a normal cell line (Bian, Mongrain et al. 2016). The assays were performed in triplicate with 1:2 serial dilutions of both cancer and normal lysate versus fixed concentration of the CID8795ATTO680 at 5  $\mu$ M, 37°C, pH 5.5 (acetate buffer). The baseline for the assay, which contained the probe and lysate buffer, was performed in duplicate. Fluorescence polarization was read on a Synergy 4™ plate reader (Biotek Inc) with settings as follows: delay of 15 minutes, followed by the addition of probe, and read for five hours with 30 minutes' interval. The filter sets as excitation at 620/40 nm, emission at 680/30 nm and, sensitivity was selected at 80. The statistical difference between samples (lysate and non-reactive probes) and background (non-reactive probes no lysate) was measure by ANOVA post-hoc Dunnett's multiple comparison test using Prism 5 software (Prism 2018, November 16). A  $p < 0.05$  was considered significant. Here we required higher probe concentration to observe changing in anisotropy due to having a complex nature of the cell environment. Therefore, to overcome the probe aggregation limitation, the probe was added within the last minute of running the experiment.

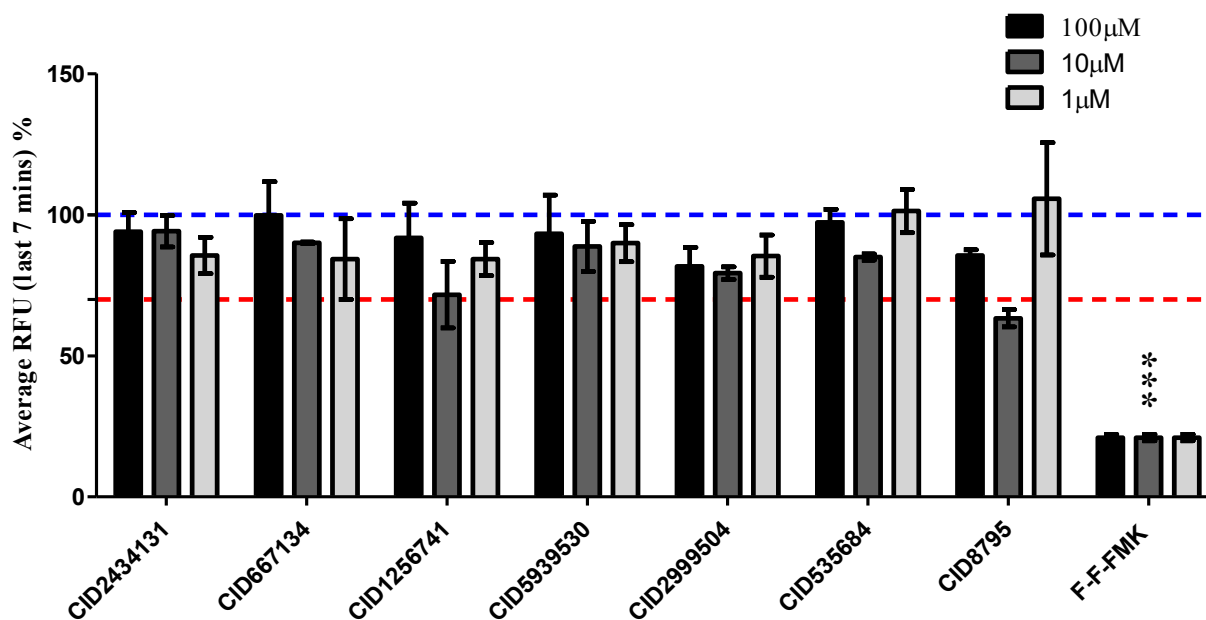
To quantify the protein concentration of the lysate, the bicinchoninic acid (BCA) protein assay kit was used and BSA with serial dilution (2, 1, 0.5, 0.25, and 0.125 mg/ml) was used as a standard curve. The BCA assay is based on colorimetric detection – depending on the conversion of  $\text{Cu}^{2+}$  to  $\text{Cu}^+$  under alkaline conditions.

### **6.3. Result for Two Novels NIR Non-Reactive Probes**

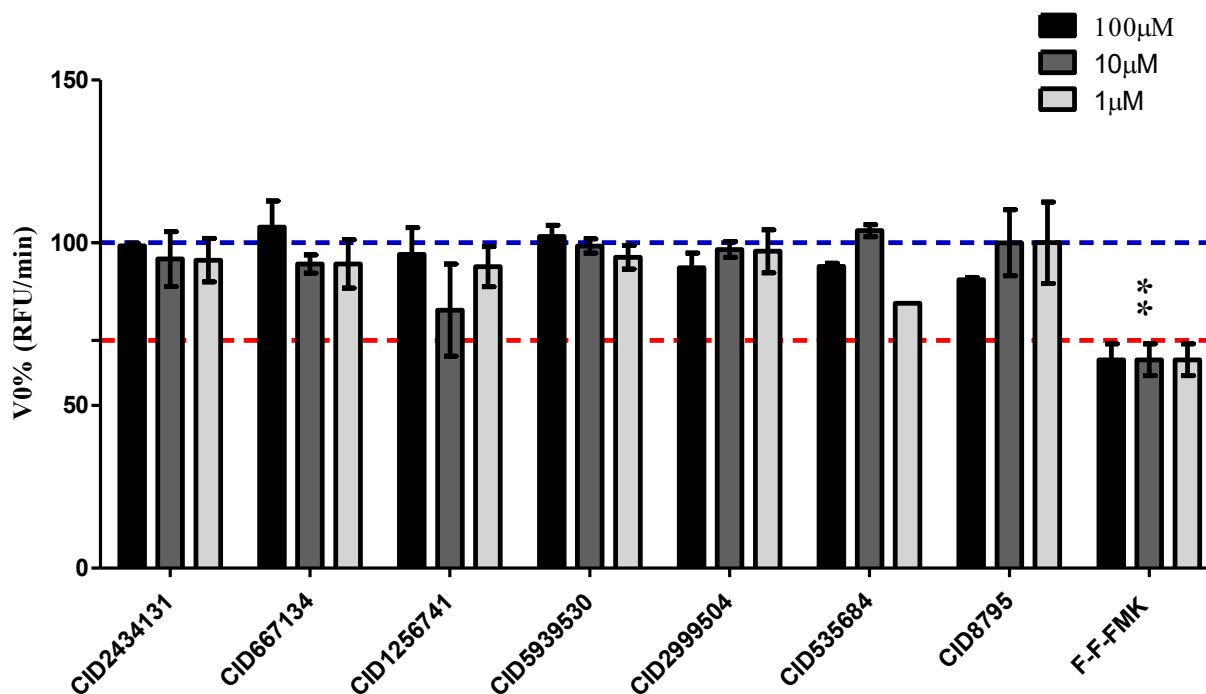
#### **6.3.1. Inhibition Assay for Seven Ligands**

To determine if the computationally discovered ligands have any inhibitory activity towards CTB, an assay was designed where all the compounds were incubated with recombinant CTB in the presence of the ligands. The effect of all seven ligands on AC-R-R-AFC activity was calculated in two ways – based on the average of the fluorescent reading of the last seven minutes which illustrates the highest signal from releasing AFC (**Figure 35**) and calculation of a slope (first few minutes of the linear portion of AFC formation with  $R^2$  value above 0.9) that represented the velocity of the enzymatic reaction (**Figure 36**).





**Figure 35.** Inhibitor screening results in the presence of ligands in three concentrations according to the average of the last seven minutes of the progression curve for the AC-R-R-AFC substrate. F-F-FMK is a known CTB inhibitor and was used as the control. The dotted line marks the 70% activity threshold and maximum activity marked as 100%. Ligands were tested at 100  $\mu$ M (black), 10  $\mu$ M (gray), and 1  $\mu$ M (light gray) in duplicates. The experiment was performed at pH 7, and a temperature of 37°C. The concentration of the substrate, F-F-FMK, and CTB was fixed at 200  $\mu$ M, 10  $\mu$ M, and 0.5  $\mu$ M, respectively. Statistical differences in inhibition activity were determined by ANOVA with post-hoc Dunnett's multiple comparison test using Prism 5 software (Prism 2018, November 16);  $p < 0.05$  was considered significant and presented by \*\*\*.



**Figure 36.** Inhibitor screening results for ligands according to the initial velocity for the first few minutes of the progression curve for the AC-R-R-AFC substrate. F-F-FMK is a known CTB inhibitor and was used as a control. The dotted line marks the 70% activity threshold and maximum activity marked as 100%. Ligands were tested at 100  $\mu$ M (black), 10  $\mu$ M (gray), and 1  $\mu$ M (light gray) in duplicates. The experiment was performed at pH 7, and a temperature of 37°C. The concentration of the substrate and, F-F-FMK, and CTB were fixed at 200  $\mu$ M, 10  $\mu$ M, and 0.5  $\mu$ M, respectively. Statistical differences in inhibition activity were determined by ANOVA post-hoc Dunnett's multiple comparison test using Prism 5 software (Prism 2018, November 16);  $p < 0.05$  was considered significant and presented by \*\*.

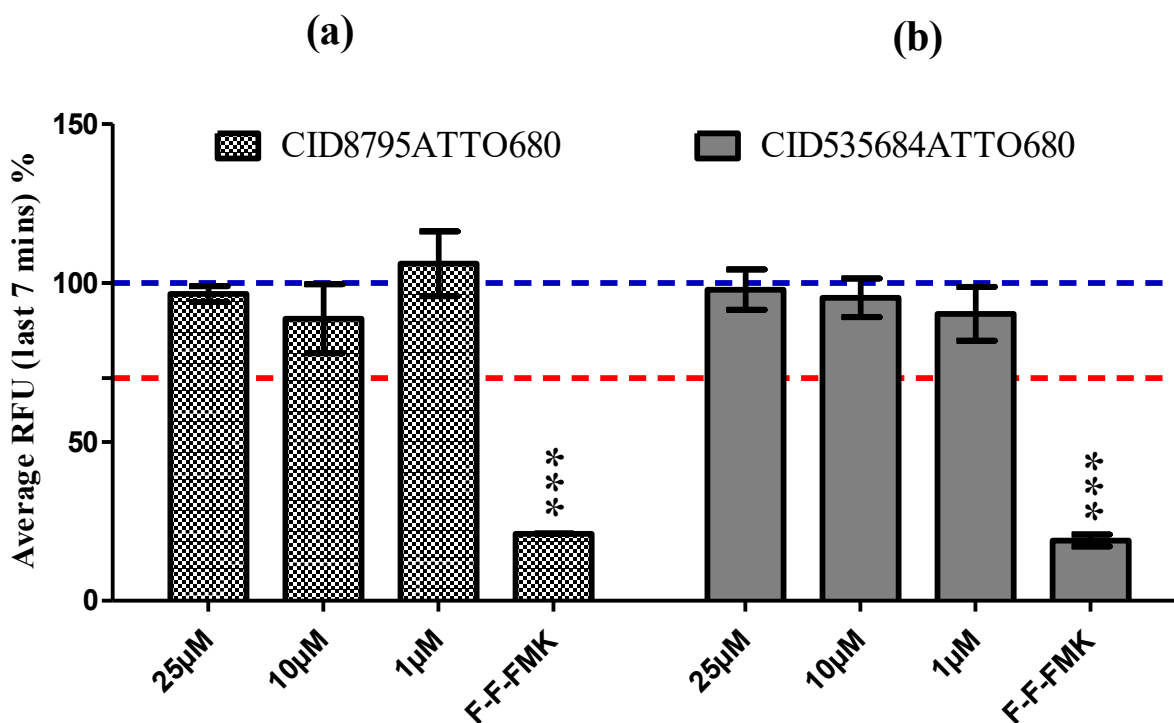
The aim of using these methods for activity percent calculation is as follows: the last 7 minutes represents the total cleavage amount of 7- Amino-4-trifluoromethyl coumarin (AFC)-based substrate to yield AFC dye, which is fluorescent. Using the initial velocity for activity calculation indicates how fast the substrate cleaved to produce the AFC. Based on these experiments, CTB activity was not strongly inhibited by any of the parent ligands screened in the assay at 1, 10, and 100  $\mu$ M. All replicates maintained CTB activity higher than 70%. However, for some graphical signals, some experimental errors resulted in wide inferential bars such as CID8795 and CID1256741. Free CTB and F-F-FMK (known CTB inhibitor) were used as controls, and as expected, F-F-FMK inhibits the enzyme over 70% according to both of two

analyses. Based on this result, there was no inhibition activity of the free ligands against CTB compared to the known inhibitor F-F-FMK. As it is evident from the result, the known inhibitor shows the maximum inhibition at the last few minutes of the assay compared to the beginning of the kinetics.

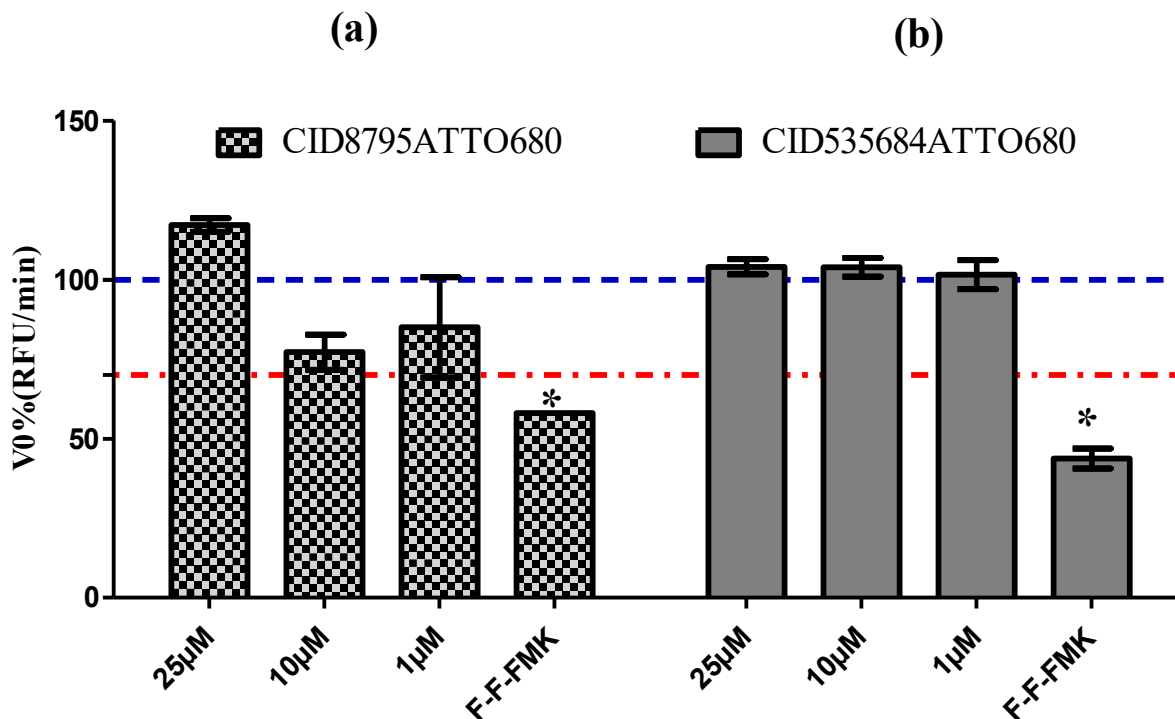
### **6.3.2. Inhibition Assay for CID8795ATTO680 and CID535684ATTO680**

Two fluorescent probe candidates (CID8795ATTO680 and CID535684ATTO680) were subjected to the inhibition screening experiment at 25, 10, and 1  $\mu\text{M}$  with Ac-RR-AFC substrate (**Figure 37** and **Figure 38**). The test was performed duplicate for both fluorescent probes and the inhibition percentage calculated in two ways; the same techniques were used for the parent ligands as described in section **6.3.1**. Inhibition assay for both NIR probes against CTB was performed under specific conditions as recommended in the manufacturer's instructions of the cathepsin B kit. The CID8795ATTO680 and CID535684ATTO680 designed to bind to CTB without inhibition effect on the enzyme. According to the results which are presented in **Figure 37** a and b, both NIR probes did not show any inhibition activity against CTB more than 30%. This inhibition percentage was assessed based on the average of the last seven minutes of the total AFC formation. The result also was calculated according to the initial velocity for the first few minutes with the  $R^2$  value above 0.99 (**Figure 38** a). Based on this result, the presence of the CID8795ATTO680 at 25  $\mu\text{M}$  increased the activity of the enzyme to induce the rate of formation of AFC compared to the background for the first few minutes. In contrast, the presence of 10  $\mu\text{M}$  of the CID8795ATTO680 for the first few minutes shows approximately 15% inhibition and at 1  $\mu\text{M}$  shows approximately 20% inhibition, although none of these changes was statistically significant by ANOVA post-hoc Dunnett's multiple comparison test. This result from the beginning of the assay could be the effect of changing the environment compared to the background. As it is shown in (**Figure 38** b) the

CID535684ATTO680 does not have any effect on the initial velocity for commercial substrate (AC-RR-AFC); therefore, the binding affinity of this fluorescent-probe towards CTB was not enough to inhibit the protein activity. The effect of F-F-FMK as a positive control for inducing inhibitory activity toward CTB was statistically different compare to free CTB in both methods of analysis.



**Figure 37.** Inhibitor screening results for ATTO 680 conjugates; CID8795ATTO680 (a) and CID535684ATTO680 (b) at 25, 10, and 1  $\mu\text{M}$  according to the average of last seven minutes of the substrate (Ac-RR-AFC) progression curve. F-F-FMK is a known CTB inhibitor and was used as a control. The dotted line marks the 70% activity threshold and maximum activity marked as 100%. The experiment runs at 37°C and pH 7. The concentration of the substrate (Ac-RR-AFC) was 200  $\mu\text{M}$ , and F-F-FMK was 10  $\mu\text{M}$ . The CTB concentration was 0.5  $\mu\text{M}$ . Statistical differences in inhibition activity of samples (3 concentrations of NIR probe or F-F-FMK with a fixed concentration of substrate and CTB) compare to the maximum activity (only substrate and CTB) were determined by ANOVA post-hoc Dunnett's multiple comparison test, using Prism 5 software (Prism 2018, November 16);  $p < 0.05$  was considered significant and presented by \*\*\*.

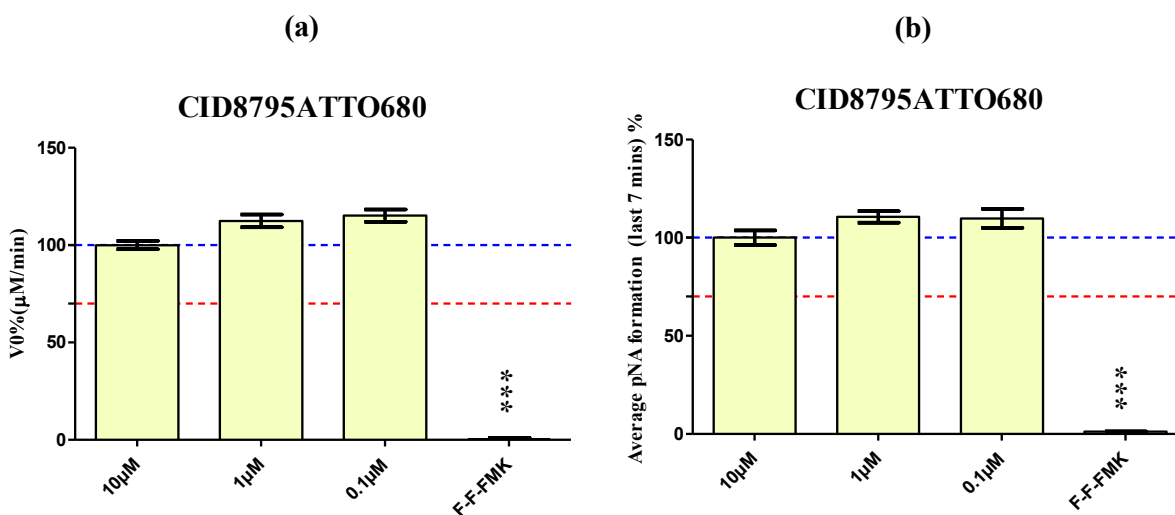


**Figure 38.** Inhibitor screening results for ATTO 680 conjugates; CID8795ATTO680 (a) and CID535684ATTO680 (b) according to the initial velocity for the first few minutes of the Ac-RR-AFC activity toward CTB. F-F-FMK is a known CTB inhibitor and was used as a control. The dotted line marks the 70% activity threshold and maximum activity marked as 100%. The experiment runs at 37°C and pH 7. The concentration of the substrate was 200 µM, and F-F-FMK was 10 µM. The CTB concentration was 0.5 µM. Statistical differences in inhibition activity of samples (3 concentrations of NIR probe or F-F-FMK with a fixed concentration of substrate) compare to the maximum activity (only substrate and CTB) were determined by ANOVA post-hoc Dunnett's multiple comparison test, using Prism 5 software (Prism 2018, November 16);  $p < 0.05$  was considered significant and presented by \*.

In inhibitor screening results **Figure 37** and **Figure 38**, we did not control the potential interference of fluorescence from CID8795ATTO680 and CID535684ATTO680 to assess the effect of both NIR probes on AFC formation either on initial velocity or on the total formation of the AFC. Therefore, the colorimetric substrate has been used for having more reliable results.

Inhibitor screening results for CID8795ATTO680 using a colorimetric substrate (Z-Arg-Arg-pNA) are presented in **Figure 39**. This result shows the effect of the NIR probe on the formation of the product (pNA) by using the average of the last seven minutes and the initial

velocity of the reading. According to this result, no significant inhibition activity of any NIR probe concentrations observed compared to the baseline (free enzyme). This result would be more reliable due to minimizing the interference of any fluorescent wavelength of the commercial substrate on CID8795ATTO680. As expected, the presence of F-F-FMK shows the highest percentage of inhibition on pNA formation.



**Figure 39.** Inhibitor screening results for ATTO 680 conjugates CID8795ATTO680 at 10, 1, and 0.1  $\mu\text{M}$ . a) Inhibitor screening results based on the initial velocity of the pNA formation at first ten minutes b) The average of the last seven minutes of the substrate (Z-Arg-Arg-pNA) progression curve (pNA formation). F-F-FMK is a known CTB inhibitor and was used as a control. The red dotted line marks the 70% activity threshold and maximum activity marked as 100%. The experiment runs at 37°C and pH 5.5. The concentration of the substrate (Z-Arg-Arg-pNA) was 200  $\mu\text{M}$ , and F-F-FMK was 10  $\mu\text{M}$ . The CTB concentration was 0.01  $\mu\text{M}$ . Statistical differences in inhibition activity of samples (3 concentrations of NIR probe or F-F-FMK with a fixed concentration of substrate) compare to the maximum activity (only substrate and CTB) were determined by ANOVA post-hoc Dunnett's multiple comparison test, using Prism 5 software (Prism 2018, November 16);  $p < 0.05$  was considered significant and presented by \*\*\*.

### 6.3.3. Fluorescence Polarization Results

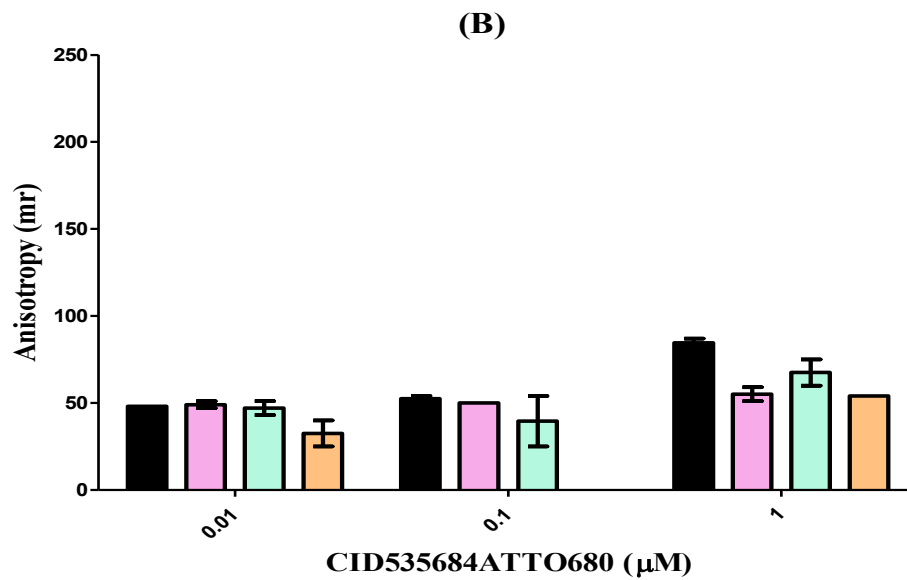
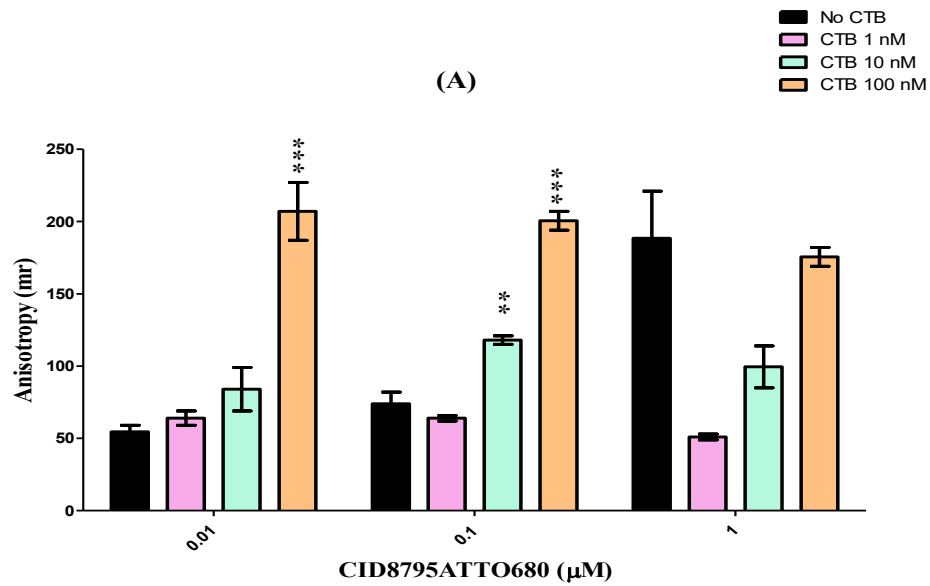
The result (**Figure 40 A and B**) shows changes in anisotropy induced by both NIR probes in the presence or absence of CTB at 100, 10, and 1 nM. According to **Figure 40 A**, the

CID8795ATTO680 at two lower concentrations (0.01  $\mu\text{M}$  and 0.1  $\mu\text{M}$ ) show a statistical increase in anisotropy with the presence of the highest concentration of CTB (at 100 nM). The statistical increase in anisotropy also observed at 0.1  $\mu\text{M}$  concentration of this probe against 10 nM of the CTB comply with the fluorescence polarization principle (**Appendix, section 2**). The rotational speed of the fluorophore molecule will affect directly the anisotropy value through dissociation/breakdown or association/binding process; therefore, at higher CTB concentration the presence of CID8795ATTO680/CTB complex induced increasing in anisotropy and this would be proof for binding interaction between CID8795ATTO680 and the target protein. The highest probe concentration at 1  $\mu\text{M}$  indicates a high level of background compare to samples (probe with different concentrations of CTB). As shown in **Figure 40 A**, increasing the probe concentration resulted in a high background (free labeled-ligand); this may confirm that the CID8795ATTO680 potentially prone to aggregation: first, the excessive concentration of the probe, which is more soluble in DMSO, leads to low solubility in acetate; second, comparing the anisotropy results of the free labeled-ligand at zero and 60 minutes indicates an increase in anisotropy over the time which is consistent with probe aggregation. For all the results, Raman scatters effect was considered by discounting the blank (only buffer). In all the results represented in **Figure 40 A**, the anisotropy at the same protein concentration was statistically the same for all probe concentrations. This observation confirmed the lower concentration of the NIR probe leads to the same anisotropy signal as the higher concentration. Although, at higher probe concentrations we observed increased in anisotropy in the background. This observation leads us to use CID8795ATTO680 at concentrations ranging from 0.01 to 0.1  $\mu\text{M}$  to build the protein concentration curves for monitoring the behavior of this probe in the presence of different concentrations of protein.

Unfortunately, from the spectrophotometric characterization results (**Figure 29**) which we selected the probe concentrations for inhibition assay, we were unable to realize the probe aggregation because the FP was only measured at zero time. Although from the emission spectra, the excessive amount of CID8795ATTO680 (40  $\mu$ M) caused self-quenching and resulted in low fluorescence intensity at 20 (a.u.) with the sensitivity of 90.

None of the preliminary FP experiments for CID535684ATTO680 lead to any changes between background (probe no protein) and samples (probe with different concentrations of protein); although, this probe did indicate some anisotropy enhancing in CTB reaction buffer (MAK200A) at pH 7.2. Since the result was not consistent with another buffer at the same pH, we aborted a further investigation for CID535684ATTO680. According to all experiments which have been done for CID535684ATTO680, there was not sufficient evidence to confirm the binding interaction of this conjugated probe towards CTB at pH 5.5 (**Figure 40 B**). Since this probe should be functional at low pH of the lysosome (CTB localization), with further analyses, the focus was on CID8795ATTO680.

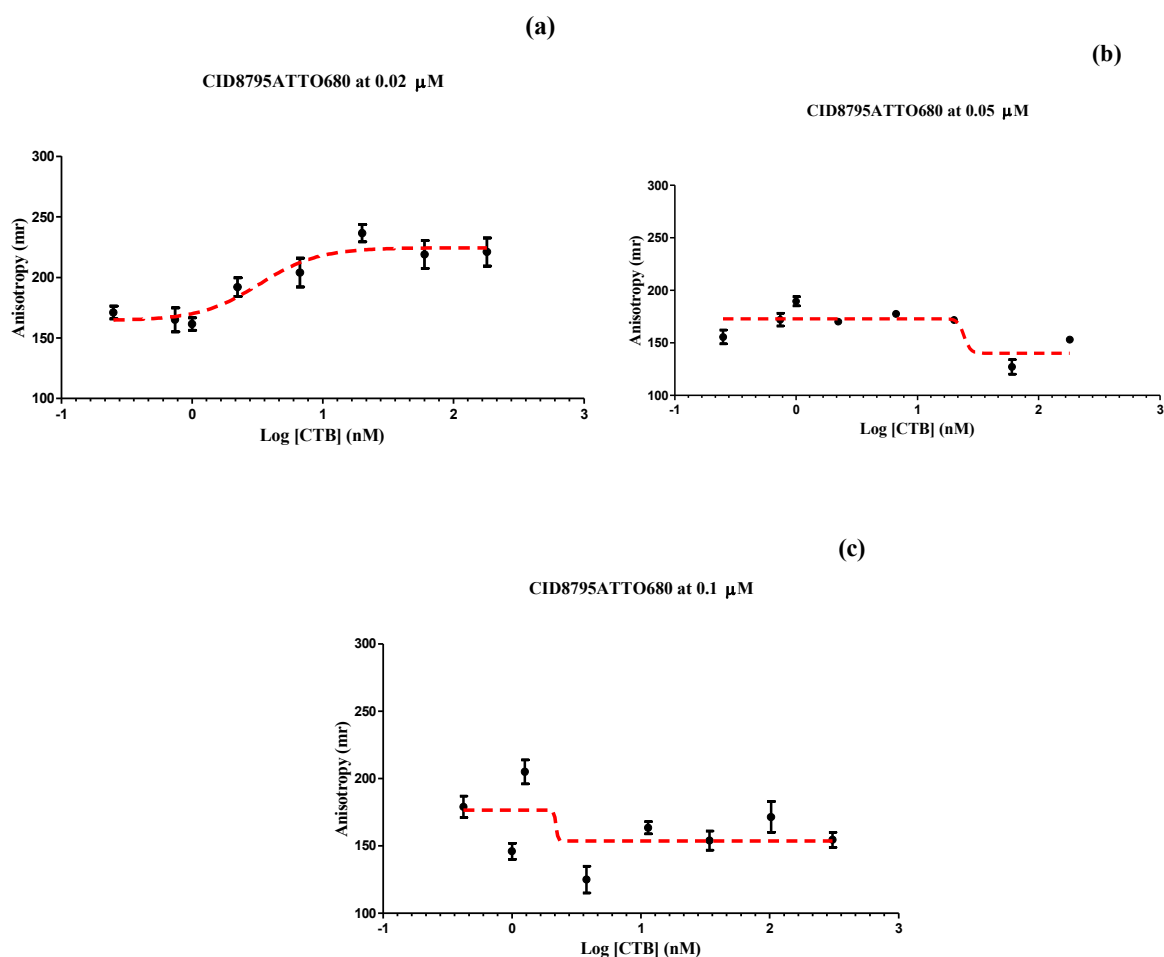




**Figure 40.** Change in anisotropy in the presence and the absence of CTB (100, 10 and 1 nM) for CID8795ATTO680 (A) and CID535684ATTO680 (B) at 1, 0.1 and 0.01 µM. Assays were performed in duplicate at pH 5.5 and read at 60 minutes which were incubated for 5 minutes at 37°C. Statistical differences in anisotropy were determined by ANOVA post-hoc Dunnett's multiple comparison test, using Prism 5 software (Prism 2018, November 16);  $p < 0.05$  was considered significant and presented by \*\*\* and \*\*.

#### 6.3.4. Fluorescence Polarization Protein Concentrations Curve

Two experiments were performed to measure anisotropy as a function of the log of CTB concentrations (three-fold serial dilution starting CTB concentration at 0.25 to 180.68 nM against 0.02  $\mu\text{M}$  and 0.05  $\mu\text{M}$  concentrations of CID8795ATTO680) (**Figure 41** (a and b)). As a final experiment, a higher concentration of the probe at 0.1  $\mu\text{M}$  and CTB (three-fold, serial dilution starting 0.42 nM to 307.16 nM) was used to indicate a better change in anisotropy compared to the background based on the preliminary results (**6.3.3**) (**Figure 41** c).



**Figure 41.** Anisotropy as a function of the log CTB concentrations started from 0.25 nM to 180.68 nM against CID8795ATTO680 at 0.02  $\mu\text{M}$  (a) and 0.05  $\mu\text{M}$  (b), and the log CTB concentrations started at 0.42 nM to 307.16 nM against CID8795ATTO680 at 0.1  $\mu\text{M}$  (c). The non-linear approximation curve fitting applied for measuring the  $\text{EC}_{50}$  of the CID8795ATTO680 by using Prism 5 software (Prism 2018, November 16). Assays were performed in triplicate at pH 5.5 and read at 60 minutes, which were incubated for 5 minutes at 37°C. The average of all results ( $\text{CV} < 10\%$ ) shows as points. The dotted-line in each dataset corresponding to the binding fit model.

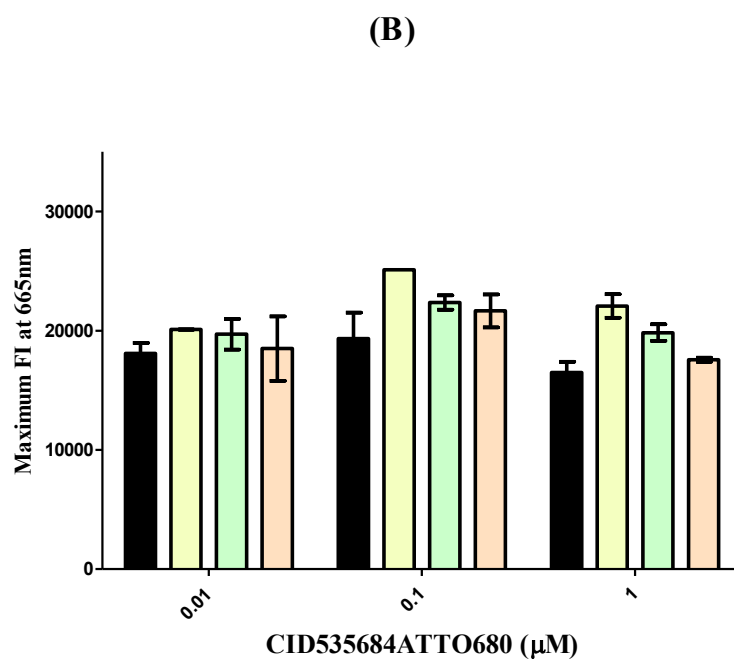
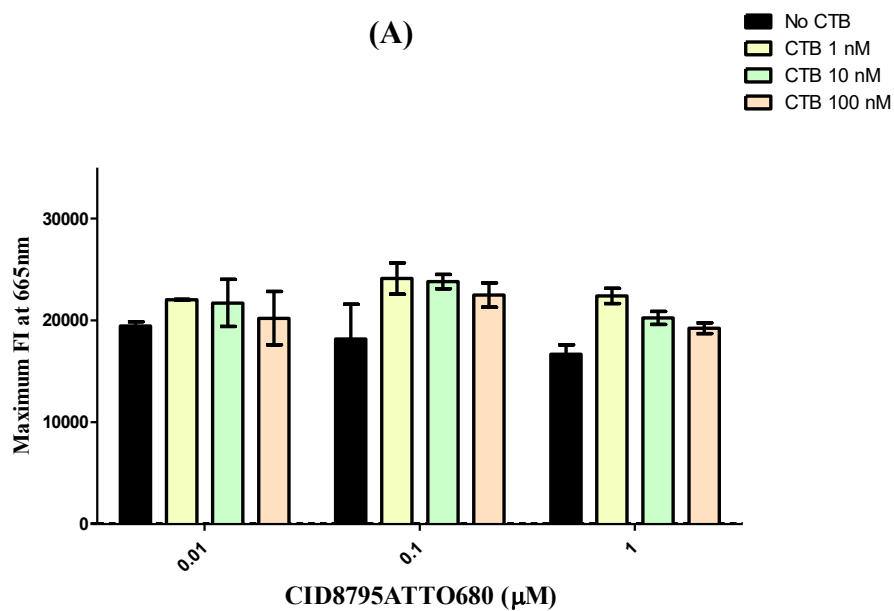
The anisotropy results as a function of log concentrations of CTB for CID8795ATTO680 at 0.02  $\mu\text{M}$  indicates  $\text{EC}_{50}$  at 3.27 nM with an  $R^2$  at 0.91 **Figure 41** (a). The  $\text{EC}_{50}$  value driven from this result is unreliable due to poor  $R^2$  value. Indeed, the increase in anisotropy in the presence of the NIR probe at 0.02  $\mu\text{M}$  with different concentrations of CTB was only low enough to give

us the sigmoidal shape of the protein curve. The anisotropy difference between free probe and samples which contains different concentrations of the protein was statistically significant by ANOVA post-hoc Dunnett's multiple comparison test (a  $p < 0.05$  was considered significant) for the three highest concentrations of the CTB (180.68, 60.22, and 20.07). However, at the two lowest CTB concentrations, no increase in anisotropy was observed as compared to the background. This result was consistent with preliminary results which increasing in anisotropy was observed at 100 nM of the CTB against 0.1 and 0.01  $\mu\text{M}$  of the CID8795ATTO680.

According to the preliminary results, the probe concentrations selected between 0.01 and 0.1  $\mu\text{M}$  to avoid probe aggregation. According to (**Figure 41** (b and c)) which is indicated changing in anisotropy of the CID8795ATTO680 at 0.05  $\mu\text{M}$  and 0.1  $\mu\text{M}$  against serial dilution of the CTB, we were not able to report the  $\text{EC}_{50}$  values due to the poor  $R^2$  (lower than 0.5). The probe at its higher concentrations resulted in a higher background in the last two experiments. As mentioned in section **6.3.3**, this probe susceptible to binding interaction to another probe like itself at high concentrations above 0.1  $\mu\text{M}$  and over time (0 to 60 minutes). For experiments a, b, and c **Figure 41**, the probe concentration was below the aggregation limit; however, preparing these experiments took at least one hour before running the assay compared to the single assay reported in section **6.3.3**. Therefore, probe prone to aggregate over time and disrupt the interpretation of the anisotropy results. Indeed, the difference in anisotropy between free probe (0.01 and 0.1  $\mu\text{M}$ ) and probe with protein (100 nM) in preliminary results (**Figure 40 A**) was almost four times higher (assay preparation was less than 15 minutes), but this difference in FP protein concentrations curve assay (**Figure 41 A**) between free probe at 0.02  $\mu\text{M}$  and probe with CTB at 180.68 nM was only two times higher (assay prepare almost in an hour before the run). This would be another interpretation for CID8795ATTO680 aggregation over time.

### 6.3.5. Probe Aggregation Results

In the same preliminary FP experiment maximum fluorescence intensity of the samples (three concentrations of the CID8795ATTO680 or CID535684ATTO680 versus three concentrations of the CTB) and background (three concentrations of the CID8795ATTO680 or CID535684ATTO680 in the acetate buffer) was measured (**Figure 42**).



**Figure 42.** Change in maximum fluorescence intensity at 665 nm in the presence and the absence of CTB (100, 10 and 1 nM) for CID8795ATTO680 (A) and CID535684ATTO680 (B) at 1, 0.1 and 0.01  $\mu\text{M}$ . Assays were performed in duplicate at pH 5.5 and read at 60 minutes which were incubated for 5 minutes at 37°C. The backgrounds correspond to free probes in acetate buffer which are represented as black columns.

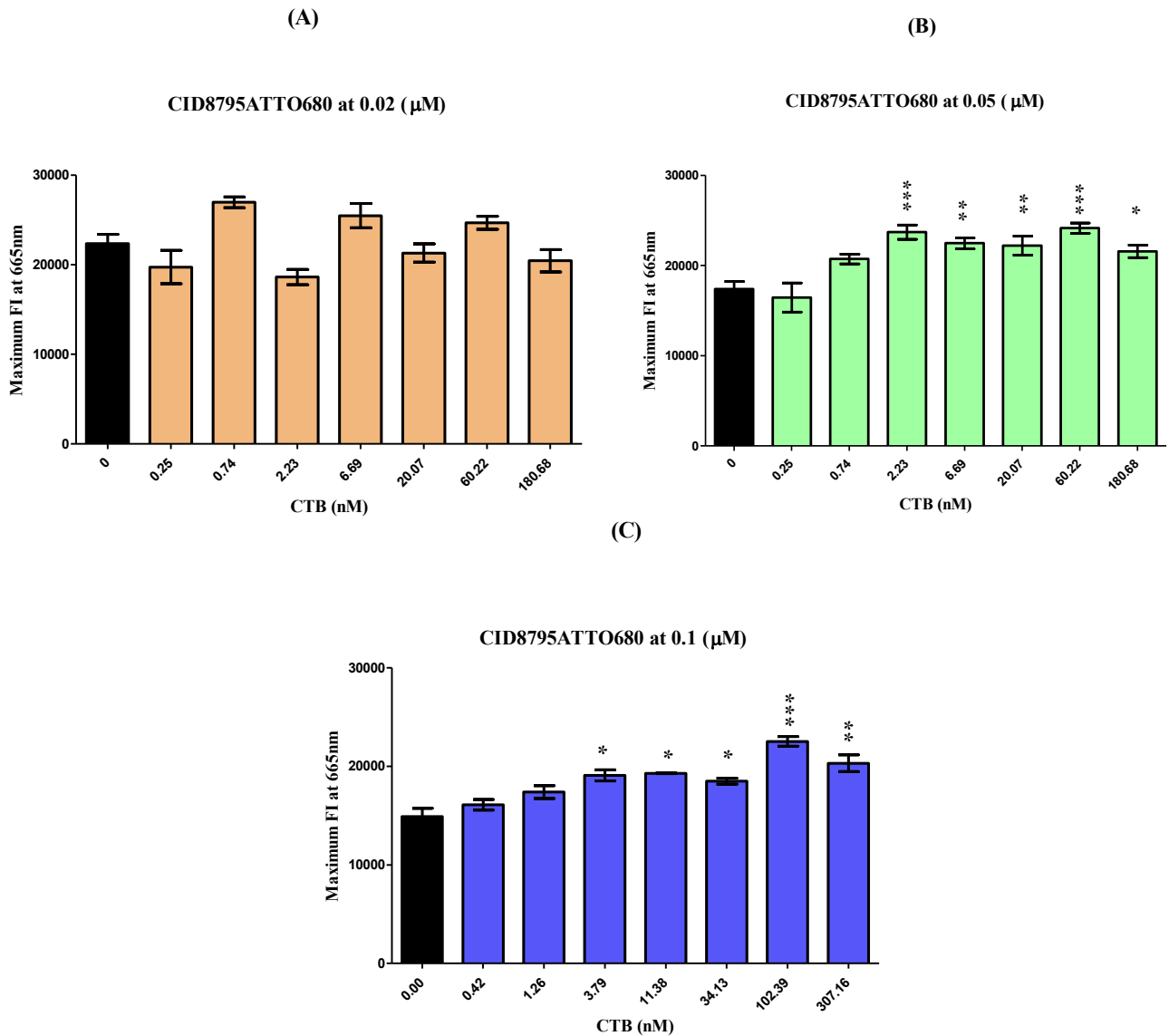
According to this result, the maximum fluorescence intensity for all samples and background was at 665 nm. For all three concentrations of both probes with protein, the increase in intensity observed compared to the background (free probes), although this was not statistically different from than background. However, at higher probes concentration (1  $\mu\text{M}$ ), there is a quenching trend of the maximum fluorescence intensity parallel to increasing the CTB concentrations, and this quenching is above the background (free probes). In other words, the higher probes concentration at 1  $\mu\text{M}$ , increased the maximum fluorescence intensity compared to the background (free probes at acetate buffer), and this number reduced along with increasing CTB concentrations. Monitoring the fluorescence intensity response of CID8795ATTO680 and CID535684ATTO680 at 1  $\mu\text{M}$  towards three concentrations of CTB compared to the background exhibits an increase in intensity at the 665 wavelengths higher than the two lower probe concentrations. This is due to the presence of an excessive amount of probes were aggregate due to its hydrophobic properties (having an aromatic chain) and result in poor solubility in acetate buffer. As a result, aggregation of the probes in the acetate holds back the non-radiation transition and then increases the fluorescence intensity (Huang, Wu et al. 2018). In this result, we also observed the fluorescence intensity quenched along with increasing the CTB concentrations at highest probes concentration (1  $\mu\text{M}$ ); this observation could be as a result of self-absorption by the fluorophore; first, the concentration of fluorophore is way too high, second increasing the protein concentration caused more congestion environment. The correlation between the hydrophobicity properties of CID8795ATTO680 and CID535684ATTO680 (having an aromatic chain in the dye structure) aligned with CTB concentration could be the reason for quenching. While the water molecules are absorbed by protein hydration at high protein concentrations, the environment may change accordingly, resulting in increasing the hydrophobicity of the external environment (Morikawa, Fujita et al.

2016). Therefore, upon increasing the concentration of the protein, solvent hydrophilicity may be reduced, which may lead to releasing the fluorophore from its bubble aggregates and reduced emissions accordingly. Therefore, we observed an increasing trend in anisotropy results (**Figure 40 A**, CID8795ATTO680 at 1  $\mu\text{M}$ ) upon increasing the protein concentration; however, this increase is below the background.

In all FP protein curve experiments, the maximum fluorescence intensity was also measured to monitor the effect of different concentrations of protein at fixed concentrations of CID8795ATTO680 (**Figure 43**). According to **Figure 43 A**, the maximum fluorescence intensity at 665 nm for all the samples (CID8795ATTO680 at 0.02  $\mu\text{M}$  with different concentrations of CTB) and background (free CID8795ATTO680) were statistically similar, except two concentrations of CTB at 2.23 nM and 20.07 nM, which were below the background; however, these changes also were not statistically different. The results represented in **Figure 41 A**, was shown statistical increasing in anisotropy of the probes with higher protein concentrations compared to the background which is consistent with fluorescence intensity results presented in **Figure 43 A**. As we observed neither any quenching along with the increase in protein concentrations, nor a statistical increase in fluorescence intensity compared to the background. This result along with anisotropy result confirmed the presence of the CID8795ATTO680 at 0.02  $\mu\text{M}$  was low enough to give us the sigmoidal shape of the protein curve for  $\text{EC}_{50}$  calculation; however, the  $R^2$  value was poor (0.91). In contrast, the results presented in **Figure 43 B** and **C**, indicate the higher probe concentrations at 0.05  $\mu\text{M}$  and 0.1  $\mu\text{M}$  induced the probe aggregation. According to the result in **Figure 43 B**, the concentration of the probe was at 0.05  $\mu\text{M}$ , which spouse to be low enough to avoid aggregation, but assay preparation was long enough to induce aggregation. This result represents the statistical changes in fluorescence intensity of the samples

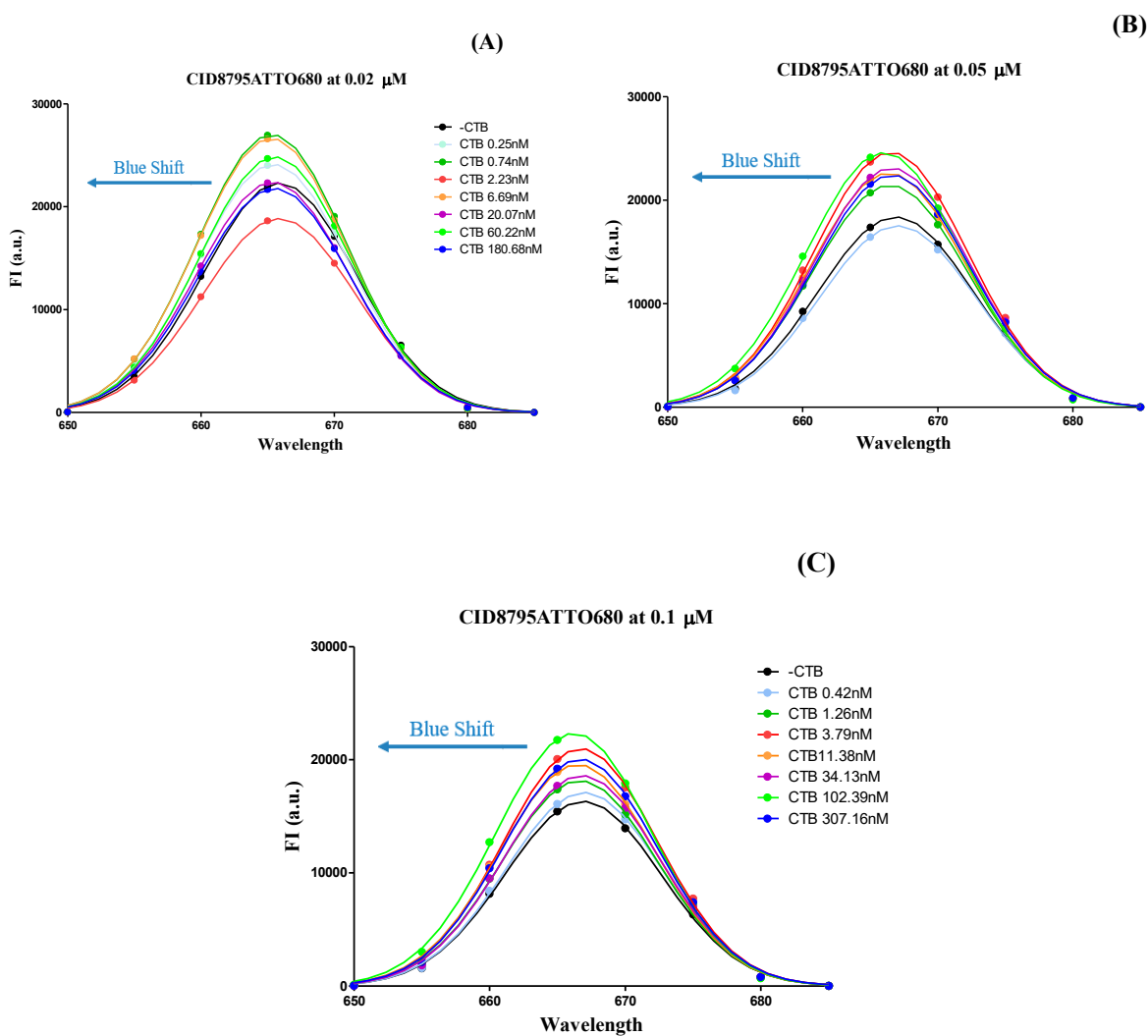


compared to the background which was almost the same in the presence of different concentrations of CTB ranging from 2.23 nM to 60.22 nM. However, at the highest CTB concentration at 180.68 nM, a decrease in intensity was observed compared to other concentrations of the protein (2.23 nM to 60.22 nM). As mentioned in section **6.3.3**, this quenching would be explained by the congestion environment caused by an excessive amount of the protein which leads to change the polarity of the solvent. The result presented in **Figure 43 C**, was also confirmed the probe aggregation hypothesis, where the fluorescence intensity of all samples (CID8795ATTO680 at 0.1  $\mu$ M with CTB at 3.79, 11.38, 34.13, 102.30 and 307.16 nM) was statistically above the background (free labeled-ligand). Besides, all the results of fluorescence intensity (preliminary and FP protein curves) confirmed the protein concentration above 100 nm reduces the intensity of fluorescence depending on the concentration of the probe. Comparing all the backgrounds with different probe concentrations in both preliminary and protein concentration curves results indicate a decrease in maximum emission caused by quenching along with increasing the probe concentrations.



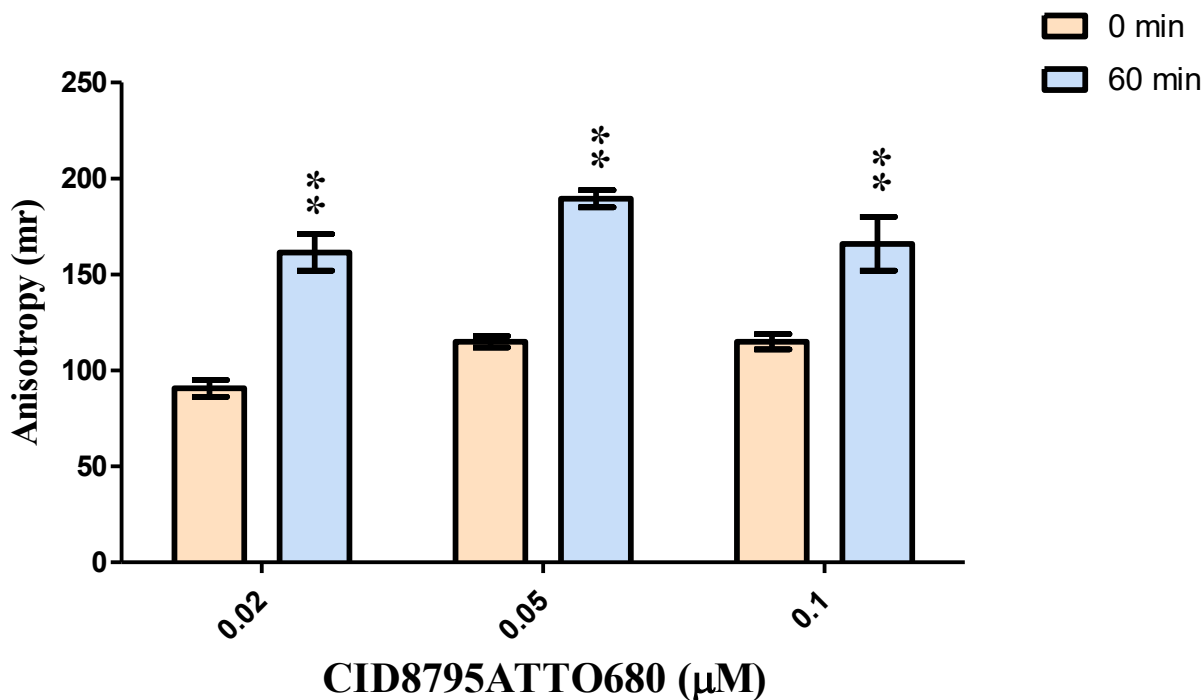
**Figure 43.** Change in maximum fluorescence intensity at 665 nm as a function of the CTB concentrations started from 0.25 nM to 180.68 nM against CID8795ATTO680 at 0.02  $\mu\text{M}$  (a) and 0.05  $\mu\text{M}$  (b), and the CTB concentrations started at 0.42 nM to 307.16 nM against CID8795ATTO680 at 0.1  $\mu\text{M}$  (c). Assays were performed in triplicate at pH 5.5 and read at 60 minutes, which were incubated for 5 minutes at 37°C. The average of all results with CV < 10% shows as a graph. The backgrounds correspond to free probes in acetate buffer which are represented as black columns. Statistical differences in anisotropy were determined by ANOVA post-hoc Dunnett's multiple comparison test, using Prism 5 software (Prism 2018, November 16);  $p < 0.05$  was considered significant and presented by \*, \*\*, and \*\*\*.

In addition to AIE, monitoring any shift of the wavelength in emission spectra induced by aggregation of the fluorophore would be another strategy. As shown in **Figure 44**, there was a spectral shift towards lower wavelength (blue shift) for all three concentrations of the probe which induced by TICT. This result also would be another interpretation for confirming the probe aggregation.



**Figure 44.** Fluorescence emission spectrums with a blue shift. CTB concentrations started from 0.25 nM to 180.68 nM against CID8795ATTO680 at 0.02  $\mu\text{M}$  (A) and 0.05  $\mu\text{M}$  (B), and the CTB concentrations started at 0.42 nM to 307.16 nM against CID8795ATTO680 at 0.1  $\mu\text{M}$  (C). Assays were performed in triplicate at pH 5.5 and read at 60 minutes, which were incubated for 5 minutes at 37°C. Nonlinear regression curve fit (Gaussian) used for fitting by using Prism 5 software (Prism 2018, November 16). All the  $R^2$  values were above 0.99.

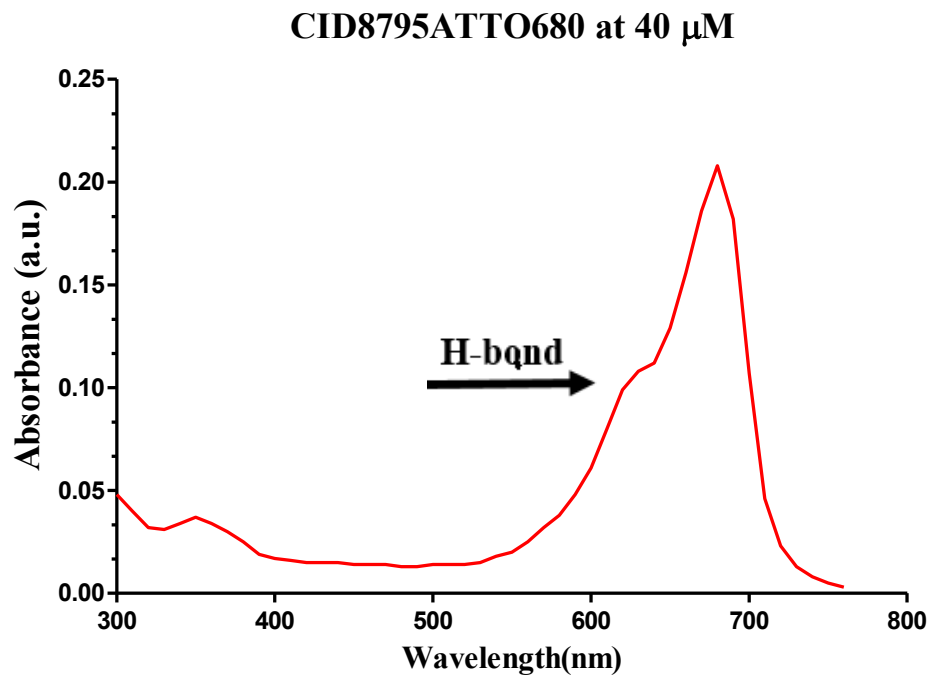
As mentioned above time would be another factor to prove the probe aggregation (Lindgren, Sorgjerd et al. 2005); therefore, the anisotropy of free probe at 0.02, 0.05, and 0.1  $\mu\text{M}$  compared at zero and 60 minutes (**Figure 45**).



**Figure 45.** Change in anisotropy of the free CID8795ATTO680 at 0.02 μM, 0.05 μM, and 0.1 μM over time (0 and 60 minutes). Assays were performed in triplicate at pH 5.5 and read at 0 and 60 minutes, which were incubated for 5 minutes at 37°C. The average of all results with CV < 10% shows as a graph. Statistical differences in anisotropy at zero and 60 minutes were determined by ANOVA post-hoc Dunnett's multiple comparison test, using Prism 5 software (Prism 2018, November 16);  $p < 0.05$  was considered significant and presented by \*\*.

As shown in the time comparison graph, there is a statistical increase in anisotropy overtime for the free fluorophore that is consistent with the aggregation of the probe over time.

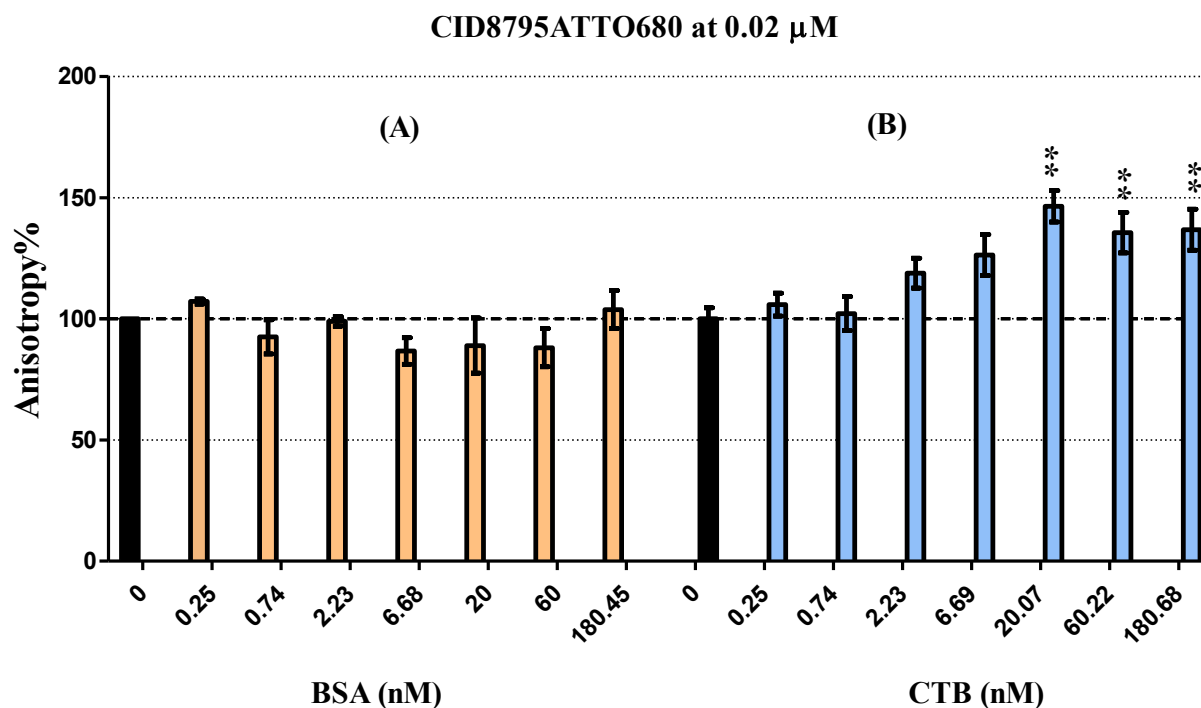
As shown in **Figure 46**, there is an H-band observed from the absorbance spectrum of the CID8795ATTO680 at 40 μM. The H-band refers to side-by-side interaction of the fluorescent probe with its closest-neighbor (in our case to another probe like itself). CID8795ATTO680 has an aromatic chain in its dye structure which makes this probe hydrophobic; therefore, it is expected in the hydrophilic environment, this probe interacts with another probe like itself through side-by-side interaction of the aromatic chain (Van der Waals interaction). This result would be another proof for probe aggregation.



**Figure 46.** Absorbance spectrum of the CID8795ATTO680 at 40  $\mu\text{M}$  in CTB reaction buffer at pH 7.2 and 37°C. The peak at lower wavelength represents the H-band induced by probe aggregation.

### 6.3.6. Selectivity Confirmation of the CID8795ATTO680 towards CTB

To confirm the selectivity of the CID8795ATTO680 toward CTB, a change in anisotropy of the probe at 0.02  $\mu\text{M}$  was measured against the serial concentrations of BSA. According to the result shown in **Figure 47**, this probe did not show any statistically significant changes despite the serial concentrations of BSA, which would confirm the lack of binding interaction of CID8795ATTO680 toward BSA compares to CTB.



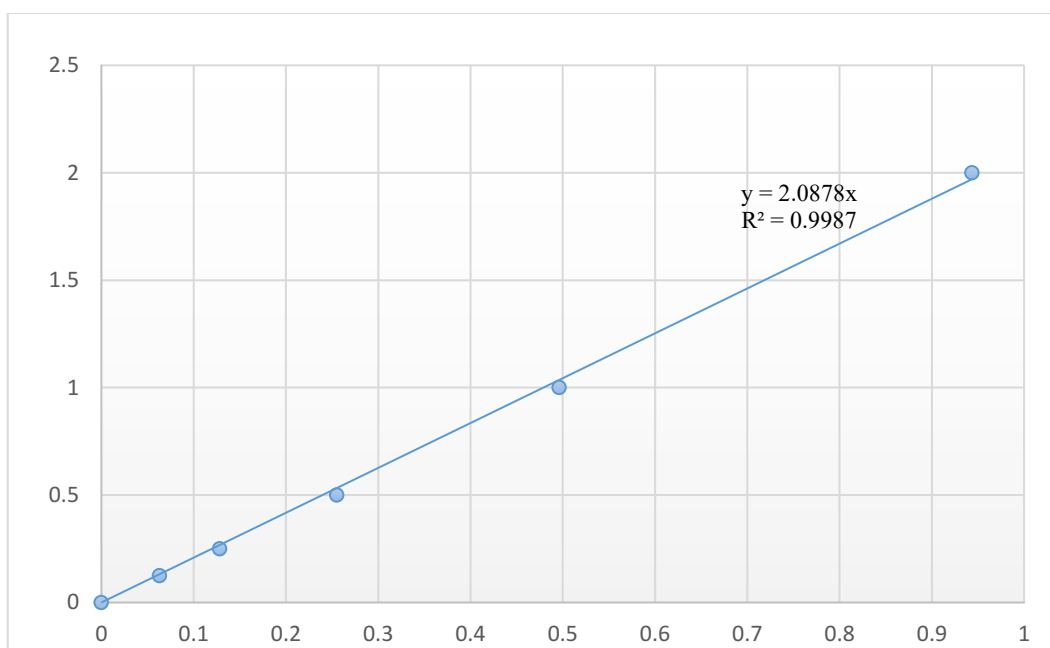
**Figure 47.** Changing in anisotropy of CID8795ATTO680 at 0.02  $\mu$ M in the presence of CTB and BSA.

A) Anisotropy of the NIR probe in the presence of BSA with two-fold serial dilution with a starting concentration at 0.25 nM to 180.45 nM in PBS buffer pH 7.4. B) Anisotropy of the NIR probe in the presence of CTB with a two-fold serial dilution with a starting concentration at 0.25 nM to 180.68 nM in acetate buffer pH 5.5 and incubated for 5 minutes at 37°C and read at 60 minutes. The result is based on the average of all triplicates (CV < 10%). Statistical differences in anisotropy were determined by ANOVA post-hoc Dunnett's multiple comparison test, using Prism 5 software (Prism 2018, November 16);  $p < 0.05$  was considered significant and presented by \*\*.

### 6.3.7. Cell Lysate Result for CID8795ATTO680

The total protein concentration was quantified based on colorimetric assay with the BCA kit. The result is shown in **Figure 48**. The total protein concentration was determined by applying the following equation:

$$\left( \text{Total Protein Concentration} \left( \frac{\text{mg}}{\text{ml}} \right) = 2.087 * \text{Average OD of the Samples} \right) \text{Equation 2}$$



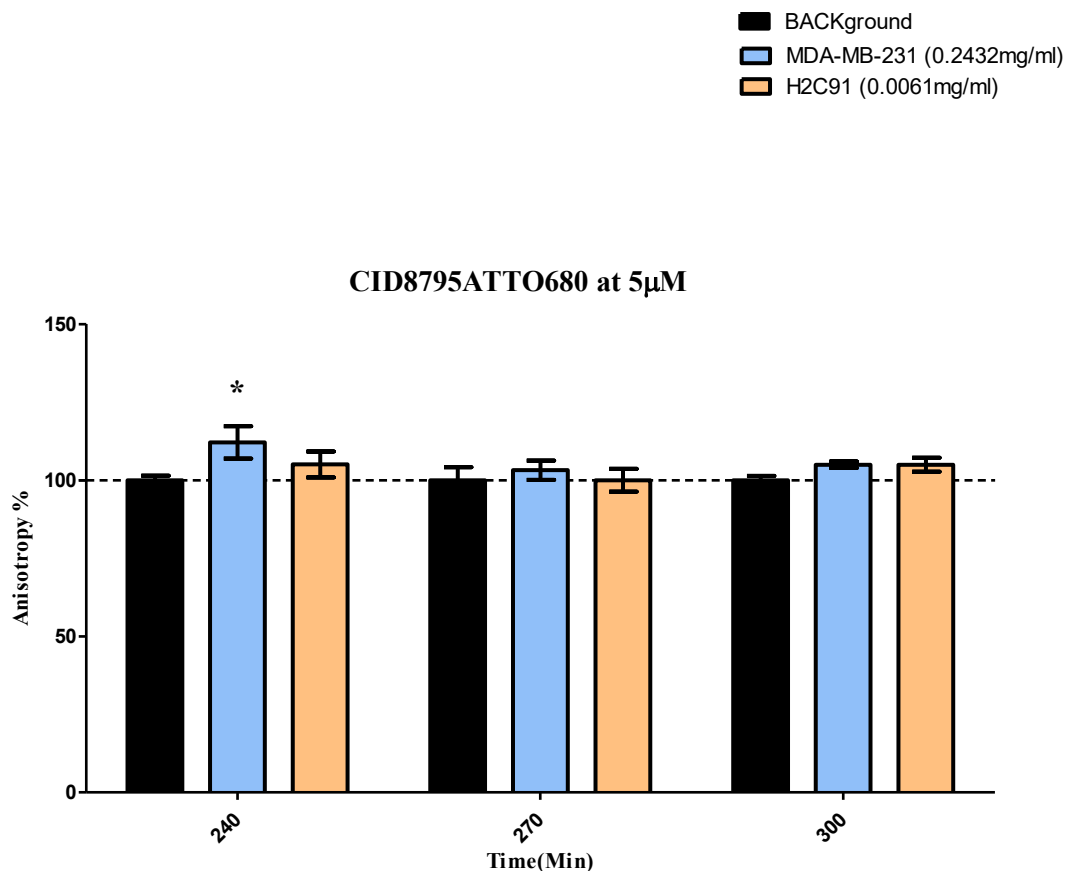
**Figure 48.** Average OD of the samples versus serial concentration of the BSA (2, 1, 0.5, 0.25, and 0.125 mg/ml) to quantify the total protein concentration of the lysate.

According to the BCA assay result, the total protein concentration in the *MDA-MB-231* cancer cells and *H9C2* rat Cardio-myocytes normal tissue cells were 0.2432 and 0.061 mg/ml respectively. The two-fold serial dilution of the lysate which was used for anisotropy assay for cancer and normal cell are as follows: *MDA-MB-231* lysate concentration (0.24, 0.12, 0.06, 0.03, 0.01, 0.007, and 0.003 mg/ml). *H9C2* lysate concentration (0.006, 0.003, 0.001, 0.0007, 0.0003, 0.0001, and 0.00009 mg/ml). As it is obvious, the total protein concentration of the cancer lysate is 40 times higher than the total protein concentration of the normal cell. Here, anisotropy was read every 30 minutes for five hours. For the first 210 minutes, no changes observed in anisotropy between samples (cancer lysate and CID8795ATTO680 at 5  $\mu$ M) and background (assay buffer and CID8795ATTO680 at 5  $\mu$ M). Hence the highest concentration of the *MDA-MB-231* cancer



cells lysate is more reliable due to having more concentrated proteins; this result was compared over the last three time points. According to the **Figure 49** result, for none of these three time points the increase in anisotropy was observed between sample (cancer lysate at 0.24 mg/ml and CID8795ATTO680 at 5  $\mu$ M) and background (assay buffer and CID8795ATTO680 at 5  $\mu$ M), except at 240 minutes

In normal lysate (*H9C2*) (**Figure 49**), the samples (normal lysate at 0.006 mg/ml and CID8795ATTO680 at 5  $\mu$ M ) were not showing any difference in the anisotropy compared to the background (assay buffer and CID8795ATTO680 at 5  $\mu$ M). However, the concentration of the normal lysate was very low. According to all the results mentioned above (**6.3.3** and **6.3.4**), we corroborated the binding interaction of the CID8795ATTO680 toward CTB. However, here the increase in anisotropy in cancer cell lysate with 5  $\mu$ M of this probe was only observed at 240 minutes and it was only 1.1 times greater than the background; therefore, stating the binding interaction of CID8795ATTO680 and CTB in cancer cell lysate is difficult to ascertain.



**Figure 49.** Change in anisotropy in the presence and the absence of *MDA-MB-231* cancer cells lysate at 0.24 mg/ml (blue graph) and *H9C2* normal cells lysate at 0.0061 mg/ml (orange graph) for CID8795ATTO680 at 5  $\mu$ M at 240, 270, and 300 minutes. Assays were performed in triplicate at pH 5.5 and read for 300 minutes, which were incubated for 15 minutes before adding probe at 37°C. The result is based on the average of all triplicates (CV < 6%). Statistical differences in anisotropy were determined by ANOVA post-hoc Dunnett’s multiple comparison test, using Prism 5 software (Prism 2018, November 16);  $p < 0.05$  was considered significant and presented by \*.

#### 6.4. Summary and Conclusions

Here, we evaluated the binding interaction of two novel NIR probes, CID8795ATTO680 and CID535684ATTO680, with CTB.

To evaluate the inhibitory activity of these two non-reactive probes and their parent ligands, activity assays with two commercially available substrates, Ac- Arg-Arg-AFC and Z-Arg-Arg-pNA, was performed. According to these results, none of the parent ligands and

CID535684ATTO680 showed inhibitory activity on AFC formation, neither on the initial velocity of the progression curve nor on the last portion of the curve, which indicated the highest AFC formation. NIR CID8795ATTO680 probe also did not show inhibitory activity against CTB. No inhibitory effect had been observed in AFC and PNA formations, neither at the initial rate nor at the end of the progression curve.

Given all the primary anisotropy results for CID535684ATTO680, no increase in anisotropy was observed between samples (probe and protein) compared to the background (free labeled-ligand), hence further investigation was aborted.

An increase in anisotropy was observed with the presence of lower concentrations of the CID8795ATTO680 (0.01  $\mu\text{M}$  and 0.1  $\mu\text{M}$ ) with CTB concentrations at 10 nM and 100 nM in preliminary results. We experimentally proved that CID8795ATTO680 was susceptible to bind to another probe like itself. First, increased anisotropy in the background (free CID8795ATTO680) at high concentration; and second, increased anisotropy over time. Therefore, to build the anisotropy curve as a function of CTB concentrations, CID8795ATTO680 concentrations were selected ranging from 0.1  $\mu\text{M}$  to 0.01  $\mu\text{M}$ . Comply to anisotropy protein concentrations curve results at lower probe concentration 0.02  $\mu\text{M}$ , the  $EC_{50}$  value defined at  $3.27 \pm 1.27$  nM with  $R^2 > 0.91$ ; however, we were not able to draw results from the higher probe concentration (0.05  $\mu\text{M}$  and 0.1  $\mu\text{M}$ ) curves due to assays preparation time-consuming which cause probe aggregation and result in high background. Although for the first anisotropy curve as a function of CTB, we had the same constraint for assay preparation (timing), the probe concentration was low enough to give us the sigmoidal shape of the protein curve. The probe aggregation was also confirmed by fluorescence intensity results, which indicated the higher intensity of the probe in the presence of CTB at different concentrations compared to the free CID8795ATTO680 and finally by observing

the blue shift in the emission spectra. The absorbance spectra of the 40  $\mu\text{M}$  concentration of free CID8795ATTO680, was also confirmed the probe aggregation due to the presence of the H-band in the spectrum. In conclusion, generally, the anisotropy value is not dependent on labeled-ligand concentration, since it is independent of the total intensity of the emission collected in both directions. In contrast, the intrinsic fluorescence intensity of a fluorescent probe may vary based on binding to another probe like itself through side-by-side interaction of the aromatic chain (Van der Waals) which induces a significant difference between the bound forms against free forms of the fluorescent probe to the total intensity of the sample. As a result, it affects the interpretation of the anisotropy measurements (Lea and Simeonov 2011).

The selectivity assay of this novel probe toward CTB was determined using anisotropy assay against BSA. These results confirm the lack of binding of the NIR probe to BSA (no change in anisotropy between sample (probe and BSA) compare to the background (free probe)). In contrast, statistical increase (ANOVA post-hoc Dunnett's multiple comparison test, using Prism 5 software) in anisotropy was observed with the presence of the same probe concentration with CTB compare to the background which corroborates binding interaction of CID8795ATTO680 to CTB.

In cell lysate assay, an increase in anisotropy was found at 240 minutes of the assay in breast cancer cell line and this increase was only 1.1 times higher than the background (free probe in lysate buffer); therefore, it is difficult to conclude the binding interaction between CID8795ATTO680 and CTB. No change in anisotropy was observed in the cardio normal cell line with CID8795ATTO680 versus background (probe with lysate buffer); however, no conclusion can be drawn due to the low concentration of the normal lysate.

From all the above, we may conclude that the CID8795ATTO680 in low concentration (below 0.02  $\mu\text{M}$ ) is likely to be an ideal probe for CTB in biochemistry assays by considering the

time factor, but most importantly, we need to evaluate the probe's behavior in intracellular protease in living cells.

## Chapter 7: Rate of Hydrolysis of Cbz-Lys-Lys-PABA-DCMF by CTB and CTL

### 7.1. Introduction

Activity assays provide a quantitative measurement of the interactions between two molecules, usually a small molecule binding to a protein. For substrate-based probes, measuring hydrolysis by the target and off-target enzymes allows us to compare the relative rates of probe turnover assessing the specificity of the probe. Although, *in vivo* environments where the concentrations of the substrate-based probe are anticipated to be much lower, these experiments can be deceptive. As a solution, identifying the Michaelis–Menten parameters  $k_{\text{cat}}$  and  $K_{\text{M}}$  provide accurate measurement of the catalytic efficiency of the selected substrate toward the target enzyme. The specificity constant ( $k_{\text{cat}}/K_{\text{M}}$ ), measuring the formation of the enzyme-substrate (E-S) complex ( $K_{\text{M}}$ ) and catalytic turnover ( $k_{\text{cat}}$ ), reflects the enzymatic efficiency at low substrate concentrations (below the  $K_{\text{M}}$  value) providing a better prediction of the substrate performance *in vivo* (Koshland 2002, Penuelas, Mazzolini et al. 2005). Here enzyme kinetic assays were applied to measure the specificity constant of the novel substrate toward CTB. The specificity constant value of this substrate toward CTB was determined using the Michaelis–Menten, and Lineweaver-Burk equations. For evaluating the selectivity of the Cbz-Lys-Lys-PABA-DCMF toward CTB over other members of the cathepsin family, CTL was selected for the experiment – due to the high similarity of the active site with cathepsin B.

### 7.2. Procedures and Materials for Cbz-Lys-Lys-PABA-DCMF

#### 7.2.1. Enzyme Kinetics

A Michaelis-Menten curve (**Appendix section 3**) was generated by plotting the initial rate of hydrolysis of Z-Lys-Lys-PABA-DCMF versus the concentration of the substrate. For initial

velocity, the enzymatic hydrolysis of the Z-Lys-Lys-PABA-DCMF was measured from the slope of the linear portion of the reaction (the first ten points of the linear portion of the reaction). Recombinant CTB at 5 nM (Novoprotein, C398) was incubated with a two-fold serial dilution of Z-Lys-Lys-PABA-DCMF with a starting concentration at 0.019 mM to 5 mM. Prior to each enzymatic reaction, CTB was activated by incubation with 5 mM DTT at 37 °C for 10 minutes. A 7-point, 10-fold serial dilution of DCMF was used to generate a standard curve (1 to  $1 \times 10^{-6}$  mM) to quantify product formation as a function of substrate concentrations. Assays were performed in triplicate at 37°C in pH 5.5 acetate buffer (acetate 50 mM with 5 mM DTT, 5 mM EDTA, and 0.05% Triton X-100). Fluorescence intensity was read on a Synergy 4™ plate reader (Biotek Inc.), every 1 minute for 170 minutes, at  $470 \pm 9$  nm excitation and  $540 \pm 9$  nm emission wavelengths. Controls containing the fluorescent substrate in the assay buffer without enzyme were used to evaluate the spontaneous hydrolysis and stability of the substrate throughout the kinetic experiments to correct for background fluorescence.

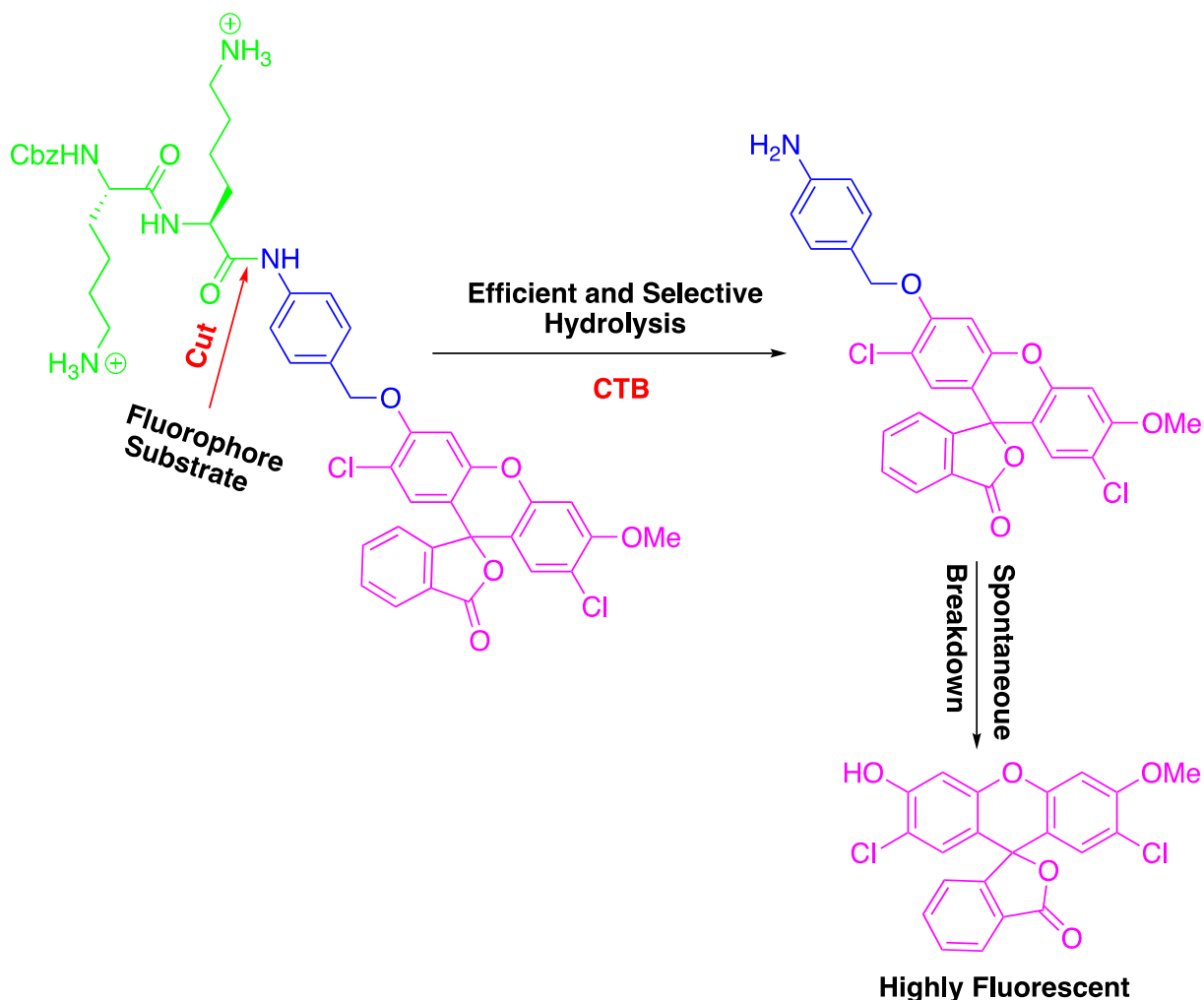
### **7.2.2. The Selectivity of the Novel Substrate towards CTB**

To evaluate the selectivity of the new substrate towards CTB over a related cysteine protease, an activity assay was performed using recombinant human cathepsin L (R&D system, 952-CY). The assay was performed as described in section 7.2.1 for CTB.

## **7.3. Results for Cbz-Lys-Lys-PABA-DCMF**

### **7.3.1. Enzyme Kinetics Assay for Novel Substrate**

The results of the kinetic assay confirmed that Z-Lys-Lys-PABA-DCMF is a substrate for CTB. The reaction steps for the hydrolysis of Cbz-Lys-Lys-PABA-DCMF are as shown in **Figure 50**.

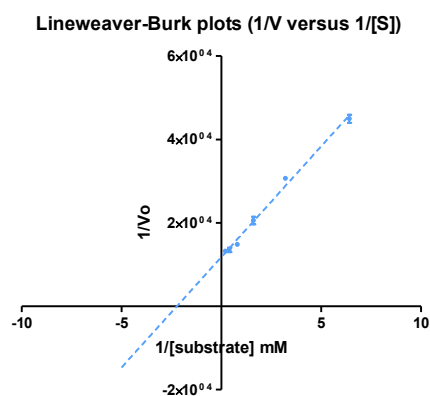


**Figure 50.** The reaction steps for hydrolysis of Cbz-Lys-Lys-PABA-DCMF, PABA immolation, and releasing DCMF.

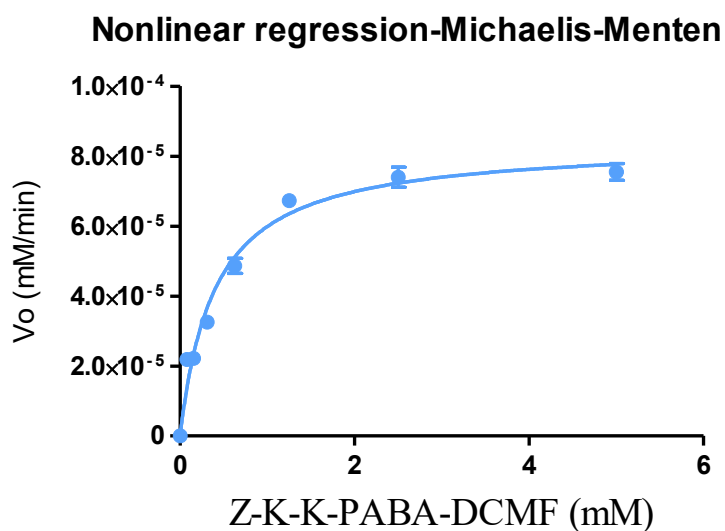
The rate of formation of the product (DCMF) was proportional to the substrate concentration at a fixed concentration of CTB at 5 nM (**Figure 51**). The slope of the first ten points of the linear portion of the curve (when the reverse reaction is not prominent) was used to calculate the initial velocity for each substrate concentration. The kinetics parameters  $K_M$ ,  $V_{max}$ , and  $k_{cat}$  were determined from the Michaelis–Menten and Lineweaver–Burk plots.



(A)



(B)



$$\begin{aligned} E_t &= 5.000\text{e-}006 \\ k_{cat} &= 16.77 \\ K_M &= 0.4006 \\ V_{max} &= 8.386\text{e-}005 \\ R^2 &= 0.97 \end{aligned}$$

**Figure 51.** Lineweaver-Burk (A) and Michaelis-Menten (B) plots applied to determine kinetic parameters for Cbz-Lys-Lys-PABA-DCMF by using Prism 5 software (Prism 2018, November 16). Experiments were performed in duplicate at 5 nM concentration of CTB, pH 5.5, and temperature of 37°C. The substrate concentration ranged from 0.019 mM to 5 mM.

As shown in **Table 9**, Cbz-Lys-Lys-PABA-DCMF has a higher  $K_M$  value ( $0.4 \pm 0.03$  mM) than probes Cbz-Phe-Lys-PABA-AMC and Cbz-Lys-Lys-PABA-AMC (Chowdhury, Moya et al.

2014). This means that Cbz-Lys-Lys-PABA-DCMF requires a higher concentration than the other two probes to reach half of  $V_{\max}$  (lower propensity to form the Michaelis-complex); however, the  $k_{\text{cat}}$  value was  $16.77 \pm 0.4$  (e.g.  $\sim 16$  substrate molecules "turned over" by each enzyme active site per second) for Cbz-Lys-Lys-PABA-DCMF. This value is  $\sim 33$  times higher than Cbz-Phe-Lys-PABA-AMC but only half of Cbz-Lys-Lys-PABA-AMC. This resulted in a modest specificity constant ( $k_{\text{cat}}/K_M$  value of  $41.9 \pm 0.1$ ) meaning that the new substrate is hydrolyzed  $\sim 5.5$  times slower than Cbz-Lys-Lys-PABA-AMC. The enzyme kinetics studies demonstrate that substrate **5** is hydrolyzed efficiently enough to proceed into future biological experiments in cell lysates and living cells.

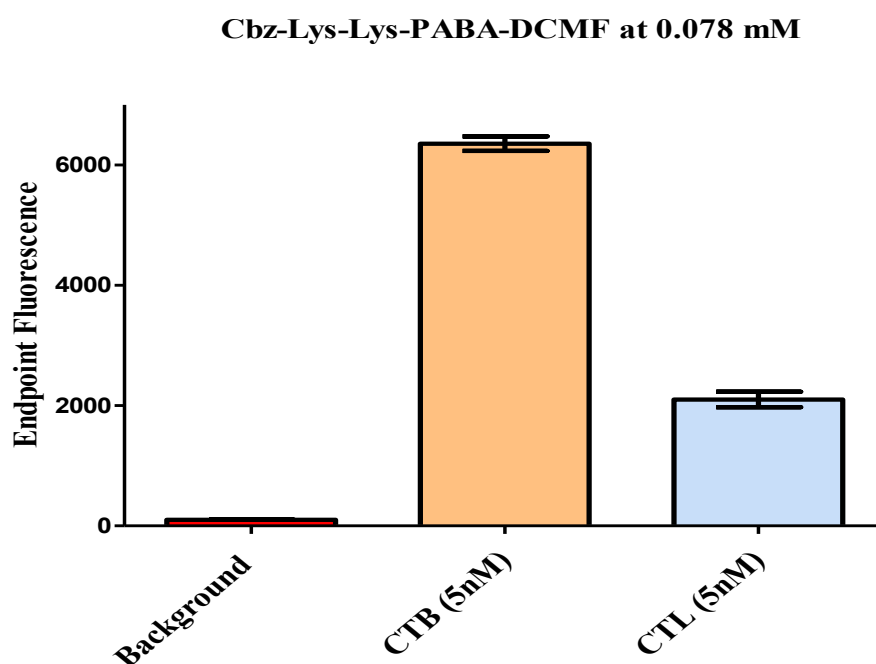
**Table 9.** Comparison of kinetic constants between Cbz-Lys-Lys-PABA-DCMF and other substrate-based probes. Nine (9) concentration of the fluorescent substrate was used to determine  $K_M$ ,  $V_{\max}$ , and  $k_{\text{cat}}$  values under the Michaelis-Menten model. Assays were performed at pH 5.5 and  $37^\circ\text{C}$ , with an enzyme concentration of 5 nM.

Substrate	$K_M$ (mM)	$k_{\text{cat}}$ ( $\text{s}^{-1}$ )	$k_{\text{cat}}/K_M$ ( $\text{mM}^{-1} \times \text{s}^{-1}$ )
<b>Cbz-Lys-Lys-PABA-DCMF</b>	$0.4 \pm 0.03$	$16.77 \pm 0.4$	$41.9 \pm 0.1$
<b>Cbz-Phe-Lys-PABA-AMC</b> (Chowdhury, Moya et al. 2014)	$0.082 \pm 0.003$	$0.50 \pm 0.04$	$6.1 \pm 7$
<b>Cbz-Lys-Lys-PABA-AMC</b> (Chowdhury, Moya et al. 2014)	$0.16 \pm 0.04$	$37 \pm 2$	$231 \pm 70$

### 7.3.2. The Selectivity of Cbz-Lys-Lys-PABA-DCMF Towards CTB and CTL

An activity assay was performed for evaluating the selectivity of the substrate probe Cbz-Lys-Lys-PABA-DCMF **5** toward CTB versus CTL. Due to the low catalytic turnover of this substrate by CTL, we were not able to calculate the initial velocity of the enzymatic reaction at high concentrations of the probe. Instead, an endpoint fluorescence assay was used, where 0.078 mM of the substrate with 5 nM of the CTB was used to compare the hydrolysis of the probe by CTB and CTL under the same biologically relevant concentrations. The well containing CTB displayed a rapid and time-dependent increase in fluorescence intensity corresponding to the

release of DCMF. Despite being a more catalytically potent enzyme, CTL hydrolysis of the Cbz-Lys-Lys-PABA-DCMF was much slower over the same time frame. This demonstrates that Cbz-Lys-Lys-PABA-DCMF is hydrolyzed much faster by CTB versus CTL at this concentration of the probe. To put this into perspective, the endpoint RFU value, which was taken after 150 minutes for CTB, was approximately three times higher than the value for CTL (**Figure 52**). This demonstrates that the probe shows selectivity towards CTB over CTL.



**Figure 52.** Selectivity assay of the Z-Lys-Lys-PABA-DCMF at a fixed concentration (0.078 mM) versus CTB or CTL at 5 nM. Fluorescence intensity at excitation and emission wavelengths  $470 \pm 9$  nm and  $540 \pm 9$  nm applied to evaluate the hydrolysis of Z-Lys-Lys-PABA-DCMF by CTB and CTL. A graph showing the endpoint RFU corresponding to releasing DCMF with CTL and CTB.

#### 7.4. Summary and Conclusions

We reported that the synthesis and preliminary kinetic evaluation of Cbz-Lys-Lys-PABA-DCMF and have shown that this compound is efficiently hydrolyzed by CTB but poorly hydrolyzed by CTL. The onset of fluorescence is consistent with the enzymatic release of DCMF.

Michaelis-Menten enzyme kinetic parameters were measured resulting in a  $K_M = 0.4 \pm 0.03$  mM and  $k_{cat} = 16.7 \pm 0.4$  s<sup>-1</sup> for CTB. These parameters translated into a modest specificity constant of  $k_{cat}/K_M = 41.9 \pm 0.07$  mM<sup>-1</sup> × s<sup>-1</sup>. Together, these experiments suggest that Cbz-Lys-Lys-PABA-DCMF may be useful for detecting CTB activity in cell lysates and living cells-based assays.

## **Chapter 8: Cell Viability and Cell Imaging Assays for CID8795ATTO680 and Cbz-Lys-Lys-PABA-DCMF**

### **8.1. Introduction**

Cell viability and proliferation can be evaluated using the 3-[4,5-dimethylthiazol-2-yl]-2,5-diphenyl tetrazolium bromide (MTT) assay. This assay is based on the principle that viable cells take up the MTT (yellow tetrazolium salt) and reduce it into an insoluble purple formazan dye by reductase enzymes. This dye is then released from the cells and viability can be calculated as a ratio of absorbance of the untreated and treated cells (Posimo, Unnithan et al. 2014). Therefore, cell death induced by the probe candidates can be quantified (Mosmann 1983).

Fluorescent microscopy is a useful technique to study living cells and obtain a better understanding of the biological function of intracellular dynamics (Ettinger and Wittmann 2014). In this study, cell imaging was performed to monitor intracellular protease activity induced by two novel candidate probes, using the "Cytation 5" fluorescent microscope.

### **8.2. Procedures and Materials**

#### **8.2.1. Cell Viability**

For evaluation of any changes in cell viability of the *MDA-MB-231* and *H9C2* cell lines incubated with Cbz-Lys-Lys-PABA-DCMF and CID8795ATTO680, an MTT assay was performed. Cell viability was quantified through endpoint absorbance reads at 540 nm. The experiments were performed three times with samples run in triplicate each time for both probes.

*MDA-MB-231* (breast cancer cells) and *H9C2* rat cardiomyocytes (normal tissue cells) were cultured in Hyclone DMEM/high-glucose media containing either 10% fetal clone III, 1%

penicillin/streptomycin, 0.1% gentamicin and 1% sodium pyruvate, or 10% fetal bovine serum and 1% penicillin/streptomycin and 1% sodium pyruvate, respectively. Cells were maintained in T75 flasks at 37°C, 5% CO<sub>2</sub>. For the MTT experiments, both cell types were seeded in a 96-well plate (Corning cat. #3596) at a concentration of 10,000 cells per well and incubated for 2 days. Media was then changed to 200 µL of media containing the probe candidates (at 1, 10, or 40 µM), followed by 20 µL of the MTT reagent (5 mg/mL stock in PBS buffer and passed through a 0.2 µM syringe filter). The plate was incubated for 4 h at 37 °C in 5% CO<sub>2</sub>. After the media was removed, 200 µL of DMSO was added to the wells and shaken for 5 minutes. The viability percentage was calculated as a ratio of absorbance of the untreated and treated cells. All procedures were performed in the dark.

### 8.2.2. Cell Imaging

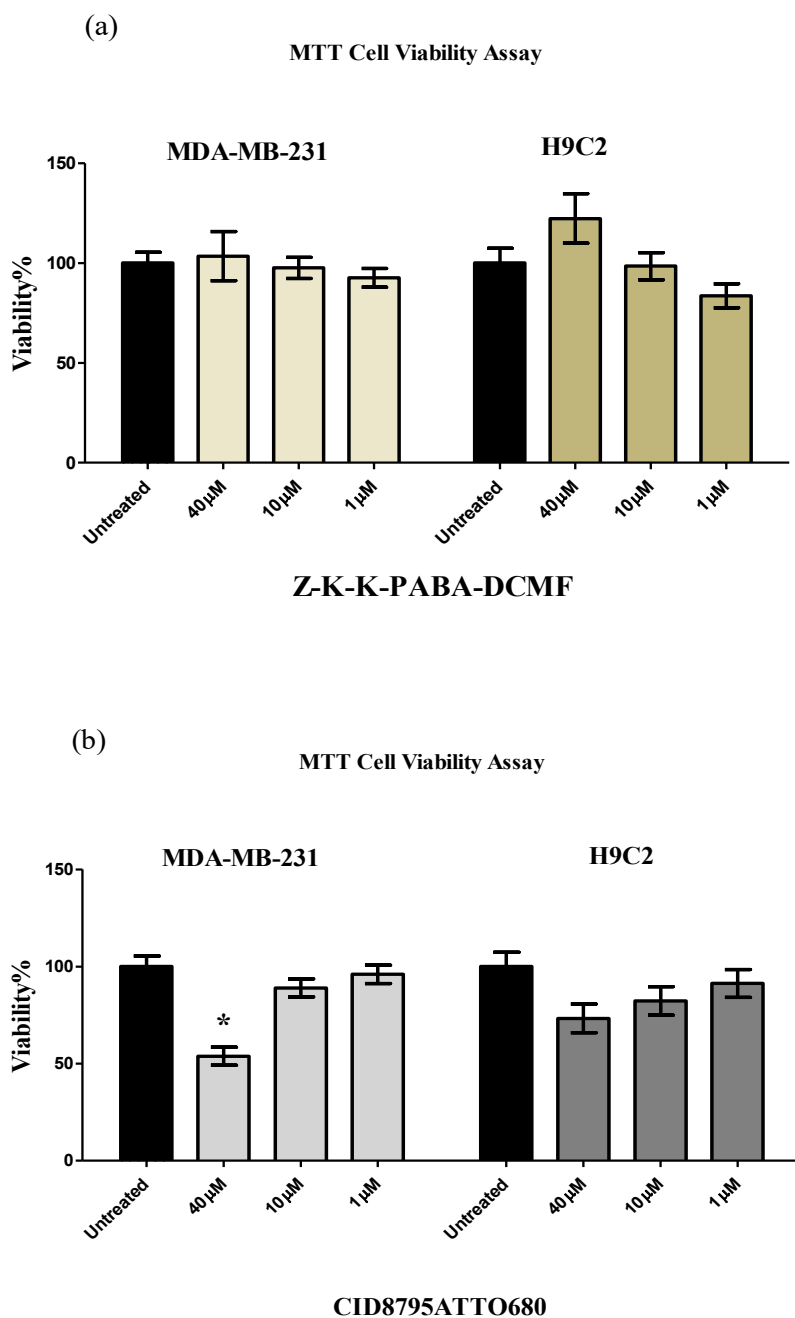
Once we determined that Cbz-Lys-Lys-PABA-DCMF and CID8795ATTO680 were chemically stable for enzymatic analysis, we evaluated the intracellular protease activity in living cells. *MDA-MB-231* and *H9C2* cells were cultured as previously described (section 9.2.1). For the experiment, cells were seeded into 24-well plates (Corning cat. #3526) at a concentration of 10,000 cells per well and incubated overnight at 37 °C in a 5% CO<sub>2</sub>. To determine if hydrolysis activity was due to CTB, cells were then incubated for 12 hours with CA-074Me (CTB inhibitor at 10 µM) before incubation with substrate-based and non-reactive probes for 4 hours. Cbz-Lys-Lys-PABA-DCMF and CID8795ATTO680 were added at concentrations of 40 and 10 µM, respectively, to each cell type. The intact and free probe was eliminated by washing the wells with PBS. Fluorescence images were then taken at ( $\lambda_{\text{ex}} = 469 \text{ nm}$  and  $\lambda_{\text{em}} = 525 \text{ nm}$ ) for Cbz-Lys-Lys-PABA-DCMF and ( $\lambda_{\text{ex}} = 628 \text{ nm}$  and  $\lambda_{\text{em}} = 685 \text{ nm}$ ) for CID8795ATTO680 of each well using the Cytation 5 by BioTek.

### 8.3. Result for Two Novels NIR Non-Reactive Probes

#### 8.3.1. Cell Viability Results

Based on the MTT results shown in **Figure 53 A**, treatment with Cbz-Lys-Lys-PABA-DCMF across all concentrations did not significantly affect cell viability in either cell type (ANOVA Tukey's multiple comparison test). However, an increase in cell viability was observed at the highest concentration of this substrate in *H9C2* cells (healthy cells). This demonstrates that the presence of Cbz-Lys-Lys-PABA-DCMF does not decrease cell viability at any concentration, making this probe useful to detect CTB activity inside living systems.

In contrast, cell viability was significantly reduced in both cell types following treatment with 40  $\mu\text{M}$  of the CID8795ATTO680 to approximately 50% in (cancer cells) and 30% in (normal cells) compared to the untreated cells. This result suggested higher toxicity of CID8795ATTO680 in cancer cell lines compared to normal cells at 40  $\mu\text{M}$ ; however, there was no toxicity observed at lower concentrations (10  $\mu\text{M}$  and 1  $\mu\text{M}$ ) of this probe toward both cell lines. Therefore, 10  $\mu\text{M}$  of the probe was used for cell imaging.



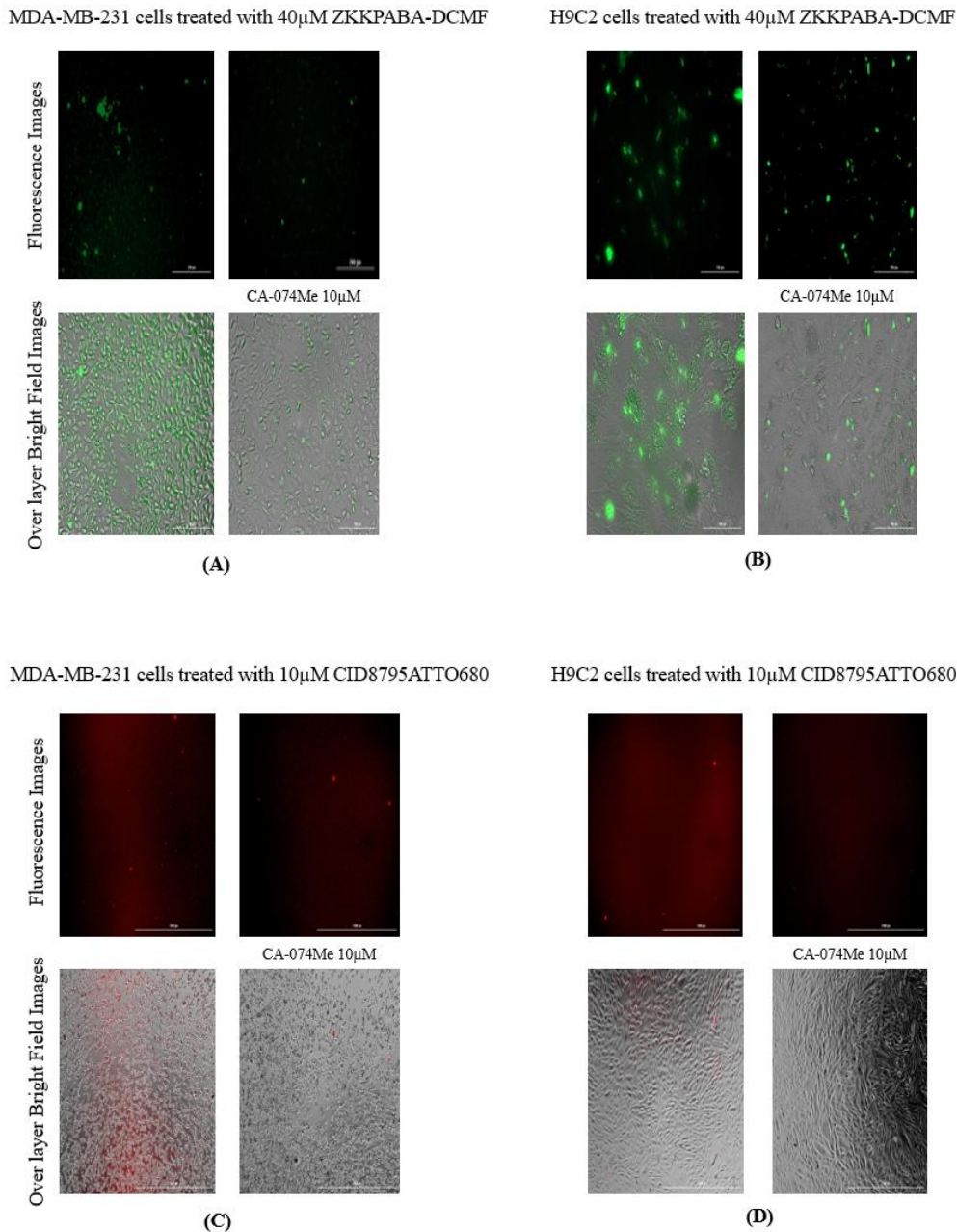
**Figure 53.** Effects of Cbz-Lys-Lys-PABA-DCMF (a) and CID8795ATTO680 (b) on cell viability. *MDA-MB-231* and *H9C2* cells were treated with candidate probes at 1, 10, and 40  $\mu$ M. The % viability of cells was quantified as a ratio of absorbance of the untreated and treated cells using MTT assay. Statistical differences in viability were determined by performing ANOVA post-hoc Dunnett's multiple comparison test, using Prism 5 software (Prism 2018, November 16);  $p < 0.05$  was considered significant and marked by \*.



### 8.3.2. Cell Imaging Results

The final step was to observe the ability of the two candidate probes to detect intracellular CTB activity. As shown in **Figure 54** A (resolution was set at zoom 200  $\mu\text{m}$  in  $x$  and  $y$ ) and B (resolution was set at zoom 100  $\mu\text{m}$  in  $x$  and  $y$ ), an increase in fluorescence intensity was visualized inside healthy cardio and cancer cells (left panel) compared to cells treated with CTB inhibitor (CA-074Me in the right panel). The increased intracellular fluorescence of the cells treated with Cbz-Lys-Lys-PABA-DCMF compared to the CA-074Me cells suggests a good CTB specificity of this probe in live cells. Evidently, this substrate enters into the cells, and clearly, the presence of the CTB inhibitor (CA-074Me) decreased the CTB activity in both cell types – leading to reduced activation of Cbz-Lys-Lys-PABA-DCMF.

As shown in **Figure 54** C and D (resolution was set at zoom 1000  $\mu\text{m}$  in  $x$  and  $y$ ), we were not able to observe fluorescence of the CID8795ATTO680 inside the cells; however, the concentration of this probe was not high enough to allow optimal fluorescent imaging and the filters ( $\lambda_{\text{ex}} = 628 \text{ nm}$  and  $\lambda_{\text{em}} = 685 \text{ nm}$ ) used for cell imaging did not cover the exact excitation maximum at 675 nm and an emission maximum at 705 nm wavelengths of the NIR probe (but high enough to cover the most spectrum part). For all the cell imaging different resolutions were used to provide a better view of the cells and detailed oriented images.



**Figure 54.** Fluorescence microscopy and overlayer "Bright filed" images of *MDA-MB-231* and *H9C2* cell lines treated with Cbz-Lys-Lys-PABA-DCMF (A (zoom at 200  $\mu$ m) and B (zoom at 100  $\mu$ m)) and CID8795ATTO680 (C (zoom at 1000  $\mu$ m) and D (zoom at 1000  $\mu$ m)). Fluorescent microscopy images taken of cells treated for 4 h with (A (cancer cells) and B (normal cells)) with 40  $\mu$ M of Cbz-Lys-Lys-PABA-DCMF ( $\lambda_{ex}$  = 469 nm and  $\lambda_{em}$  = 525 nm) (C (cancer cells) and D (normal cells)) 10  $\mu$ M of CID8795ATTO680 ( $\lambda_{ex}$  = 628 nm and  $\lambda_{em}$  = 685 nm). To evaluate probe specificity cells were treated with 10  $\mu$ M CA-074Me for 12 hours before incubation with each probe for 4 h. In all images, the left side panel shows uninhibited cells versus inhibited in the right panel.

#### 8.4. Summary and Conclusions

In conclusion, we evaluated the effects of Cbz-Lys-Lys-PABA-DCMF and CID8795ATTO680 on cell viability using two cell lines: breast cancer (*MDA-MB-231*) and healthy cardio (*H9C2*). No statistically significant differences in cell viability were observed following treatment with Cbz-Lys-Lys-PABA-DCMF in either cell type. In contrast, the cells treated with CID8795ATTO680 at 40  $\mu\text{M}$  resulted in approximately 50% and 30% cell death in cancer and normal cells, respectively. CID8795ATTO680 did not induce any significant cell death at lower concentrations (10 and 1  $\mu\text{M}$ ). These data confirm that both probes may be used for cell-based assays, except CID8795ATTO680 which must be used at concentrations below 10  $\mu\text{M}$ . Additionally, cell imaging results suggest Cbz-Lys-Lys-PABA-DCMF would be useful for detecting intercellular CTB activity using fluorescence microscopy. However, more detailed experiments are required to determine the intercellular location of probe hydrolysis (e.g. lysosomes, mitochondria, and other organelles). In contrast, no change was observed in fluorescence from uninhibited cells treated with CID8795ATTO680 compare to inhibited one, neither in breast cancer cells nor in the cardio normal cell. However, there were two main limitations for cell imaging of this probe; 1) using a lower concentration of the probe (10  $\mu\text{M}$ ) due to toxicity, 2) the emission and excitation wavelength used for cell imaging were almost cover the maximum emission and excitation of the probe. Therefore, these experiments suggest that the CID8795ATTO680 probe may not be an ideal probe for evaluating intracellular protease activity in living cells.

## Chapter 9: Summary and Conclusions

The present study aimed to synthesize and experimentally test novel, efficient, selective, and cell-permeable fluorescence probes for detecting human protease cathepsin B, a biomarker for cancer. This enzyme was studied from nucleotide sequence to three-dimensional structure using computational methods. We also investigated genetic variations of the *CTSB* gene based on the information stored in the 1,000 Genomes database. In this study, data from the 1,000 Genomes database suggests that identified genetic variations should not affect the binding of probes to the CTB active site (no SNPs found in these positions). The results also identified variants, Leu26Val and Ser53Gly located on the signal peptide part of the protein (which is removed before forming mature enzyme). Leu26Val, which this study found to be present in 39.5% of the 1,000 Genomes population, has been associated with chronic pancreatic cancer.

Another achievement of this study was reporting the synthesis of two novel fluorescent probes: CID8795ATTO680 and CID535684ATTO680. Both have fluorescence emission in the near-infrared (705 nm and 690 nm, respectively), above the autofluorescence characteristic of live cells. Given that these probes are already fluorescent; they do not require CTB for activation. Although CID535684ATTO680 did not indicate a significant change in FP and anisotropy in the presence of CTB, future work may involve synthesizing other derivatives to increase the binding affinity toward CTB. Neither probe demonstrated inhibitory activity toward CTB as determined experimentally using the commercially available fluorogenic substrate (Ac-RR-AFC) and colorimetric substrate (Z-Arg-Arg-pNA).

The NIR probe CID8795ATTO680 demonstrates no change in anisotropy toward BSA; in contrast, statistical increase (ANOVA post-hoc Dunnett's multiple comparison test with  $p < 0.05$

in anisotropy was observed with the same probe concentration against CTB. The preliminary fluorescence anisotropy result for CID8795ATTO680 indicated an increase in anisotropy in the presence of lower concentrations of the probe (0.01  $\mu\text{M}$  and 0.1  $\mu\text{M}$ ) against CTB at 10 nM and 100 nM. Compatible with the principle of fluorescence polarization (**Appendix section 2**), changing in anisotropy directly affected by the rotational speed of the fluorescent molecule through dissociation/breakdown or association/binding process; therefore, observing an increase in anisotropy between sample (probe with target protein) and background (free probe) confirmed the existence of CID8795ATTO680/CTB complex. From preliminary anisotropy and fluorescence intensity results, we proved that CID8795ATTO680 aggregation behavior induced high background in anisotropy (caused by excessive probe concentration and time), which was also consistent with fluorescence intensity, emission spectrum (blue shift observation), and anisotropy of the protein concentrations curve results (probe at 0.05  $\mu\text{M}$  and 0.1  $\mu\text{M}$ ). From the anisotropy protein concentrations curve for CID8795ATTO680 at 0.02  $\mu\text{M}$ , an  $\text{EC}_{50}$  value of  $3.27 \pm 1.27$  nM was estimated with  $R^2 > 0.91$ . Although this probe gave us the sigmoidal shape of the protein curve at 0.02  $\mu\text{M}$ , the  $R^2$  was low for  $\text{EC}_{50}$  evaluation. Increased the anisotropy was observed in human invasive breast cancer cell line *MDA-MB-231* treated with 5  $\mu\text{M}$  of CID8795ATTO680. However, this was observed only at 240 minutes out of 10-time points (read every 30 minutes for 300 minutes) and only 1.1 times higher compared to the background; therefore, no conclusion could be drawn from it. According to the cell viability results, the presence of CID8795ATTO680 at 40  $\mu\text{M}$  induced statistically significant cell death in normal (30%) and cancer (50%) cells but no toxicity was found at lower concentrations (10 and 1  $\mu\text{M}$ ). From the cell imaging results, we were not able to observe any fluorescence induced by CID8795ATTO680. Even though the CID8795ATTO680 probe that we reported was not as successful as we expected for the evaluation

of intracellular protease activity in living cells, these shortcomings will help us to create a better version of it in the future.

In this study, a substrate-based probe Cbz-Lys-Lys-PABA-DCMF was successfully synthesized and evaluated for CTB hydrolysis. The selectivity of this substrate to CTB, over a related proteases CTL, was confirmed using fluorescence endpoint assay. The  $k_{\text{cat}}/K_M$  ( $41.9 \pm 0.07 \text{ mM}^{-1} \times \text{s}^{-1}$ ) indicates that Cbz-Lys-Lys-PABA-DCMF is a good substrate for CTB. Although the specificity constant ( $k_{\text{cat}}/K_M$ ) value of the substrate was lower than Cbz-Lys-Lys-PABA-AMC fluorogenic substrate ( $231 \pm 70 \text{ mM}^{-1} \times \text{s}^{-1}$ ) reported by Dr. Phenix's group previously. Cell viability assays did not indicate any toxicity of this probe in both *MDA-MB-231* (breast cancer cells) and *H9C2* (healthy cardio cells). DCMF released by CTB was easily detected inside cells using fluorescence microscopy, suggesting that this fluorophore is a better choice for imaging activity in living cells. This is due to longer excitation and emission wavelengths characteristic of DCMF.

## Chapter 10: Future Work

One of the most concerning diseases that attract so many scientists' attention is cancer. Diagnosing cancer at initial steps of progression would be a promising method to control or even prevent this disease. Cathepsin B is a biomarker for cancer; therefore, monitoring its overexpression may be a promising strategy for cancer diagnosis (Gondi and Rao 2013). In this project, we synthesized, evaluated, and reported two fluorescent probes intended to target CTB. Recently, the combination of NIR optical imaging with fluorescent probes has become attractive for sensitive detection of cancers like colon, breast, and lung cancer and identification of tumors margins during surgeries (Habibollahi, Figueiredo et al. 2012). The next step of this project would be testing the substrate-based probe in different cell lines. It is also would be good to evaluate this substrate-based probe in an animal model for localization of cancer tumors. Modeling a new substrate-based probe based on the Cbz-Lys-Lys-PABA scaffold modified with NIR dyes would be another approach for future work. Regarding the CID8795ATTO680 probe, improving water solubility to prevent probe aggregation, perhaps adding DMSO (1% (v/v)) would be effective. Time as an important factor for probe aggregation was another limitation; therefore for future assays reduce the time of reading for anisotropy to 30 minutes may optimize the results. This probe show toxicity to the cells (*MDA-MB-231* and *H9C2*) and aggregation in high concentration, due the presence of aromatic chain in its dye structure To assess toxicity, it is advisable to evaluate the effect of the lower concentrations of the probe ranging from 10 to 30  $\mu\text{M}$  in MTT assay. For optimizing cell imaging assay using ideal emission and excitation wavelength will be required.

## Appendix:

### 1. Reagents, Buffers, and Enzymes

All buffers and reagents were obtained from Fisher Scientific or Sigma-Aldrich unless otherwise noted and used without further purification. Seven compounds selected from (Kamstra, Dadgar et al. 2014) for the experimental testing (**Table 7**) are as follows: CID2434131; CID5939530; CID667134; CID1256741; CID2999504; CID8795, and CID535684. The commercially available dye used in this study called ATTO 680 NHS-ester (MW (g/mol): 622.68866). A commercial kit used for the activity assay (Sigma Aldrich, Catalog number: MAK200) contained CTB reaction buffer (Sigma Aldrich, Catalog Number MAK200A) at pH 7.2, CTB Reagent (Sigma Aldrich, catalog number MAK200B), cathepsin B, human (Sigma Aldrich, catalog number MAK200C), CTB Substrate, Ac-RR-AFC (Sigma Aldrich, catalog number MAK200D) and CTB Inhibitor, F-F-FMK (Sigma Aldrich, catalog number MAK200E). Cathepsin B (CTB) was obtained from Novoprotein, C398 - 50 $\mu$ g at 1 mg/mL concentration. Supplied as a 0.2  $\mu$ M filtered solution of 20 mM Tris-HCl, 150 mM NaCl, pH 8.0 with C-6His tag. Arg18-Ile339 and each vial contains 1 mg/mL = 27.1023  $\mu$ M. acetate buffer at pH 5.5 (50 mM acetate with 5 mM DTT, 5 mM EDTA, and 0.05% Tween® 20). Cathepsin L (CTL) Recombinant Human Cathepsin L/CTSL (C-6His) and Recombinant Human Cathepsin L (R&D system, 4952-CY), supplied as a 0.2  $\mu$ M filtered solution of 20 mM HAc-NaAc, 150 mM NaCl, pH 4.5. Bovine Serum Albumin (Brudal, Lampe et al.) used as standard and control for the experiment (Sigma-Aldrich catalog number P5619) at 2, 1.5, 1, 0.75, 0.5, 0.25 and 0.125 mg/ml. BSA working solution; BSA mixed with PBS (R&D Systems, catalog number WA126), pH 7.2 with 0.05% Tween® 20.



2', 7'-dichloro-6'-methoxy-fluorescein (DCMF) was synthesized by mixing one part of phthalic anhydride (CAS; 85-44-9, MW; 148.117 g/mol and MF; C<sub>8</sub>H<sub>4</sub>O<sub>3</sub>) to two parts of 4-Chlororesorcinol (CAS; 95-88-5, MW; 144.554 g/mol and MF; C<sub>6</sub>H<sub>5</sub>ClO<sub>2</sub>) under the presence of 4N Sulfuric acid (CAS; 7664-93-9, MW; 98.072 g/mol and MF; H<sub>2</sub>SO<sub>4</sub>) as the catalyzer. The 2', 7'-dichloro-6'-methoxy-fluorescein with the chemical formula of C<sub>21</sub>H<sub>12</sub>Cl<sub>2</sub>O<sub>5</sub> and molecular weight of 415.223 was conjugated with the Cbz-Lys-Lys-PABA under the presence of 4N H<sub>2</sub>SO<sub>4</sub>, 2-(6'-Methoxy-3'-oxo-3'H-xanthene-9'-yl)-benzoic acid methyl ester. Cbz-Lys-Lys-PABA-DCMF (MW of 932.29 g/mol; maximum excitation at 470 nm, maximum emission at 565 nm). Cathepsin B commercial substrate, Z-Arg-Arg-pNA, as 10 mg lyophilized powder (Sigma-Aldrich SCP0108; MW 584.63; Abs=410 nm;  $K_M$  (pH 7.7) = 0.78 mM and  $K_M$  (pH 6.0) = 1.19 mM (Hasnain et al., 1993 J Bio Chem 268:235-240). *p-nitroaniline* (pNA) (Aldrich-CAS; 100-01-6, MW; 138.126 g/mol and MF; O<sub>2</sub>NC<sub>6</sub>H<sub>4</sub>NH<sub>2</sub>) for the standard curve. Protease inhibitor cocktail (Sigma-Aldrich catalog number P8340) containing two cysteine protease inhibitors, E-64 (MW 357.41 Da) at 1.4 mM, and Leupeptin (hemisulfate salt, MW 475.59 Da) at 2 mM, in DMSO. Clear, flat bottom, non-binding, 96-well microplate (Fisher 12565501). Black, flat bottom, non-binding 96-well microplate (Costar catalog number 3915).

## 2. Fluorescence Polarization and Anisotropy

Fluorescence polarization is based on the differential rotational speed between the free small molecules fluorescent ligand in solution versus ligand bound to the protein of interest (Moerke 2009). A larger molecule, including the ligand-protein complex, leads to enhanced polarization of light compared to the free ligand and protein.

The equation for FP measurement is as follows:

$$P = (FIU_{par} - G * FIU_{per}) / (FIU_{par} + G * FIU_{per}) \quad \text{Equation 3}$$

$$mP = P/1000$$

Where the  $FIU_{par}$  defined as fluorescent intensity is evaluated in the parallel plane, whereas  $FIU_{per}$  is measured in the perpendicular plane. The G factor is the intensity ratio of the vertical to the horizontal component which depends on the instrument.

Fluorescence polarization is influenced by environmental factors, such as pH-level; solution viscosity; absolute temperature; molecular volume, and the gas constant **Equation 4** (Jameson and Ross 2010). Care must be taken for the recording of the experimental conditions.

$$\mu = \eta V / RT \quad \text{Equation 4}$$

where R refers to the gas constant, V is the molecular weight of the rotating molecule, T is the absolute temperature, and  $\eta$  refers to the solvent viscosity, and  $\mu$  is the rotational correlation time of the fluorophore molecule. When all the factors remind constant, the rotational speed of the fluorophore will affect directly the FP value; through dissociation/breakdown or association/binding process.

Anisotropy ( $A_n$ ) is the same principle as FP; however, anisotropy is calculated using a different equation, and it is more accurate than FP; due to revealing the mean angular shift of the fluorescent molecule that induced between absorption followed by emission of the photon. This angular displacement influenced by the rotational diffusion rate during the lifetime of the excited state. The rotational diffusion rate correlated to the viscosity of the solvent, resulting in changes in anisotropy.

The equation for measurement is as follows:

$$r (\text{Anisotropy}) = (FIU_{par} - G * FIU_{per}) / (FIU_{par} + 2 * G * FIU_{per}) \quad \text{Equation 5}$$
$$mA = r / 1000$$

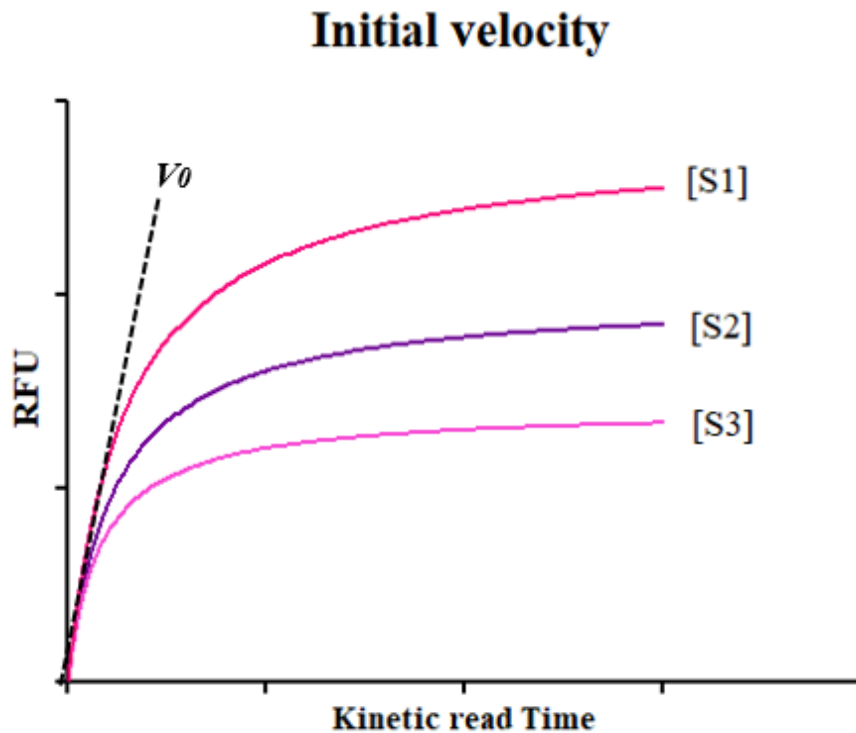
Where the  $FIU_{par}$  and  $FIU_{per}$  are the fluorescent intensity measured in the parallel and perpendicular plane respectively. The  $(FIU_{par} + 2 * G * FIU_{per})$  correlates to the total intensity of the sample.

### 3. Enzyme Kinetic

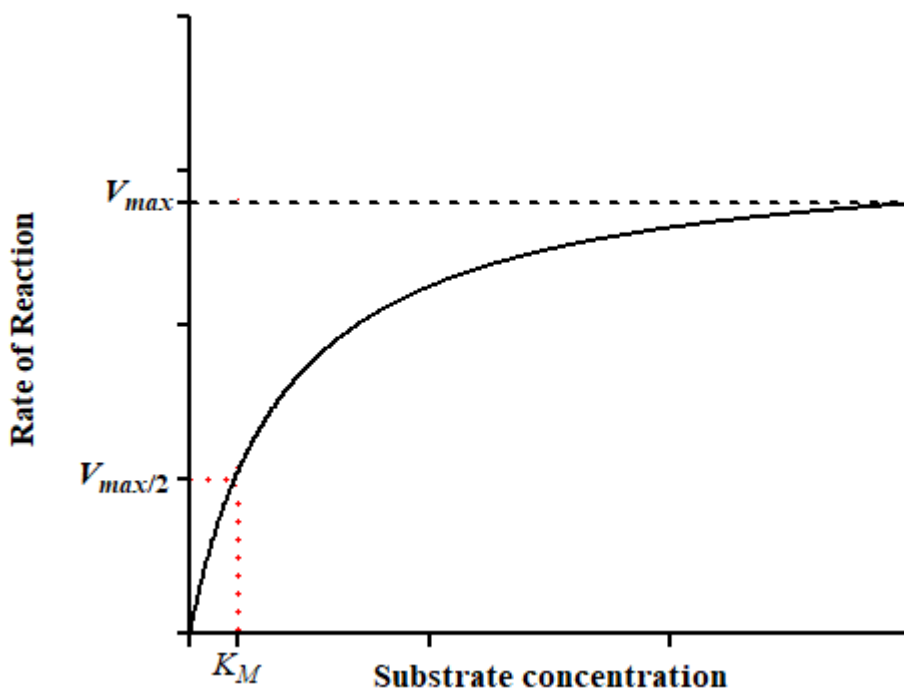
The two most common methods for determining enzyme kinetic parameters are an endpoint assay and continuous assay. The endpoint assay occurs when the enzymatic reaction is stopped after a given period and the substrate or product concentration is quantified. In contrast, a continuous assay refers to the quantification of substrate turnover or the formation of a product during the enzymatic reaction. In either case, monitoring hydrolase activity generally uses artificial substrates that generate a colorimetric or fluorescent signal as the reaction progresses. The kinetic properties of an enzymatic reaction catalyzed can be characterized by the Michaelis-Menten equation. The  $K_M$  (mM) value refers to the substrate concentration where the reaction reaches half of the  $V_{max}$ . The lower  $K_M$ , the more propensity the enzyme must form the Michaelis complex reaching half  $V_{max}$  at lower concentrations of substrate. The  $V_{max}$  value represents the maximum rate of the reaction when the enzyme is saturated by the substrate at a specific enzyme concentration. And finally, the catalytic constant of the reaction ( $k_{cat}$ ) illustrates a number of substrate molecules turned over by each enzyme per unit time (Cutillas-Lozano and Gimenez 2013). All the parameters can be defined by the following equation:

$$V_0 = \frac{V_{max} [S]}{(K_M + [S])} \quad \text{Equation 6}$$

Where  $V_0$  refers to the initial velocity (concentration/time) of an enzymatic reaction, S is the substrate concentration in molar units,  $V_{\max}$  is the maximum velocity, and  $K_M$  is the substrate concentration at half  $V_{\max}$ . Generally, the amount of substrate should vary at least seven points above and below the  $K_M$  value, while the amount of enzyme is constant during the experiment. The initial velocity (slope of the initial curve) is calculated as the linear portion of the beginning of the reaction progress versus time curve at steady-state and typically has units of RFU/time or Abs/time (**Figure 55**). A standard plot of the reporter group in assay buffer is then used to convert units of RFU or Abs to a concentration. Then on the XY data graph, the  $V_0$  of reaction (Y) is plotted against the substrate concentration (X) to generate a Michaelis-Menten plot **Figure 56**.



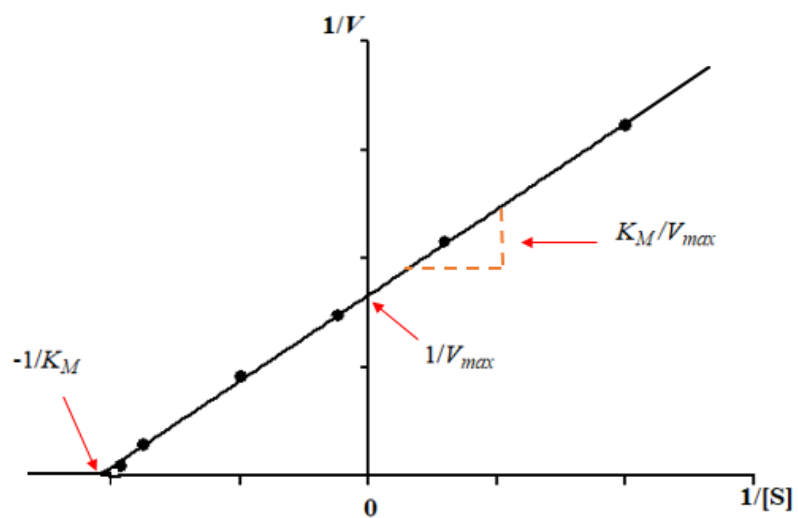
**Figure 55.** Linear regression fitting model for initial velocity (RFU/Time) calculation.



**Figure 56.** Michaelis-Menten kinetics graph. The X-axis is the substrate concentration, and the reaction velocity is shown on the Y-axis.  $K_M$  and  $V_{max}$  could be defined with the following equation

$$Y = V_{max} * X / (K_M + X) \quad \text{Equation 7}$$

An alternative way to determine  $K_M$ ,  $V_{max}$ , and  $k_{cat}$  is the Lineweaver–Burk equation represented by Hans Lineweaver and Dean Burk in 1934 (Burk, Lineweaver et al. 1934). Historically, Lineweaver-Burk plots are the graphical representations of the Lineweaver-Burk equation. In Lineweaver-Burk plots (**Figure 57**), the slope is the ratio of  $K_M/V_{max}$ ; X-intercept is  $-1/K_M$  and Y-intercept equal to  $1/V_{max}$ . Curve fitting programs are typically used in modern enzyme kinetic studies.



**Figure 57.** The double-reciprocal plot of Michaelis-Menten equation, Lineweaver-Burk plots ( $1/V$  versus  $1/[S]$ ), where  $-1/K_M$  is X-intercept,  $1/V_{max}$  is Y-intercept, and  $K_M/V_{max}$  is the slop.

## References

1. Aggarwal, N. and B. F. Sloane (2014). "Cathepsin B: Multiple roles in cancer." Proteomics Clinical Applications **8**(5-6): 427-437.
2. Ahmed, S. and R. K. Shahid (2012). "Disparity in cancer care: a Canadian perspective." Current Oncology **19**(6): E376-E382.
3. Alcalay, N. I., M. Sharma, D. Vassmer, B. Chapman, B. Paul, J. Zhou, J. G. Brantley, D. P. Wallace, R. L. Maser and G. B. V. Heuvel (2008). "Acceleration of polycystic kidney disease progression in cpk mice carrying a deletion in the homeodomain protein Cux1." American Journal of Physiology-Renal Physiology **295**(6): F1725-F1734.
4. Alford, R., M. Ogawa, P. L. Choyke and H. Kobayashi (2009). "Molecular probes for the in vivo imaging of cancer." Molecular Biosystems **5**(11): 1279-1291.
5. Andersson, H., T. Baechi, M. Hoechl and C. Richter (1998). "Autofluorescence of living cells." Journal of Microscopy-Oxford **191**: 1-7.
6. Angles, R., M. Arenas-Salinas, R. Garcia, J. A. Reyes-Suarez and E. Pohl (2020). "GSP4PDB: a web tool to visualize, search and explore protein-ligand structural patterns." BMC Bioinformatics **21**(Suppl 2): 85.
7. Atlas of Genetics and Cytogenetics in Oncology and Haematology. (2008, Aug). "CTSB (cathepsin B)." from [http://atlasgeneticsoncology.org/Genes/GC\\_CTSB.html](http://atlasgeneticsoncology.org/Genes/GC_CTSB.html).
8. Bakst, R. L., H. Z. Xiong, C. H. Chen, S. Deborde, A. Lyubchik, Y. Zhou, S. Z. He, W. McNamara, S. Y. Lee, O. C. Olson, I. M. Leiner, A. R. Marcadis, J. W. Keith, H. A. Al-Ahmadie, N. Katabi, Z. Gil, E. Vakiani, J. A. Joyce, E. Pamer and R. J. Wong (2017). "Inflammatory Monocytes Promote Perineural Invasion via CCL2-Mediated Recruitment and Cathepsin B Expression." Cancer Research **77**(22): 6400-6414.
9. Baldwin, E. T., T. N. Bhat, S. Gulnik, M. V. Hosur, R. C. Sowder, 2nd, R. E. Cachau, J. Collins, A. M. Silva and J. W. Erickson (1993). "Crystal structures of native and inhibited forms of human cathepsin D: implications for lysosomal targeting and drug design." Proc Natl Acad Sci U S A **90**(14): 6796-6800.
10. Bian, B., S. Mongrain, S. Cagnol, M. J. Langlois, J. Boulanger, G. Bernatchez, J. C. Carrier, F. Boudreau and N. Rivard (2016). "Cathepsin B promotes colorectal tumorigenesis, cell invasion, and metastasis." Mol Carcinog **55**(5): 671-687.
11. Blum, G., S. R. Mullins, K. Keren, M. Fonovic, C. Jedeszko, M. J. Rice, B. F. Sloane and M. Bogoy (2005). "Dynamic imaging of protease activity with fluorescently quenched activity-based probes." Nature Chemical Biology **1**(4): 203-209.
12. Blum, G., G. von Degenfeld, K. Keren, H. M. Blau and M. Bogoy (2005). "Imaging protease activity using novel fluorescently quenched probes." Biopolymers **80**(4): 501-501.

13. Blum, G., G. von Degenfeld, M. J. Merchant, H. M. Blau and M. Bogyo (2007). "Noninvasive optical imaging of cysteine protease activity using fluorescently quenched activity-based probes." Nature Chemical Biology **3**(10): 668-677.
14. Blum, G., R. M. Weimer, L. E. Edgington, W. Adams and M. Bogyo (2009). "Comparative Assessment of Substrates and Activity Based Probes as Tools for Non-Invasive Optical Imaging of Cysteine Protease Activity." Plos One **4**(7).
15. Bohley, P. and P. O. Seglen (1992). "Proteases and proteolysis in the lysosome." Experientia **48**(2): 151-157.
16. Briggs, M. S., D. D. Burns, M. E. Cooper and S. J. Gregory (2000). "A pH sensitive fluorescent cyanine dye for biological applications." Chemical Communications(23): 2323-2324.
17. Brudal, E., E. O. Lampe, L. Reubsæet, N. Roos, I. K. Hegna, I. M. Thrane, E. O. Koppang and H. C. Winther-Larsen (2015). "Vaccination with outer membrane vesicles from *Francisella noatunensis* reduces development of francisellosis in a zebrafish model." Fish Shellfish Immunol **42**(1): 50-57.
18. Burk, D., H. Lineweaver and C. K. Horner (1934). "The Specific Influence of Acidity on the Mechanism of Nitrogen Fixation by *Azotobacter*." J Bacteriol **27**(4): 325-340.
19. Buschmann, V., K. D. Weston and M. Sauer (2003). "Spectroscopic study and evaluation of red-absorbing fluorescent dyes." Bioconjugate Chemistry **14**(1): 195-204.
20. Buschmann, V., K. D. Weston and M. Sauer (2003). "Spectroscopic study and evaluation of red-absorbing fluorescent dyes." Bioconjug Chem **14**(1): 195-204.
21. Campo, E., J. Munoz, R. Miquel, A. Palacin, A. Cardesa, B. F. Sloane and M. R. Emmert-Buck (1994). "Cathepsin B expression in colorectal carcinomas correlates with tumor progression and shortened patient survival." Am J Pathol **145**(2): 301-309.
22. Celli, J. P., B. Q. Spring, I. Rizvi, C. L. Evans, K. S. Samkoe, S. Verma, B. W. Pogue and T. Hasan (2010). "Imaging and Photodynamic Therapy: Mechanisms, Monitoring, and Optimization." Chemical Reviews **110**(5): 2795-2838.
23. Chen, S., H. Dong, S. Yang and H. Guo (2017). "Cathepsins in digestive cancers." Oncotarget **8**(25): 41690-41700.
24. Chen, S. Y., H. Dong, S. M. Yang and H. Guo (2017). "Cathepsins in digestive cancers." Oncotarget **8**(25): 41690-41700.
25. Chen, X., D. Lee, S. Yu, G. Kim, S. Lee, Y. Cho, H. Jeong, K. T. Nam and J. Yoon (2017). "In vivo near-infrared imaging and phototherapy of tumors using a cathepsin B-activated fluorescent probe." Biomaterials **122**: 130-140.
26. Chen, X. Q., T. Pradhan, F. Wang, J. S. Kim and J. Yoon (2012). "Fluorescent Chemosensors Based on Spiroring-Opening of Xanthenes and Related Derivatives." Chemical Reviews **112**(3): 1910-1956.



27. Chenna, R., H. Sugawara, T. Koike, R. Lopez, T. J. Gibson, D. G. Higgins and J. D. Thompson (2003). "Multiple sequence alignment with the Clustal series of programs." Nucleic Acids Research **31**(13): 3497-3500.
28. Chowdhury, M. A., I. A. Moya, S. Bhilocha, C. C. McMillan, B. G. Vigliarolo, I. Zehbe and C. P. Phenix (2014). "Prodrug-inspired probes selective to cathepsin B over other cysteine cathepsins." J Med Chem **57**(14): 6092-6104.
29. Cingolani, P., A. Platts, L. Wang le, M. Coon, T. Nguyen, L. Wang, S. J. Land, X. Lu and D. M. Ruden (2012). "A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3." Fly (Austin) **6**(2): 80-92.
30. Clustal: Multiple Sequence Alignment. (AUG,2012). "Multiple alignment of nucleic acid and protein sequences".
31. Cock, P. J. A., B. A. Gruning, K. Paszkiewicz and L. Pritchard (2013). "Galaxy tools and workflows for sequence analysis with applications in molecular plant pathology." Peerj **1**.
32. Couto, M. F., L. A. Peternelli and M. H. P. Barbosa (2013). "Classification of the coefficients of variation for sugarcane crops." Ciencia Rural **43**(6): 957-961.
33. Cutillas-Lozano, J. M. and D. Gimenez (2013). "Determination of the kinetic constants of a chemical reaction in heterogeneous phase using parameterized metaheuristics." 2013 International Conference on Computational Science **18**: 787-796.
34. Cygler, M., J. Sivaraman, P. Grochulski, R. Coulombe, A. C. Storer and J. S. Mort (1996). "Structure of rat procathepsin B: model for inhibition of cysteine protease activity by the proregion." Structure **4**(4): 405-416.
35. Dai, N. and E. T. Kool (2011). "Fluorescent DNA-based enzyme sensors." Chem Soc Rev **40**(12): 5756-5770.
36. Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. McVean, R. Durbin and G. Genomes Project Analysis (2011). "The variant call format and VCFtools." Bioinformatics **27**(15): 2156-2158.
37. Devuyst, O. (2015). "The 1000 Genomes Project: Welcome to a New World." Peritoneal Dialysis International **35**(7): 676-U677.
38. Dheer, D., J. Nicolas and R. Shankar (2019). "Cathepsin-sensitive nanoscale drug delivery systems for cancer therapy and other diseases." Adv Drug Deliv Rev.
39. Du, Y., K. Moulick, A. Rodina, J. Aguirre, S. Felts, R. Dingledine, H. Fu and G. Chiosis (2007). "High-throughput screening fluorescence polarization assay for tumor-specific Hsp90." J Biomol Screen **12**(7): 915-924.
40. Dubach, J. M., C. Vinegoni, R. Mazitschek, P. Fumene Feruglio, L. A. Cameron and R. Weissleder (2014). "In vivo imaging of specific drug-target binding at subcellular resolution." Nature Communications **5**.

41. Duffy, M. J. (1992). "The Role of Proteolytic-Enzymes in Cancer Invasion and Metastasis." Clinical & Experimental Metastasis **10**(3): 145-155.
42. Emmert-Buck, M. R., M. J. Roth, Z. Zhuang, E. Campo, J. Rozhin, B. F. Sloane, L. A. Liotta and W. G. Stetler-Stevenson (1994). "Increased gelatinase A (MMP-2) and cathepsin B activity in invasive tumor regions of human colon cancer samples." Am J Pathol **145**(6): 1285-1290.
43. Ettinger, A. and T. Wittmann (2014). "Fluorescence live cell imaging." Quantitative Imaging in Cell Biology **123**: 77-94.
44. Feixas, F., E. Matito, J. Poater and M. Sola (2011). "Understanding conjugation and hyperconjugation from electronic delocalization measures." J Phys Chem A **115**(45): 13104-13113.
45. Fox, J. G., M. Batchelder, R. Marini, L. Yan, L. Handt, X. Li, B. Shames, A. Hayward, J. Campbell and J. C. Murphy (1995). "Helicobacter Pylori-Induced Gastritis in the Domestic Cat." Infection and Immunity **63**(7): 2674-2681.
46. Fox, T., E. Demiguel, J. S. Mort and A. C. Storer (1992). "Potent Slow-Binding Inhibition of Cathepsin-B by Its Propeptide." Biochemistry **31**(50): 12571-12576.
47. French, T., W. Burton and J. C. Owicki (2000). "Enhanced fluorescence-lifetime methods for high-throughput screening." Biophysical Journal **78**(1): 443a-443a.
48. Frohlich, E., G. Schaumburg-Lever and C. Klessen (1993). "Immunoelectron microscopic localization of cathepsin B in human exocrine glands." J Cutan Pathol **20**(1): 54-60.
49. Fujii, T., M. Kamiya and Y. Urano (2014). "In Vivo Imaging of Intraperitoneally Disseminated Tumors in Model Mice by Using Activatable Fluorescent Small-Molecular Probes for Activity of Cathepsins." Bioconjugate Chemistry **25**(10): 1838-1846.
50. Galaxy. (2019). "Lakehead University High Performance Computing Centre (LUHPCC)." from <http://wesley.lakeheadu.ca:8800/>.
51. Garcia-Cattaneo, A., F. X. Gobert, M. Muller, F. Toscano, M. Flores, A. Lescure, E. Del Nery and P. Benaroch (2012). "Cleavage of Toll-like receptor 3 by cathepsins B and H is essential for signaling." Proceedings of the National Academy of Sciences of the United States of America **109**(23): 9053-9058.
52. Garland, M., J. J. Yim and M. Bogyo (2016). "A Bright Future for Precision Medicine: Advances in Fluorescent Chemical Probe Design and Their Clinical Application." Cell Chemical Biology **23**(1): 122-136.
53. Gibson, G. (2012). "Rare and common variants: twenty arguments." Nature Reviews Genetics **13**(2): 135-145.
54. Gocheva, V., W. Zeng, D. Ke, D. Klimstra, T. Reinheckel, C. Peters, D. Hanahan and J. A. Joyce (2006). "Distinct roles for cysteine cathepsin genes in multistage tumorigenesis." Genes Dev **20**(5): 543-556.

55. Gocheva, V., W. Zeng, D. X. Ke, D. Klimstra, T. Reinheckel, C. Peters, D. Hanahan and J. A. Joyce (2006). "Distinct roles for cysteine cathepsin genes in multistage tumorigenesis." Genes & Development **20**(5): 543-556.
56. Gondi, C. S. and J. S. Rao (2013). "Cathepsin B as a cancer target." Expert Opinion on Therapeutic Targets **17**(3): 281-291.
57. Gopinathan, A., G. M. Denicola, K. K. Frese, N. Cook, F. A. Karreth, J. Mayerle, M. M. Lerch, T. Reinheckel and D. A. Tuveson (2012). "Cathepsin B promotes the progression of pancreatic ductal adenocarcinoma in mice." Gut **61**(6): 877-884.
58. Goulet, B., A. Baruch, N. S. Moon, M. Poirier, L. L. Sansregret, A. Erickson, M. Bogoy and A. Nepveu (2004). "A cathepsin L isoform that is devoid of a signal peptide localizes to the nucleus in S phase and processes the CDP/Cux transcription factor." Molecular Cell **14**(2): 207-219.
59. Guo, M., P. A. Mathieu, B. Linebaugh, B. F. Sloane and J. J. Reiners (2002). "Phorbol ester activation of a proteolytic cascade capable of activating latent transforming growth factor-beta - A process initiated by the exocytosis of cathepsin B." Journal of Biological Chemistry **277**(17): 14829-14837.
60. Guo, Z., S. Park, J. Yoon and I. Shin (2014). "Recent progress in the development of near-infrared fluorescent probes for bioimaging applications." Chem Soc Rev **43**(1): 16-29.
61. Habibollahi, P., J. L. Figueiredo, P. Heidari, A. M. Dulak, Y. Imamura, A. J. Bass, S. Ogino, A. T. Chan and U. Mahmood (2012). "Optical Imaging with a Cathepsin B Activated Probe for the Enhanced Detection of Esophageal Adenocarcinoma by Dual Channel Fluorescent Upper GI Endoscopy." Theranostics **2**(2): 227-234.
62. Hamilton, D. F., M. Ghert and A. H. R. W. Simpson (2015). "Interpreting regression models in clinical outcome studies." Bone & Joint Research **4**(9): 152-153.
63. Haraksingh, R. R. and M. P. Snyder (2013). "Impacts of variation in the human genome on gene regulation." J Mol Biol **425**(21): 3970-3977.
64. Harlan, F. K., J. S. Lusk, B. M. Mohr, A. P. Guzikowski, R. H. Batchelor, Y. Jiang and J. J. Naleway (2016). "Fluorogenic Substrates for Visualizing Acidic Organelle Enzyme Activities." PLoS One **11**(5).
65. Hasnain, S., T. Hiram, C. P. Huber, P. Mason and J. S. Mort (1993). "Characterization of cathepsin B specificity by site-directed mutagenesis. Importance of Glu245 in the S2-P2 specificity for arginine and its role in transition state stabilization." J Biol Chem **268**(1): 235-240.
66. HGMD. (2020, June). "The Human Gene Mutation Database." from <http://www.hgmd.cf.ac.uk/ac/all.php>.
67. Hook, G., J. S. Jacobsen, K. Grabstein, M. Kindy and V. Hook (2015). "Cathepsin B is a New Drug Target for Traumatic Brain injury Therapeutics: evidence for E64d as a Promising Lead Drug Candidate." Frontiers in Neurology **6**.

68. Huang, S., Y. Wu, F. Zeng, J. Chen and S. Wu (2018). "A turn-on fluorescence probe based on aggregation-induced emission for leucine aminopeptidase in living cells and tumor tissue." Anal Chim Acta **1031**: 169-177.
69. Hughes, L. D., R. J. Rawle and S. G. Boxer (2014). "Choose Your Label Wisely: Water-Soluble Fluorophores Often Interact with Lipid Bilayers." Plos One **9**(2).
70. Iacobuzio-Donahue, C. A., S. Shuja, J. Cai, P. Peng and M. J. Murnane (1997). "Elevations in cathepsin B protein content and enzyme activity occur independently of glycosylation during colorectal tumor progression." J Biol Chem **272**(46): 29190-29199.
71. IGSR: The International Genome Sample Resource (2015, June 30). "IGSR and the 1000 Genomes Project."
72. Im, E., A. Venkatakrisnan and A. Kazlauskas (2005). "Cathepsin B regulates the intrinsic angiogenic threshold of endothelial cells." Mol Biol Cell **16**(8): 3488-3500.
73. Jameson, D. M. and J. A. Ross (2010). "Fluorescence polarization/anisotropy in diagnostics and imaging." Chem Rev **110**(5): 2685-2708.
74. Jedeszko, C. and B. F. Sloane (2004). "Cysteine cathepsins in human cancer." Biol Chem **385**(11): 1017-1027.
75. Kamstra, R. L., S. Dadgar, J. Wigg, M. A. Chowdhury, C. P. Phenix and W. B. Floriano (2014). "Creating and virtually screening databases of fluorescently-labelled compounds for the discovery of target-specific molecular probes." Journal of Computer-Aided Molecular Design **28**(11): 1129-1142.
76. Kamstra, R. L. and W. B. Floriano (2014). "Identifying potential selective fluorescent probes for cancer-associated protein carbonic anhydrase IX using a computational approach." J Mol Graph Model **54**: 184-193.
77. Kato, D., K. M. Boatright, A. B. Berger, T. Nazif, G. Blum, C. Ryan, K. A. H. Chehade, G. S. Salvesen and M. Bogyo (2005). "Activity-based probes that target diverse cysteine protease families." Nature Chemical Biology **1**(1): 33-38.
78. Khouri, H. E., C. Plouffe, S. Hasnain, T. Hirama, A. C. Storer and R. Menard (1991). "A Model to Explain the Ph-Dependent Specificity of Cathepsin B-Catalyzed Hydrolyzes." Biochemical Journal **275**: 751-757.
79. Kisin-Finfer, E., S. Ferber, R. Blau, R. Satchi-Fainaro and D. Shabat (2014). "Synthesis and evaluation of new NIR-fluorescent probes for cathepsin B: ICT versus FRET as a turn-ON mode-of-action." Bioorg Med Chem Lett **24**(11): 2453-2458.
80. Kisin-Finfer, E., S. Ferber, R. Blau, R. Satchi-Fainaro and D. Shabat (2014). "Synthesis and evaluation of new NIR-fluorescent probes for cathepsin B: ICT versus FRET as a turn-ON mode-of-action." Bioorganic & Medicinal Chemistry Letters **24**(11): 2453-2458.
81. Kobayashi, H., N. Moniwa, M. Sugimura, H. Shinohara, H. Ohi and T. Terao (1993). "Effects of Membrane-Associated Cathepsin-B on the Activation of Receptor-Bound Prourokinase and

- Subsequent Invasion of Reconstituted Basement-Membranes." Biochimica Et Biophysica Acta **1178**(1): 55-62.
82. Koshland, D. E. (2002). "The application and usefulness of the ratio  $k(\text{cat})/K(\text{M})$ ." Bioorg Chem **30**(3): 211-213.
83. Kostoulas, G., A. Lang, H. Nagase and A. Baici (1999). "Stimulation of angiogenesis through cathepsin B inactivation of the tissue inhibitors of matrix metalloproteinases." Febs Letters **455**(3): 286-290.
84. Kramer, L., M. Renko, J. Završnik, D. Turk, M. A. Seeger, O. Vasiljeva, M. G. Grutter, V. Turk and B. Turk (2017). "Non-invasive in vivo imaging of tumour-associated cathepsin B by a highly selective inhibitory DARPIn." Theranostics **7**(11): 2806-2821.
85. Kryczka, J., I. Papiewska-Pajak, M. A. Kowalska and J. Boncela (2019). "Cathepsin B Is Upregulated and Mediates ECM Degradation in Colon Adenocarcinoma HT29 Cells Overexpressing Snail." Cells **8**(3).
86. Lai, W. F. T., C. H. Chang, Y. Tang, R. Bronson and C. H. Tung (2004). "Early diagnosis of osteoarthritis using cathepsin B sensitive near-infrared fluorescent probes." Osteoarthritis and Cartilage **12**(3): 239-244.
87. Laine, D., M. Palovich, B. McClelland, E. Petitjean, I. Delhom, H. Xie, J. Deng, G. Lin, R. Davis, A. Jolit, N. Nevins, B. Zhao, J. Villa, J. Schneck, P. McDevitt, R. Midgett, C. Kmett, S. Umbrecht, B. Peck, A. B. Davis and D. Bettoun (2011). "Discovery of novel cyanamide-based inhibitors of cathepsin C." ACS Med Chem Lett **2**(2): 142-147.
88. Lea, W. A. and A. Simeonov (2011). "Fluorescence polarization assays in small molecule screening." Expert Opin Drug Discov **6**(1): 17-32.
89. Lee, H., J. Kim, H. Kim, Y. Kim and Y. Choi (2014). "A folate receptor-specific activatable probe for near-infrared fluorescence imaging of ovarian cancer." Chem Commun (Camb) **50**(56): 7507-7510.
90. Li, C. S., L. W. Chen, J. L. Wang, L. Y. Zhang, P. Z. Tang, S. Q. Zhai, W. W. Guo, N. Yu, L. D. Zhao, M. B. Liu and S. M. Yang (2011). "Expression and clinical significance of cathepsin B and stefin A in laryngeal cancer." Oncology Reports **26**(4): 869-875.
91. Lindgren, M., K. Sorgjerd and P. Hammarstrom (2005). "Detection and characterization of aggregates, prefibrillar amyloidogenic oligomers, and protofibrils using fluorescence spectroscopy." Biophysical Journal **88**(6): 4200-4212.
92. Lindkvist, B., I. Fajardo, G. Pejler and A. Borgstrom (2006). "Cathepsin B activates human trypsinogen 1 but not proelastase 2 or procarboxypeptidase B." Pancreatology **6**(3): 224-231.
93. Liu, W. L., D. Liu, K. Cheng, Y. J. Liu, S. Xing, P. D. Chi, X. H. Liu, N. Xue, Y. Z. Lai, L. Guo and G. Zhang (2016). "Evaluating the diagnostic and prognostic value of circulating cathepsin S in gastric cancer." Oncotarget **7**(19): 28124-28138.

94. Los, M., H. Walczak, K. Schulze-Osthoff and J. C. Reed (2000). "Fluorogenic substrates as detectors of caspase activity during natural killer cell-induced apoptosis." Methods Mol Biol **121**: 155-162.
95. Loser, R. and J. Pietzsch (2015). "Cysteine cathepsins: their role in tumor progression and recent trends in the development of imaging probes." Front Chem **3**: 37.
96. Loser, R. and J. Pietzsch (2015). "Cysteine cathepsins: their role in tumor progression and recent trends in the development of imaging probes." Frontiers in Chemistry **3**.
97. Ma, J. X., R. L. Finley, D. M. Waisman and B. F. Sloane (2000). "Human procathepsin B interacts with the annexin II tetramer on the surface of tumor cells." Journal of Biological Chemistry **275**(17): 12806-12812.
98. Mach, L., K. Stuwe, A. Hagen, C. Ballaun and J. Glossl (1992). "Proteolytic Processing and Glycosylation of Cathepsin-B - the Role of the Primary Structure of the Latent Precursor and of the Carbohydrate Moiety for Cell-Type-Specific Molecular-Forms of the Enzyme." Biochemical Journal **282**: 577-582.
99. Mahurkar, S., M. M. Idris, D. N. Reddy, S. Bhaskar, G. V. Rao, V. Thomas, L. Singh and G. R. Chandak (2006). "Association of cathepsin B gene polymorphisms with tropical calcific pancreatitis." Gut **55**(9): 1270-1275.
100. Mahurkar, S., M. M. Idris, D. N. Reddy, S. Bhaskar, G. V. Rao, V. Thomas, L. Singh and G. R. Chandak (2006). "Association of cathepsin B gene polymorphisms with tropical calcific pancreatitis." Gut **55**(9): 1270-1275.
101. Mertens, M. D., J. Schmitz, M. Horn, N. Furtmann, J. Bajorath, M. Mares and M. Gutschow (2014). "A Coumarin- Labeled Vinyl Sulfone as Tripeptidomimetic Activity- Based Probe for Cysteine Cathepsins." Chembiochem **15**(7): 955-959.
102. Micalizzi, D. S., S. M. Farabaugh and H. L. Ford (2010). "Epithelial-Mesenchymal Transition in Cancer: Parallels Between Normal Development and Tumor Progression." Journal of Mammary Gland Biology and Neoplasia **15**(2): 117-134.
103. Miyamoto, K., M. Iwadate, Y. Yanagisawa, E. Ito, J. I. Imai, M. Yamamoto, N. Sawada, M. Saito, S. Suzuki, I. Nakamura, S. Ohki, Z. Saze, M. Kogure, M. Gotoh, K. Obara, H. Ohira, K. Tasaki, M. Abe, N. Goshima, S. Watanabe, S. Waguri and S. Takenoshita (2011). "Cathepsin L is highly expressed in gastrointestinal stromal tumors." International Journal of Oncology **39**(5): 1109-1115.
104. Moerke, N. J. (2009). "Fluorescence Polarization (FP) Assays for Monitoring Peptide-Protein or Nucleic Acid-Protein Binding." Curr Protoc Chem Biol **1**(1): 1-15.
105. Molecular Operating Environment MOE. (March,2020). from <https://www.chemcomp.com/Products.htm>.
106. Molport. (2018, November 16). "molport." from <https://www.molport.com/shop/index>.
107. Morikawa, T. J., H. Fujita, A. Kitamura, T. Horio, J. Yamamoto, M. Kinjo, A. Sasaki, H. Machiyama, K. Yoshizawa, T. Ichimura, K. Imada, T. Nagai and T. M. Watanabe (2016).

- "Dependence of fluorescent protein brightness on protein concentration in solution and enhancement of it." Sci Rep **6**: 22342.
108. Mort, J. S. and D. J. Buttle (1997). "Cathepsin B." Int J Biochem Cell Biol **29**(5): 715-720.
109. Mosmann, T. (1983). "Rapid Colorimetric Assay for Cellular Growth and Survival - Application to Proliferation and Cyto-Toxicity Assays." Journal of Immunological Methods **65**(1-2): 55-63.
110. Mugerli, L., O. N. Burchak, F. Chatelain and M. Y. Balakirev (2006). "Fluorogenic ester substrates to assess proteolytic activity." Bioorg Med Chem Lett **16**(17): 4488-4491.
111. Muntener, K., R. Zwicky, G. Csucs, J. Rohrer and A. Baici (2004). "Exon skipping of cathepsin B: mitochondrial targeting of a lysosomal peptidase provokes cell death." J Biol Chem **279**(39): 41012-41017.
112. Murphy, G., R. Ward, J. Gavrilovic and S. Atkinson (1992). "Physiological mechanisms for metalloproteinase activation." Matrix Suppl **1**: 224-230.
113. Musil, D., D. Zucic, D. Turk, R. A. Engh, I. Mayr, R. Huber, T. Popovic, V. Turk, T. Towatari, N. Katunuma and et al. (1991). "The refined 2.15 Å X-ray crystal structure of human liver cathepsin B: the structural basis for its specificity." EMBO J **10**(9): 2321-2330.
114. NCBI. (2013, May). "NCBI, 1000 Genome Browser." from [https://www.ncbi.nlm.nih.gov/variation/tools/1000genomes/?assm=GCF\\_000001405.25](https://www.ncbi.nlm.nih.gov/variation/tools/1000genomes/?assm=GCF_000001405.25).
115. NCBI. (2018, Nov 13). "cathepsin B isoform X1 [Homo sapiens]." from [https://www.ncbi.nlm.nih.gov/protein/XP\\_024302839.1](https://www.ncbi.nlm.nih.gov/protein/XP_024302839.1).
116. NCBI (2018, OCT 08). "CTSB cathepsin B [ Homo sapiens (human) ]."
117. NCBI. (2019, Aug). "National Center for Biotechnology Information." from <https://www.ncbi.nlm.nih.gov/guide/genomes-maps/>.
118. Nguyen, H. H., S. H. Lee, U. J. Lee, C. D. Fermin and M. Kim (2019). "Immobilized Enzymes in Biosensor Applications." Materials (Basel) **12**(1).
119. Ofori, L. O., N. P. Withana, T. R. Prestwood, M. Verdoes, J. J. Brady, M. M. Winslow, J. Sorger and M. Bogyo (2015). "Design of Protease Activated Optical Contrast Agents That Exploit a Latent Lysosomotropic Effect for Use in Fluorescence-Guided Surgery." Acs Chemical Biology **10**(9): 1977-1988.
120. Oliveira, V., E. S. Ferro, M. D. Gomes, M. E. Oshiro, P. C. Almeida, M. A. Juliano and L. Juliano (2000). "Characterization of thiol-, aspartyl-, and thiol-metallo-peptidase activities in Madin-Darby canine kidney cells." J Cell Biochem **76**(3): 478-488.
121. Oortveld, M. A. W., I. M. J. J. van Vlijmen-Willems, F. F. J. Kersten, T. Cheng, M. Verdoes, P. E. J. van Erp, S. Verbeek, T. Reinheckel, W. J. A. J. Hendriks, J. Schalkwijk and P. L. J. M. Zeeuwen (2017). "Cathepsin B as a potential cystatin M/E target in the mouse hair follicle." Faseb Journal **31**(10): 4286-4294.

122. Owicki, J. C. (2000). "Fluorescence polarization and anisotropy in high throughput screening: Perspectives and primer." Journal of Biomolecular Screening **5**(5): 297-306.
123. Patel, S., A. Homaei, H. R. El-Seedi and N. Akhtar (2018). "Cathepsins: Proteases that are vital for survival but can also be fatal." Biomedicine & Pharmacotherapy **105**: 526-532.
124. Pavlova, A. and I. Bjork (2003). "Grafting of features of cystatins C or B into the N-terminal region or second binding loop of cystatin A (stefin A) substantially enhances inhibition of cysteine proteinases." Biochemistry **42**(38): 11326-11333.
125. Penuelas, I., G. Mazzolini, J. F. Boan, B. Sangro, J. Marti-Climent, M. Ruiz, J. Ruiz, N. Satyamurthy, C. Qian, J. R. Barrio, M. E. Phelps, J. A. Richter, S. S. Gambhir and J. Prieto (2005). "Positron emission tomography imaging of adenoviral-mediated transgene expression in liver cancer patients." Gastroenterology **128**(7): 1787-1795.
126. Phan, H. T., S. Bartelt-Hunt, K. B. Rodenhausen, M. Schubert and J. C. Bartz (2015). "Investigation of Bovine Serum Albumin (BSA) Attachment onto Self-Assembled Monolayers (SAMs) Using Combinatorial Quartz Crystal Microbalance with Dissipation (QCM-D) and Spectroscopic Ellipsometry (SE)." PLoS One **10**(10): e0141282.
127. Piston, D. W. (2010). "Fluorescence anisotropy of protein complexes in living cells." Biophys J **99**(6): 1685-1686.
128. Podgorski, I., B. E. Linebaugh, J. E. Koblinski, D. L. Rudy, M. K. Herroon, M. B. Olive and B. F. Sloane (2009). "Bone Marrow-Derived Cathepsin K Cleaves SPARC in Bone Metastasis." American Journal of Pathology **175**(3): 1255-1269.
129. Podgorski, I. and B. F. Sloane (2003). "Cathepsin B and its role(s) in cancer progression." Proteases and the Regulation of Biological Processes **70**: 263-276.
130. Podobnik, M., R. Kuhelj, V. Turk and D. Turk (1997). "Crystal structure of the wild-type human procathepsin B at 2.5 angstrom resolution reveals the native active site of a papain-like cysteine protease zymogen." Journal of Molecular Biology **271**(5): 774-788.
131. Posimo, J. M., A. S. Unnithan, A. M. Gleixner, H. J. Choi, Y. R. Jiang, S. H. Pulugulla and R. K. Leak (2014). "Viability Assays for Cells in Culture." Love-Journal of Visualized Experiments(83).
132. Prism. (2018, November 16). "GraphPad Prism." from <https://www.graphpad.com/>.
133. Public Health Agency of Canada. (2010, March 29). "CANADIAN CANCER SOCIETY." from <https://www.canada.ca/en/public-health.html>.
134. Puri, N. and P. A. Roche (2008). "Mast cells possess distinct secretory granule subsets whose exocytosis is regulated by different SNARE isoforms." Proceedings of the National Academy of Sciences of the United States of America **105**(7): 2580-2585.
135. RCSB PDB. (2018, OCT 30). "A Structural View of Biology." from <http://www.rcsb.org/>.



136. Redzyna, I., A. Ljunggren, M. Abrahamson, J. S. Mort, J. C. Krupa, M. Jaskolski and G. Bujacz (2008). "Displacement of the occluding loop by the parasite protein, chagasin, results in efficient inhibition of human cathepsin B." J Biol Chem **283**(33): 22815-22825.
137. Reiser, J., B. Adair and T. Reinheckel (2010). "Specialized roles for cysteine cathepsins in health and disease." J Clin Invest **120**(10): 3421-3431.
138. Reiser, J., J. Oh, I. Shirato, K. Asanuma, A. Hug, T. M. Mundel, K. Honey, K. Ishidoh, E. Kominami, J. A. Kreidberg, Y. Tomino and P. Mundel (2004). "Podocyte migration during nephrotic syndrome requires a coordinated interplay between cathepsin L and alpha3 integrin." J Biol Chem **279**(33): 34827-34832.
139. Rottiers, P., F. Saltel, T. Daubon, B. Chaigne-Delalande, V. Tridon, C. Billottet, E. Reuzeau and E. Genot (2009). "TGF beta-induced endothelial podosomes mediate basement membrane collagen degradation in arterial vessels." Journal of Cell Science **122**(23): 4311-4318.
140. Rowan, A. D., P. Mason, L. Mach and J. S. Mort (1992). "Rat procathepsin B. Proteolytic processing to the mature form in vitro." J Biol Chem **267**(22): 15993-15999.
141. Ruan, H., S. Hao, P. Young and H. Zhang (2015). "Targeting Cathepsin B for Cancer Therapies." Horiz Cancer Res **56**: 23-40.
142. Ryu, J. H., J. H. Na, H. K. Ko, D. G. You, S. Park, E. Jun, H. J. Yeom, D. H. Seo, J. H. Park, S. Y. Jeong, I. S. Kim, B. S. Kim, I. C. Kwon, K. Choi and K. Kim (2014). "Non-invasive optical imaging of cathepsin B with activatable fluorogenic nanoprobe in various metastatic models." Biomaterials **35**(7): 2302-2311.
143. Sameiro, M. and T. Goncalves (2009). "Fluorescent Labeling of Biomolecules with Organic Probes." Chemical Reviews **109**(1): 190-212.
144. Sever, S., M. M. Altintas, S. R. Nankoe, C. C. Moller, D. Ko, C. L. Wei, J. Henderson, E. C. del Re, L. Hsing, A. Erickson, C. D. Cohen, M. Kretzier, D. Kerjaschki, A. Rudensky, B. Nikolic and J. Reiser (2007). "Proteolytic processing of dynamin by cytoplasmic cathepsin L is a mechanism for proteinuric kidney disease." Journal of Clinical Investigation **117**(8): 2095-2104.
145. Somanna, A., V. Mundodi and L. Gedamu (2002). "Functional analysis of cathepsin B-like cysteine proteases from *Leishmania donovani* complex - Evidence for the activation of latent transforming growth factor beta." Journal of Biological Chemistry **277**(28): 25305-25312.
146. Sobic, I., B. Mirkovic, K. Arenz, B. Stefane, J. Kos and S. Gobec (2013). "Development of New Cathepsin B Inhibitors: Combining Bioisosteric Replacements and Structure-Based Design To Explore the Structure-Activity Relationships of Nitroxoline Derivatives." Journal of Medicinal Chemistry **56**(2): 521-533.
147. Spano, F. C. and C. Silva (2014). "H- and J-aggregate behavior in polymeric semiconductors." Annu Rev Phys Chem **65**: 477-500.
148. Sui, H. Y., C. X. Shi, Z. P. Yan and M. Wu (2016). "Overexpression of Cathepsin L is associated with chemoresistance and invasion of epithelial ovarian cancer." Oncotarget **7**(29): 45995-46001.

149. Sun, R. L., Y. Zhang, Q. S. Lv, B. Liu, M. Jin, W. J. Zhang, Q. He, M. J. Deng, X. T. Liu, G. C. Li, Y. H. Li, G. H. Zhou, P. L. Xie, X. M. Xie, J. Y. Hu and Z. J. Duan (2011). "Toll-like Receptor 3 (TLR3) Induces Apoptosis via Death Receptors and Mitochondria by Up-regulating the Transactivating p63 Isoform alpha (TAP63 alpha)." Journal of Biological Chemistry **286**(18).
150. Szpaderska, A. M. and A. Frankfater (2001). "An intracellular form of cathepsin B contributes to invasiveness in cancer." Cancer Res **61**(8): 3493-3500.
151. Tan, G. J., Z. K. Peng, J. P. Lu and F. Q. Tang (2013). "Cathepsins mediate tumor metastasis." World J Biol Chem **4**(4): 91-101.
152. Tannock, I. F. and D. Rotin (1989). "Acid Ph in Tumors and Its Potential for Therapeutic Exploitation." Cancer Research **49**(16): 4373-4384.
153. Terasawa, Y., T. Hotani, Y. Katayama, M. Tachibana, H. Mizuguchi and F. Sakurai (2015). "Activity levels of cathepsins B and L in tumor cells are a biomarker for efficacy of reovirus-mediated tumor cell killing." Cancer Gene Therapy **22**(4): 188-197.
154. The International Genome Sample Resource. (2008). "IGSR and the 1000 Genomes Project." from <https://www.internationalgenome.org/>.
155. Thiery, J. P., H. Acloque, R. Y. J. Huang and M. A. Nieto (2009). "Epithelial-Mesenchymal Transitions in Development and Disease." Cell **139**(5): 871-890.
156. Thornberry, N. A., K. T. Chapman and D. W. Nicholson (2000). "Determination of caspase specificities using a peptide combinatorial library." Apoptosis **322**: 100-110.
157. Tian, R., M. Li, J. Wang, M. Yu, X. Kong, Y. Feng, Z. Chen, Y. Li, W. Huang, W. Wu and Z. Hong (2014). "An intracellularly activatable, fluorogenic probe for cancer imaging." Org Biomol Chem **12**(29): 5365-5374.
158. Turk, D., M. Podobnik, R. Kuhelj, M. Dolinar and V. Turk (1996). "Crystal structures of human procathepsin B at 3.2 and 3.3 angstrom resolution reveal an interaction motif between a papain-like cysteine protease and its propeptide." Febs Letters **384**(3): 211-214.
159. Turk, D., M. Podobnik, T. Popovic, N. Katunuma, W. Bode, R. Huber and V. Turk (1995). "Crystal structure of cathepsin B inhibited with CA030 at 2.0-A resolution: A basis for the design of specific epoxy-succinyl inhibitors." Biochemistry **34**(14): 4791-4797.
160. Turk, V., J. Kos and B. Turk (2004). "Cysteine cathepsins (proteases)--on the main stage of cancer?" Cancer Cell **5**(5): 409-410.
161. Turk, V., V. Stoka, O. Vasiljeva, M. Renko, T. Sun, B. Turk and D. Turk (2012). "Cysteine cathepsins: from structure, function and regulation to new frontiers." Biochim Biophys Acta **1824**(1): 68-88.
162. Turk, V., V. Stoka, O. Vasiljeva, M. Renko, T. Sun, B. Turk and D. Turk (2012). "Cysteine cathepsins: From structure, function and regulation to new frontiers." Biochimica Et Biophysica Acta-Proteins and Proteomics **1824**(1): 68-88.

163. UCSC Genes. (2018, NOV 24). "Human Gene CTSB ", from [https://genome.ucsc.edu/cgi-bin/hgTracks?hgtgroup\\_map\\_close=1&hgtgroup\\_genes\\_close=0&hgtgroup\\_phenDis\\_close=0&hgtgroup\\_rna\\_close=0&hgtgroup\\_expression\\_close=0&hgtgroup\\_regulation\\_close=0&hgtgroup\\_compGeno\\_close=0&hgtgroup\\_neandertal\\_close=1&hgtgroup\\_denisova\\_close=1&hgtgroup\\_varRep\\_close=0&hgtgroup\\_rep\\_close=0&hgsid=695285183\\_aSXeAbtWwDaGcalhfmaqZF50keRA&position=CTSB&hgt.positionInput=CTSB&hgt.jump=go&hgt.suggestTrack=knownGene&db=hg19&c=chr21&l=33031596&r=33041570&pix=950&dinkL=2.0&dinkR=2.0](https://genome.ucsc.edu/cgi-bin/hgTracks?hgtgroup_map_close=1&hgtgroup_genes_close=0&hgtgroup_phenDis_close=0&hgtgroup_rna_close=0&hgtgroup_expression_close=0&hgtgroup_regulation_close=0&hgtgroup_compGeno_close=0&hgtgroup_neandertal_close=1&hgtgroup_denisova_close=1&hgtgroup_varRep_close=0&hgtgroup_rep_close=0&hgsid=695285183_aSXeAbtWwDaGcalhfmaqZF50keRA&position=CTSB&hgt.positionInput=CTSB&hgt.jump=go&hgt.suggestTrack=knownGene&db=hg19&c=chr21&l=33031596&r=33041570&pix=950&dinkL=2.0&dinkR=2.0).
164. UCSC Genome Browser. (2013, Dec). "Gene Expression in 53 tissues from GTEx RNA-seq of 8555 samples (570 donors) (CTSB)." from [https://genome.ucsc.edu/cgi-bin/hgc?hgsid=748683097\\_9kDdKtdWORlvjobwAOVK1EKpsFA1&c=chr8&l=11842523&r=1868229&o=11842523&t=11869448&g=gtexGene&i=CTSB](https://genome.ucsc.edu/cgi-bin/hgc?hgsid=748683097_9kDdKtdWORlvjobwAOVK1EKpsFA1&c=chr8&l=11842523&r=1868229&o=11842523&t=11869448&g=gtexGene&i=CTSB)
165. UniProtKB - P07858 (CATB\_HUMAN). (2018, May 25). "UniProtKB." from <https://www.uniprot.org/uniprot/P07858>.
166. UniProtKB. (2018, Sep 09). from <https://www.uniprot.org/>.
167. Verdoes, M., K. O. Bender, E. Segal, W. A. van der Linden, S. Syed, N. P. Withana, L. E. Sanman and M. Bogyo (2013). "Improved Quenched Fluorescent Probe for Imaging of Cysteine Cathepsin Activity." Journal of the American Chemical Society **135**(39): 14726-14730.
168. Verdoes, M., K. Oresic Bender, E. Segal, W. A. van der Linden, S. Syed, N. P. Withana, L. E. Sanman and M. Bogyo (2013). "Improved quenched fluorescent probe for imaging of cysteine cathepsin activity." J Am Chem Soc **135**(39): 14726-14730.
169. Vinegoni, C., J. M. Dubach, P. F. Feruglio and R. Weissleder (2016). "Two-photon Fluorescence Anisotropy Microscopy for Imaging and Direct Measurement of Intracellular Drug Target Engagement." IEEE J Sel Top Quantum Electron **22**(3).
170. Vizovisek, M., M. Fonovic and B. Turk (2019). "Cysteine cathepsins in extracellular matrix remodeling: Extracellular matrix degradation and beyond." Matrix Biology **75-76**: 141-159.
171. VMD. (APR,2008). "Visual molecular dynamics." from <https://www.ks.uiuc.edu/Research/vmd/>.
172. Wang, Y., J. Li, L. Feng, J. Yu, Y. Zhang, D. Ye and H. Y. Chen (2016). "Lysosome-Targeting Fluorogenic Probe for Cathepsin B Imaging in Living Cells." Anal Chem **88**(24): 12403-12410.
173. Wang, Y. Q., J. B. Li, L. D. Feng, J. F. Yu, Y. Zhang, D. J. Ye and H. Y. Chen (2016). "Lysosome-Targeting Fluorogenic Probe for Cathepsin B Imaging in Living Cells." Analytical Chemistry **88**(24): 12403-12410.
174. Wartmann, T., J. Mayerle, T. Kahne, M. Sahin-Toth, M. Ruthenburger, R. Matthias, A. Kruse, T. Reinheckel, C. Peters, F. U. Weiss, M. Sendler, H. Lippert, H. U. Schulz, A. Aghdassi, A. Dummer, S. Teller, W. Halangk and M. M. Lerch (2010). "Cathepsin L inactivates human trypsinogen, whereas cathepsin L-deletion reduces the severity of pancreatitis in mice." Gastroenterology **138**(2): 726-737.

175. Wei, B., J. Gunzner-Toste, H. Yao, T. Wang, J. Wang, Z. Xu, J. Chen, J. Wai, J. Nonomiya, S. P. Tsai, J. Chuh, K. R. Kozak, Y. Liu, S. F. Yu, J. Lau, G. Li, G. D. Phillips, D. Leipold, A. Kamath, D. Su, K. Xu, C. Eigenbrot, S. Steinbacher, R. Ohri, H. Raab, L. R. Staben, G. Zhao, J. A. Flygare, T. H. Pillow, V. Verma, L. A. Masterson, P. W. Howard and B. Safina (2018). "Discovery of Peptidomimetic Antibody-Drug Conjugate Linkers with Enhanced Protease Specificity." J Med Chem **61**(3): 989-1000.
176. Whitcomb, D. C. (2004). "Inflammation and cancer - V. Chronic pancreatitis and pancreatic cancer." American Journal of Physiology-Gastrointestinal and Liver Physiology **287**(2): G315-G319.
177. Zhang, X., S. Bloch, W. Akers and S. Achilefu (2012). "Near-infrared molecular probes for in vivo imaging." Curr Protoc Cytom **Chapter 12**: Unit12 27.
178. Zhu, L., C. Yang and J. Qin (2008). "An aggregation-induced blue shift of emission and the self-assembly of nanoparticles from a novel amphiphilic oligofluorene." Chem Commun (Camb)(47): 6303-6305.