

**The Stabilisation of the Cysteine Protease of *Carica papaya* (Papain) and  
the Catalytic Domain of the Cathepsin L-like Cysteine Protease of  
*Trypanosoma congolense* (TcoCATL)**

by

**Ryan Chetty**

**BSc (Hons)**

**Submitted in the fulfilment of the**

**academic requirements for the**

**MSc degree**

in

Biochemistry

School of Life Sciences

College of Agriculture, Engineering and Science

University of KwaZulu-Natal

Pietermaritzburg

South Africa

2019

## **PREFACE**

This dissertation represents the experimental work performed in the School of Life Sciences, University of KwaZulu-Natal, Pietermaritzburg. The data and findings in this dissertation have been viewed at the 25<sup>th</sup> Congress of the SASBMB Conference which was held in East London from the 10<sup>th</sup> July to the 14<sup>th</sup> July 2016.



---

Ryan Chetty

20 June 2019

As the Supervisor I agree to the submission of this dissertation

---

Dr Raymond Hewan

20 June 2019

As the Co-supervisor I agree to the submission of this dissertation

---

Professor Theresa Coetzer

20 June 2019

## **DECLARATION – PLAGIARISM**

I, Ryan Chetty, declare that

1. The research reported in this dissertation, except where otherwise indicated, is my original research.
2. This dissertation has not been submitted for any degree or examination at any other university.
3. This dissertation does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.
4. This dissertation does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:
  - a. Their words have been re-written, but the general information attributed to them has been referenced
  - b. Where their exact words have been used, then their writing has been placed in italics and inside quotation marks and referenced.
5. This dissertation does not contain text, graphics or tables copied and pasted from the Internet, unless specifically acknowledged, and the source being detailed in the dissertation and in the References section.



---

Ryan Chetty

20 June 2019

## ABSTRACT

Proteases play an intricate role in the numerous functions of a living organism. Proteases are responsible for the cleavage of proteins into smaller fragments by catalysing the hydrolysis of peptide bonds. The class of cysteine proteases have a cysteine thiol group in their active site have been found in lower and higher organisms. They have been investigated as promising drug targets for various diseases due to their fundamental functions in catabolism and protein processing. The thermal stability of a protease is a key characteristic feature that is largely dependent on its amino acid sequence and composition and quantified through the determination of its melting temperature ( $T_M$ ). Papain is the most well characterised cysteine protease and is commonly used as a model for other cysteine proteases. Congopain is the major cysteine protease of *Trypanosoma congolense* which has been identified as the main causative agent of trypanosomiasis in livestock. The thermal stability for papain and congopain were investigated in this study *via* the thermal shift assay. Papain was purchased and the catalytic domain of congopain was expressed using the *Pichia pastoris* yeast expression system. The thermal stability of the proteases were determined under neutral pH conditions and the effect of pH and ligand binding were evaluated to determine if the proteases could be further stabilised. The most stable forms of papain and the catalytic domain of congopain in its monomeric form was observed at pH 5.0 with 50  $\mu$ M chymostatin. The thermal stability of both cysteine proteases was successfully evaluated *via* the thermal shift assay and conditions to further stabilise papain and the catalytic domain of congopain were determined. The thermal shift assay has been proven to be a reliable technique in identifying factors which increase the stability of a protein. More specifically, the technique serves as a simple and primary diagnostic tool to screen potential inhibitors of a protein and detect changes in the  $T_M$  of a protein.

## **ACKNOWLEDGEMENTS**

I would firstly like to thank the Lord Jesus Christ because everything is possible with him. I am grateful for every opportunity which has led me to this point in my life. I would like to express my gratitude and love to my wonderful family for their support and love during this stressful time. To my parents for their unwavering belief in my abilities and providing me with an amazing blessed life, I am forever grateful. I love you and hope I can continue to make you proud. To my amazing little sister who always wanted to help me whenever she knew I had work to do and to my older sisters for always wanting to help me destress whenever they got a chance. To my girlfriend, Darian Rajah, thank you for your support, understanding and patience and allowing me to follow my dreams, love you now and forever.

I would like to thank Dr Raymond Hewer for his guidance and patience for the duration of this Masters project. Your belief in me has helped me more than you could ever know. Thank you for the time and effort you have put into not only this dissertation but my growth as a scientist and as a human being. You are not only an awesome supervisor but also a wonderful human being and I am thankful for your guidance and allowing me to be a part of your laboratory team. I look forward to working with you in the future.

I would like to express my gratitude to Prof Theresa Coetzer for offering me the opportunity to work under her supervision for this study. I appreciate the time, effort and resources that you generously granted me.

I would like to thank my fellow lab members of Lab 46. It has been a pleasure working with each and every one of you and wish you all everything of the best in the future. I would also like to thank Jessica, our biochemistry technician, you have always been so kind and helpful, and I appreciate all of your contributions.

Finally, I would like to thank the National Research Foundation (NRF) for the funding of this study.

# TABLE OF CONTENTS

<b>PREFACE</b> .....	i
<b>DECLARATION – PLAGIARISM</b> .....	ii
<b>ABSTRACT</b> .....	iii
<b>ACKNOWLEDGEMENTS</b> .....	iv
<b>LIST OF FIGURES</b> .....	viii
<b>LIST OF TABLES</b> .....	xi
<b>LIST OF ABBREVIATIONS</b> .....	xii
<b>1. LITERATURE REVIEW</b> .....	1
<b>1.1. Function, classification and nomenclature of proteases</b> .....	1
<b>1.2. Cysteine Proteases</b> .....	2
<b>1.3. Papain</b> .....	3
<b>1.3.1. Inhibitors of Papain</b> .....	5
<b>1.3.2. Applications and uses of Papain</b> .....	5
<b>1.4. A cysteine protease from <i>Trypanosoma congolense</i> (<i>TcoCATL</i>)</b> .....	5
<b>1.4.1. Inhibitors of <i>TcoCATL</i></b> .....	9
<b>1.4.2. Applications and uses of <i>TcoCATL</i></b> .....	9
<b>1.5. Protein stability</b> .....	9
<b>1.5.1. Approaches to increase protein stability</b> .....	11
<b>1.5.2. Mutations</b> .....	11
<b>1.5.3. Buffer composition</b> .....	12
<b>1.5.4. Ligands</b> .....	12
<b>1.6. Techniques used to determine protein stability</b> .....	13
<b>1.6.1. Thermal shift assay</b> .....	18
<b>1.7. Hypothesis, Aims and Objectives of the study</b> .....	22
<b>2. MATERIALS AND METHODS</b> .....	23
<b>2.1. Materials</b> .....	23
<b>2.1.1. Expression of recombinant <i>TcoCATL</i><sub>cat</sub></b> .....	23
<b>2.1.2. Analysis of expressed <i>TcoCATL</i><sub>cat</sub></b> .....	23
<b>2.1.3. Stability determination <i>via</i> thermal shift assay</b> .....	23
<b>2.2. Methods</b> .....	24
<b>2.2.1. Analysis of the amino acid sequence for papain and the catalytic domain of <i>TcoCATL</i></b> .....	24

2.2.2. Recombinant expression of the catalytic domain of <i>TcoCATL</i> from <i>T. congolense</i> .....	25
2.2.3. Analysis of expressed <i>TcoCATL<sub>cat</sub></i> .....	26
2.2.3.1. Determination of protein concentration using the Bradford assay .....	26
2.2.3.2. Determination of the molecular weight of the expressed protein using reducing and non-reducing SDS-PAGE .....	27
2.2.3.3. Confirmation of expressed protein as the catalytic domain of <i>TcoCATL</i> via western Blot .....	27
2.2.3.4. Gelatin substrate containing SDS-PAGE zymogram .....	28
2.2.3.5. Dimerisation of the catalytic domain of <i>TcoCATL</i> .....	29
2.2.4. Control thermal shift assay using a Protein Thermal Shift™ Starter Kit .....	29
2.2.5. Optimisation of the thermal shift assay .....	30
2.2.6. Determining the stability of papain using the optimised thermal shift assay .....	30
2.2.6.1. Determining the effect of pH on the stability of papain .....	30
2.2.6.2. Determining the effect of ligand binding on the stability of papain .....	30
2.2.6.3. Determining the combined effect of pH and ligand binding on the stability of papain .....	31
2.2.7. Determining the stability of <i>TcoCATL<sub>cat</sub></i> via the thermal shift assay .....	31
2.2.7.1. Determining the effect of pH on the stability of <i>TcoCATL<sub>cat</sub></i> .....	31
2.2.7.2. Determining the effect of ligand binding on the stability of <i>TcoCATL<sub>cat</sub></i> .....	32
2.2.7.3. Determining the combined effect of pH and ligand binding on the stability of <i>TcoCATL<sub>cat</sub></i> .....	32
2.2.8. Statistical analysis .....	33
3. RESULTS .....	34
3.1. Hydrophobicity analysis of papain .....	34
3.2. Hydrophobicity analysis of <i>TcoCATL<sub>cat</sub></i> .....	37
3.3. The expression of recombinant <i>TcoCATL<sub>cat</sub></i> .....	38
3.4. Evaluation of the dimerisation of <i>TcoCATL<sub>cat</sub></i> .....	42
3.5. Thermal shift assay .....	43
3.5.1. Thermal shift kit .....	43
3.6. Determining the stability of papain via the TSA .....	45
3.6.1. Determining the effect of pH on the stability of papain .....	45
3.6.2. Determining the effect of ligand binding on the stability of papain .....	48
3.6.3. Determining the combined effect of pH and ligand binding on the stability of papain .....	53
3.7. Determining the stability of <i>TcoCATL<sub>cat</sub></i> via the thermal shift assay .....	54

3.7.1. Determining the effect of pH on the stability of <i>TcoCATL<sub>cat</sub></i> .....	54
3.7.2. Determining the effect of ligand binding on the stability of <i>TcoCATL<sub>cat</sub></i> .....	57
3.7.3. Determining the combined effect of pH and ligand binding on the stability of <i>TcoCATL<sub>cat</sub></i> .....	61
4. DISCUSSION.....	64
4.1. Papain and <i>TcoCATL<sub>cat</sub></i> are suitable for use in the thermal shift assay .....	65
4.2. Biochemical characterisation of the cysteine proteases selected for thermal shift analysis .....	66
4.3. The stability of the cysteine proteases under neutral pH conditions.....	67
4.4. The factors contributing towards the stabilisation of the two cysteine proteases.....	69
4.4.1. Increasing stability of the cysteine proteases by decreasing pH.....	69
4.4.2. Stabilisation of the selected proteases <i>via</i> the introduction of a ligand.....	70
4.4.3. Increased protease stabilisation by pH manipulation and ligand binding .	73
4.5. The robustness and limitations of the thermal shift assay .....	73
4.6. Conclusion.....	75
4.7. Future Studies.....	76
5. REFERENCES .....	77
APPENDIX A .....	89



## LIST OF FIGURES

Figure 1.1:	Nomenclature of the substrate specificity of a protease.....	2
Figure 1.2:	The mechanism of peptide hydrolysis by cysteine proteases.....	3
Figure 1.3:	A ribbon representation for a cysteine protease from the papain family...	4
Figure 1.4:	A homology model (ribbon representation) for <i>TcoCATL<sub>cat</sub></i> based on using papain, actinidin, papaya protease omega and <i>TcrCATL</i> as template proteins.....	7
Figure 1.5:	A homology-based model for the catalytic domain of <i>TcoCATL<sub>cat</sub></i> emphasising the S2 subsite substitution found in <i>TcoCATL</i> .....	8
Figure 1.6:	Thermal denaturation curves for porcine odorant-binding protein (OBP-I) in the presence and absence of ligands.....	14
Figure 1.7:	The progressive appearance of unfolded human $\alpha$ -lactalbumin as the number of NMR signals increase with an increase in urea concentration from panel a-d.....	16
Figure 1.8:	Differential scanning calorimetry (DSC) experiment for the two-state unfolding of a globular protein.....	18
Figure 1.9:	Melt curve plot of hen egg-white lysozyme at two different concentrations.....	19
Figure 1.10:	Derivate plot of hen egg-white lysozyme at two different concentrations.....	20
Figure 1.11:	The increase in stability of a protein due to ligand-protein interaction, as detected through the thermal shift assay.....	21
Figure 1.12:	Melt curve plot for porcine citrate synthase (1 $\mu$ M) screened against three different buffers in the presence and absence of 100 $\mu$ M oxaloacetate.....	22
Figure 2.1:	Standard curve for the Bradford protein assay.....	26

Figure 3.1:	The amino acid sequence for papain with hydrophobic residues highlighted.....	34
Figure 3.2:	Analysis of the hydrophobicity of papain.....	36
Figure 3.3:	The amino acid sequence for <i>TcoCATL<sub>cat</sub></i> with hydrophobic residues highlighted.....	37
Figure 3.4:	Analysis of the hydrophobicity of <i>TcoCATL<sub>cat</sub></i> .....	39
Figure 3.5:	SDS-PAGE analysis for the expressed <i>TcoCATL<sub>cat</sub></i> on 12% TGX FastCast acrylamide gels.....	38
Figure 3.6:	The analysis of the expressed recombinant <i>TcoCATL<sub>cat</sub></i> <i>via</i> western blot.....	40
Figure 3.7:	Qualitative determination of enzymatic activity of <i>TcoCATL<sub>cat</sub></i> <i>via</i> non-reducing SDS-PAGE and gelatin-containing SDS-PAGE analysis with papain enzymatic activity on a gelatin-containing SDS-PAGE. ....	41
Figure 3.8:	An SDS-PAGE displaying the dimerisation of <i>TcoCATL<sub>cat</sub></i> due to pH...42	
Figure 3.9:	The evaluation of observed dimer formation of <i>TcoCATL<sub>cat</sub></i> under neutral pH conditions using western blot.....	43
Figure 3.10:	Raw fluorescence obtained for the model protein in the absence and presence of the control ligand at various concentrations.....	44
Figure 3.11:	Differential scanning fluorimetry analysis obtained for the thermal shift control protein in the absence and presence of the control ligand at various concentrations.....	45
Figure 3.12:	Differential scanning fluorimetry analysis obtained for papain under different pH conditions.....	47
Figure 3.13:	A representative differential scanning fluorimetry analysis obtained for papain with compounds from the ChemBridge library.....	52
Figure 3.14:	Differential scanning fluorimetry analysis of papain with increasing concentrations of chymostatin.....	52

Figure 3.15:	Differential scanning fluorimetry analysis obtained demonstrating the effectiveness of a buffer-ligand combination on the stability of papain.....	54
Figure 3.16:	A representative differential scanning fluorimetry analysis obtained for the melting point of <i>TcoCATL<sub>cat</sub></i> in different Acetate-MES-Tris (AMT) buffers.....	55
Figure 3.17:	A representative differential scanning fluorimetry analysis obtained for the melting point of <i>TcoCATL<sub>cat</sub></i> at different pH using the Reinhard buffer library.....	56
Figure 3.18:	A representative differential scanning fluorimetry analysis of <i>TcoCATL<sub>cat</sub></i> using compounds from the ChemBridge library and using chymostatin as a positive control.....	61
Figure 3.19	Differential scanning fluorimetry analysis obtained for the combined effect of pH and ligand binding on the stability of <i>TcoCATL<sub>cat</sub></i> .....	62

## LIST OF TABLES

Table 3.1:	The total number of each hydrophobic amino acid present in the papain sequence with relative hydrophobicity scores as per the Kyte and Doolittle index.....	34
Table 3.2:	The total number of each hydrophobic amino acid present in the <i>TcoCATL<sub>cat</sub></i> sequence with relative hydrophobicity scores as per the Kyte and Doolittle index.....	37
Table 3.3:	The average melting temperature obtained for papain under various buffer conditions.....	46
Table 3.4:	The average melting temperature obtained for papain in the presence of distinct compounds and within phosphate buffered saline (pH 7.4) by means of thermal shift analysis.....	48
Table 3.5:	The average $T_M$ obtained for papain when testing combinations of pH and 50 $\mu$ M chymostatin.....	53
Table 3.6:	The average melting temperature obtained for <i>TcoCATL<sub>cat</sub></i> under various buffer conditions.....	55
Table 3.7:	The average melting temperature obtained for <i>TcoCATL<sub>cat</sub></i> in the presence of distinct compounds and within phosphate buffered saline (pH 7.4) as calculated through thermal shift analysis.....	57
Table 3.8:	The average $T_M$ obtained for <i>TcoCATL<sub>cat</sub></i> in the presence of 50 $\mu$ M chymostatin and varying pH.....	62
Table A.1:	The chemical structures of the 50 compounds obtained from the ChemBridge library of compounds.....	78

## LIST OF ABBREVIATIONS

1D	One dimensional
$\beta$ -Me	$\beta$ -mercaptoethanol
BMGY	Buffered medium glycerol yeast
BMM	Buffered minimal medium
BSA	Bovine serum albumin
CD	Circular dichroism
CP	Cysteine protease
Cp	Excess heat capacity
dH <sub>2</sub> O	Distilled water
DSC	Differential scanning calorimetry
DSF	Differential scanning fluorimetry
DTT	Dithiothreitol
E-64	L-transepoxy succinyl-4-guanidinobutane
EDTA	Ethylenediaminetetraacetic acid
FTIR	Fourier-transform infrared spectroscopy
HCA	Hydrophobicity cluster analysis
NMR	Nuclear magnetic resonance
PDB	Protein data bank
RT-PCR	Real time – polymerisation chain reaction
TBS	Tris buffered saline
<i>Tco</i> CATL	Congopain
<i>Tco</i> CATL <sub>cat</sub>	Catalytic domain of <i>Tco</i> CATL
<i>Tcr</i> CATL	Cruzipain
TLCK	Tosyl lysyl chloromethyl ketone
TPCK	Tosyl phenylalanyl chloromethyl ketone
TPP	Three phase partitioning
TSA	Thermal shift assay
VSG	Variable surface glycoprotein
YP	Yeast extract, peptone
YPD	Yeast extract, peptone, dextrose

# 1. LITERATURE REVIEW

---

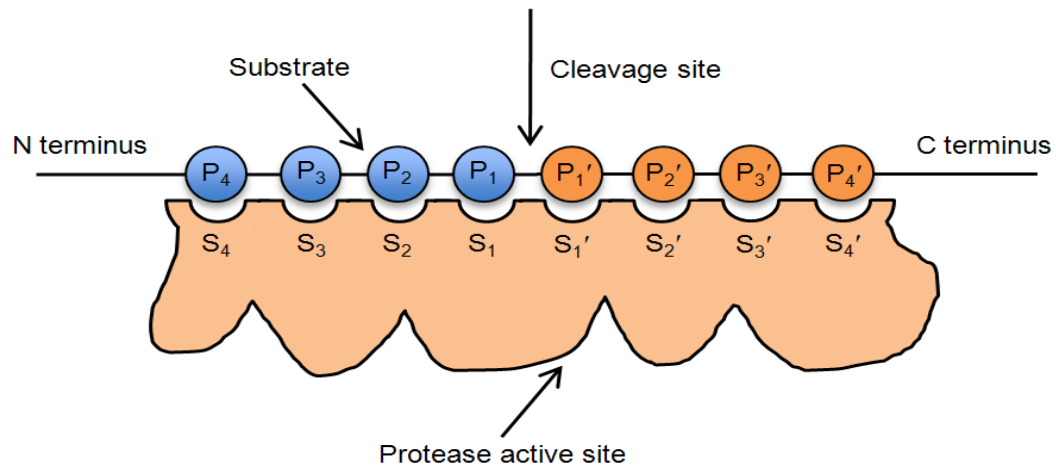
## 1.1. Function, classification and nomenclature of proteases

Proteases (also known as proteinases, proteolytic enzymes or peptidases) are responsible for the cleavage of proteins into peptides and amino acids as they catalyse the hydrolysis of peptide bonds. Numerous biological processes, in organisms ranging from bacteria and parasites to higher organisms such as mammals, are fundamentally dependent on this catalytic action (Otto and Schirmeister, 1997; Grzonka *et al*, 2001). Proteases are classified as either endoproteases or exoproteases – dependent on the site of hydrolysis in the peptide chain. Endoproteases catalyse the cleavage of peptide bonds within the polypeptide chain while exoproteases catalyse the cleavage of peptide bonds at the N-terminal or C-terminal ends (Otto and Schirmeister, 1997; Mótyán *et al*, 2013). Proteases are further classified - according to the reactive groups at the active site of the enzyme - as aspartic, glutamic, metallo, serine, threonine and cysteine proteases (CPs) or asparagine peptide lyase.

For catalysis, the aspartic, glutamic and metallo proteases activate a water molecule which then undergoes a nucleophilic attack on a peptide bond. Serine, threonine and cysteine proteases employ a two-step hydrolysis mechanism whereby a residue within the enzyme active site is first activated to act as a nucleophile to attack the substrate peptide bond. The nucleophilic attack results in the formation of an intermediate comprising of the enzyme covalently linked to the N-terminal half of the substrate. Hydrolysis is completed *via* the activation of water to hydrolyse the intermediate. These classes of proteases can then be classified into different families based on their sequence similarities and within each family they can be categorised into different clans based on structure-based classification (Rawlings and Barrett, 1993; Rawlings *et al*, 2009; Rawlings *et al*, 2018).

The cleavage site specificity of proteases is described using nomenclature proposed by Schechter and Berger (1967). As illustrated in Figure 1.1, the subsites on the protease that interact with the substrate amino acid residues are labelled S and the substrate amino acid residues are labelled P (Schechter and Berger, 1967; Hooper, 2002). The subsites are numbered outward in both directions from the catalytic site where S1, S2, S3, etc are numbered towards the N-terminal and S1', S2', S3', etc are numbered towards

the C-terminal. The substrate amino acid residues are numbered from the scissile bond outward to match the subsites of the protease (Schechter and Berger, 1967; Hooper, 2002; Laskar and Chatterjee, 2009).

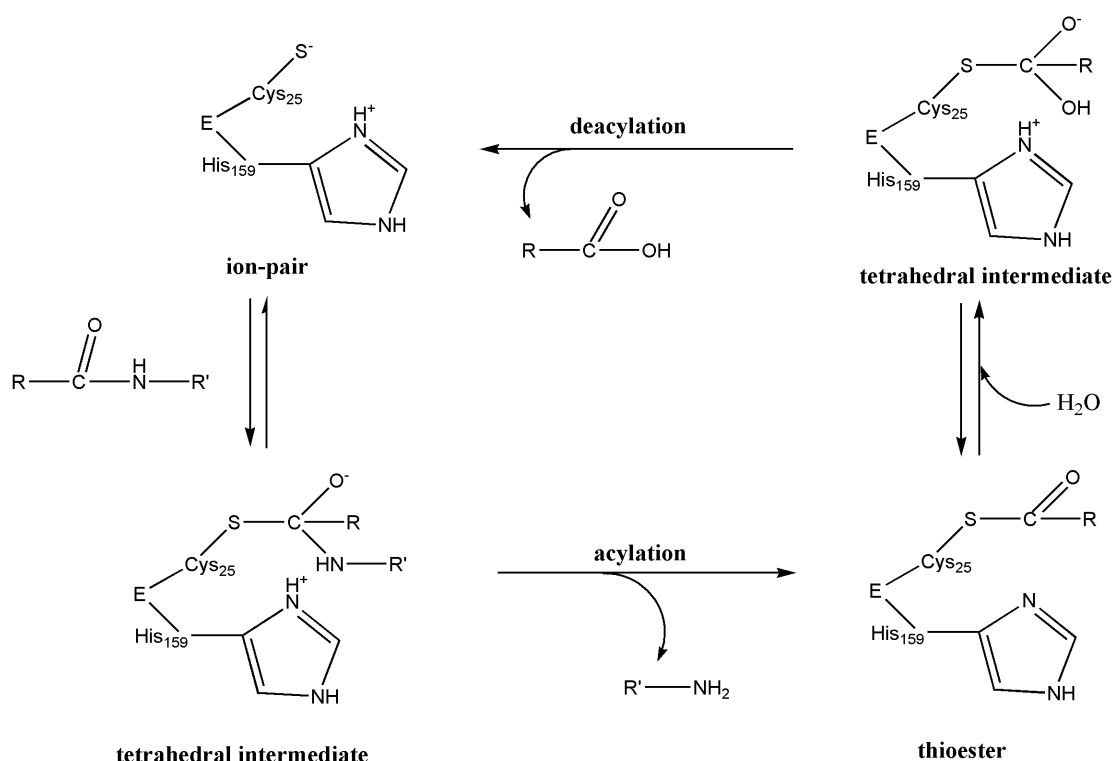


**Figure 1.1: Nomenclature of the substrate specificity of a protease.** The cleavage site and the location of the substrate scissile bond are between P1 and P1'. The protease active site shows the specific subsites numbered to correspond with the substrate amino acid residue in the substrate sequence. Taken from Song *et al.*, 2011.

## 1.2. Cysteine Proteases

Cysteine proteases, initially termed thiol proteases, are characterised by the presence of an active site cysteine thiol group (Otto and Schirmeister, 1997; Grzonka *et al*, 2001) and have been found in plants, mammals, viruses, bacteria, protozoa and most recently in fungi (Otto and Schirmeister, 1997; Mótyán *et al*, 2013). Cysteine proteases vary greatly in molecular weight with papain, congopain and cathepsin K possessing molecular weights between 20 - 33 kDa while the molecular weight of calpains can exceed 80 kDa. Cysteine proteases yield maximum hydrolytic activity at a pH range between 4 and 6.5 (Otto and Schirmeister, 1997).

During the hydrolysis of a peptide (Figure 1.2), the sulfhydryl of the active site cysteine (Cys-25 in Figure 1.2) initiates a nucleophilic attack on the carbonyl carbon of the scissile bond of the bound substrate and forms a tetrahedral intermediate that is stabilised by the oxyanion hole (Lecaille *et al*, 2002). The stabilised tetrahedral intermediate is then transformed into an enzyme-substrate thiol ester while releasing the C-terminal segment of the substrate. The enzyme-substrate thiol ester is hydrolysed, resulting in the formation of a second tetrahedral intermediate that undergoes deacylation which separates the free enzyme and the N- terminal segment of the substrate (Lecaille *et al*, 2002).



**Figure 1.2: The mechanism of peptide hydrolysis by cysteine proteases.** The cysteine of the catalytic triad initiates a nucleophilic attack on the bound substrate resulting in a tetrahedral intermediate which undergoes acylation to form a thioester that is hydrolysed to form the second tetrahedral intermediate which undergoes deacylation. Taken from Lecaille *et al*, 2002.

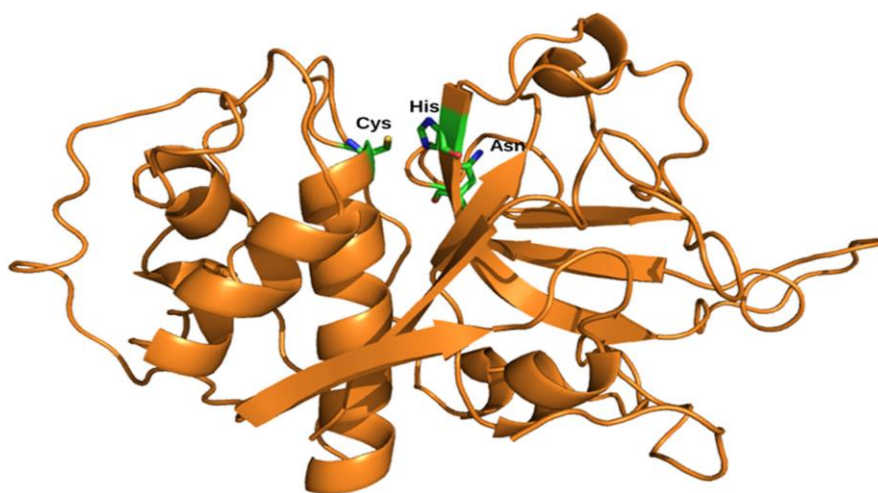
### 1.3. Papain

The papain-like cysteine protease family is the best characterised of the cysteine proteases and has been duly classified as the C1 family. The C1 family contains proteases with a wide range of activities (including those that have no catalytic activity) (Balls and Lineweaver, 1939; Patel *et al*, 2012; Liu *et al*, 2018). Of these, papain was first isolated and purified in its native crystalline state in 1937 from the latex of the green fruit of *Carica papaya* (Balls *et al*, 1937; Monti *et al*, 2000). The method was later modified (Kimmel and Smith, 1954), which then became the standard for the purification of papain from papaya latex. This method involved the extraction of the latex from the papaya, removal of insoluble material in the extract at pH 9.0, followed by ammonium sulfate precipitation and three recrystallisation steps (Monti *et al*, 2000; Malek *et al*, 2016). The method yielded three forms of papain - active papain, activatable papain and non-activatable papain. In the active papain the thiol group is fully reduced whereas the activatable papain is inactive but can be converted to active papain *via* reaction with thiols. The non-activatable papain form, however, cannot be



converted to active papain by thiol reactions (Monti *et al*, 2000; Malek *et al*, 2016). This method has been modified over time to increase the quality and the yield of active papain (Monti *et al*, 2000; Malek *et al*, 2016). The final purified papain product sought is white in colour and is completely soluble in water and glycerol. The final quality of the purified papain for industrial use is determined by colour, moisture content, total ash content, the absence of contaminants and proteolytic activity (Monti *et al*, 2000; Malek *et al*, 2016).

Papain has a molecular weight of 23.41 kDa and is comprised of 345 amino acid residues with three disulfide bridges and an active site sulfhydryl group (Amri and Mamboya, 2012). The 345 amino acid residues constitute a signal sequence (18 amino acids), a propeptide region (115 amino acids) and the mature peptide (212 amino acids) (Amri and Mamboya, 2012). The three dimensional structure of papain consists of two distinct structural domains (Figure 1.3). The two domains are defined as the L-domain, which is characterised mainly by the presence of  $\alpha$  helices, and the R-domain, which is characterised by the presence of antiparallel  $\beta$  sheets (Turk *et al*, 1998; Turk *et al*, 2012). The catalytic triad along with the residues that define the substrate binding site are situated between the L and R domains (Turk *et al*, 1998; Turk *et al*, 2012). The catalytic triad forms the catalytic site and is composed of cysteine, histidine and asparagine (Cys-25, His-159 and Asn-17) residues which are highly conserved amongst papain-like cysteine proteases (Turk *et al*, 1998; Turk *et al*, 2012). The Cys-25 and His-159 form an ion pair which is stabilised *via* a hydrogen bond by Asn-175.



**Figure 1.3:** A ribbon representation for a cysteine protease from the papain family. The catalytic triad, shown in ball and stick, is located between the L-domain and the R-domain. Taken from Verma *et al*, 2016.

### 1.3.1. Inhibitors of Papain

Papain is inhibited by antipain, chymostatin, L-transepoxy succinyl-4-guanidinobutane (E-64), leupeptin, tosyl lysyl chloromethyl ketone (TLCK) and tosyl phenylalanyl chloromethyl ketone (TPCK) (Klein and Kirsch, 1969; Tomkinson *et al*, 1992; Farady and Craik, 2010). L-transepoxy succinyl-4-guanidinobutane was first isolated from *Aspergillus japonicus* and has been proven to be a potent irreversible inhibitor of many cysteine proteases by binding to the  $S_n$  ( $n = 1\sim3$ ) subsites of the papain family of cysteine proteases (Matsumoto *et al*, 1999; Varughese *et al*, 1989). The L-trans-epoxy succinic acid of E-64 is crucial for inhibition as the C-2 atom of the epoxy ring reacts with the Cys-25 of papain and papain like cysteine proteases (Varughese *et al*, 1989). This irreversible binding occurs with the active thiol group of many cysteine proteases but will not react with the functional thiol group of non-protease enzymes. Leupeptin, TLCK and TPCK are broad spectrum inhibitors and inhibit cysteine and also serine proteases (Matsumoto *et al*, 1999; Farady and Craik, 2010).

### 1.3.2. Applications and uses of Papain

Papain has been investigated in different biotechnology processes and has found application particularly in the food and medical industry (Amri and Mamboya, 2012). The protease was first used as a natural remedy to heal wounds without harmful effects (Flindt, 1979). Papain has also been used in the drug discovery sector due to similarities in structure with cathepsin L proteases (Amri and Mamboya, 2012) which has allowed papain to act as a model enzyme for the design of specific cathepsin L inhibitors (Amri and Mamboya, 2012). In the food industry, papain has become a key ingredient in meat tenderisers, is a common ingredient in the brewery industry and papain has also been used in the confectionary, dairy, textile and tanning industries (Amri and Mamboya; Chaplins, 2002; Lambri *et al*, 2014).

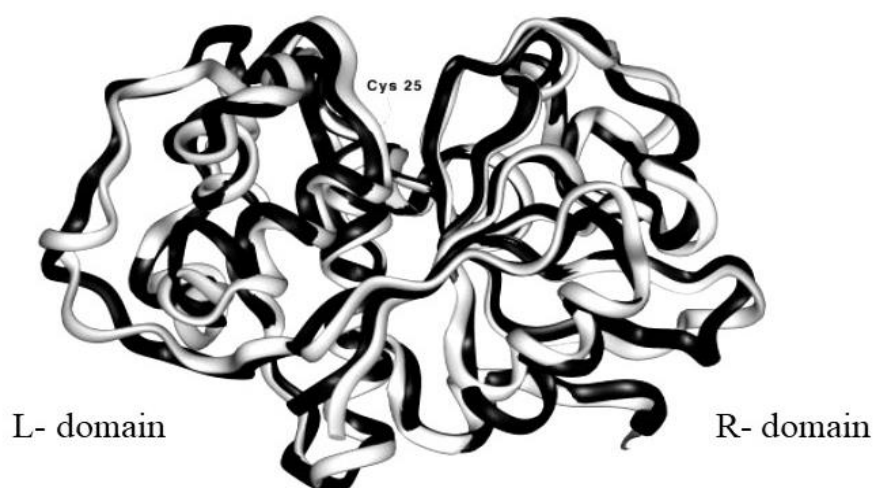
## 1.4. A cysteine protease from *Trypanosoma congolense* (TcoCATL)

Congopain (TcoCATL) is a 33 kDa cysteine protease that was first purified from bloodstream forms of *Trypanosoma congolense* by monoclonal antibody affinity chromatography (Authié *et al*, 1992) and validated by N-terminal sequencing. The protein was subsequently expressed in *Escherichia coli* and in a baculovirus system,

however, the bacterial system yielded an expressed protein which was incorrectly folded and inactive while the baculovirus system gave a poor yield of protein (a maximum of 1 mg per 1L of culture supernatant) (Boulangé *et al*, 2001). Numerous other expression systems were tested due to these challenges and it was shown by Boulangé and co-workers that using the methylotrophic yeast *Pichia pastoris* as an expression system yielded high levels of the protein (Boulangé *et al*, 2011). The recombinant full length and catalytic domain expressed in *P. pastoris* showed unusual dimerisation of the recombinant protein at a neutral pH (Boulangé *et al*, 2011). The recombinant catalytic domain of congopain (*TcoCATL<sub>cat</sub>*) appeared as a dimer with an apparent size of 40 kDa under neutral pH conditions, however, once incubated at an acidic pH reverted to a monomer with an apparent size of 26 kDa (Boulangé *et al*, 2011).

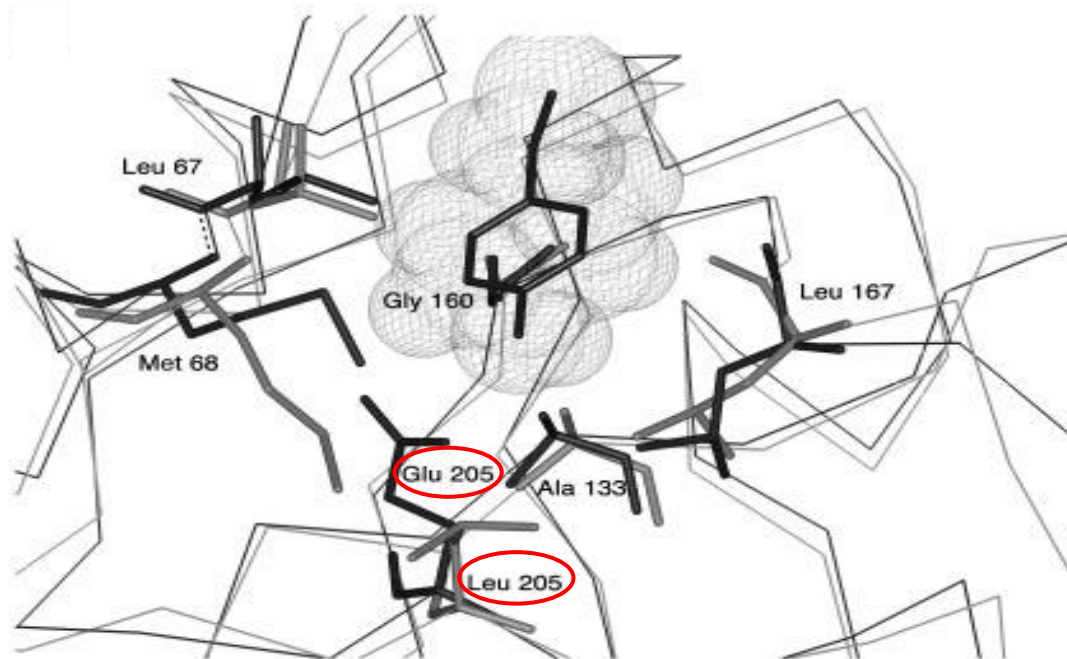
*TcoCATL* is a member of the papain family of cysteine proteases and has been identified as a cathepsin L-like enzyme. The protease is expressed as a 444 amino acid long protein which is composed of a signal peptide of 20 amino acid residues, a propeptide region of 105 amino acids, a catalytic domain of 215 amino acids and a unique C-terminal extension of 104 amino acids (Boulangé *et al*, 2001). The propeptide region contains an inhibitory domain of 56 amino acids which controls proteolytic activity by blocking the active site during the folding of the protein (Lalmanach *et al*, 1998). This propeptide region is also responsible for the production of mature protein. The C-terminal extension is linked to the catalytic domain by a proline-rich hinge region (Huson *et al*, 2009). This C-terminal extension has been found to occur in other trypanosomal cysteine proteases such as cruzipain from *T. cruzi* (*TcrCATL*), however, it has not been observed in mammalian cysteine proteases (Lalmanach *et al*, 1998; Boulangé *et al*, 2001; Huson *et al*, 2009).

A homology-based model for *TcoCATL<sub>cat</sub>* was constructed by Lecaille *et al* (2001) due to the absence of a resolved crystal structure. Papain, actinidin, papaya protease omega and *TcrCATL* were used as template proteins for the construction of the homology model (Lecaille *et al*, 2001). A ribbon plot homology-based model for *TcoCATL<sub>cat</sub>* (Figure 1.4) reveals that the protein folds into two domains – a characteristic commonly exhibited by most cysteine proteases of the papain family (Lecaille *et al*, 2001; Lalmanach *et al*, 2002).



**Figure 1.4: A homology model (ribbon representation) for *TcoCATL<sub>cat</sub>* based on using papain, actinidin, papaya protease omega and *TcrCATL* as template proteins.** The catalytic domain of congopain is represented in grey and cruzain is represented in black. The proteins fold into two domains which are the L-domain and the R-domain. The model displays structural features which are similar to *TcrCATL* and can be accounted for by the 68% sequence identity that *TcoCATL* shares with *TcrCATL* in the catalytic domain. Taken from Lecaille *et al*, 2001.

The homology model supported a hypothesis by Chagas *et al* (1997) wherein they postulated that *TcoCATL* would have a restricted specificity as compared to other papain like cysteine proteases due to the presence of leucine (Leu205, using papain numbering) at the bottom of the S2 subsite instead of a glutamic acid (Glu205, using papain numbering) residue that is present in *TcrCATL* (Lecaille *et al*, 2001). This substitution at position 205 can be observed in the homology-based model shown in Figure 1.5. The interdomain electrostatic interactions between the side chains of glutamic acid (Glu35 and Glu50, using papain numbering) and the positive charged lysine (Lys17 and Lys174, using papain numbering) residues have been conserved through evolution and this is observed in the 3D homology model constructed for *TcoCATL<sub>cat</sub>*. These interdomain electrostatic interactions are essential for the proper folding and stability of cysteine proteases in the papain family (Lecaille *et al*, 2001).



**Figure 1.5:** A homology-based model for the *TcoCATL<sub>cat</sub>* emphasising the S2 subsite substitution found in *TcoCATL*. The S2 subsite for *TcoCATL* is represented in grey with the S2 subsite of *TcrCATL* superimposed and displayed in black. The glutamic acid at position 205 (papain numbering) for *TcrCATL* that is encircled in red is substituted by a leucine at position 205 (papain numbering) for *TcoCATL*. Adapted from Lecaille *et al*, 2001.

The substrate specificity of cathepsin L-like cysteine proteases is dependent on the P2 subsites (Lecaille *et al*, 2001). Cysteine proteases of the papain family show an affinity for bulky hydrophobic residues such as phenylalanine in P2 and for arginyl residues in P1. This affinity for specific residues is due to the presence of a glutamic acid residue (Glu-205 in Figure 1.5) located at the bottom of the S2 subsite (Lecaille *et al*, 2001). This glutamic acid is substituted by a leucyl residue in *TcoCATL<sub>cat</sub>*. This substitution results in the affinity of phenylalanine in the P2 subsite but not an affinity for both phenylalanine and arginyl residues as observed with *TcrCATL* (Lecaille *et al*, 2001). The S2 subsite of *TcoCATL* has an amino acid composition that is remarkably similar to that of cathepsin K (Lecaille *et al*, 2001; Lalmanach *et al*, 2002). *TcoCATL* has a leucine at position 57 (papain numbering) which is not present in cathepsins L and B (Lecaille *et al*, 2001; Lalmanach *et al*, 2002). The leucine at position 57 in cathepsin K is responsible for the size restriction and depth of the S2 pocket which suggests its presence plays a similar role in *TcoCATL* and therefore may affect the S2 specificity (Lecaille *et al*, 2001; Lalmanach *et al*, 2002). *TcoCATL*, similarly to *TcrCATL*, is

activated by reducing agents such as L-cysteine, dithiothreitol (DTT) or  $\beta$ -mercaptoethanol ( $\beta$ -Me) (Chagas *et al*, 1997).

#### **1.4.1. Inhibitors of *Tco*CATL**

*Tco*CATL is inhibited by E-64, cystatin, T-kininogens, leupeptin and peptidyl diazomethane which are common cathepsin L-like protease inhibitors (Chagas *et al*, 1997; Lalmanach *et al*, 2002). In an attempt to generate novel inhibitors, Lalmanach and co-workers (Lalmanach *et al*, 1998) investigated the possibility of inhibiting congopain by using its prodomain. It was revealed, through this study, that peptides containing a 5-mer sequence YHNGA were competitive inhibitors for *Tco*CATL (Lalmanach *et al*, 1998). This YHNGA motif is highly conserved in trypanosomal cysteine proteases (Lalmanach *et al*, 1998; Lalmanach *et al*, 2002).

#### **1.4.2. Applications and uses of *Tco*CATL**

*Tco*CATL is the major lysosomal protease of *Trypanosoma congolense*, a parasite responsible for trypanosomiasis in livestock (nagana) and are mainly transmitted by the tsetse fly (Gitonga *et al*, 2017). Although the parasite has been known for more than a century, the control of the disease remains elusive as the current methods being utilised rely on vector control, chemoprophylaxis (Giordani *et al*, 2016) and chemotherapy (Sahin *et al*, 2014) which has been largely ineffective (Lalmanach *et al*, 2002). Observations made from inhibiting the activity of *Tco*CATL have led to *Tco*CATL being used as a target in vaccine development. Analysis of the two major domains of congopain have revealed that the C-terminal domain is the most antigenic but considered unlikely to produce antibodies which would inhibit enzyme activity (Lalmanach *et al*, 2002; Huson *et al*, 2009). The unique feature of the recombinant protein dimerising at a neutral pH may play a critical role in vaccine development given that the epitopes recognised by the sera of trypanosome-infected trypanotolerant cattle appear dimer-specific (Boulangé *et al*, 2001; Katereggga *et al*, 2012).

### **1.5. Protein stability**

The linear chain of amino acids that is translated from mRNA by the ribosome forms a polypeptide or random coil. As translation and synthesis continues, the growing

polypeptide folds into a three-dimensional structure and ultimately assumes its native state (Dill, 1990; Magliery, 2015). The correctly folded structure, which can contain flexible random coil regions, is essential for function. Proteins in their native state can be unfolded by environmental changes in pH, temperature and ionic strength as well as through the introduction of chemical denaturants (Dill, 1990; Hinz *et al*, 1993; Magliery, 2015). The resulting unfolded or denatured protein is rendered biologically inactive.

Protein stability refers to the net balance of forces that determine whether the protein exists within its active, native state or in its inactive, unfolded state. While electrostatic interactions, van der Waals forces and disulfide bonding contribute to the stability of a correctly-folded state of a globular protein, it is hydrophobic interactions and hydrogen bonding that contribute most (Dill, 1990; Pace, 1990; Moosavi-Movahedi *et al*, 2016). Analysis of the different forces that impart stability revealed that hydrophobic interactions contribute 50.8% and hydrogen bonding contribute 27.1% to overall protein stability while van der Waals forces, electrostatic interactions and disulphide bonding contribute 14.6%, 6.4% and 1.1% respectively (Pace, 1990). Data from the same study suggested that hydrophobic interactions provide the initial drive towards the folded state while the other forces then emerge to resist unfolding (Pace, 1990). The primary destabilising force, which acts to maintain the polypeptide in the unfolded state, is conformational entropy. Each amino acid within a polypeptide can exist in a number of different conformations due to the rotation allowed around the bonds of the polypeptide backbone and the amino acid side-chain (Pace, 1990; Pace, 1992; Moosavi-Movahedi *et al*, 2016). The multitude of possible conformations, and the coinciding loss of conformational entropy that would arise, decreases the likelihood that of a protein folding into a specific conformation.

Under physiological conditions, the folding of most globular proteins is a reversible, two state process, between the folded and unfolded protein (Dill, 1990; Pace, 1990). Typically, the folding of the protein is thermodynamically favoured (i.e.  $\Delta G < 0$ ) and the destabilising force of conformational entropy, calculated to add  $\sim 7\text{kJ/mol}$  per amino acid residue to the reaction, is overcome by the accumulative contributions of the noncovalent interactions (Pace *et al*, 2009). By this thermodynamic definition, the

folding of the protein occurs spontaneously and, accordingly, the folded protein is more stable than any of the unfolded forms (Pace *et al*, 2009).

### **1.5.1. Approaches to increase protein stability**

Proteins are extensively investigated for research, medical and industrial purposes. The folded state of the protein offers the most functionality and application and is therefore of most practical relevance and interest (van den Burg and Eijssink, 2002; Scott *et al*, 2013). Despite the spontaneous nature of folding and the complexity of the three-dimensional structure of the active protein, the folded state of a protein is only marginally more stable than the unfolded forms (Pace, 1990; Fu *et al*, 2010). For most proteins at physiological pH and temperature, the  $\Delta G$  of folding, and therefore the net stability of a protein, falls within the range of  $-20$  to  $-60$  kJ/mol. This modest thermodynamic stability presents a common challenge for research and a general limitation to their industrial application. Substantial effort has therefore been directed towards increasing the net stability of the folded form of the protein to yield practical benefits of increased production yield, longer shelf-life and greater tolerance to environmental stresses (Pace *et al*, 2009; Fu *et al*, 2010). The most effective methods derived to date include the introduction of mutations, the manipulation of buffer conditions and the use of small molecular weight ligands as will be explained in the following sections.

### **1.5.2. Mutations**

Altering the protein primary structure is a well-documented approach aimed at increasing stability (Scott *et al*, 2013; Magliery, 2015). A plethora of computational tools have been developed to predict the effect of mutations (including established tools such as Protherm, Rossetta and FoldX) and facilitate experimental methods such as site-specific mutations and random point mutations (Gromiha *et al*, 1999; Gromiha, 2007). Efforts to increase stability in this manner have proven challenging with low success rates. As measured through phenotypic functionality assays, random substitutions are estimated to yield stabilising mutations in  $< 2\%$  of cases (Broom *et al*, 2017). Computational tools largely avoid functional residues; however, the over-emphasis



towards increasing surface hydrophobicity typically generates stabilised proteins with poor solubility profiles (Broom *et al*, 2017).

### **1.5.3. Buffer composition**

The manipulation of buffer conditions is a widely accepted and extensively used method to increase protein stability. Arguably the most common condition that can affect protein stability is the pH of the solution (Yang and Honig, 1993). The concentration of the H<sup>+</sup> ions present in a solution affects the charge of the amino acid side chains and, consequently, the interactions that these side chains are involved in (Yang and Honig, 1993). Accordingly, protein folding and stability is largely pH dependent and typically follows a bell-shaped curve with either a single stability maximum or multiple stability maxima across a pH range. Prior studies have also established that the pH optima for stability is correlated with the pH optima for activity (Talley and Alexov, 2010). Equally, ionic species have been observed to impact protein stability at physiological concentrations (i.e. 1 to 200 mM). Although not precisely defined, it is accepted that ionic species affect protein stability through multiple, complex, mechanisms. In the predominate mechanism, increasing salt concentration induces point charges on the protein that yields unfavourable, repulsive forces between the medium (i.e. the buffer solution) and the protein leading to aggregation and precipitation (i.e. the principle behind ‘salting-out’) (Talley and Alexov, 2010; Vedadi *et al*, 2006). As the correctly folded protein has less surface area than the unfolded forms, the increase in salt concentration (salting out) destabilises the unfolded forms to a greater extent than the folded protein. As per the thermodynamic equation of stability, this action leads to a net increase in stability of the folded protein over the unfolded forms. Additionally, the increase in salt ions is also known to increase the hydrophobic effect (Thomas *et al*, 2002) and solvation energy; both of which contribute towards increasing the net stability of a folded protein.

### **1.5.4. Ligands**

As a characteristic paradigm of drug discovery, small-molecule ligands bind to proteins in order to inhibit the protein’s activity (especially in the case of enzymes and receptors), activate the protein (i.e. the activation of membrane receptors) or disrupt the

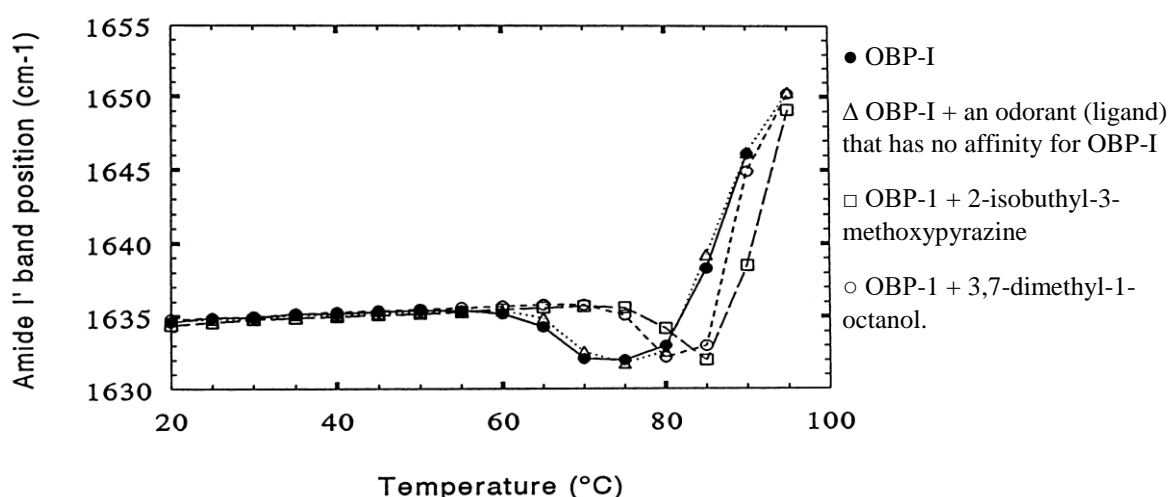
interaction of a protein with another protein or in a complex (as in the case of protein-protein interaction inhibitors) (Celej *et al*, 2003; Waldron and Murphy, 2003). Over the past two decades, it has been conclusively determined that when ligands bind to a protein, they also effectively increase the net stability of that protein. In binding, ligands reduce the inherent flexibility of a protein and decrease the surface area of the protein exposed to solvent. Taken together, these actions reduce conformational entropy (the force favouring unfolding) and decrease the Gibbs free energy of the ligand – protein complex (Celej *et al*, 2003; Waldron and Murphy, 2003). As such, when a ligand binds preferentially to a folded protein it yields a net increase in stability of the folded protein over the unfolded forms. Equally, ligands bind to and stabilise unfolded forms of the protein. Ligands that bind preferentially to an unfolded form of a protein, shift the equilibrium away from the folded protein thereby establishing the unfolded form of the protein as the dominate species within the solution (New *et al*, 2014). In theory, ligands may bind equally to both the folded and unfolded forms of a protein; however, in practice it has been observed that the high affinity binding to the folded protein is thermodynamically favoured over the multiple, low affinity binding to the unfolded forms (Waldron and Murphy, 2003).

## **1.6. Techniques used to determine protein stability**

Stabilisation of a protein has numerous practical purposes. Amongst others, protein stabilisation facilitates expression, purification and storage as stabilised proteins are of higher homogeneity and less likely to unfold, aggregate or succumb to proteolysis (Bloom *et al*, 2006). Similarly, a more stable protein holds increased value as an immunogen as immune responses to aggregates and unfolded forms of the protein are diminished (Rantanji *et al*, 2014; Scheiblhofer *et al*, 2017). The higher yields, increased homogeneity and reduced flexibility induced through protein stabilisation have also significantly improved the likelihood of successful protein crystallisation (Bloom *et al*, 2006; Chen *et al*, 2008). Finally, as protein stability is often a consequence of ligand binding, this measure has been utilised effectively in drug discovery efforts to identify ligands and characterise binding. A large number of experimental techniques based on the principles of, amongst others, Fourier transform infrared (FTIR) spectroscopy, nuclear magnetic resonance (NMR), circular dichroism (CD) spectroscopy, differential

scanning calorimetry (DSC) and differential scanning fluorometry (DSF) have subsequently been designed to determine protein stability.

Fourier transform infrared spectroscopy is an improvement on infrared (IR) spectroscopy, which is one of the oldest and well-established techniques used primarily for structural studies (Kong and Yu, 2007). The method is defined as a measurement of wavelength and intensity of the absorption of IR radiation by a sample (Dzwolak *et al*, 2002). The IR radiation of a sample is interpreted by vibrations of a structural repeat unit (Garidel and Schott, 2006). The polypeptide and protein repeat units give rise to nine characteristic IR absorption bands which are then analysed to determine the secondary structure of the protein (Kong and Yu, 2007). Fourier transformed infrared spectroscopy has been used successfully to determine the secondary structure of proteins along with any conformational changes (due to ligand binding, temperature, pH and pressure changes), structural stability and aggregation of proteins (Kong and Yu, 2007). The structural stability of proteins are determined by measuring the amide I band position as a function of the temperature as the transitions of the proteins are observed over a temperature range (Fernandez-Ballester *et al*, 1992; Paolini *et al*, 1999). The stability of porcine odorant-binding protein (OBP-I) in the presence and absence of ligands is observed in Figure 1.6. The shift of both transitions observed for samples of OBP-I with 2-isobutyl-3-methoxypyrazine and 3,7-dimethyl-1-octanol indicate that the protein is more stable with the addition of these ligands.

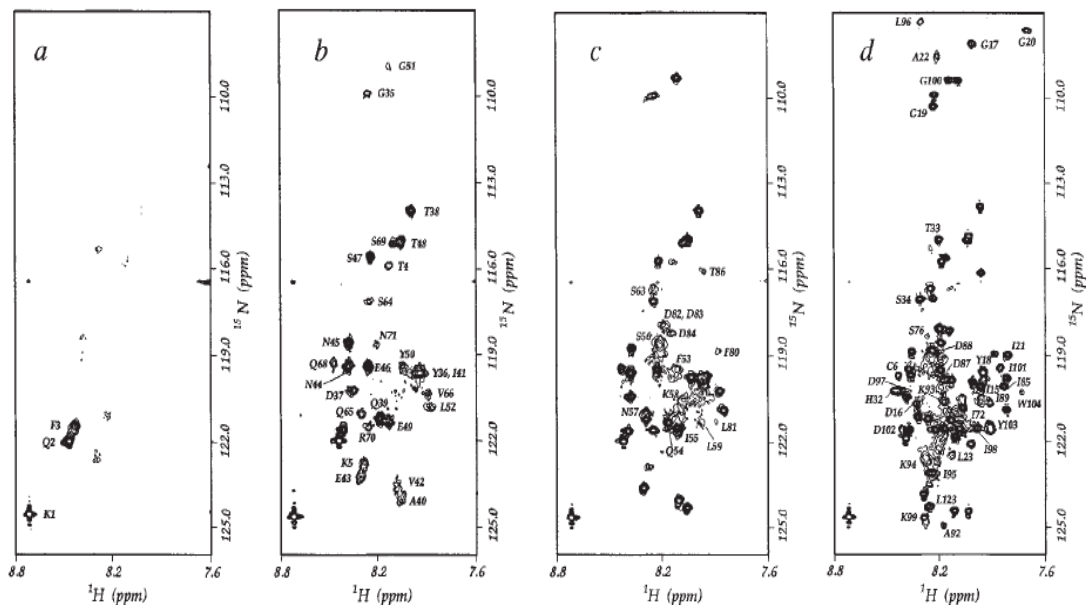


**Figure 1.6: Thermal denaturation curves for porcine odorant-binding protein (OBP-I) in the presence and absence of ligands.** There are two transitions observed in the plot for all of the samples. The first is a slight decrease which is followed by a steady increase in the amide I band position. The samples are: ● OBP-I, △ OBP-I with the addition of an odorant (ligand) that has no affinity for OBP-I, □ OBP-I with the addition of 2-isobutyl-3-methoxypyrazine and ○ OBP-I with the addition of 3,7-dimethyl-1-octanol. Taken from Paolini *et al*, 1999.

Raman spectroscopy was discovered in 1928 by Sir C.V Raman and K.S. Krishnan when they observed the inelastic light scattering of molecules (Wen, 2007). Raman spectroscopy has been used for decades in academic environments to depict secondary and tertiary structure conformations of proteins' by using laser-based technology (de la Cuesta *et al*, 2014). Raman spectroscopy accomplishes this in the same manner as FTIR as it detects molecular vibrations. These molecular vibrations of the polypeptide backbone of a protein sample are used to determine positions and intensities of amide bands which provide estimations of the secondary structure of a protein (Tuma, 2005; Bunaciu *et al*, 2015). The amide bands are further analysed using methods such as spectral decompositions or simplifying the complex and broad envelopes into component bands to generate more detailed information of the secondary structure (Bunaciu *et al*, 2015). The decomposed bands are then related to a set of reference spectra collected from proteins with defined three-dimensional structures and secondary structures (Bunaciu *et al*, 2015). A change in the secondary structure of a protein occurs due to denaturation, misfolding and aggregation and this technique is capable of characterising these changes in protein structure (Wen, 2007).

Nuclear magnetic resonance is a powerful tool used for the analysis of protein structure and function and has been the basis for approximately 8300 derived protein structures being deposited into the Protein Data Bank by the year 2011 (Kwan *et al*, 2011). This technique is also very useful for studying protein-ligand interactions and protein dynamics (Li and Kang, 2017). Nuclear magnetic resonance spectroscopy is similar to the other forms of spectroscopy described here as the spectra are determined by observing changes between different energy states created by atomic nuclei found in the protein (Keeler, 2002; Kwan *et al*, 2011). The atomic nuclei of many isotopes such as  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  carry magnetic dipoles. These dipoles have different orientations and this property has been termed as the "spin" of the nuclei (Keeler, 2002). A key principle of NMR spectroscopy is the fact that the atomic nuclei have a nuclear spin and the nuclei are characterised by a nuclear spin quantum number (Keeler, 2002). The spin quantum number reveals the energy levels that will be generated when the nuclei are exposed to a magnetic field (Keeler, 2002). An NMR spectrometer is equipped with a powerful magnet capable of generating a magnetic field and then records the NMR signal of the atomic nuclei based on their dipole orientation or spin (Kwan *et al*, 2011). A plot of NMR signal intensity against resonance frequency is the simplest of NMR

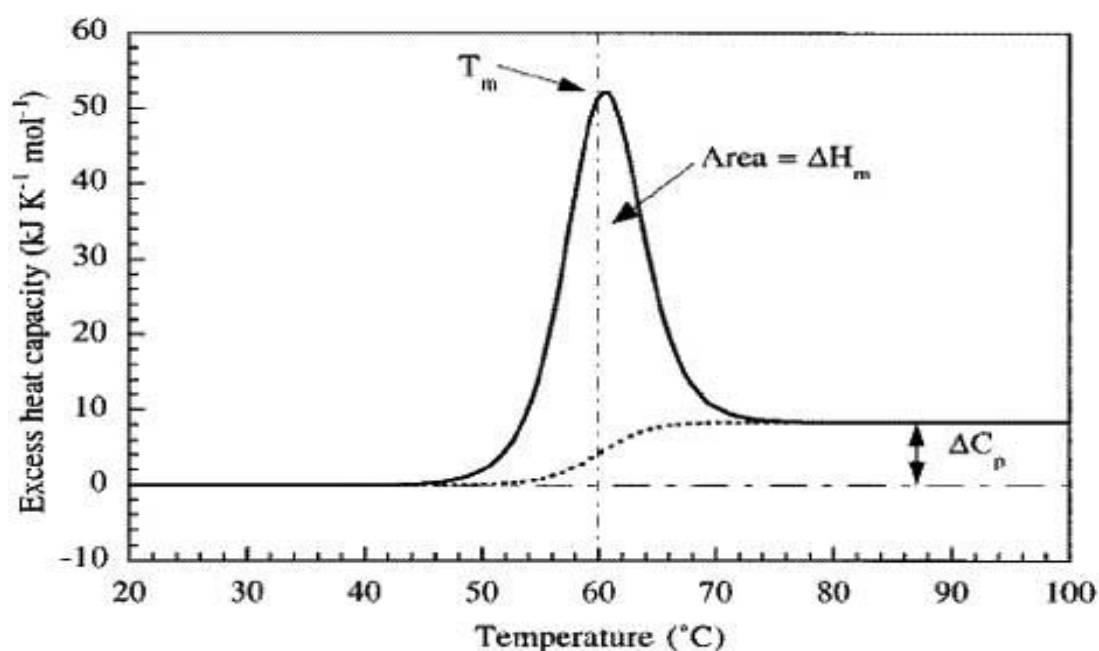
spectra that can be detected and is defined as a one-dimensional (1D) NMR spectrum (Kwan *et al*, 2011). Two of the most useful and sensitive NMR spectra are the 1D  $^1\text{H}$ -NMR spectrum (this spectrum determines the signal for the hydrogen protons in the protein) and 2D  $^{15}\text{N}$ -HSQC (heteronuclear single quantum coherent) spectrum which shows a signal for each  $^1\text{H}$ - $^{15}\text{N}$  covalently bonded group (Hoffman *et al*, 2005). A series of 1D  $^1\text{H}$ -NMR spectra over a long period of time can determine the stability of a protein when it is exposed to different factors such as pH, ionic strength or temperature. The unfolding of the protein is observed by the appearance of new signals in the central part of the spectrum (Shulman *et al*, 1997; Hoffman *et al*, 2005; Kwan *et al*, 2011). This degradation and increase in signal intensity can be observed for human  $\alpha$ -lactalbumin in Figure 1.7 as the concentration of urea increases from panel a to panel d which causes the protein to unfold (Shulman *et al*, 1997).



that protein's ability to differentially absorb clockwise (right handed) or counter-clockwise (left handed) circularly polarised light components of a plane-polarised wave (Kelly and Price, 2000). A CD instrument can measure the difference in absorbance between the left and right handed circularly polarised light components (Kelly and Price, 2000; van Mierlo and Steensma, 2000). This signal can also be represented as ellipticity (Kelly and Price, 2000; van Mierlo and Steensma, 2000). The method can successfully examine the secondary structure of a protein because the peptide bond is asymmetric and thus show the phenomenon CD. The CD spectra obtained for a protein is then compared to characteristic CD spectra of a folded protein to determine the secondary structure (Kelly and Price, 2000; van Mierlo and Steensma, 2000; Li *et al*, 2011). The ellipticity observed at 222 nm is correlated with the  $\alpha$ -helical content of a protein and is an example of characteristic CD spectra (Van Mierlo and Steensma, 2000). Proteins which are folded in their correct state often have asymmetric secondary structural elements such as  $\alpha$ -helices and  $\beta$ -sheets (van Mierlo and Steensma, 2000; Greenfield, 2006). When proteins begin to unfold, they lose these secondary structural elements and their CD bands change. As a function of temperature these changes in CD bands under characteristic wavelengths can be used to determine the thermal stability of the protein (Greenfield, 2006).

Differential scanning calorimetry (DSC) has been reported as one of the most frequently used techniques in determining the thermal stability of a protein due to it being a label free technique (Huynh and Partch, 2016). A calorimeter measures the heat entering a sample or the heat exiting a sample. The technique incorporates the use of a differential scanning calorimeter which measures the heat of a sample relative to a reference sample over a defined time and temperature profile (Chen and Oakley, 1995; Gill *et al*, 2014; Kodre *et al*, 2014). A basic DSC experiment requires a sample chamber/cell which contains the protein of interest in solution and a reference cell which contains only the solvent and serves as the control (Chen and Oakley, 1995; Kodre *et al*, 2014). The technique measures the difference in the heat required to maintain a sample's temperature as close as possible to that of the reference sample as it is exposed to a range of defined temperatures over a specific time profile (Chen and Oakley, 1995). This measurement is reflected as the excess heat capacity ( $C_p$ ) of the sample and is quantified in units of J/g or J/mol (Gill *et al*, 2014; Kodre *et al*, 2014). Changes in the  $C_p$  of a sample originate due to the forces which disrupt the protein's

native structure (Gill *et al*, 2014). The ability to measure the  $C_p$  of a protein sample along a temperature gradient allows the determination of physical changes such as phase transitions that occur during exposure to heating or cooling conditions (Gill *et al*, 2014). These phase transitions include melting, decomposition and crystallisation of a protein sample. Differential scanning calorimetry experiments are also capable of providing information on protein structure as the thermodynamic parameters obtained from DSC experiments are sensitive to the structural state of the protein. A graph of excess heat capacity against temperature may reveal the melting point of the sample as a peak or as a dip depending on the DSC equipment used (Figure 1.8).

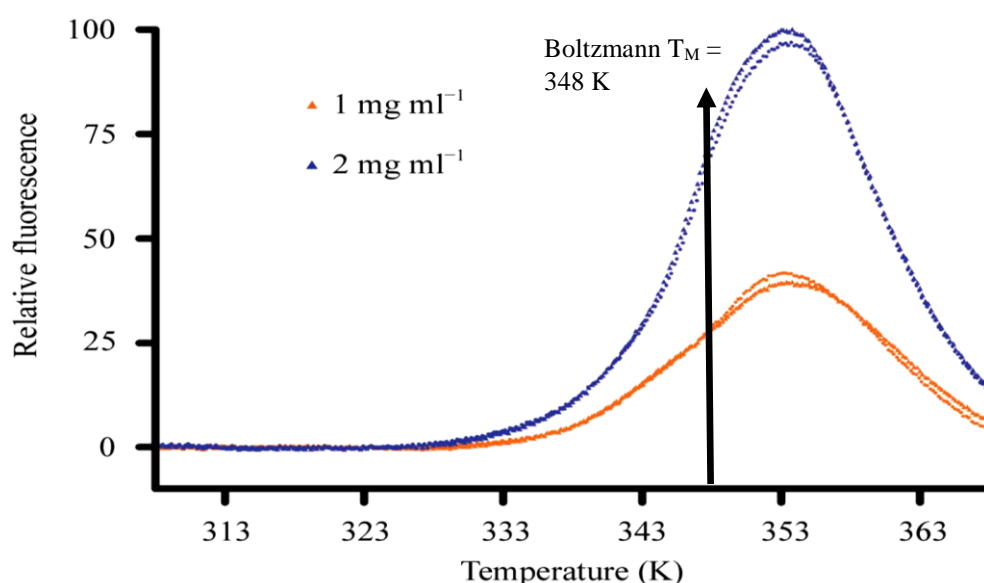


**Figure 1.8: Differential scanning calorimetry (DSC) experiment for the two-state unfolding of a globular protein.** The graph of excessive heat capacity against a temperature range is plotted from measurements taken by the calorimeter and the melting point ( $T_M$ ) is depicted as the highest point i.e. the greatest heat capacity observed for the run. Taken from Bruylants *et al*, 2005.

### 1.6.1. Thermal shift assay

The TSA, which is also known as thermal scanning (Lavinder *et al*, 2009) or differential scanning fluorometry (DSF, Niesen *et al*, 2007), was first described by Pantoliano and co-workers as a technique that could aid in drug discovery (Pantoliano *et al*, 2001) and is seen as an improvement of the DSC technique (Reinhard *et al*, 2012; Vivoli *et al*, 2014; Bruce *et al*, 2019). The TSA is a method which incorporates the use of environmentally safe dyes, which bind specifically to the hydrophobic regions of a

protein, and a real-time polymerase chain reaction (RT-PCR) machine which creates a temperature profile and records the fluorescence emitted by the dye. The dyes which are commonly used in TSA are highly fluorescent in non-polar environments such as the hydrophobic regions of an unfolded protein as compared to aqueous solutions where the fluorescence of the dye is quenched (Niesen *et al*, 2007). The RT-PCR machine therefore measures fluorescence of the dye as it binds to the hydrophobic regions of the protein. These hydrophobic regions are typically not exposed in the protein's native conformation and thus the dye is unable to bind. As the temperature increases these hydrophobic regions become exposed due to unfolding or denaturation of the protein. The assay allows for the determination of the temperature at which the  $\Delta G_u$  of the protein reaches zero and this temperature will reflect the melting point of the protein. The point at which the  $\Delta G_u$  reaches zero for hen egg-white lysozyme at two different concentrations can be observed in Figure 1.9. This graph shows that the concentration of the protein does not affect the melting point of the protein as there is no difference in the melting point obtained. However, as can be seen, an increase in fluorescence is observed in response to an increase in concentration.



**Figure 1.9: Melt curve plot of hen egg-white lysozyme at two different concentrations.** The plot depicts the fluorescence obtained for samples of hen egg-white lysozyme during a temperature gradient in Kelvin. Each of the concentrations were run in duplicate. The red line indicates the hen egg-white lysozyme at a concentration of 1 mg/ml and the blue line indicates the hen egg-white lysozyme at a concentration of 2 mg/ml. The arrow indicates the melting point of the protein which is 348 K based on the Boltzmann equation. Taken from Yeh *et al*, 2006.



The data used to construct a melt curve as in Figure 1.9 is a measure of the real time fluorescence obtained at a specific temperature. To identify the melting point, it is necessary to determine the temperature at which there are equal quantities of folded and unfolded protein. This is referred to as the inflection point and is represented by the midpoint of the slope. However, this point is not easily visible on the graph and can lead to improper interpretation of data. There are two ways to calculate the melting point based on the data obtained from the melt curve. The first method utilises the Boltzmann equation,

$$y = LL + \frac{(UL - LL)}{1 + \exp\left(\frac{T_m - x}{a}\right)}, \text{ where}$$

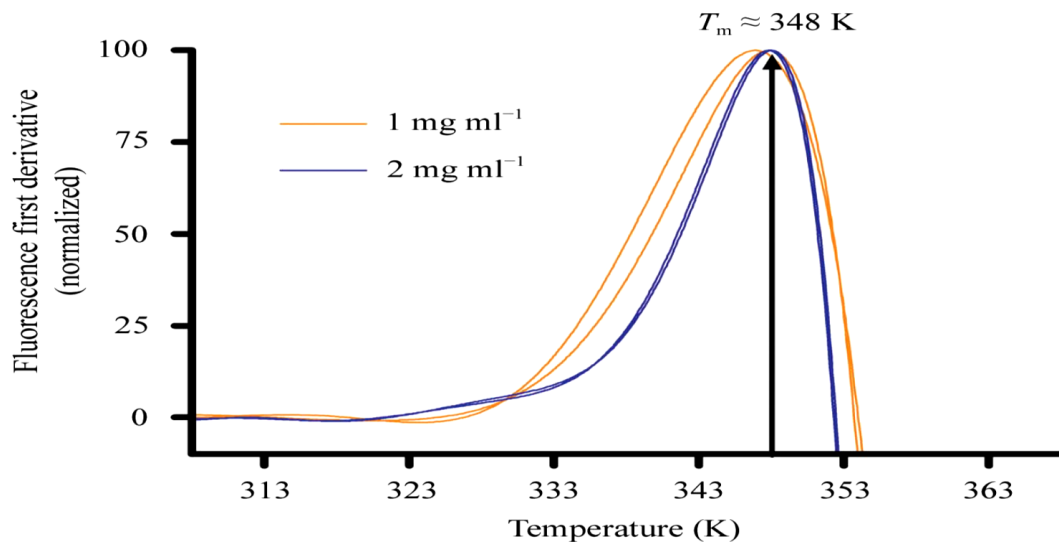
$y$  is the fluorescence intensity at temperature  $x$ ,

$LL$  represents the minimum intensity observed,

$UL$  represents the maximum intensity observed and

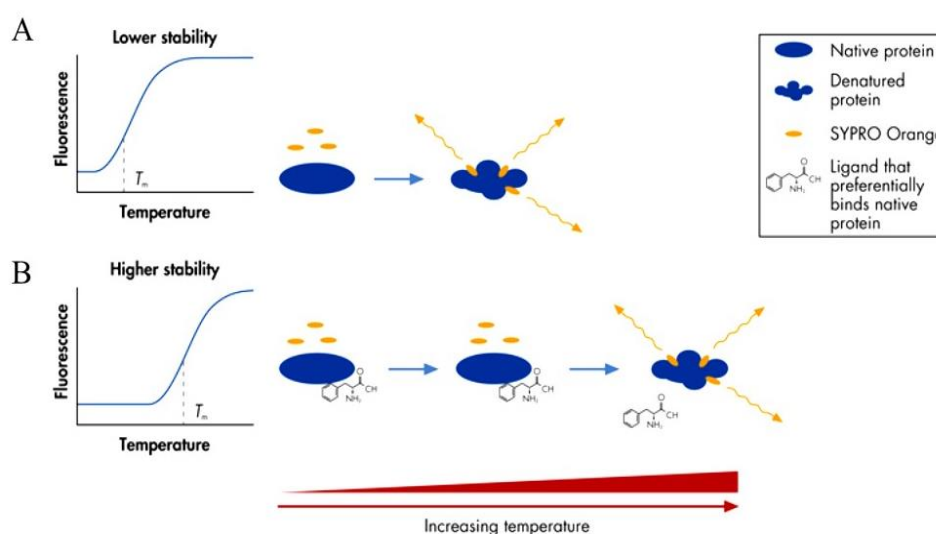
$a$  denotes the slope of the curve within  $T_M$  (Ericsson *et al*, 2006).

The second method is the simpler of the two and can be performed by normalising the data from the melt curve and this is achieved by calculating the first derivative of the data. The melting point of the protein of interest is then determined by identifying the maximum first derivative point (Niesen *et al*, 2007). An example of the derivative graph using the data from the melt curve in Figure 1.9 can be observed in Figure 1.10.



**Figure 1.10: Derivate plot of hen egg-white lysozyme at two different concentrations.** The plot depicts the first derivative for the raw fluorescence obtained for samples of hen egg-white lysozyme during a temperature gradient in Kelvin. The arrow indicates the melting point observed for the protein and is represented by the peaks of the derivative curve. The melting point for the derivative curve is 348 K which was the same melting temperature determined by the Boltzmann equation. Taken from Yeh *et al*, 2006.

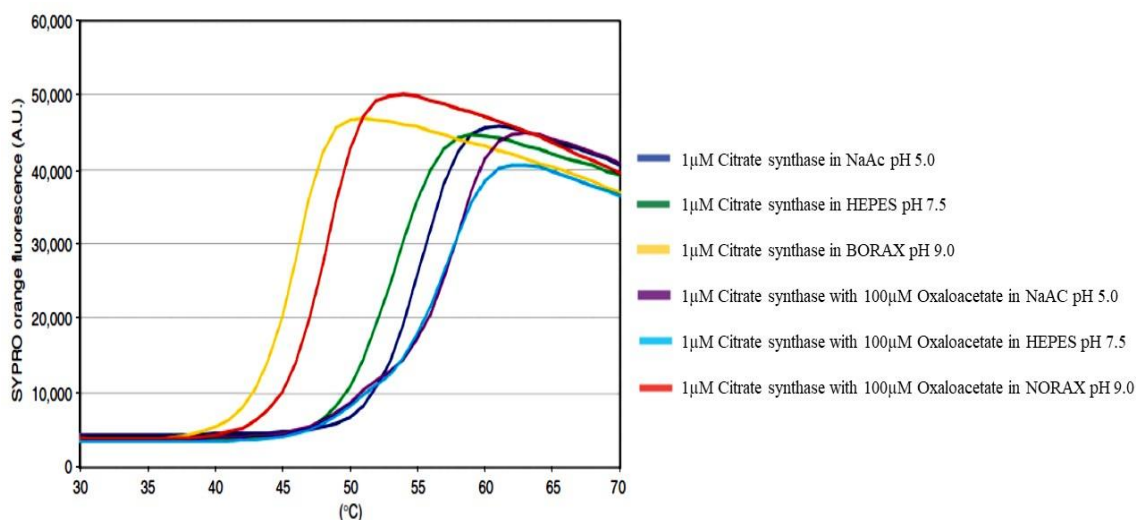
Numerous factors which can affect protein stability such as pH, salt concentration, aggregation, mutation and ligand binding can be studied using the thermal shift assay (Huynh and Partch, 2016; Miyazaki *et al*, 2017; Bruce *et al*, 2019). The method aids in determining the optimum conditions such as pH and ligand interaction that can increase the  $T_M$  of the protein thus causing the protein to become more stable. The principle of whether a ligand can increase stability of a protein and the detection thereof using the thermal shift assay is shown in Figure 1.11 where there is a shift observed in the melt plot of a protein with the presence of a ligand attached as compared to the absence of a ligand.



**Figure 1.11: The increase in stability of a protein due to ligand-protein interaction, as detected through the thermal shift assay.** (A) depicts how SYPRO Orange binds to a protein as the hydrophobic regions become exposed and (B) shows the addition of a ligand will cause the protein to become more stable and thus the protein starts unfolding at a higher temperature. Qiagen-Literature: News, 2010.

The ability to study the effect of more than one factor at a time on protein stability has been a limitation for many of the previously described techniques, however, TSA does not share this limitation (Ericsson *et al*, 2006; Niesen *et al*, 2007; Ragoonanan and Aksan, 2007). The technique can be used to monitor the stability of a protein in the presence of numerous buffers and can also be used to monitor the effect of introducing a ligand on protein stability. Thus, the technique is able to determine the combined effect of a range of buffers and ligands on protein stability (Eriksson *et al*, 2006; Niesen *et al*, 2007; Ragoonanan and Aksan, 2007; New *et al*, 2014; Miyazaki *et al*, 2017). An example of this combined screening can be observed in Figure 1.12. The TSA has a number of advantages over previously described methods as this method does not require a significantly large quantity of protein sample, the method does not require

numerous expensive reagents and the time required for data acquisition is drastically reduced (Niesen *et al*, 2007; Ragoonanan and Aksan, 2007; Huynh and Partch, 2016).



**Figure 1.12: Melt curve plot for porcine citrate synthase (1  $\mu$ M) screened against three different buffers in the presence and absence of 100  $\mu$ M oxaloacetate.** The graph shows the melt curve plot of porcine citrate synthase in NaAc buffer (pH 5.0) represented by a blue line without 100  $\mu$ M oxaloacetate and by a purple line in the presence of 100  $\mu$ M oxaloacetate, HEPES buffer (pH 7.5) represented in green without 100  $\mu$ M oxaloacetate and by an aqua line in the presence of 100  $\mu$ M oxaloacetate and BORAX buffer (pH 9.0) represented by a yellow line without 100  $\mu$ M oxaloacetate and by a red line in the presence of 100  $\mu$ M oxaloacetate. Taken from Niesen *et al*, 2007.

### 1.7. Hypothesis, Aims and Objectives of the study

The hypothesis of this study proposes that cysteine proteases can be stabilised through the manipulation of buffer and/or ligand conditions. The aim of this study was, therefore, to deduce the precise experimental conditions that would favour and enhance the stabilisation of two representative cysteine proteases; namely papain and the catalytic domain of congoain. To this end, protein thermal stability was selected as the means to determine the extent of stabilisation conferred by the conditions evaluated. To realise this aim, the following objectives were defined and undertaken:

- 1) Assessment of the suitability of papain and congoain for use in thermal stability assays.
- 2) Determination of the thermal stability of the two selected cysteine proteases under standard conditions.
- 3) Comparative evaluation of experimental conditions and the identification of factors that stabilise the two cysteine proteases.

## 2. MATERIALS AND METHODS

---

### 2.1. Materials

#### 2.1.1. Expression of recombinant *TcoCATL<sub>cat</sub>*

The catalytic domain of congopain in pPIC9 (*TcoCATL<sub>cat</sub>*) (Pillay, 2010) was sourced from previous transformations performed in our laboratory. Yeast extract, peptone, glucose, tetracycline, ampicillin, bacteriological agar, yeast nitrogen base without amino acids, glycerol, biotin and Whatman #4 filter paper were purchased from Sigma-Aldrich (USA). Buffer salts and other common chemicals were purchased from Sigma-Aldrich (USA) and Merck (Germany).

#### 2.1.2. Analysis of expressed *TcoCATL<sub>cat</sub>*

Bovine serum albumin (BSA) and Bradford reagent stock were obtained from Sigma-Aldrich (USA). The TGX FastCast Kit and the Precision Plus protein standard were purchased from BioRad (USA). SDS-PAGE reagents using the Laemmli system and staining reagents were purchased from Sigma-Aldrich (USA) and Merck (Germany). Papain from papaya latex was obtained from Sigma-Aldrich (USA). The SeeBlue Plus2 pre-stained protein was obtained from ThermoFisher Scientific (USA) and BioTrace polyvinylidene fluoride transfer (PVDF) membrane was purchased from PALL Corp (USA). The low-fat milk powder was purchased from Sigma-Aldrich (USA). The affinity purified chicken anti-*TcoCATL<sub>cat</sub>* IgY primary antibody was previously prepared in our lab (Pillay, 2010). Goat-anti-chickenn IgY (IgG) (H + L), horseradish peroxidase enzyme (HRP) conjugate and the WesternBright ECL chemiluminescent HRP substrate were purchased from Advansta (USA).

#### 2.1.3. Stability determination *via* thermal shift assay

The Protein Thermal Shift<sup>TM</sup> Starter Kit and SYPRO orange dye were purchased from ThermoFisher Scientific (USA). The extra clear PCR tubes were obtained from Whitehead Scientific (USA). E-64 and chymostatin were obtained from Sigma-Aldrich (USA). A total of 50 compounds was extracted from the DIVERSet ChemBridge Library of compounds and obtained from ChemBridge Corp (USA). Buffer salts and other common chemicals were purchased from Sigma-Aldrich (USA) and Merck (Germany).

## 2.2. Methods

### 2.2.1. Analysis of the amino acid sequence for papain and the catalytic domain of *TcoCATL*.

The UniProt protein database (<https://www.uniprot.org/>) was used for obtaining the amino acid sequence for papain (accession number: P00784) and *TcoCATL*<sub>cat</sub> (accession number: Q26895) and was further used to predict the overall hydrophobicity and the theoretical melting temperature of each protein. The Kyte-Doolittle hydrophobicity plot was generated using the online tool available at the Expert Protein Analysis System (ExPASy) Bioinformatics Resource portal ([www.expasy.org](http://www.expasy.org)). The amino acid sequence for papain was pasted into the ProtScale tool and the Kyte-Doolittle hydrophobicity plot was selected as the desired output. The parameters for the plot were set to default with a window size of 9, the relative weight of the window edges compared to the window centre set to 100%, the weight variation model set to linear while the option to normalise the scale from 0 to 1 was not selected. This sequence of steps was repeated for *TcoCATL*<sub>cat</sub>.

A hydrophobicity cluster analysis (HCA) was generated using the HCA v1.0.2. tool available at the Ressource Parisienne en BioInformatique Structurale (RPBS) portal (<http://mobyli.rpbs.univ-paris-diderot.fr/cgi-bin/portal.py?form=HCA#forms::HCA>).

A helical wheel was generated *via* the online helical wheel projection tool available at <http://heliquiest.ipmc.cnrs.fr/>. The sequence of each protein was inserted into the designated window and the parameters were then set. The helix type was set to alpha helix and the option to have one letter code size proportional to amino acid volume was not selected. The window size was set to full, allowing for a maximum of three rotations for the helical wheel output while the rotation of the helix to align vertically to the <μH> vector was not selected. The helical wheel generated was saved as an image and was analysed by using the index available *via* the helical wheel tool. A theoretical melting point was generated using the online tool available at <http://tm.life.nthu.edu.tw/>. The amino acid sequence was inserted in the window and submitted for analysis. The tool then generated an index value with its associated melting point.

### **2.2.2. Recombinant expression of the catalytic domain of *TcoCATL* from *T. congolense***

The catalytic domain of *TcoCATL* had previously been cloned into a pPIC 9 yeast expression vector and transformed into GS115 yeast cells in the laboratory (Pillay *et al*, 2010). Glycerol stocks of these transformed cells were used for the expression of *TcoCATL*<sub>cat</sub>. Briefly, the glycerol stock was streaked onto yeast extract peptone dextrose (YPD) plates [10% (w/v) 10x yeast extract peptone (YP), 5% (w/v) 2 M glucose, 10 µg/ml tetracycline (Sigma-Aldrich, USA), 50 µg/ml ampicillin, 2% bacteriological agar) and the plates were incubated at 30°C for 72 hours. A single colony was selected from the plate and inoculated into YPD broth [10% (w/v) 10x YP, 5% (w/v) 2 M glucose, 10 µg/ml tetracycline, 50 µg/ml ampicillin]. The inoculated YPD broth was incubated in baffled flasks at 30°C with agitation (200 rpm) for 48 hours. The culture in YPD broth was transferred to buffered medium glycerol yeast [BMGY, 100 mM potassium phosphate buffer, pH 6.5, 1.34% (w/v) yeast nitrogen base without amino acids (Sigma-Aldrich, USA), 10% (w/v) 10x YP, 50% glycerol, 0.0004% (w/v) biotin (Sigma-Aldrich, USA), 10 µg/ml tetracycline and 50 µg/ml ampicillin] in a 1:9 ratio and was grown at 30°C for 72 hours with agitation.

The culture was centrifuged at 2000  $\times$  g for 10 minutes at 4°C. The pellet was resuspended in buffered minimal medium [BMM, 100 mM potassium phosphate buffer, pH 6.5, 1.34% (w/v) yeast nitrogen base without amino acids, 0.0004% (w/v) biotin, 10 µg/ml tetracycline, 50 µg/ml ampicillin] and the supernatant was discarded. The resuspended cells were transferred into baffled flasks and covered with three layers of sterile cheesecloth to promote aeration. The first round of expression of the recombinant protein was achieved over the duration of a 7-day incubation at 30°C with agitation (200 rpm) and induction by the addition of 0.5% (v/v) methanol daily. The samples were centrifuged at 2000  $\times$  g for 10 minutes at 4°C as before however the supernatant containing the expressed protein was collected. The pellet was resuspended in BMM as before, transferred into baffled flasks and a second round of expression done exactly as the first round.

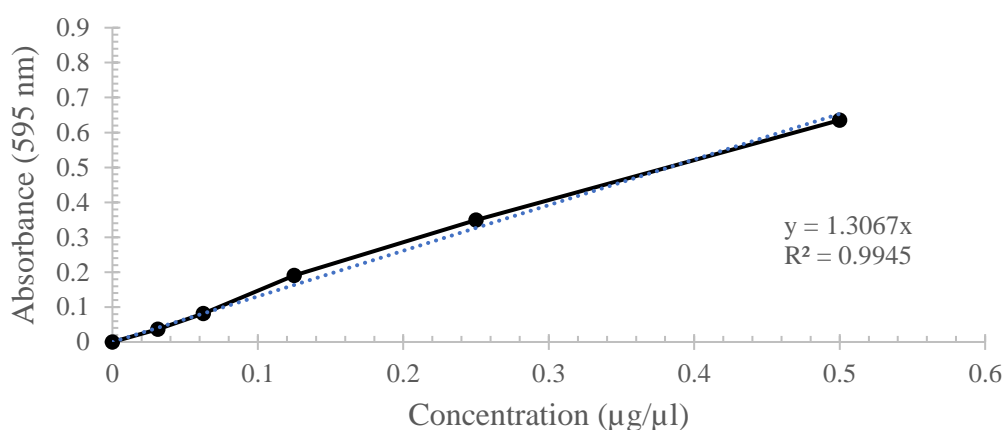
The supernatant containing the expressed protein was collected as described. The recombinant *TcoCATL*<sub>cat</sub> in the supernatant collected from both rounds of expression was isolated using three phase partitioning (TPP) (Dennison and Lovrien, 1997). The

supernatant was filtered through Whatman #4 filter paper, before tertiary butanol was added to the filtered supernatant at a 30% (v/v) final concentration. To this mixture, ammonium sulfate was added at a 30% (w/v) final concentration and mixed until it had entirely dissolved. The resulting mixture was centrifuged at  $6000 \times g$  at  $4^{\circ}\text{C}$  for 10 minutes. The centrifugation resulted in the formation of three phases. The protein precipitate layer formed in the middle and therefore the upper tertiary butanol layer and aqueous bottom layer were removed. The protein precipitate was resuspended in the least amount of phosphate buffered saline (PBS), pH 7.2 and thereafter aliquoted into 2 ml microtubes.

### 2.2.3. Analysis of expressed *TcoCATL<sub>cat</sub>*

#### 2.2.3.1. Determination of protein concentration using the Bradford assay

A Bovine serum albumin (BSA, Sigma-Aldrich, USA) solution (1 mg/ml) was used for the generation of the standard curve. A two-fold serial dilution was performed to obtain five concentrations of BSA that would be tested from 0.03125 to 0.5 mg/ml working concentration. The Bradford reagent stock (Sigma-Aldrich, USA) was used in a 9:1 ratio. The samples were incubated at room temperature for 5 minutes and absorbance was measured at 595 nm using a Biowave DNA WPA V1.8.0 spectrophotometer. The data was collected using the Resolution V2.4.5.0 program. A standard curve was produced by plotting the average absorbance of the samples against concentration and the expressed *TcoCATL<sub>cat</sub>* samples concentrations were derived from the standard curve.



**Figure 2.1: Standard curve for the Bradford protein assay.** Bovine serum albumin standards ranging from 0.03125 to 0.5 mg/ml were added to Bradford reagent and the resulting absorbance values measured at 595 nm after incubation at room temperature for 5 minutes. The equation of the trendline is  $y = 1.3067x$  with a correlation coefficient of 0.994.

### **2.2.3.2. Determination of the molecular weight of the expressed protein using reducing and non-reducing SDS-PAGE**

FastCast acrylamide gels (12%) were prepared using the TGX FastCast Kit (BioRad, USA) as per the manufacturer's instructions. The TGX FastCast gels were loaded with 5 µl of the expressed *TcoCATL<sub>cat</sub>* mixed with either 5 µl of reducing [125 mM Tris-HCL, 4% (w/v) SDS, 20% (v/v) glycerol, 10% (v/v) 2-mercaptoethanol, pH 6.8] or non-reducing sample buffer [125 mM Tris-HCL, 4% (w/v) SDS, 20% (v/v) glycerol, pH 6.8]. The samples containing the reducing sample buffer were placed in a boiling water bath for 90 seconds prior to being loaded onto the gel. The Precision Plus protein standard (BioRad, USA) was used as a molecular weight marker for the gels. Electrophoresis was conducted at 20 mA per gel for approximately 90 minutes. The gel was thereafter stained in Coomassie blue R-250 staining solution [0.025% Coomassie Brilliant blue R-250, 40% methanol, 7% glacial acetic acid] overnight. The stain was discarded, and the gel was placed in Destaining Solution 1 [40% methanol, 7% glacial acetic acid] for 30 minutes followed by Destaining Solution 2 [5% methanol, 7% glacial acetic acid] until background was visibly reduced. The gel image was captured using the SYNGENE G:BOX and analysed using the GeneSys V1.2.4.0 program.

An SDS-PAGE was also performed using the Laemmli system (Laemmli, 1970) under reducing conditions. A 12.5% running gel was prepared with a 4% stacking gel (Laemmli, 1970). The gel was prepared in duplicate and these gels were loaded with 2 µg of the expressed protein, 5 µg of papain from papaya latex (Sigma-Aldrich, USA) as a control and 2 µl of SeeBlue Plus2 pre-stained protein standard (ThermoFisher Scientific, USA) and run at 20 mA per gel for approximately 90 minutes. One gel was thereafter stained in Coomassie blue R-250 staining solution and destained as previously described. The gel image was captured using the SYNGENE G:BOX and analysed using the GeneSys V1.2.4.0 program. The duplicate gel was used for the western blot analysis.

### **2.2.3.3. Confirmation of expressed protein as the catalytic domain of *TcoCATL* via western Blot**

The SDS-PAGE gel prepared as described in section 2.2.3.2 was placed into a electro-transfer sandwich with the following components in order; a layer of Scotchbrite foam, two layers of Whatman No.4 filter paper, the SDS-PAGE gel, BioTrace polyvinylidene



fluoride transfer (PVDF) membrane (PALL Corp, USA), two layers of Whatman No.4 filter paper, a layer of Scotchbrite foam. The SDS-PAGE gel was carefully placed onto the PVDF membrane to avoid any air bubbles. The sandwich was placed into a western blot tank, filled with blotting buffer and the blot was run at 40 mA for 16 hours. The gel was removed and stained using the Coomassie staining procedure as described in section 2.2.3.2 while the unoccupied sites on the PVDF membrane was blocked for 1 hour using a 5% (w/v) low fat milk powder in tris buffered saline (TBS: 20 mM Tris, 200 mM NaCl, pH 7.4). The membrane was washed in TBS (3 x 5 minutes), followed by a two-hour incubation with the affinity purified chicken-anti-*TcoCATL<sub>cat</sub>* IgY primary antibody (in-house preparation, Pillay *et al*, 2010) prepared in a 0.5% (w/v) BSA-TBS at a 10 000 x dilution. The membrane was washed in TBS (2 x 5 minutes) followed by a one hour incubation with the horseradish peroxidase enzyme (HRP)-conjugated secondary antibody [goat-anti-chicken IgY (IgG) (H + L), HRP conjugate, Advansta, USA] prepared in 0.5% (w/v) BSA-TBS solution at a 40 000 x dilution. The membrane was washed in TBS (3 x 5 minutes) and incubated for two minutes in WesternBright ECL chemiluminescent HRP substrate (Advansta, USA). The image was captured using the SYNGENE G:BOX and analysed using the GeneSys V1.2.4.0 program.

#### **2.2.3.4. Gelatin substrate containing SDS-PAGE zymogram**

The protocol for a standard SDS-PAGE gel was modified by incorporating 0.1% (w/v) gelatin substrate into the 12.5% running gel. Additionally, an overlay solution [0.05% (w/v) gelatin] was poured over the running gel and allowed to set and a standard 4% stacking gel was prepared. The gel was loaded with 2 µl of SeeBlue Plus2 pre-stained protein standard (ThermoFisher Scientific, USA) and recombinantly expressed *TcoCATL<sub>cat</sub>* was loaded in concentrations ranging from 0.01 µg/ul to 0.1 µg/µl. The gel was run at 20 mA per gel for approximately 90 minutes and soaked in two changes of 2.5% (v/v) Triton X-100 at room temperature for 60 minutes to replace the SDS with Triton X-100 to restore proteolytic activity. The gel was incubated in assay buffer [PBS pH 7.4] at 37°C for 120 minutes. The gel was stained using 0.1% (w/v) amido black solution (amido black was dissolved in methanol: acetic acid: distilled water in the proportions 30:10:60) for 1 hour and destained by using a solution of methanol, acetic acid and distilled water (30:10:60). The gel was placed in several changes of destain solution until the clear areas of the proteolytic digestion of the dark stained gelatin were

visible. An image of the gel was captured using the SYNGENE G:BOX and analysed using the GeneSys V1.2.4.0 program.

#### **2.2.3.5. Dimerisation of the catalytic domain of *Tco*CATL**

An SDS-PAGE was performed using the Laemmli system (Laemmli, 1970) under non-reducing conditions. A 12.5% running gel was used with a 4% stacking gel. The catalytic domain of *Tco*CATL was loaded at 2 µg. Three buffers of different pH (4.5, 7.4 and 8.0) were used. At each pH, the catalytic domain of *Tco*CATL sample was heated at 30°C, 60°C or 90°C for a period of 90 seconds. The gel was loaded with the prepared samples and 2 µl of SeeBlue Plus2 pre-stained protein standard (ThermoFisher Scientific, USA) and run at 20 mA per gel for approximately 90 minutes. The gel was stained in Coomassie blue R-250 staining solution and destained as described in section 2.2.3.2. An image of the gel was captured using the SYNGENE G:BOX and analysed using the GeneSys V1.2.4.0 program.

#### **2.2.4. Control thermal shift assay using a Protein Thermal Shift™ Starter Kit**

The Protein Thermal Shift™ Starter Kit was used as per manufacturer's instructions (ThermoFisher Scientific, USA). Briefly, test samples were prepared through the addition of control protein at a final concentration of 0.1 µg, followed by the control ligand at 0.1 mM or 1 mM and protein thermal shift buffer to an extra clear PCR tube (Whitehead Scientific, USA). Preparation of each sample was done on ice. The thermal shift dye was diluted in the protein thermal shift buffer and added at a final concentration of 8x directly before the tubes were loaded into the Rotor-Gene™ 6000 2-Plex RT-PCR machine (Corbett Research, USA). The Rotor-Gene™ 6000 2-Plex RT-PCR machine was used in conjunction with the Rotor-Gene™ 6000 series software version 1.7 to analyse the data obtained for each run. The run conditions programmed for the experiment were to create a temperature profile between 25°C and 99°C with the temperature increasing by 0.1°C increments after a period of 2 seconds. At this point, the temperature was held in order to detect the fluorescence through the HRM channel using an emission wavelength at 460 nm and an excitation wavelength of 510 nm, with the option for gain optimisation selected.

### **2.2.5. Optimisation of the thermal shift assay**

Papain, at concentrations ranging from 0.08 to 0.4  $\mu\text{g}/\mu\text{l}$ , was evaluated in conjunction with different SYPRO orange dye (ThermoFisher Scientific, USA) concentrations ranging from 10x to 40x diluted in PBS (pH 7.4). The run conditions designed for the optimisation experiments recorded fluorescence for a temperature profile between 25°C and 99°C with the temperature increasing by various increments, whereby the minimum tested was 0.1°C and the maximum increment tested was 0.5°C. A range of holding times were also tested where the minimum holding time tested was 1 second and the maximum tested was 5 seconds at each temperature point.

### **2.2.6. Determining the stability of papain using the optimised thermal shift assay**

#### **2.2.6.1. Determining the effect of pH on the stability of papain**

Two separate buffer libraries were prepared to test the effect of pH on the stability of papain. The first buffer library was adapted from the buffer library of Reinhard *et al* (2012). The buffer library consisted of the following 50mM buffers: sodium citrate, pH 4.0; sodium citrate, pH 5.0; sodium citrate, pH 5.5; sodium phosphate, pH 6.0; MES, pH 6.7; PIPES, pH 6.7; MOPS, pH 7; sodium phosphate, pH 7.0; PBS, pH 7.4; bicine, pH 8.0 and HEPES, pH 8.0. The second buffer library prepared made use of a single Acetate-MES-Tris (AMT) buffer (Ellis and Morrison, 1982). The buffer consisted of 100 mM acetate, 100 mM MES and 200 mM Tris and 4 mM ethylenediaminetetraacetic acid (EDTA), and the pH adjusted to a pH range between 4.0 and 8.5 using HCl. The buffer libraries were tested on papain using the optimised TSA. The volume of each buffer, the concentration of papain (0.2  $\mu\text{g}/\mu\text{l}$ ) tested and the concentration of SYPRO orange dye (10X) were kept constant. The conditions for each run were set as per the optimised conditions determined as per section 2.2.5. The fluorescence was measured by using the HRM channel on the Rotor-Gene RT-PCR machine with the gain optimisation option selected. The temperature profile tested was set from 25°C to 99°C. The temperature was set to increase in 0.5°C increments with the holding time set to 5 seconds. The data obtained was analysed using the Rotor-Gene<sup>TM</sup> 6000 series software version 1.7.

#### **2.2.6.2. Determining the effect of ligand binding on the stability of papain**

The compound library consisting of 50 compounds was extracted from the DIVERSet ChemBridge Library of compounds (ChemBridge Corp, USA). The 50 compounds

were tested against papain along with chymostatin, an inhibitor for papain and other cysteine proteases. The compounds from the library were tested at a concentration of 10  $\mu$ M in PBS whilst chymostatin was tested at concentrations from 10  $\mu$ M to 50  $\mu$ M all in PBS, pH 7.2. The compounds were incubated with 0.2  $\mu$ g/ $\mu$ l of papain and 10X SYPRO orange dye. The run conditions were set to create a temperature range from 25°C to 99°C with the temperature set to increase in increments of 0.5°C after holding at each temperature point for 5 seconds. The data obtained from testing the compounds were thereafter analysed by using the Rotor-Gene<sup>TM</sup> 6000 series software version 1.7.

### **2.2.6.3. Determining the combined effect of pH and ligand binding on the stability of papain**

The combination of applying a pH change and a ligand interaction with papain was then tested by using papain and the inhibitor, chymostatin, with a variety of different buffers. The buffers that were tested were sodium citrate at pH 4.0, pH 5.0, pH 5.5 and sodium phosphate pH 6.0. Papain (0.2  $\mu$ g/ $\mu$ l), 10X SYPRO orange dye and chymostatin (10  $\mu$ M) were added to each buffer. The run conditions used were set to create a temperature range between 25°C and 99°C whereby the temperature was set to increase at 0.5°C increments after holding at a temperature for 5 seconds and thereafter recording the fluorescence. The data obtained from testing the compounds with different combinations of buffers were thereafter analysed by using the Rotor-Gene<sup>TM</sup> 6000 series software version 1.7.

### **2.2.7. Determining the stability of *TcoCATL*<sub>cat</sub> via the thermal shift assay**

#### **2.2.7.1. Determining the effect of pH on the stability of *TcoCATL*<sub>cat</sub>**

Two separate buffer libraries were prepared to test the effect of pH on the stability of *TcoCATL*<sub>cat</sub>. The first buffer library was adapted from the buffer library by Reinhard *et al* (2012). The buffer library consisted of the following buffers each at 50 mM; sodium acetate, pH 4.5; sodium citrate, pH 4.0, pH 5.0 and pH 5.5; sodium phosphate, pH 6.0; PBS, pH 7.4; bicine, pH 8.0 and CHES, pH 9.5. The second buffer library prepared made use of a single AMT buffer adjusted to different pH-values. The buffer consisted of 100 mM acetate, 100 mM MES and 200 mM Tris and 4 mM EDTA. AMT buffers were prepared at pH 4.0, 4.5, 5.5, 6.0, 6.5, 7.0, 7.5 and 8. The effect of each buffer on the stability of *TcoCATL*<sub>cat</sub> was tested using the TSA. The volume of each buffer, the concentration of *TcoCATL*<sub>cat</sub> (0.1  $\mu$ g/ $\mu$ l) and SYPRO orange dye (5X) was

kept constant. The conditions for each run were set as per the optimised conditions in section 2.2.5. The fluorescence was read by using the HRM channel on the RT-PCR with the gain optimisation option selected. The temperature profile tested was set from 25°C to 99°C. The temperature was set to increase in 0.5°C increments with the holding time set to 5 seconds. The data obtained was thereafter analysed using the Rotor-Gene™ 6000 series software version 1.7.

#### **2.2.7.2. Determining the effect of ligand binding on the stability of *TcoCATL<sub>cat</sub>***

The compound library consisting of 50 compounds were extracted from the DIVERSet ChemBridge Library of compounds (ChemBridge Corp, USA). The structures of these compounds are found in appendix A. The 50 compounds were tested on *TcoCATL<sub>cat</sub>* along with chymostatin as it is an inhibitor of cysteine proteases. The compounds from the library were tested at a concentration of 10 µM at a constant pH of 7.4 (PBS) whilst chymostatin was tested at concentrations from 10 µM to 50 µM. The compounds were tested against 0.1 µg/µl of *TcoCATL<sub>cat</sub>* with 5X SYPRO orange dye. The run conditions were set to create a temperature range from 25°C to 99°C with the temperature set to increase in increments of 0.5°C after holding at each temperature point for 5 seconds. The data obtained from testing the compounds were thereafter analysed by using the Rotor-Gene™ 6000 series software version 1.7.

#### **2.2.7.3. Determining the combined effect of pH and ligand binding on the stability of *TcoCATL<sub>cat</sub>***

The combination of applying a pH change and a ligand interaction with *TcoCATL<sub>cat</sub>* was then tested by using *TcoCATL<sub>cat</sub>* and the inhibitor chymostatin with a variety of different buffers. The buffers that were tested were sodium citrate at pH 4.0, pH 5.0, pH 5.5 and sodium phosphate pH 6.0. The buffers were tested with *TcoCATL<sub>cat</sub>* (0.1 µg/µl) using 5X SYPRO orange dye and chymostatin (10 µM). The run conditions used were set to create a temperature range between 25°C and 99°C whereby the temperature was set to increase at 0.5°C increments after holding at a temperature for 5 seconds and thereafter recording the fluorescence. The data obtained from testing the compounds with different combinations of buffers were thereafter analysed by using the Rotor-Gene™ 6000 series software version 1.7.

#### **2.2.8. Statistical analysis**

The experiments were performed in triplicate with the averages and standard deviations calculated using Microsoft Excel from Microsoft Office (Microsoft Corporation, Version 18.1903.1152.0). The p-values of the data sets were determined by performing T-tests. The T-tests were calculated using the Simple Interactive Statistical Analysis (SISA) online tool (<http://www.quantitativeskills.com/sisa/statistics/t-test.htm>) using a 95% confidence interval.

### 3. RESULTS

#### 3.1. Hydrophobicity analysis of papain

The amino acid sequence and composition of papain is displayed in Figure 3.1 with the amino acids which contribute to the protein's overall hydrophobicity highlighted in different colours. The amino acids highlighted in Figure 3.1 were then scored based on the Kyte and Doolittle index as per Table 3.1. Specifically, the amino acids which contribute to a protein's hydrophobicity according to the Kyte and Doolittle index were alanine, cysteine, isoleucine, leucine, methionine, phenylalanine and valine.

MAMTPSISKLLIFVAICLFVYMGLSFGDFSIVGYSONDITSTERLIQIFESWMIKHNKIYK 60  
 NIDEKIYRFEIFKDNLKYIDETNKKNSYWLGLNVFADMNDEFKEKYTGSIAGNYTTTEL 120  
 SYEEVLNDGDVNIPEYVDWRQKGAVTPVKNQSGCGSCWAFSAVVTIEGIIKIRTGNIENEY 180  
 SEQELIDCDRRSYGCNGGYPWSALQLVAQYGIHYRNTYPYEGVQRYCRSREKGPYAAKTD 240  
 GVRQVQPYNEGALLYSIANQPVSVVLEAAGKDFQLYRGGIFVGPCKGNKVDHAAVAVGYGP 300  
 NYILIKNSWGTGWGENGYIRIKRGTGNSYGVCGLYTSSFYPVKN 345

**Figure 3.1: The amino acid sequence for papain with hydrophobic residues highlighted.** The following amino acids were identified from the papain sequence as contributing to the hydrophobicity of the protein: alanine (green), cysteine (cyan), isoleucine (purple), leucine (red), methionine (yellow), phenylalanine (dark aqua) and valine (blue).

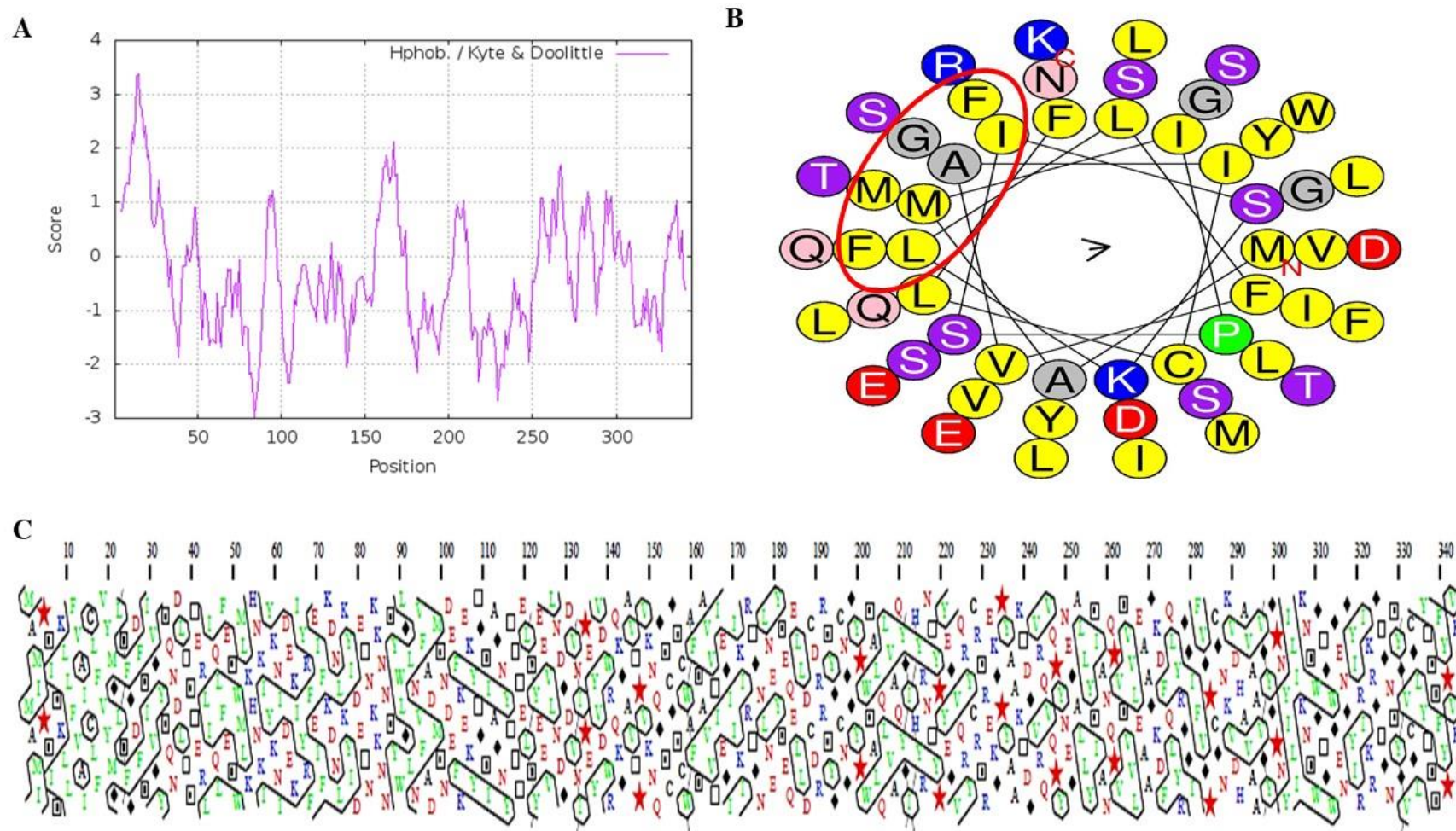
**Table 3.1: The total number of each hydrophobic amino acid present in the papain sequence with relative hydrophobicity scores as per the Kyte and Doolittle index.**

Hydrophobic Amino Acid	Symbol	Total number of Amino Acids	Kyte and Doolittle Score
Alanine (Ala)	A	18	1.8
Cysteine (Cys)	C	8	2.5
Isoleucine (Ile)	I	23	4.5
Leucine (Leu)	L	24	3.8
Methionine (Met)	M	5	1.9
Phenylalanine (Phe)	F	13	2.8
Valine (Val)	V	24	4.2

Of the 345 amino acids comprising papain, 115 were scored as hydrophobic residues as per the Kyte and Doolittle scale (Table 3.1) with isoleucine and valine possessing the highest hydrophobicity scores of 4.5 and 4.2, respectively. As such, isoleucine accounted for 20% while valine accounted for 20.87% of the hydrophobicity observed in papain. Thereafter, papain was visualised through a Kyte and Doolittle hydrophobicity plot, a helical wheel form and a hydrophobicity cluster analysis (HCA) (Figure 3.2.)

A Kyte and Doolittle hydrophobicity plot (Figure 3.2A) was used to evaluate and visualise the hydrophobicity of papain. The hydrophobicity plot revealed that hydrophilic residues dominated the protein sequence, indicative of a cytosolic protein. The hydrophobic residues present comprised the N-terminal and C-terminal and spanned periodically through the full-length sequence. A helical wheel analysis (Figure 3.2B) and HCA (Figure 3.2C) further visualised regions of the protein which were rich in hydrophobic residues. The amino acids encircled in Figure 3.2.B represents the hydrophobic face of the helical wheel projection. The HCA grouped the hydrophobic amino acids within the sequence and it was observed that there are numerous hydrophobic clusters within the sequence of papain. The amino acid sequence of papain was then used to generate a theoretical melting point. The  $T_M$  index calculated for papain was 1.46, which corresponded to a predicted melting temperature greater than 65°C (Ku *et al*, 2009).





**Figure 3.2: Analysis of the hydrophobicity of papain.** (A) Plot of hydrophobicity score for the amino acid sequence of papain using the Kyte and Doolittle index. The plot displays the assigned Kyte-Doolittle index score assigned to a specific amino acid plotted against the amino acid position. (B) Helical wheel analysis for papain with the hydrophobic face encircled. (C) Hydrophobicity cluster analysis of papain where the hydrophobic residues are marked in green and are grouped together by a black outline.

### 3.2. Hydrophobicity analysis of *TcoCATL<sub>cat</sub>*

The amino acid composition for *TcoCATL<sub>cat</sub>* is displayed in Figure 3.3 with the amino acids which contribute to the hydrophobicity score highlighted. These highlighted amino acids were scored based on the Kyte and Doolittle index listed in Table 3.2. The hydrophobicity of *TcoCATL<sub>cat</sub>* was further evaluated and visualised using three different methods, i.e. Kyte and Doolittle hydrophobicity plot, helical wheel analysis and hydrophobicity cluster analysis (Figure 3.4).

**Figure 3.3: The amino acid sequence for *TcoCATL<sub>cat</sub>* with hydrophobic residues highlighted.** The following amino acids were identified from the *TcoCATL<sub>cat</sub>* sequence as contributing to the hydrophobicity of the protein: alanine (green), cysteine (cyan), isoleucine (purple), leucine (red), methionine (yellow), phenylalanine (dark aqua) and valine (blue).

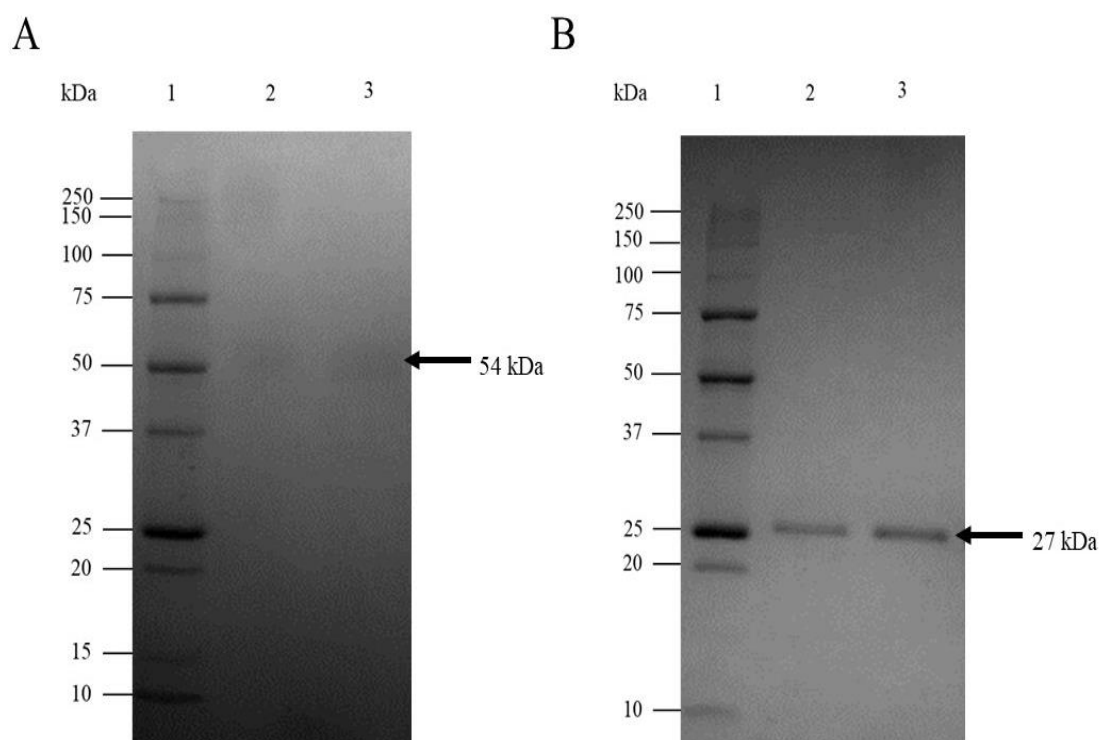
**Table 3.2: The total number of each hydrophobic amino acid present in the *TcoCATL<sub>cat</sub>* sequence with relative hydrophobicity scores as per the Kyte and Doolittle index**

Hydrophobic Amino Acid	Symbol	Total number of Amino Acids	Kyte and Doolittle Score
Alanine (Ala)	A	17	1.8
Cysteine (Cys)	C	7	2.5
Isoleucine (Ile)	I	9	4.5
Leucine (Leu)	L	13	3.8
Methionine (Met)	M	3	1.9
Phenylalanine (Phe)	F	5	2.8
Valine (Val)	V	18	4.2

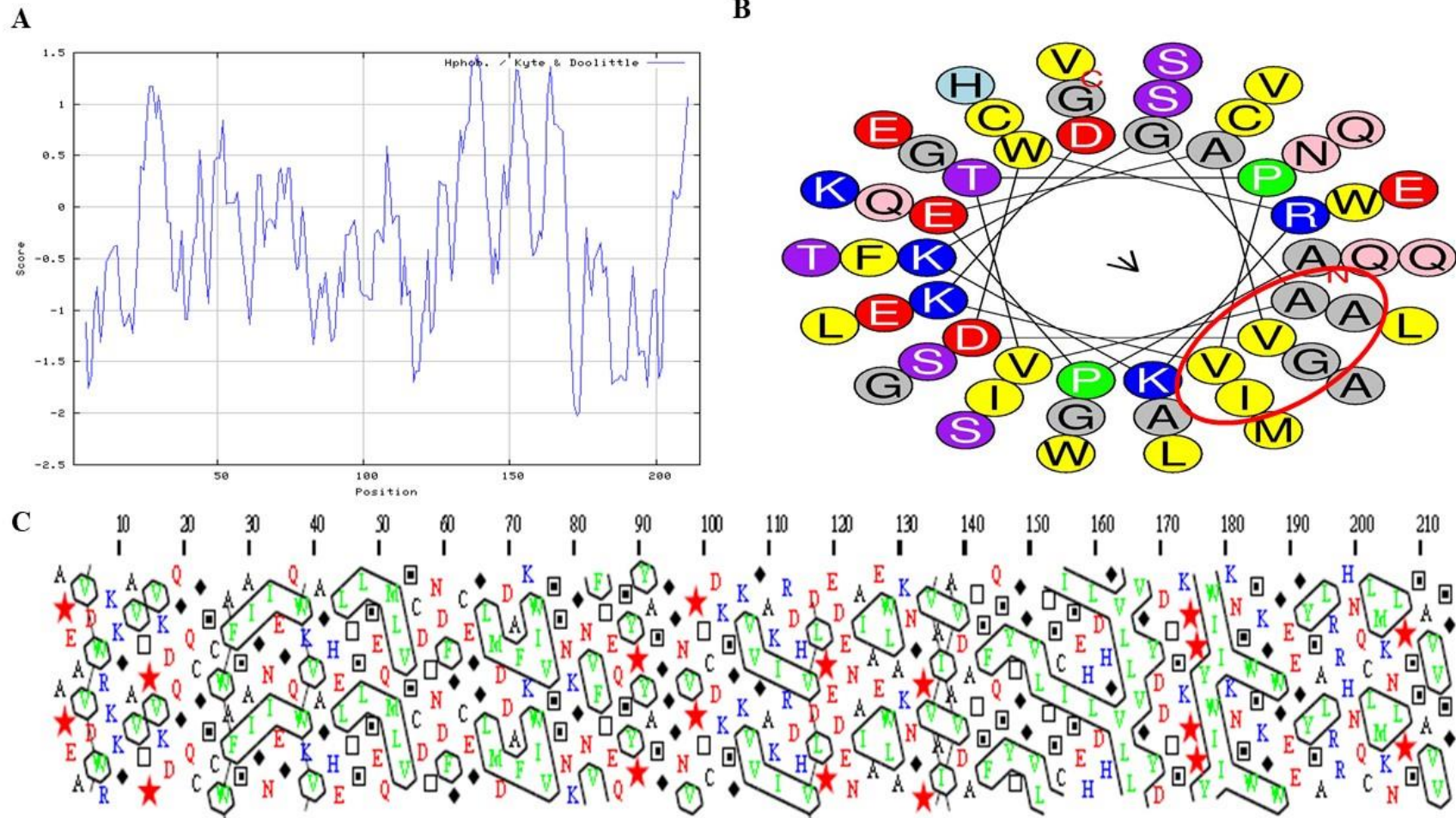
The hydrophobicity for *TcoCATL<sub>cat</sub>*, based on the Kyte and Doolittle index, revealed that the amino acid sequence had a higher degree of hydrophilicity (Figure 3.4A) as the sequence contains 72 hydrophobic amino acids (33.38% of the amino acid residues for *TcoCATL<sub>cat</sub>*). The helical wheel plot and the HCA (Figure 3.4B and C) revealed several regions of hydrophobicity within the sequence and a distinct hydrophobic face. The  $T_M$  index of *TcoCATL<sub>cat</sub>* was predicted to be 1.11 and the melting temperature was calculated to be higher than 65°C.

### 3.3. The expression of recombinant *TcoCATL<sub>cat</sub>*

The pPIC 9 yeast (*Pichia pastoris*) expression vector with the *TcoCATL<sub>cat</sub>* insert, previously been transformed into GS115 yeast cells was used to express the recombinant *TcoCATL<sub>cat</sub>*. Three-phase partitioning was used to isolate the expressed protein from the medium as per standard protocol (Dennison and Lovrien, 1997). The concentration of the expressed *TcoCATL<sub>cat</sub>* was determined *via* the Bradford assay and an average concentration of 5 µg/µl / expression was recorded. The expressed *TcoCATL<sub>cat</sub>* was separated on a 12% polyacrylamide gel (Figure 3.5).



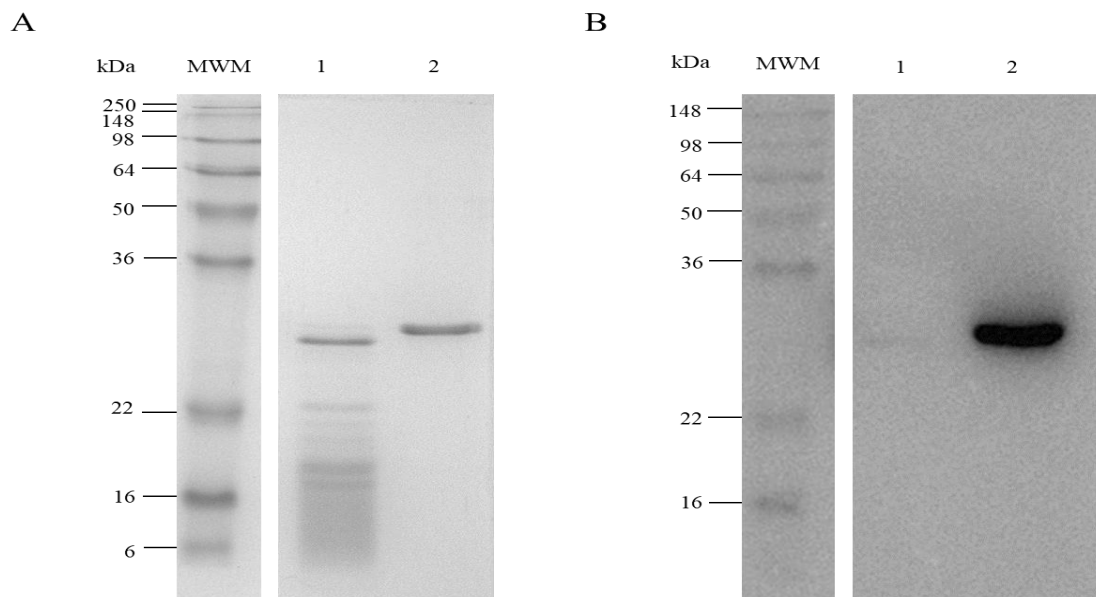
**Figure 3.5: SDS-PAGE analysis for the expressed *TcoCATL<sub>cat</sub>* on 12% TGX FastCast acrylamide gels.** (A) Non-reducing SDS-PAGE gel loaded with Precision Plus protein standard in lane 1, 1.5 µg *TcoCATL<sub>cat</sub>* loaded in lane 2 and 2 µg *TcoCATL<sub>cat</sub>* loaded in lane 3. (B) Reducing SDS-PAGE gel loaded Precision Plus protein standard in lane 1, 1.5 µg *TcoCATL<sub>cat</sub>* loaded in lane 2 and 2 µg *TcoCATL<sub>cat</sub>* loaded in lane 3.



**Figure 3.4: Analysis of the hydrophobicity of *TcoCATL<sub>cat</sub>*.** (A) Hydrophobicity score for the amino acid composition of *TcoCATL<sub>cat</sub>* using the Kyte and Doolittle index. The plot displays the assigned Kyte-Doolittle index score assigned to a specific amino acid plotted against the amino acid position. (B) Helical wheel analysis for *TcoCATL<sub>cat</sub>* where the hydrophobic face is encircled. (C) Hydrophobicity cluster analysis of *TcoCATL<sub>cat</sub>* where the hydrophobic residues are marked in green and are grouped together by a black outline

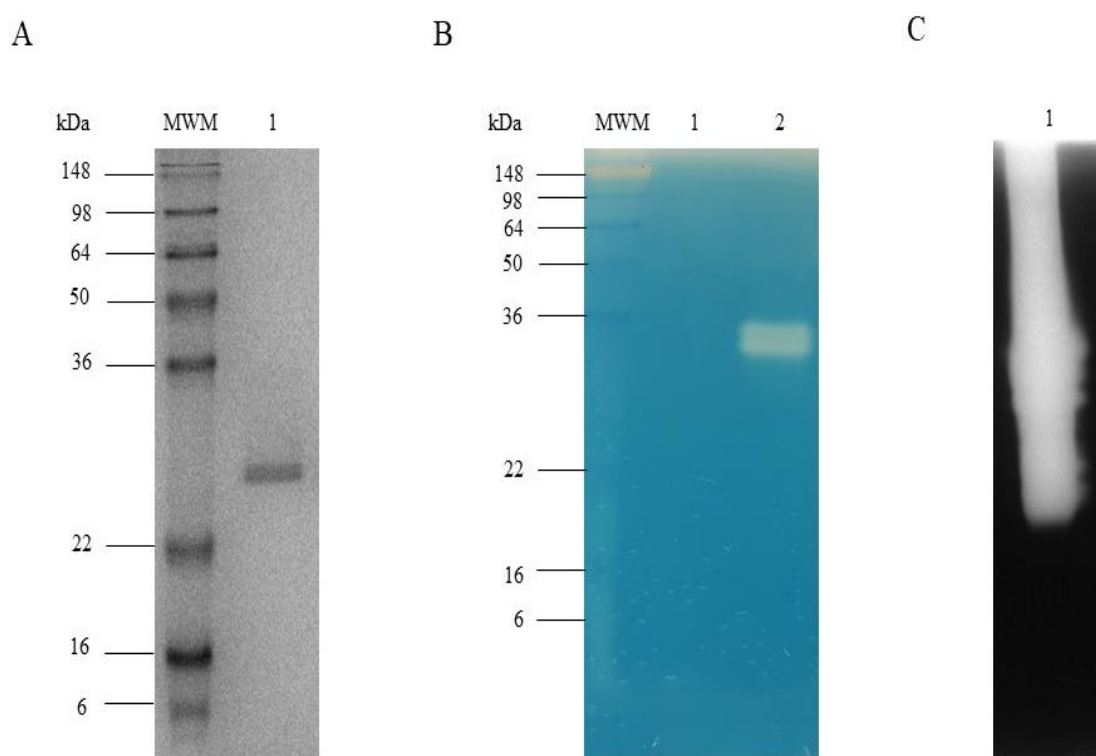


When analysed on an SDS-PAGE gel under non-reducing conditions, the over-expressed protein appeared as a single band with a molecular weight of 54 kDa (Figure 3.5A) which correlated with the expected size of the *TcoCATL<sub>cat</sub>* dimer. The expressed protein appeared as a single band with a molecular weight of 27 kDa (Figure 3.5B) under reducing conditions, which agreed with the expected size of *TcoCATL<sub>cat</sub>* monomer (Authié *et al*, 1992). The protein sample appeared as a single band in both the non-reducing gel and the reducing gel which suggested that the expressed protein was free of contaminating proteins under the staining conditions employed. The expressed protein was further evaluated by means of a western blot. The protein sample was probed and detected by the affinity purified chicken-anti-*TcoCATL<sub>cat</sub>* IgY primary antibody specific to *TcoCATL<sub>cat</sub>* (previously isolated, Pillay *et al*, 2010) as evident by the single band detected on the PVDF membrane (Figure 3.6B) following exposure of the membrane in WesternBright ECL chemiluminescent HRP substrate. The band detected on PVDF membrane in Figure 3.6B and the band observed for the expressed protein in the reducing SDS-PAGE gels in Figures 3.5B and 3.6A were of an apparent similar molecular weight. The affinity purified chicken-anti- *TcoCATL<sub>cat</sub>* IgY primary antibody did not detect the papain sample loaded in lane 1 (Figure 3.6A and B).



**Figure 3.6: The analysis of the expressed recombinant *TcoCATL<sub>cat</sub>* via western blot. (A)** Reducing SDS-PAGE of the recombinantly expressed *TcoCATL<sub>cat</sub>*. Lane MWM, SeeBlue Plus2 pre-stained protein standard; lane 1, 5 µg of papain and lane 2, 2 µg *TcoCATL<sub>cat</sub>* in PBS buffer (pH 7.4). **(B)** Western blot of the recombinantly expressed *TcoCATL<sub>cat</sub>*. Lane MWM, SeeBlue Plus2 pre-stained protein standard; lane 1, 5 µg of papain and lane 2, 2 µg *TcoCATL<sub>cat</sub>* in PBS buffer (pH 7.4). The images of both the SDS-PAGE gel and the western blot were spliced to remove non-relevant wells containing samples not relevant to this section of work.

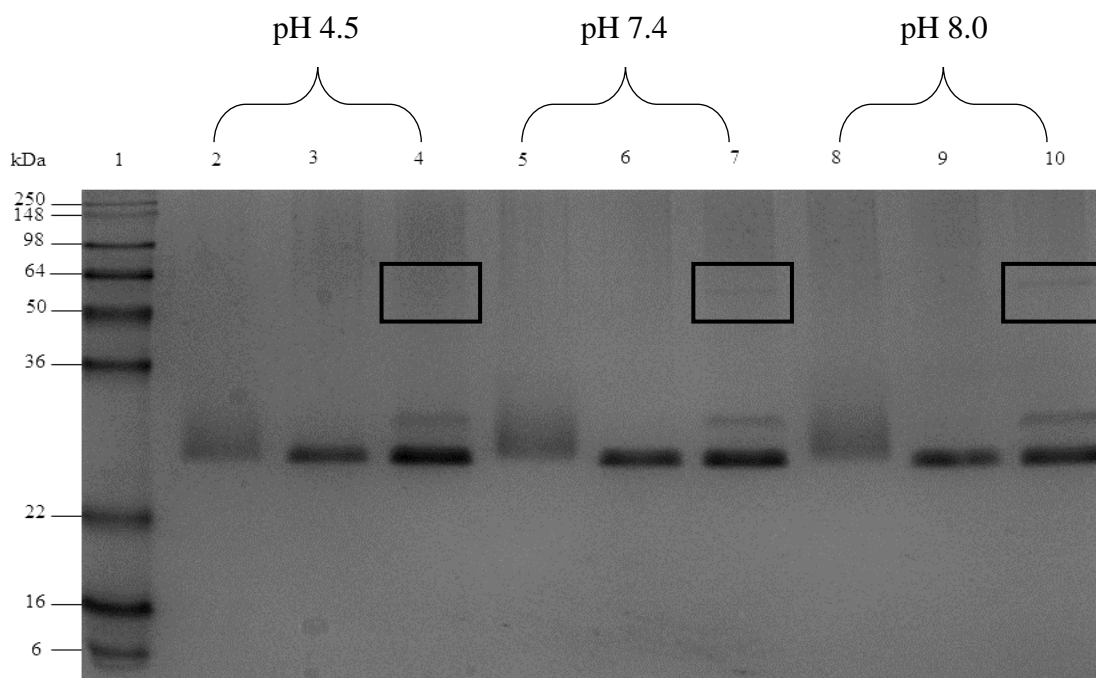
The gelatinolytic activity of the expressed *TcoCATL<sub>cat</sub>* was determined after separating the expressed protein samples on a gelatin substrate containing gel (Figure 3.7). The gelatin-based zymogram can be observed in Figure 3.7B alongside the control SDS-PAGE gel in Figure 3.7A and a control gelatin-based zymogram showing the activity of papain in Figure 3.7C. The image revealed gelatin degradation in the lane loaded with the expressed *TcoCATL<sub>cat</sub>*. A single band was observed for the sample in both the SDS-PAGE and the zymogram, however, it appeared at different molecular weights with the degradation band present in the zymogram (Figure 3.7B) at a higher molecular weight (32 kDa) than the band in the corresponding SDS-PAGE (27 kDa) (Figure 3.7A). The banding pattern of the SeeBlue Plus2 pre-stained protein standard appeared to follow the same trend. The molecular weight marker bands migrated a shorter distance from the wells in Figure 3.7B when compared to migration observed in Figure 3.7A.



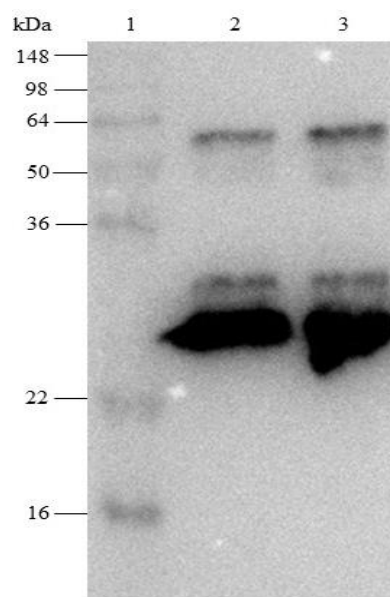
**Figure 3.7: Qualitative determination of enzymatic activity of *TcoCATL<sub>cat</sub>* via non-reducing SDS-PAGE and gelatin-containing SDS-PAGE analysis with papain enzymatic activity on a gelatin-containing SDS-PAGE.** (A) A standard SDS-PAGE 12.5% acrylamide gel used as a control with SeeBlue Plus2 pre-stained protein standard loaded in the lane marked MWM and 2 µg of expressed *TcoCATL<sub>cat</sub>* in lane 1. (B) A gelatin containing substrate SDS-PAGE gel with SeeBlue Plus2 pre-stained protein standard loaded in the lane marked MWM, non-reducing buffer loaded in lane 1 and 2 µg of recombinantly expressed *TcoCATL<sub>cat</sub>* in lane 2. (C) A gelatin containing substrate SDS-PAGE gel with 2 µg of papain in lane 1.

### 3.4. Evaluation of the dimerisation of *TcoCATL<sub>cat</sub>*

The recombinantly-expressed *TcoCATL<sub>cat</sub>* protein was exposed to three different pH values and three distinct temperature conditions prior to separation via SDS-PAGE (Figure 3.8). A single visible band was observed in lanes 2, 3, 5, 6, 8 and 9 which represents *TcoCATL<sub>cat</sub>* as a monomer with an apparent molecular weight of 27 kDa. A change in banding pattern for *TcoCATL<sub>cat</sub>* was observed in lanes 4, 7 and 10. These lanes represented the samples that were exposed to a temperature of 90°C for 90 seconds. The monomer was still present in these lanes however a band with a greater molecular weight was observed in lanes 7 and 10 (pH 7.4 and pH 8.0) which was not observed under acid conditions (pH 4.5) in lane 4. Previously, Boulangé and co-workers demonstrated a unique characteristic of this protein under acidic conditions through the transition from a dimer presents at neutral pH to a monomer under acidic pH conditions (Boulangé *et al*, 2011). The protein bands that resolved at the apparent molecular weight of the *TcoCATL<sub>cat</sub>* dimer (Figure 3.8) were detected when probed with protein-specific antibody in a western blot analysis (Figure 3.9). Additionally, the bands that resolved on the membrane at the higher molecular weight were distinct and not indicative of the banding pattern of an aggregated protein (Figure 3.9).



**Figure 3.8: An SDS-PAGE displaying the dimerisation of *TcoCATL<sub>cat</sub>* due to pH.** The SeeBlue Plus2 pre-stained protein standard was loaded in lane 1. *TcoCATL<sub>cat</sub>* in pH 4.5 was held for 90 seconds at 30°C in lane 2, 60°C in lane 3 and 90°C in lane 4 before being loaded. This was repeated at pH 7.4 in lanes 5-7 and at pH 8.0 in lanes 8-10. The concentration of *TcoCATL<sub>cat</sub>* was kept constant at 2 µg.



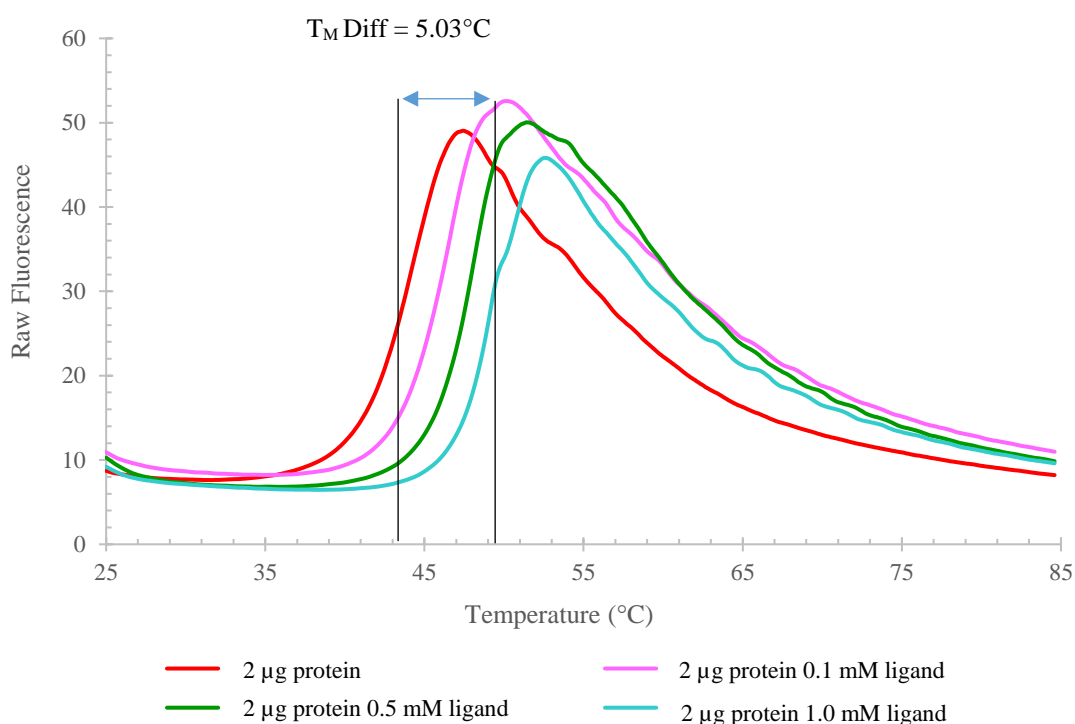
**Figure 3.9: The evaluation of observed dimer formation of *TcoCATL<sub>cat</sub>* under neutral pH conditions using western blot.** The SeeBlue Plus2 pre-stained protein standard was loaded in lane 1. *TcoCATL<sub>cat</sub>* was kept at 90°C for 90 seconds in PBS pH 7.4. A total of 2 µg was loaded in lane 2 and 3 µg loaded in lane 3.

### 3.5. Thermal shift assay

#### 3.5.1. Thermal shift kit

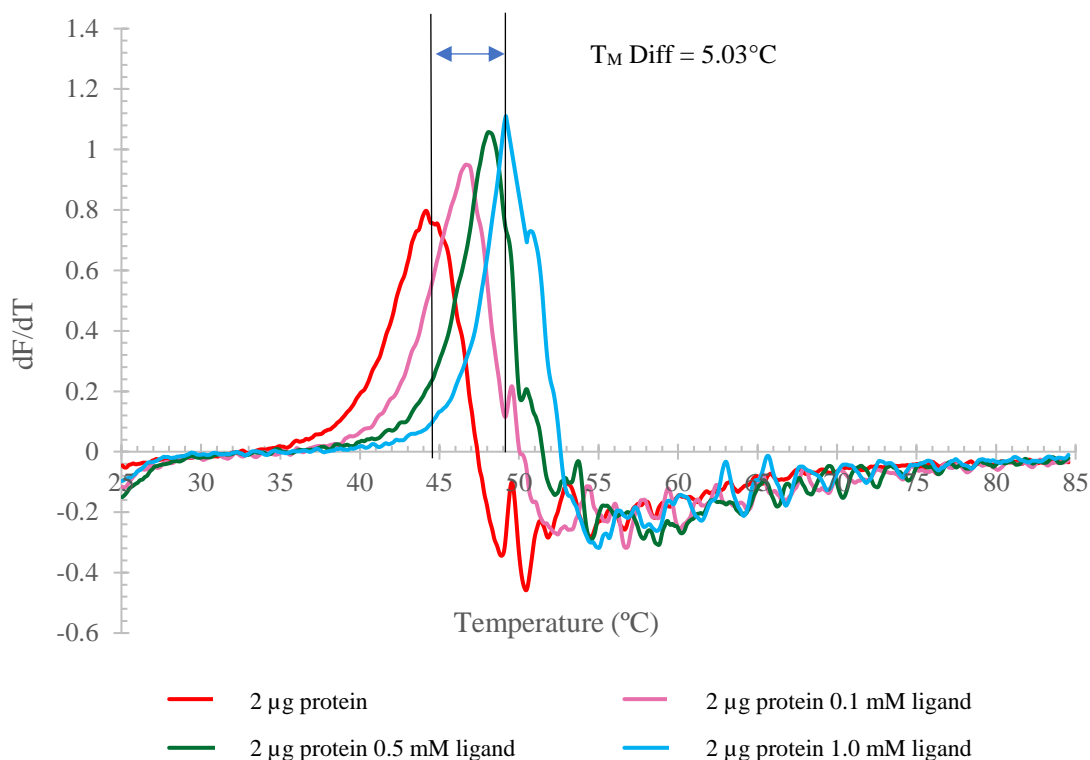
The thermal shift kit (ThermoFisher Scientific, USA) was used to validate the use of the thermal shift assay for the identification of factors that influence a protein's stability. The model protein, provided with the kit, was used at a final concentration of 0.08 µg/µl while the ligand was tested at concentrations of 0.1 mM, 0.5 mM and 1.0 mM. A comparison of the raw fluorescence obtained for the model protein alone was compared to the raw fluorescence obtained when the model protein was combined with the thermal shift kit control ligand at different concentrations is shown in Figure 3.10. The raw fluorescence was recorded as the temperature increased from 25°C to 85°C.





**Figure 3.10: Raw fluorescence obtained for the model protein in the absence and presence of the control ligand at various concentrations.** The real time fluorescence detected for the model protein was plotted against temperature ( $^\circ\text{C}$ ). The fluorescence of the model protein with the ligand absent is displayed in red, followed by the model protein with the addition of 0.1 mM ligand displayed in pink. Thereafter, the protein with the addition of 0.5 mM ligand is displayed in green and lastly the protein with the highest ligand concentration (1.0 mM) represented in blue. The Boltzmann  $T_M$  for the protein with the ligand absent was calculated to be  $44.15^\circ\text{C}$  and the Boltzmann  $T_M$  for the protein with the ligand added at a final concentration of 1mM was calculated to be  $49.18^\circ\text{C}$ , resulting in a  $T_M$  shift of  $5.03^\circ\text{C}$ .

The melting point of the model protein was determined by calculating the derivative ( $dF/dT$ ) of the raw fluorescence obtained. The calculated derivative of the raw fluorescence obtained against temperature was plotted in Figure 3.11. By calculating the derivate ( $dF/dT$ ), the Boltzmann  $T_M$  was clearly visible and represented by the peaks shown in Figure 3.11 as compared to the Boltzmann  $T_M$  (represented by the midpoint of the increase in fluorescence) obtained in Figure 3.10. The  $T_M$  of the model protein increased as the concentration of the ligand increased. The 1 mM concentration of ligand effectively increased the  $T_M$  of the protein from  $44.15^\circ\text{C}$  to  $49.18^\circ\text{C}$  ( $p < 0.5$ ) and this was the greatest shift observed as compared to lower ligand concentrations. The derivative of the fluorescence obtained increased in intensity as the concentration of the ligand increased.



**Figure 3.11: Differential scanning fluorimetry analysis obtained for the thermal shift control protein in the absence and presence of the control ligand at various concentrations.** The calculated derivative of raw fluorescence ( $dF/dT$ ) was plotted against temperature ( $^{\circ}\text{C}$ ). The derivative of the model protein displayed in red, followed by the model protein with the addition of 0.1 mM ligand displayed in pink, thereafter the protein with the addition of 0.5 mM ligand displayed in green and lastly the protein with the highest ligand concentration (1.0 mM) tested represented in blue. The melting point for the protein without the addition of the ligand was  $44.15^{\circ}\text{C}$  and the melting point of the protein with 1.0 mM ligand was  $49.18^{\circ}\text{C}$ , with a  $T_M$  difference of  $5.03^{\circ}\text{C}$ .

### 3.6. Determining the stability of papain *via* the TSA

#### 3.6.1. Determining the effect of pH on the stability of papain

The TSA was implemented to determine the stability of papain under various pH conditions by a measure of the melting point of the protein in combination with the Reinhard *et al* (2012) or AMT buffer libraries. Through this analysis, it was clearly observed that the melting temperature of papain was dependent on the buffer used (Table 3.3). Specifically, the melting temperature of papain increased as the pH decreased from pH 8 to 4.5 - at which point the highest melting point ( $83.40 \pm 0.10^{\circ}\text{C}$ ) of the protein was obtained (Table 3.3). This would suggest that papain stability is favoured and optimal within an acidic environment.

**Table 3.3: The average melting temperature obtained for papain under various buffer conditions**

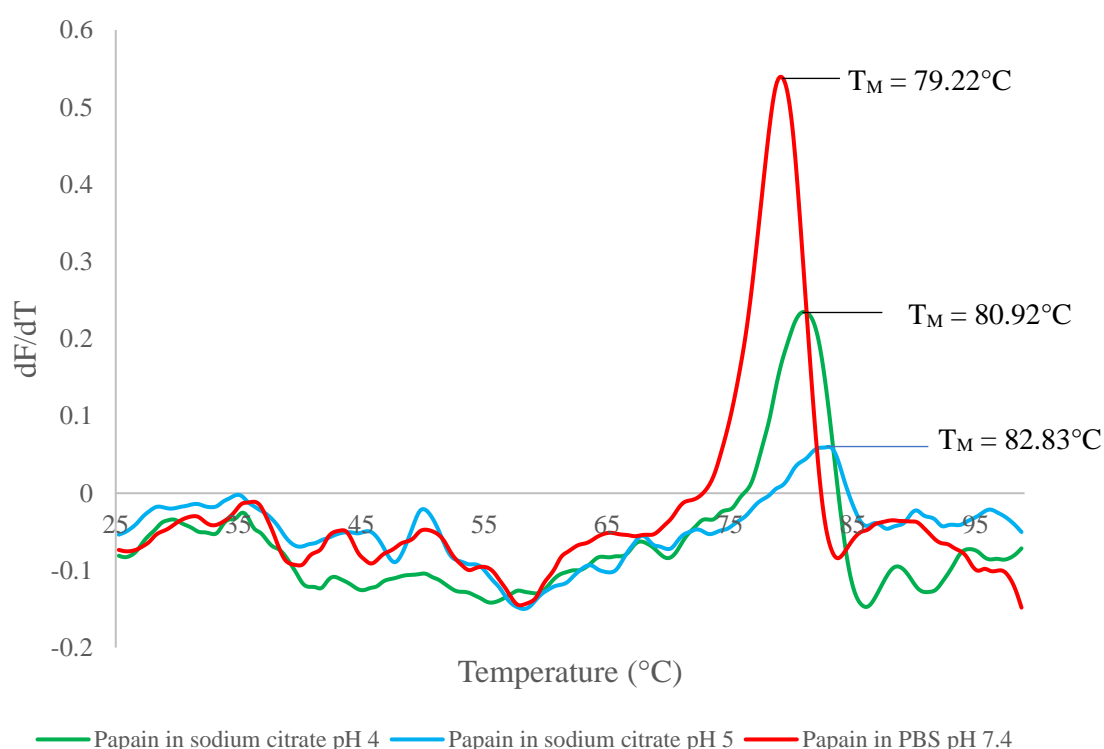
<b>Reinhard buffer library</b>	<b>Papain <math>T_M \pm \text{Std Dev}^a</math> (°C)</b>
Sodium citrate pH 4.0	80.92 $\pm$ 0.14
Sodium citrate pH 5.0	82.83 $\pm$ 0.14
Sodium citrate pH 5.5	82.13 $\pm$ 0.13
Sodium phosphate pH 6.0	81.95 $\pm$ 0.09
MES pH 6.7	80.33 $\pm$ 0.08
PIPES pH 6.7	80.65 $\pm$ 0.09
MOPS pH 7	80.52 $\pm$ 0.08
Sodium phosphate pH 7	79.50 $\pm$ 0.43
PBS pH 7.4	79.22 $\pm$ 0.06
Bicine pH 8	78.83 $\pm$ 0.14
HEPES pH 8	78.95 $\pm$ 0.09
<b>AMT<sup>b</sup> buffer library</b>	
AMT pH 4.5	83.40 $\pm$ 0.10
AMT pH 5.5	82.12 $\pm$ 0.13
AMT pH 8.0	80.92 $\pm$ 0.14

<sup>a</sup> Std Dev; standard deviation, n = 3

<sup>b</sup> AMT, Acetate-MES-Tris

Further decrease in the pH level, however, served to decrease the stability of papain. As shown in Table 3.3, the melting temperature decreased from 83.40°C at pH 4.5 to 80.92°C at pH 4.0). These findings correlated with a previous report in which papain was determined to undergo acid denaturation and unfold at pH 2.0 (Fosado-Quiroz and Rojo-Dominguez, 2011). The unfolding of papain was further explained by Huet and co-workers as they described papain undergoing state changes when exposed to acidic pH's (Huet *et al*, 2006). The authors concluded that papain appeared in its native state between pH 4-10 and transitioned to a molten globule state between pH 2-4 (Huet *et al*, 2006). The molten globule state was less stable than the native state. Accordingly, this can account for a decrease in stability for papain when solvated within the sodium citrate pH 4.0 buffer.

A representative derivative plot of papain exposed to three different pH conditions revealed that there was a significant  $T_M$  difference of  $3.85^\circ\text{C}$  ( $p < 0.05$ ) between the  $T_M$  obtained using PBS, pH 7.4, and that obtained when using sodium citrate, pH 5.0 (Figure 3.12). A significant difference ( $p < 0.05$ ) in melting point was also observed for papain in sodium citrate buffers of pH 4.0 and pH 5.0. It was also noted that the derivative of fluorescence against temperature obtained decreased as the melting point of papain increased. This decrease in fluorescence has been accounted for previously by the effect that the pH has on the binding of SYPRO orange to the hydrophobic regions of the protein (Ericsson *et al*, 2006; Reinhard *et al*, 2012; Huynh and Partch, 2016). Additionally, this trend has been described by Bai *et al* (2019) and Cimmerman *et al* (2008) as a reduction in available hydrophobic regions for the SYPRO orange dye to bind due to an increase in the stability of the protein. The decrease in the intensities of fluorescence observed, however, does not influence the melting point of the protein (Ericsson *et al*, 2006; Reinhard *et al*, 2012).



**Figure 3.12: Differential scanning fluorimetry analysis obtained for papain under different pH conditions.** The melting point for papain using three different buffers from the Reinhard *et al*, 2012 buffer library was obtained from the derivative plot.

### 3.6.2. Determining the effect of ligand binding on the stability of papain

The compounds obtained from the ChemBridge library were analysed at a final concentration 10  $\mu$ M and at pH 7.4 using PBS as the standard buffer. The average  $T_M$  obtained for the library of compounds with their standard deviation are shown in Table 3.4. The compound chymostatin, which was not part of the ChemBridge library, was used as a control. The compounds from the ChemBridge library were unable to shift the  $T_M$  of papain by a margin greater than 1°C when used at a final concentration of 10  $\mu$ M. The observed shifts in  $T_M$  of papain when in the presence of these compounds were negligible ( $p > 0.05$ ). The lack of binding may be due to the structure of the compounds or inadequate concentration of ligands tested.

**Table 3.4: The average melting temperature obtained for papain in the presence of distinct compounds and within phosphate buffered saline (pH 7.4) by means of thermal shift analysis.**

Compounds <sup>a</sup>	Compound ID (ChemBridge library)	Papain $T_M \pm$ Std Dev <sup>b</sup> (°C)
4-(3,4-dimethoxyphenyl)-1-phenyl-1H-pyrazol-5-amine	5118317	78.95 $\pm$ 0.09
N, N'-(1-methyl-4,4-piperidinediyl) diacetamide	5118841	79.17 $\pm$ 0.08
2-butyl-5-methylisophthalic acid	5119061	79.05 $\pm$ 0.18
N-[1-(1-adamantyl) ethyl]-N'-propylurea	5142981	79.03 $\pm$ 0.06
2-{[3-(trifluoromethyl)phenyl] amino} benzamide	5144439	79.00 $\pm$ 0.00
N-(tert-butyl)-2-hydroxy-3-(4-methoxyphenoxy)-1-propanaminium chloride	5144520	78.97 $\pm$ 0.06
3-[(1,1-dioxidotetrahydro-3-thienyl) amino]-1,2-propanediol	5155819	78.92 $\pm$ 0.08
3-(4-bromophenoxy) tetrahydrothiophene 1,1-dioxide	5155858	78.95 $\pm$ 0.18
4-(4-morpholinylsulfonyl) benzoic acid	5156995	79.10 $\pm$ 0.09
2-(3-chlorophenoxy) propanohydrazide	5189169	79.15 $\pm$ 0.00
1,6-diethyltetrahydroimidazo[4,5-d] imidazole-2,5(1H,3H)-dione	5233951	78.95 $\pm$ 0.09

1-(2-bromo-4,5-dimethoxybenzyl)-4-phenylpiperazine	5411732	79.10 ± 0.00
1-(3-chlorophenyl)-4-(2,5-difluorobenzyl)piperazine	5414621	78.97 ± 0.06
N-(2,6-difluorobenzyl)-N', N'-diethyl-N-methyl-1,2-ethanediamine	5415800	78.97 ± 0.06
1-(3-methylcyclopentyl)-4-(3-phenylprop-2-en-1-yl) piperazine dihydrochloride	5417271	79.00 ± 0.00
methyl 4-{[4-(2,3-dimethylphenyl)-1-piperazinyl] methyl} benzoate	5418608	79.18 ± 0.06
1-(2,5-difluorobenzyl)-4-methylpiperazine	5419034	79.08 ± 0.18
1-ethyl-4-(3-phenylbutyl) piperazine	5430819	79.03 ± 0.06
N-ethyl-N', N'-dimethyl-N-[2-(trifluoromethyl)benzyl]-1,2-ethanediamine	5430906	79.12 ± 0.13
4-(2-fluorobenzyl) morpholine	5431331	79.08 ± 0.08
N-[3-(dimethylamino) propyl]-N', N'-dimethyl-N-(2-naphthylmethyl)-1,3-propanediamine	5431334	79.00 ± 0.10
2-[(2,4-difluorobenzyl) (propyl)amino] ethanol	5431798	79.07 ± 0.06
1-[(2-methoxy-1-naphthyl) methyl]-4-methylpiperazine	5431935	78.95 ± 0.09
6-bromo-5-methyl-3H-imidazo[4,5-b] pyridine	5722012	79.00 ± 0.00
4-benzoyl-3-hydroxy-1-methyl-5-phenyl-1,5-dihydro-2H-pyrrol-2-one	5787093	79.07 ± 0.06
4-acetyl-5-(4-chlorophenyl)-1-[2-(diethylamino) ethyl]-3-hydroxy-1,5-dihydro-2H-pyrrol-2-one	5791634	78.88 ± 0.13
2-(benzylthio)-6,7-dihydro-5H-cyclopenta [4,5] thieno[2,3-d] pyrimidin-4-amine	6136720	79.03 ± 0.06
2-(4-methoxyphenyl)-3-[(2-phenylethyl) amino]-1H-inden-1-one	6194607	79.12 ± 0.13
1-allyl-4-benzoyl-3-hydroxy-5-phenyl-1,5-dihydro-2H-pyrrol-2-one	6197483	79.12 ± 0.03

2-(2-chloro-4-fluorobenzyl)-5-(2-methylphenyl)-2H-tetrazole	6271753	79.00 ± 0.15
5-chloro-N-ethyl-2-methoxy-N-phenylbenzamide	6612023	79.03 ± 0.06
4-chloro-1-(2-ethoxybenzoyl)-1H-pyrazole	6660132	79.08 ± 0.14
<b>10-(1-methylethylidene)-4-phenyl-4-azatricyclo [5.2.1.0~2,6~] dec-8-ene-3,5-dione</b>	<b>6714292<sup>c</sup></b>	<b>79.17 ± 0.14</b>
1-(2-chloro-4-nitrobenzoyl)-4-ethylpiperazine	6800944	79.00 ± 0.00
5-chloro-2-[(2-fluorobenzyl) oxy] benzaldehyde oxime	6832681	79.00 ± 0.00
5-chloro-2-[(3-fluorobenzyl) oxy] benzaldehyde oxime	6843219	78.87 ± 0.03
<b>1-[cyclohexyl(methyl)amino]-3-(4-methoxyphenoxy)-2-propanol hydrochloride</b>	<b>6943696<sup>d</sup></b>	<b>78.87 ± 0.13</b>
4-[4-(4-chlorophenyl)-1,3-thiazol-2-yl]-N, N-dimethyl-1-piperazinecarboxamide	6950620	78.3 ± 1.13
1-(4-chlorophenyl)-4-(tetrahydro-2-furanylcarbonyl) piperazine	7275637	79.12 ± 0.13
3-cyclopentyl-N-(4-pyridinylmethyl) propanamide	7293667	79.03 ± 0.06
ethyl 5-[(dimethylamino)carbonyl]-2-[(methoxycarbonyl)amino]-4-methyl-3-thiophenecarboxylate	7294926	78.93 ± 0.06
1-(2-ethylhexanoyl)-3,5-dimethyl-1H-pyrazole	7299305	78.98 ± 0.13
3-(4-chlorophenyl)-N-(2-methoxy-1-methylethyl) acrylamide	7364709	78.95 ± 0.13
2-{4-[3-(methylamino)-4-nitrophenyl]-1-piperazinyl} ethanol	7951028	79.08 ± 0.08
2-(cyclopentylamino)-N-(4-methoxy-2,5-dimethylphenyl)-2-thioxoacetamide	7952968	78.90 ± 0.09
N-([2-chloro-5-(trifluoromethyl) phenyl] amino) carbonyl)-2,2,2-trifluoroacetamide	7953011	78.98 ± 0.13

N-(4-chloro-2,5-dimethoxyphenyl)-7-(2,4-dichlorophenyl)-5-methyl-4,7-dihydro [1,2,4] triazolo[1,5-a] pyrimidine-6-carboxamide	7953379	79.03 ± 0.06
6-(1,3-benzodioxol-5-yl)-1,3,5-triazinane-2,4-dithione	7953623	78.88 ± 0.13
4-bromo-1-(2,6-dichlorobenzyl)-1H-pyrazole	7955822	79.03 ± 0.06
4,5-dimethyl-2-(4-nitrophenyl)-1H-imidazol-1-ol 3-oxide	7959564	79.03 ± 0.20
50 uM Chymostatin		84.38 ± 0.23
60 uM Chymostatin		83.48 ± 0.20

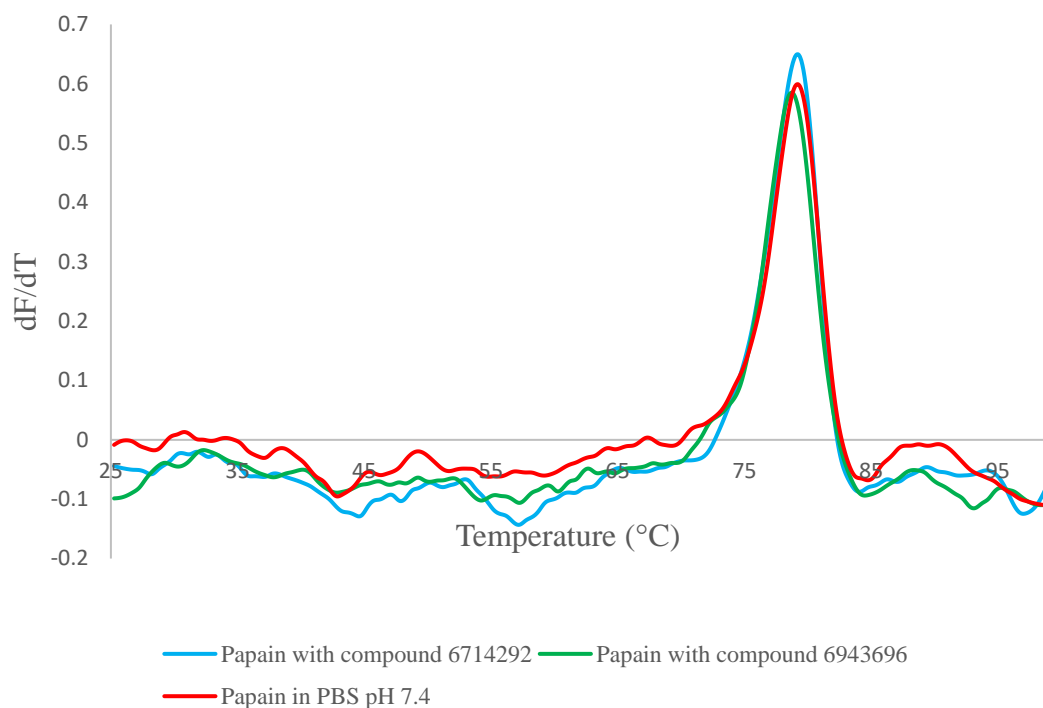
<sup>a</sup> Ligands tested at 10 µM

<sup>b</sup> Std Dev; Standard deviation, where n = 3.

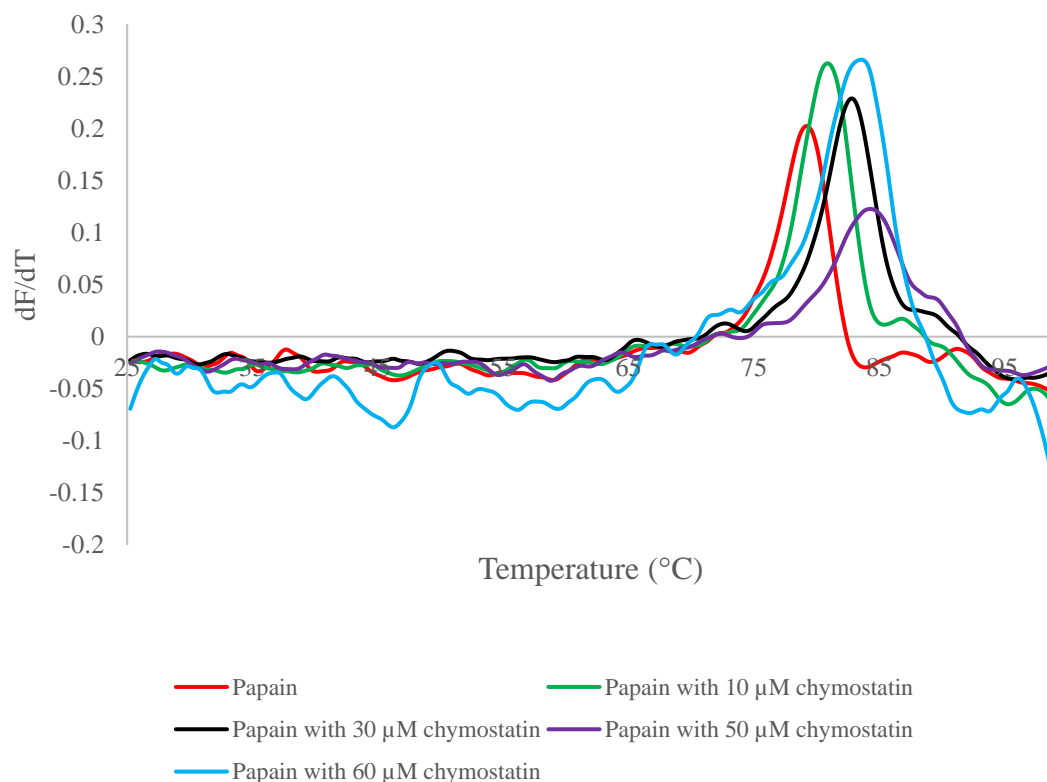
<sup>c-d</sup> Compounds used for representative dF/dT vs temperature (°C) plot

Compound 6714292 and compound 6943696 listed Table 3.4 were used to generate the representative plot (Figure 3.13). The control for the experiment was papain dissolved in PBS pH 7.4 in the absence of a compound which returned a  $T_M$  of 79.25°C. The  $T_M$  remained unchanged when compound 6714292 and compound 6943696 ( $p > 0.05$ ) were introduced. Furthermore, these two compounds did not affect the intensities of the fluorescence obtained. The TSA obtained for papain in the presence of increasing concentrations of chymostatin was used to generate the derivative plot (Figure 3.14). The melting point of the protein increased significantly ( $p < 0.05$ ) from 79.25°C to a maximum of 84.35°C as the concentration of chymostatin increased from 10 µM to 50 µM (Table 3.4). The addition of 60 µM chymotrypsin served to decrease the melting point which then decreased to 83.25°C.





**Figure 3.13: A representative differential scanning fluorimetry analysis obtained for papain with compounds from the ChemBridge library.** The  $dF/dT$  vs temperature was plotted for papain in the absence of a ligand and in the presence of ligands 6714292 and 6943696 at 10  $\mu\text{M}$ . The  $T_M$  of the samples are represented by the peaks in the graph.



**Figure 3.14: Differential scanning fluorimetry analysis of papain with increasing concentrations of chymostatin.** The  $dF/dT$  vs temperature for papain in the absence and the presence of chymostatin at concentrations ranging from 10  $\mu\text{M}$  to 60  $\mu\text{M}$  was plotted.

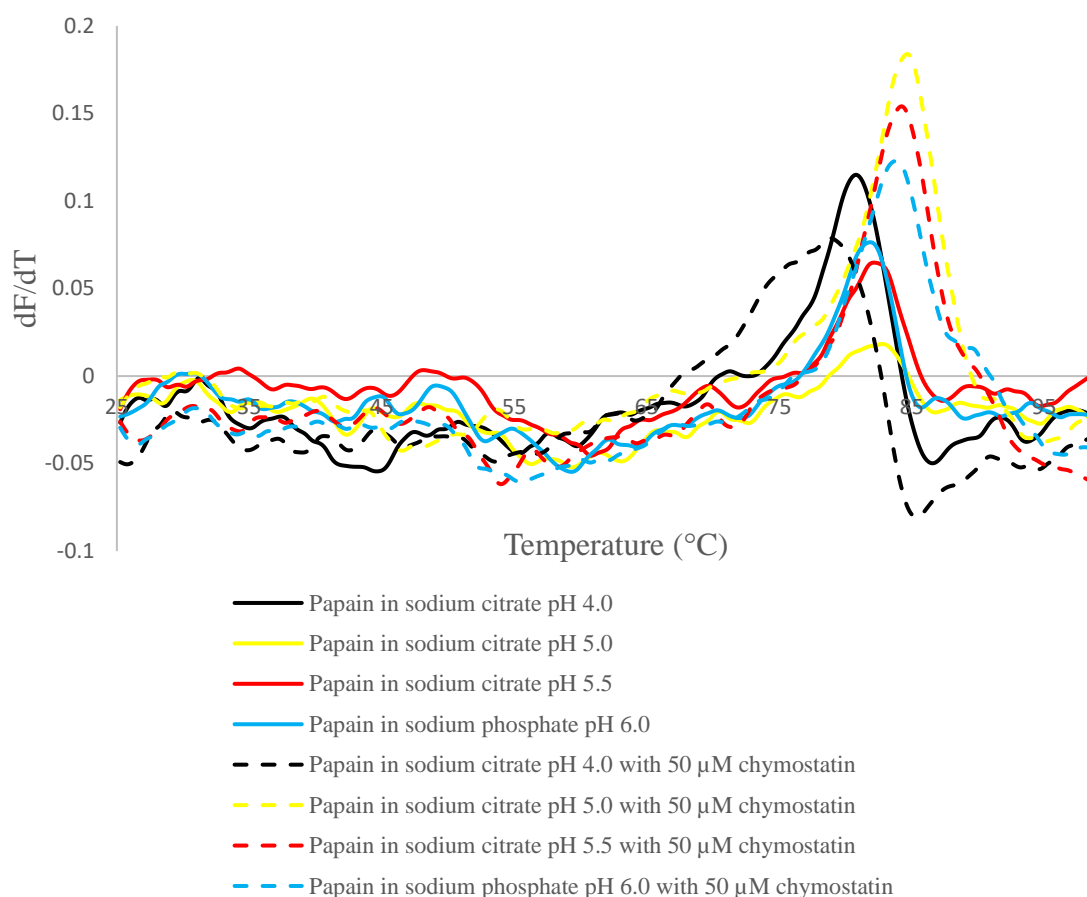
### 3.6.3. Determining the combined effect of pH and ligand binding on the stability of papain.

A combination of the buffer library and the control ligand, chymostatin, was used to evaluate if the melting temperature of papain could be increased further. The average  $T_M$  for each of the combinations tested is shown in Table 3.5. The derivative plot for these combinations and the observed  $T_M$  for a representative run are shown in Figure 3.15. The  $T_M$  of papain was shown to increase with a decrease in the pH of the buffers in which the protein was dissolved. The  $T_M$  for papain is increased further by the addition of 50  $\mu\text{M}$  chymostatin for each of the buffers except for sodium citrate pH 4.0. In this solitary instance in which a decrease in melting temperature was observed in the presence of chymostatin, the  $T_M$  decreased significantly from 80.68°C to 78.78°C ( $p < 0.05$ ). This decrease is supported by the state change explained by Huet *et al* (2006) and may be due to the inability of chymostatin to bind to the globular state of papain. The largest increase in  $T_M$  due to the addition of chymostatin was observed for papain in sodium citrate pH 5.5 in which the  $T_M$  increased from  $82.22 \pm 0.06^\circ\text{C}$  to  $84.35 \pm 0.09^\circ\text{C}$  - a statistically significant increase of  $2.13^\circ\text{C}$  ( $p < 0.05$ ). Taken together, the highest melting point obtained for papain was  $84.45 \pm 0.26^\circ\text{C}$ , which was obtained through the combination of chymostatin and a pH of 5.0 (Table 3.5).

**Table 3.5: The average  $T_M$  obtained for papain when testing combinations of pH and 50  $\mu\text{M}$  chymostatin**

Buffer and compound	Papain $T_M \pm \text{Std Dev}^a$ ( $^\circ\text{C}$ )
Sodium citrate pH 4.0	$80.68 \pm 0.12$
Sodium citrate pH 4.0 + 50 $\mu\text{M}$ chymostatin	$78.78 \pm 0.12$
Sodium citrate pH 5.0	$82.68 \pm 0.12$
Sodium citrate pH 5.0 + 50 $\mu\text{M}$ chymostatin	$84.45 \pm 0.26$
Sodium citrate pH 5.5	$82.22 \pm 0.06$
Sodium citrate pH 5.5 + 50 $\mu\text{M}$ chymostatin	$84.35 \pm 0.09$
Sodium phosphate pH 6.0	$81.71 \pm 0.15$
Sodium phosphate pH 6.0 + 50 $\mu\text{M}$ chymostatin	$83.78 \pm 0.06$

<sup>a</sup> Std Dev; Standard deviation, where  $n = 3$ .



**Figure 3.15: Differential scanning fluorimetry analysis obtained demonstrating the effectiveness of a buffer-ligand combination on the stability of papain.** The  $dF/dT$  vs temperature observed for papain under different combinations of buffer and ligand. The papain samples dissolved in different buffers in the absence of chymostatin are represented by solid lines and samples with the same pH conditions with the addition of 50  $\mu\text{M}$  chymostatin are represented by dashed lines.

### 3.7. Determining the stability of $TcoCATL_{cat}$ via the thermal shift assay

#### 3.7.1. Determining the effect of pH on the stability of $TcoCATL_{cat}$

The buffer libraries prepared were tested at constant volume with a constant concentration of  $TcoCATL_{cat}$  (0.1  $\mu\text{g} / \mu\text{l}$ ) and SYPRO orange dye (5X). The respective average melting temperatures obtained for  $TcoCATL_{cat}$  in each buffer was recorded (Table 3.6). It can be observed that the  $T_M$  for  $TcoCATL_{cat}$  reached its highest point when exposed to an acidic pH and gradually decreases as the pH is increased. This trend was observed in both the AMT buffer library (Figure 3.16) and the Reinhard *et al* (2012) library (Figure 3.17).

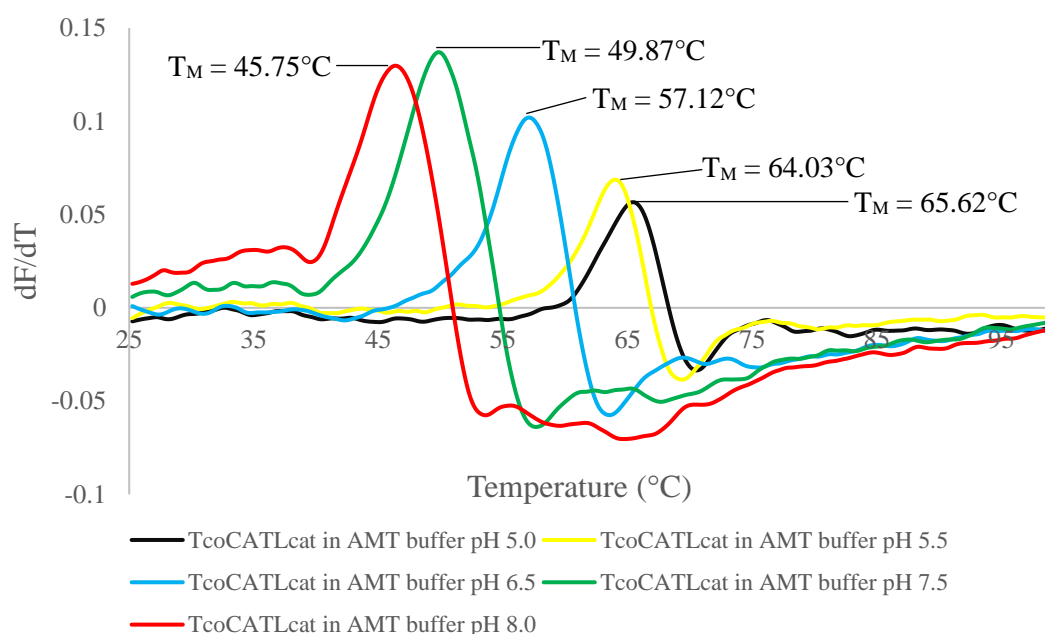
**Table 3.6: The average melting temperature obtained for *TcoCATL<sub>cat</sub>* under various buffer conditions.**

Reinhard buffer library	<i>TcoCATL<sub>cat</sub></i> <sup>a</sup>
	$T_M \pm \text{Std Dev}^b$ (°C)
Sodium citrate pH 4.0	65.27 $\pm$ 0.13
Sodium acetate pH 4.5	64.57 $\pm$ 0.20
Sodium citrate pH 5.0	63.45 $\pm$ 0.18
Sodium citrate pH 5.5	60.32 $\pm$ 0.12
Sodium phosphate PH 6.0	58.17 $\pm$ 0.08
Phosphate buffered saline pH 7.4	47.12 $\pm$ 0.16
Bicine pH 8	45.75 $\pm$ 0.10
CHES pH 9.5	37.17 $\pm$ 0.29
<b>AMT<sup>c</sup> buffer library</b>	
AMT pH 4.0	65.58 $\pm$ 0.29
AMT pH 4.5	65.80 $\pm$ 0.18
AMT pH 5.0	65.62 $\pm$ 0.20
AMT pH 5.5	64.03 $\pm$ 0.20
AMT pH 6.0	59.45 $\pm$ 0.43
AMT pH 6.5	57.12 $\pm$ 0.13
AMT pH 7.0	53.90 $\pm$ 0.09
AMT pH 7.5	49.87 $\pm$ 0.13
AMT pH 8.0	45.75 $\pm$ 0.35

<sup>a</sup> *TcoCATL<sub>cat</sub>*, catalytic domain of congopain

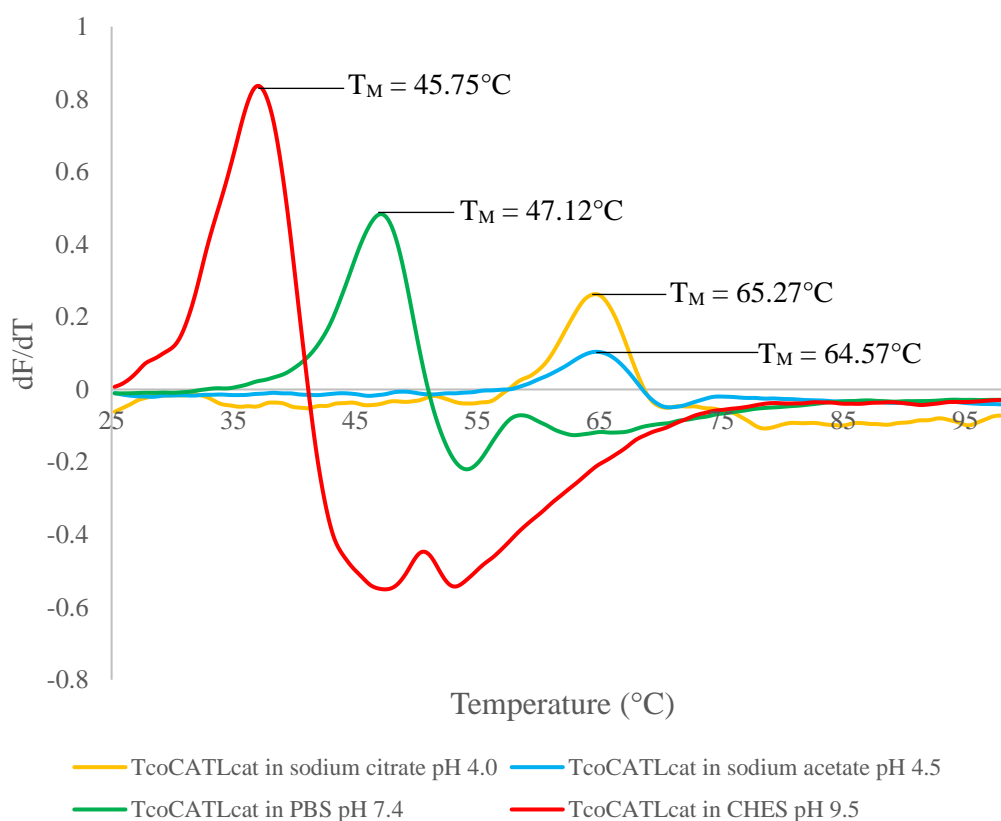
<sup>b</sup> Std Dev; standard deviation, n = 3

<sup>c</sup> AMT, Acetate-MES-Tris



**Figure 3.16: A representative differential scanning fluorimetry analysis obtained for the melting point of the catalytic domain of congopain (*TcoCATL<sub>cat</sub>*) in different Acetate-MES-Tris (AMT) buffers.** The  $dF/dT$  vs temperature observed for *TcoCATL<sub>cat</sub>* under various pH conditions using the AMT buffer library. The pH range displayed was between 5.0 and 8.0 whilst keeping the concentration of protein and dye constant (0.1  $\mu\text{g}/\mu\text{l}$  *TcoCATL* with 5X SYPRO Orange dye).

In the AMT buffers, *TcoCATL<sub>cat</sub>* yielded a minimum  $T_M$  of 45.75°C when solvated within AMT buffer pH 8.0 and a maximum  $T_M$  of 65.80°C when tested with the AMT buffer pH 4.5 (Table 3.6). The difference in  $T_M$  between the maximum and minimum  $T_M$  obtained in this buffer library was 20.05°C ( $p < 0.05$ ). The highest  $T_M$  for *TcoCATL<sub>cat</sub>* in the Reinhart library was observed when sodium citrate pH 4.0 and sodium acetate pH 4.5 were used. The melting temperatures recorded were 65.15°C and 64.75°C, respectively. The lowest melting temperature for *TcoCATL<sub>cat</sub>* was observed when using the CHES pH 9.5 buffer which yielded a  $T_M$  at 37.17°C. The  $T_M$  difference between the maximum and minimum observed was calculated to be 28.15°C ( $p < 0.05$ ). The catalytic domain of *TcoCATL* was previously described by Boulangé *et al* (2001) to be stable and demonstrate enzyme activity over an extensive range of pH with acidic pH conditions favoured. This correlates with the results shown in Figures 3.16 and 3.17 which demonstrated that the stability of the protein is favoured in acidic conditions.



**Figure 3.17: A representative differential scanning fluorimetry analysis obtained for the melting point of the catalytic domain of congopain (*TcoCATL<sub>cat</sub>*) at different pH using the Reinhart buffer library.** The dF/dT vs Temperature obtained for *TcoCATL<sub>cat</sub>* (0.1 µg / µl) under different pH conditions with 5X SYPRO orange dye. PBS, phosphate buffered saline.

### 3.7.2. Determining the effect of ligand binding on the stability of *TcoCATL<sub>cat</sub>*

The ChemBridge library of compounds was used to determine the effect of ligand binding on the  $T_M$  of *TcoCATL<sub>cat</sub>* at pH 7.4 and the average  $T_M$  obtained for each of these compounds are shown in Table 3.7. The compounds from the ChemBridge Library were tested at 10  $\mu$ M while chymostatin was tested at 50  $\mu$ M.

**Table 3.7: The average melting temperature obtained for *TcoCATL<sub>cat</sub>* in the presence of distinct compounds and within phosphate buffered saline (pH 7.4) as calculated through thermal shift analysis.**

Compound <sup>a</sup>	Compound ID (ChemBridge library)	<i>TcoCATL<sub>cat</sub></i> <sup>b</sup> $T_M \pm \text{Std Dev}^c$ ( $^{\circ}\text{C}$ )
4-(3,4-dimethoxyphenyl)-1-phenyl-1H-pyrazol-5-amine	5118317	$47.30 \pm 0.83$
N, N'-(1-methyl-4,4-piperidinediyl) diacetamide	5118841	$47.42 \pm 0.64$
2-butyl-5-methylisophthalic acid	5119061	$47.30 \pm 0.75$
N-[1-(1-adamantyl) ethyl]-N'-propylurea	5142981	$47.75 \pm 0.80$
2-{[3-(trifluoromethyl) phenyl] amino} benzamide	5144439	$47.33 \pm 0.88$
N-(tert-butyl)-2-hydroxy-3-(4-methoxyphenoxy)-1-propanaminium chloride	5144520	$47.27 \pm 0.78$
3-[(1,1-dioxidotetrahydro-3-thienyl) amino]-1,2-propanediol	5155819	$47.33 \pm 0.80$
3-(4-bromophenoxy) tetrahydrothiophene 1,1-dioxide	5155858	$47.22 \pm 0.85$
4-(4-morpholinylsulfonyl) benzoic acid	5156995	$47.30 \pm 0.84$
2-(3-chlorophenoxy) propanohydrazide	5189169	$47.25 \pm 0.69$
1,6-diethyltetrahydroimidazo[4,5-d]imidazole-2,5(1H,3H)-dione	5233951	$47.23 \pm 0.66$
1-(2-bromo-4,5-dimethoxybenzyl)-4-phenylpiperazine	5411732	$47.36 \pm 1.00$

1-(3-chlorophenyl)-4-(2,5-difluorobenzyl) piperazine	5414621	46.95 ± 0.69
N-(2,6-difluorobenzyl)-N', N'-diethyl-N-methyl-1,2-ethanediamine	5415800	47.15 ± 0.43
1-(3-methylcyclopentyl)-4-(3-phenylprop-2-en-1-yl) piperazine dihydrochloride	5417271	47.05 ± 0.88
methyl 4- {[4-(2,3-dimethylphenyl)-1-piperazinyl] methyl} benzoate	5418608	46.93 ± 0.72
1-(2,5-difluorobenzyl)-4-methylpiperazine	5419034	46.88 ± 0.40
1-ethyl-4-(3-phenylbutyl) piperazine	5430819	47.03 ± 0.64
N-ethyl-N', N'-dimethyl-N-[2-(trifluoromethyl) benzyl]-1,2-ethanediamine	5430906	47.17 ± 0.72
4-(2-fluorobenzyl) morpholine	5431331	47.03 ± 0.92
N-[3-(dimethylamino) propyl]-N', N'-dimethyl-N-(2-naphthylmethyl)-1,3-propanediamine	5431334	47.08 ± 0.69
2-[(2,4-difluorobenzyl) (propyl)amino] ethanol	5431798	46.85 ± 0.69
1-[(2-methoxy-1-naphthyl) methyl]-4-methylpiperazine	5431935	46.87 ± 0.34
6-bromo-5-methyl-3H-imidazo[4,5-b]pyridine	5722012	46.82 ± 0.72
4-benzoyl-3-hydroxy-1-methyl-5-phenyl-1,5-dihydro-2H-pyrrol-2-one	5787093	47.25 ± 1.17
4-acetyl-5-(4-chlorophenyl)-1-[2-(diethylamino) ethyl]-3-hydroxy-1,5-dihydro-2H-pyrrol-2-one	5791634	47.17 ± 0.72
2-(benzylthio)-6,7-dihydro-5H-cyclopenta[4,5] thieno[2,3-d] pyrimidin-4-amine	6136720	46.95 ± 0.93
2-(4-methoxyphenyl)-3-[(2-phenylethyl) amino]-1H-inden-1-one	6194607	47.55 ± 0.52

1-allyl-4-benzoyl-3-hydroxy-5-phenyl-1,5-dihydro-2H-pyrrol-2-one	6197483	47.17 ± 0.62
2-(2-chloro-4-fluorobenzyl)-5-(2-methylphenyl)-2H-tetrazole	6271753	46.88 ± 0.32
5-chloro-N-ethyl-2-methoxy-N-phenylbenzamide	6612023	46.83 ± 0.33
4-chloro-1-(2-ethoxybenzoyl)-1H-pyrazole	6660132	46.92 ± 0.29
10-(1-methylethylidene)-4-phenyl-4-azatricyclo [5.2.1.0~2,6~] dec-8-ene-3,5-dione	6714292	46.92 ± 0.21
1-(2-chloro-4-nitrobenzoyl)-4-ethylpiperazine	6800944	47.08 ± 0.38
5-chloro-2-[(2-fluorobenzyl) oxy] benzaldehyde oxime	6832681	46.97 ± 0.38
5-chloro-2-[(3-fluorobenzyl) oxy] benzaldehyde oxime	6843219	47.05 ± 0.84
1-[cyclohexyl(methyl)amino]-3-(4-methoxyphenoxy)-2-propanol hydrochloride	6943696	46.85 ± 0.35
4-[4-(4-chlorophenyl)-1,3-thiazol-2-yl]-N,N-dimethyl-1-piperazinecarboxamide	6950620	46.70 ± 0.13
<b>1-(4-chlorophenyl)-4-(tetrahydro-2-furanylcarbonyl) piperazine</b>	<b>7275637<sup>d</sup></b>	<b>47.37 ± 0.65</b>
3-cyclopentyl-N-(4-pyridinylmethyl) propanamide	7293667	47.13 ± 0.54
ethyl 5-[(dimethylamino)carbonyl]-2-[(methoxycarbonyl)amino]-4-methyl-3-thiophenecarboxylate	7294926	47.30 ± 1.04
1-(2-ethylhexanoyl)-3,5-dimethyl-1H-pyrazole	7299305	47.13 ± 0.78
3-(4-chlorophenyl)-N-(2-methoxy-1-methylethyl) acrylamide	7364709	46.75 ± 0.4



2-{4-[3-(methylamino)-4-nitrophenyl]-1-piperazinyl} ethanol	7951028	46.80 ± 0.40
2-(cyclopentylamino)-N-(4-methoxy-2,5-dimethylphenyl)-2-thioxoacetamide	7952968	46.68 ± 0.20
N-([2-chloro-5-(trifluoromethyl) phenyl] amino) carbonyl)-2,2,2-trifluoroacetamide	7953011	46.63 ± 0.25
N-(4-chloro-2,5-dimethoxyphenyl)-7-(2,4-dichlorophenyl)-5-methyl-4,7-dihydro [1,2,4] triazolo[1,5-a] pyrimidine-6-carboxamide	7953379	46.67 ± 0.21
<b>6-(1,3-benzodioxol-5-yl)-1,3,5-triazinane-2,4-dithione</b>	<b>7953623<sup>e</sup></b>	<b>46.25 ± 0.25</b>
4-bromo-1-(2,6-dichlorobenzyl)-1H-pyrazole	7955822	46.70 ± 0.49
4,5-dimethyl-2-(4-nitrophenyl)-1H-imidazol-1-ol 3-oxide	7959564	46.83 ± 0.60
50 μM Chymostatin		62.43 ± 0.27

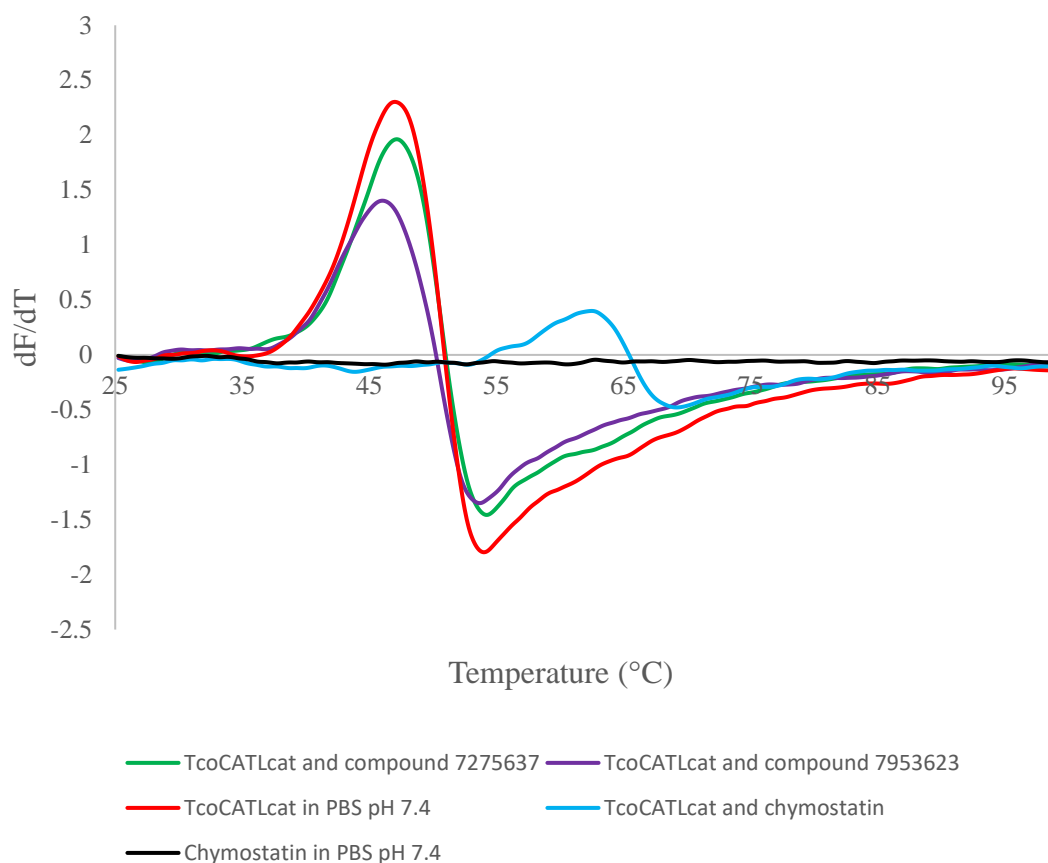
<sup>a</sup> Compounds tested at 10 μM

<sup>b</sup> *TcoCATL<sub>cat</sub>*, catalytic domain of congopain

<sup>c</sup> Std Dev; standard deviation, n = 3

<sup>d-e</sup> Compounds used for representative dF/dT vs temperature (°C) plot

The compounds from the ChemBridge library had little to no effect on the stability of *TcoCATL<sub>cat</sub>* as non-significant shifts ( $p > 0.05$ ) in  $T_M$  observed. The compounds which yielded the largest difference in  $T_M$  for *TcoCATL<sub>cat</sub>* were used to construct a derivative plot represented in Figure 3.18. The  $T_M$  observed for *TcoCATL<sub>cat</sub>* in the absence of a ligand was 47°C. An increase in  $T_M$  was observed with the addition of compound 7275637 as the  $T_M$  increased to 47.15°C and the largest decrease in  $T_M$  was observed with compound 7953623 as the  $T_M$  decreased to 46.00°C. These negligible shifts suggest that the compounds tested are incapable of binding to the protein at 10 μM. In contrast, the addition of chymostatin influenced a significant increase ( $p < 0.05$ ) in the melting temperature of *TcoCATL<sub>cat</sub>* by 15.31°C to yield a final maximum  $T_M$  of 62.43°C.



**Figure 3.18: A representative differential scanning fluorimetry analysis of congopain using compounds from the ChemBridge library and using chymostatin as a positive control.** The  $dF/dT$  vs temperature for  $TcoCATL_{cat}$  in absence and presence of two selected compounds from the ChemBridge library with chymostatin (50  $\mu M$ ) serving as a positive control and chymostatin in the absence of  $TcoCATL_{cat}$  as a negative control when using 5X SYPRO orange dye.

### 3.7.3. Determining the combined effect of pH and ligand binding on the stability of $TcoCATL_{cat}$

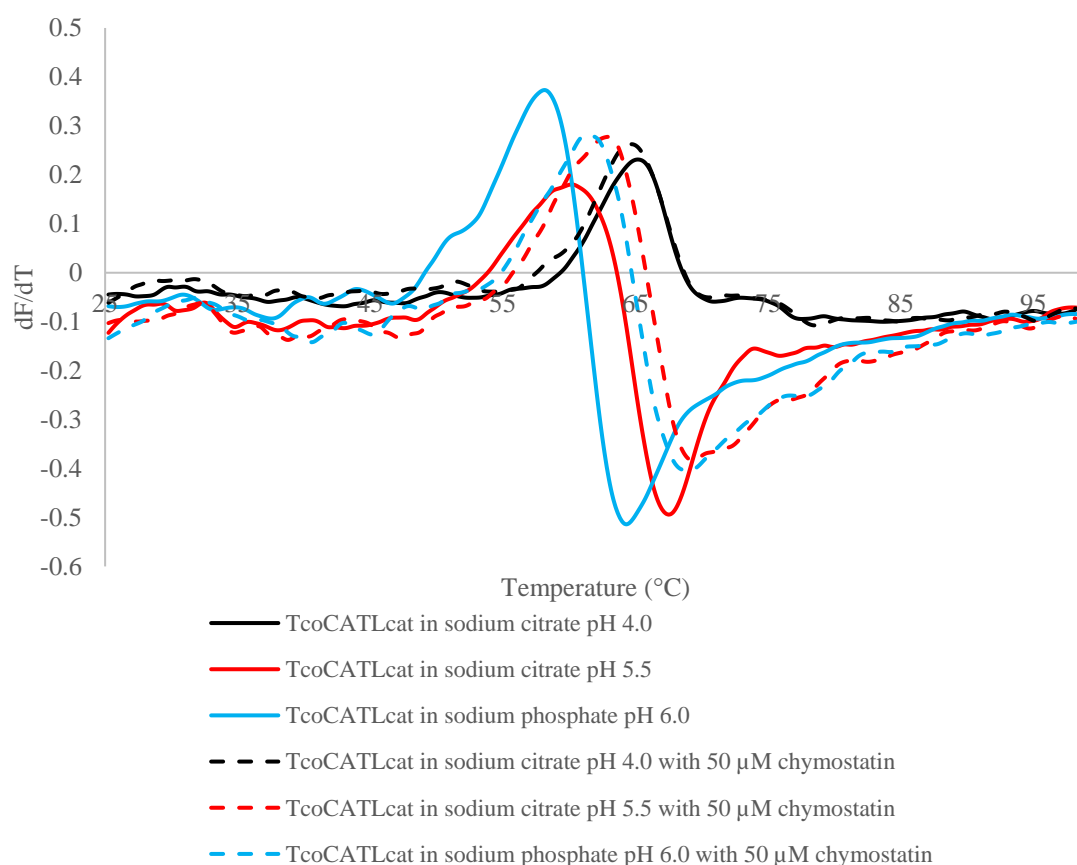
A combination of the buffer library and chymostatin was used to further stabilise  $TcoCATL_{cat}$ . For each analysis, the buffer was kept at the same volume, the concentration of chymostatin was evaluated at 50  $\mu M$  and the final concentration of  $TcoCATL_{cat}$  was kept constant at 0.1  $\mu g / \mu l$ . The average  $T_M$  recorded for each of the pH-ligand combinations tested are shown in Table 3.8. The derivative plot for these pH-ligand combinations and the observed  $T_M$  for a representative run are shown in Figure 3.19.

**Table 3.8: The average  $T_M$  obtained for  $TcoCATL_{cat}$  in the presence of 50  $\mu M$  chymostatin and varying pH.**

Buffer and compound	$TcoCATL_{cat}^a$ $T_M \pm Std Dev^b$ ( $^{\circ}C$ )
Sodium citrate pH 4.0	$65.32 \pm 0.21$
Sodium citrate pH 4.0 + 50 $\mu M$ chymostatin	$64.72 \pm 0.12$
Sodium citrate pH 5.0	$63.48 \pm 0.20$
Sodium citrate pH 5.0 + 50 $\mu M$ chymostatin	$64.53 \pm 0.10$
Sodium citrate pH 5.5	$60.52 \pm 0.25$
Sodium citrate pH 5.5 + 50 $\mu M$ chymostatin	$63.05 \pm 0.09$
Sodium phosphate pH 6.0	$58.18 \pm 0.06$
Sodium phosphate pH 6.0 + 50 $\mu M$ chymostatin	$61.35 \pm 0.09$

<sup>a</sup>  $TcoCATL_{cat}$ , catalytic domain of congopain

<sup>b</sup> Std Dev; standard deviation,  $n = 3$



**Figure 3.19: Differential scanning fluorimetry analysis obtained for the combined effect of pH and ligand binding on the stability of the catalytic domain of congopain ( $TcoCATL_{cat}$ ). The  $dF/dT$  vs Temperature observed for  $TcoCATL_{cat}$  under different combinations of buffer and ligand. The solid lines represent the  $TcoCATL_{cat}$  samples which were made up of different buffers without the presence of chymostatin and the broken lines represent the same buffer with the addition of 50  $\mu M$  chymostatin.**

The combination of using the buffer library with chymostatin on *TcoCATL<sub>cat</sub>* resulted in further stabilisation of the protein. The  $T_M$  for *TcoCATL<sub>cat</sub>* was increased further by the addition of 50  $\mu$ M chymostatin for each of the buffers except for sodium citrate pH 4.0 as the  $T_M$  decreased from 65.32°C to 64.72°C. This decrease was, however, negligible ( $p > 0.05$ ). The lack of a shift was likely due to the inability of chymostatin to bind to the protein at that pH. The greatest increase in  $T_M$  due to the combination of pH and ligand was observed for sodium citrate pH 6.0 as the  $T_M$  increased by 3.17°C from  $58.18 \pm 0.06^\circ\text{C}$  to  $61.35 \pm 0.09^\circ\text{C}$ .

## 4. DISCUSSION

---

Proteases are of significant scientific importance and are subject to extensive investigation for research purposes and industrial applications (Mótyán *et al*, 2013). Cysteine proteases are ubiquitous enzymes with multifaceted roles in a range of biological organisms and have been previously exploited for industrial purposes (Mótyán *et al*, 2013). This study was designed with the aim of determining the precise experimental conditions that would favour the stabilisation of two cysteine proteases; namely papain and *TcoCATL<sub>cat</sub>* (commonly known as congopain). For this study, papain was selected as it is the founder enzyme of the large C1 family of papain-like cysteine proteases and widely used as a model representative of cysteine proteases (Kimmel and Smith, 1954). A wealth of knowledge exists for this enzyme as it has been extensively studied and successfully crystallised (Kamphuis *et al*, 1984). The second cysteine protease selected for this study was *TcoCATL<sub>cat</sub>*. As the major cysteine protease of *Trypanosoma congolense*, *TcoCATL<sub>cat</sub>* is considered a potential drug target and diagnostic. While much is known about this enzyme, *TcoCATL<sub>cat</sub>* is yet to be successfully crystallised and the 3-D structure is yet to be resolved.

Protein crystallisation typically involves large-scale screening of the protein in buffers of different composition and subsequent observation of crystal formation (Uson and Sheldrick, 1999; Wlodawer *et al*, 2007). A more recent, empirical approach to protein crystallisation involves the optimisation of solution parameters prior to attempted crystallisation, such as the determination of buffers and buffer components that serve to enhance protein stability (Wlodawer *et al*, 2007). A useful technique for determining protein stability, and one which is amenable to high throughput screening, is differential scanning fluorimetry (Pantoliano *et al*, 2001; Niesen *et al*, 2007). This technique is effectively applied in the thermal shift assay, in which the melting temperature or  $T_M$  of a protein in monomeric form can be precisely determined following addition of a hydrophobic-binding fluorescent dye and detection *via* RT-PCR (Pantoliano *et al*, 2001; Niesen *et al*, 2007). As melting point is inherently dependent on the stability of the protein, conditions that increase stability can be determined through the identification of conditions that significantly increase the  $T_M$  (Pantoliano *et al*, 2001; Ericsson *et al*, 2006; Niesen *et al*, 2007). Similarly, this assay can be employed to determine stoichiometry of proteins as the separation of a protein dimer will produce a fluorescent peak at a lower melting temperature than the unfolding of the monomers.

Additionally, the thermal shift assay has been effectively utilised to identify small molecule protein binders as ligand binding similarly serves to increase the  $T_M$  of the target protein (Vedadi *et al*, 2006).

Accordingly, the thermal shift assay was considered the ideal method to determine the optimal conditions required to stabilise papain and  $TcoCATL_{cat}$ . In order to achieve this aim, a number of objectives were addressed. Specifically, it was imperative to determine whether the two cysteine proteases were viable candidates for the thermal shift assay and then to obtain protein in sufficient concentration and purity to execute the planned experiments. Additionally, it was necessary to confirm the proteins' identity and activity to ensure that a correctly-folded form of the protein was used in the thermal shift assay. Thereafter, two distinct buffer libraries and a small compound library were prepared and evaluated for their effect on the melting temperature of two cysteine proteases. Finally, it was sought to determine whether the melting temperature of a buffer-stabilised protein could be further stabilised through the addition of a weakly-binding small protease inhibitor, chymostatin.

As discussed below in more detail, each of these objectives were successfully fulfilled, culminating in the identification of buffer conditions that optimally stabilised both proteins and the conclusion that chymostatin serves to stabilise both proteins. The thermal shift assay also neatly demonstrated the presence of  $TcoCATL_{cat}$  in distinct dimer and monomer forms and clearly illustrated that chymostatin binds to  $TcoCATL_{cat}$  in monomeric form. Taken together, this study details the experimentally-determined optimal conditions for enhancing the stability of papain and  $TcoCATL_{cat}$ . The findings can therefore be used to facilitate further biochemical research of these proteins and, in particular, guide crystallisation efforts of  $TcoCATL_{cat}$ . The findings may also provide a broad guideline for optimising the stability of other members of the cysteine protease family.

#### **4.1. Papain and $TcoCATL_{cat}$ are suitable for use in the thermal shift assay**

A literature search indicated that neither papain nor  $TcoCATL_{cat}$  have been subjected to thermal shift assays to determine the proteins' thermal stability. Accordingly, the cysteine proteases selected for this study were analysed to determine their suitability for use within the thermal shift assay. In general, proteins that are suitable for thermal shift analysis require low surface hydrophobicity, a hydrophobic core and a melting

temperature within the range of the instrument used (i.e. >90°C in this case) (Pantoliano *et al*, 2001; Ericsson *et al*, 2006; Niesen *et al*, 2007). Accordingly, the hydrophobicity of both proteins were determined through amino acid analysis, hydrophobicity plots, helical wheels and hydrophobicity cluster analysis. This analysis revealed that, although papain and *TcoCATL<sub>cat</sub>* both contain approximately 33% hydrophobic residues, papain has higher overall hydrophobicity owing to a larger content of the amino acids which score highly according to the Kyte and Doolittle index (i.e. isoleucine, leucine and valine) (Kyte and Doolittle, 1982). Additionally, papain has a larger hydrophobic face than *TcoCATL<sub>cat</sub>* and a greater number of hydrophobic clusters. Interestingly, the hydrophobicity cluster analysis revealed numerous common clusters supporting published homology models showing that the two cysteine proteases may have secondary structure similarities (Lalmanach *et al*, 2002). Papain was observed to have greater surface hydrophobicity than *TcoCATL<sub>cat</sub>*, however, both proteins displayed prominent hydrophobic cores which rendered them suitable for the thermal shift assay. Finally, theoretical analysis predicted that both proteins would produce  $T_M$  values higher than 65°C. Taken together, the two cysteine proteases selected for this study were deemed viable candidates for analysis *via* the thermal shift assay.

#### **4.2. Biochemical characterisation of the cysteine proteases selected for thermal shift analysis**

Prior to undertaking the thermal shift assay, it was essential to ensure the proteins were purified to homogeneity and in active forms. Papain was purchased from a commercial source as a purified product and its molecular weight and activity were confirmed through SDS-PAGE and a gelatin-containing zymogram, respectively. Following recombinant expression, *TcoCATL<sub>cat</sub>* was purified and resolved into a 54 kDa dimer and 26 kDa monomer through non-reducing and reducing SDS-PAGE, respectively. The identity of the protein was confirmed through western blot and its proteolytic activity using a gelatin-containing zymogram. Notably, a single band was observed for the protein in both the SDS-PAGE and the zymogram, however, it appeared at a higher molecular weight in the zymogram. This distinction has been previously observed by Heussen and Dowdle (1980) who pioneered using substrate-containing polyacrylamide gel to show proteolytic activity of plasminogen activators. Similarly, Hummel and co-workers observed that a metalloprotease from the basidiomycete fungus *Schizophyllum*

*commune* of 57 kDa appeared at a molecular weight of 70 kDa when separated on a gelatin containing SDS-PAGE (Hummel *et al.*, 1996). The slower migration observed was attributed to an interaction between the protease evaluated and the substrate by Hummel *et al* (1996). This may account for the higher position of the band observed within the zymogram obtained in this study.

#### **4.3. The stability of the cysteine proteases under neutral pH conditions.**

The thermal stability of a protein has always been a key factor towards the evaluation of a protein as it is linked with the functional activity of a protein and the structural folding of a protein (Chen *et al*, 2008; Scheiblhofer *et al*, 2017). The melting temperature of a protein is a popular measure of a protein's thermal stability. The melting temperature of a protein may be easily determined if there is an established 3-D structure available for the required protein. There is currently no available 3-D structure available for *TcoCATL<sub>cat</sub>* and therefore a precise theoretical melting point could not be determined based on the proteins structure.

There are over 10 million amino acid sequences deposited in the UniProtKB/TrEMBL protein databases with less than 1% of these entries having verified protein structures. This led to numerous attempts at determining the thermal stability of a protein from the verified amino acid sequence (Ku *et al*, 2009). A theoretical melting point range was determined for papain and *TcoCATL<sub>cat</sub>* based on the amino acid sequence of the protein. This determination of the melting point was derived from using the online  $T_M$  index program created by Ku *et al* (2009) which assigns a score for each amino acid in the sequence. The scores are generated based on the identity, position of the amino acid in the sequence and the identity and position of the amino acid before and after in the sequence, under neutral pH conditions. This method for determining a theoretical  $T_M$  is capable of identifying a  $T_M$  range rather than identifying a precise melting point (Ku *et al*, 2009). The theoretical  $T_M$  range for papain and *TcoCATL<sub>cat</sub>* were both predicted to be greater than 65°C and this  $T_M$  is regarded as a high  $T_M$  based on the program index. The range determined from the program was limited, however, as it can only take the amino acid composition into consideration and does not factor in different transitional states of a protein or any other factors that can affect the stability of a protein such as a change in pH or the introduction of a ligand (Ku *et al*, 2009). The program is able to detect changes in  $T_M$  due to mutations. Stability achieved due to changes in the amino



acid composition may be observed, however, it will only be detected if these changes are sufficient to increase the  $T_M$  from a  $T_M$  index of  $<1$  which correlates to a  $T_M$  of  $<55^\circ\text{C}$  to a  $T_M$  index of  $>1$  which correlates to a  $T_M$  of  $>65^\circ\text{C}$ .

The TSA was implemented to derive the experimental  $T_M$  of papain and  $TcoCATL_{cat}$  under neutral pH conditions where the pH was kept at 7.4 by using PBS buffer. The Protein Thermal Shift<sup>TM</sup> Starter Kit (ThermoFisher Scientific, USA) was first used to optimise the parameters of the TSA for use with the Rotor-Gene<sup>TM</sup> 6000 2-Plex RT-PCR machine. The kit contained a control ligand capable of stabilising the control protein and it was observed that increasing the concentration of ligand resulted in an increase in  $T_M$  and an increase in the fluorescence obtained. The concentration of the SYPRO orange dye and the protein of interest does not affect the  $T_M$  obtained, but rather affects the intensity of the fluorescence obtained (Ericsson *et al*, 2006; Niesen *et al*, 2007; Lavinder *et al*, 2009).

The initial experimental parameters were determined by following those outlined when using the commercial thermal shift kit. These parameters were then optimised for papain and  $TcoCATL_{cat}$ . The experimental  $T_M$  observed under neutral pH conditions for papain was  $79.22 \pm 0.06^\circ\text{C}$  and the  $T_M$  for  $TcoCATL_{cat}$  was  $47.12 \pm 0.16^\circ\text{C}$ . The experimental  $T_M$  observed for papain correlates with the predicted  $T_M$  range  $>65^\circ\text{C}$  however this correlation was not observed for  $TcoCATL_{cat}$  as the experimental  $T_M$  obtained from the TSA revealed a  $T_M < 65^\circ\text{C}$ . However, this is accounted for by the dimerisation of  $TcoCATL_{cat}$  under physiological conditions (Boulangé *et al*, 2011). The online  $T_M$  predictor is incapable of incorporating this structural change into its algorithm. The  $T_M$  of  $TcoCATL_{cat}$  obtained from the TSA at neutral pH possibly represents the dimer dissociating at a lower temperature rather than protein unfolding. Papain was not affected by this structural change as it appears as a monomer under neutral pH conditions and does not transition to a dimer through pH manipulation (Sorrentino *et al*, 1981; Boulangé *et al*, 2001). Rather, the dimerisation of papain can be effected by the introduction of cross linkage reagents such as dimethyl suberimidate (Sorrentino *et al*, 1981).

#### **4.4. The factors contributing towards the stabilisation of the two cysteine proteases**

The stability of the two cysteine proteases were evaluated under neutral pH conditions and it was observed that *TcoCAT<sub>cat</sub>* had a lower melting point than papain. The effect of pH manipulation and ligand binding were evaluated to determine if the melting temperature, and therefore the stability, of the cysteine proteases could be increased further.

##### **4.4.1. Increasing stability of the cysteine proteases by decreasing pH**

The buffer library designed by Reinhard *et al* (2012) was used in initial thermal shift assay experiments to test the effect of pH on the stability of papain and *TcoCAT<sub>Lcat</sub>*. The library prepared consisted of the buffers A1 - B12 from Reinhard *et al* (2012) as this provided the required pH range. The thermal shift assays performed using the Reinhard buffer library revealed that both cysteine proteases favoured an acidic pH as both showed general trends of decreasing stability as the pH increased above 5.5. The Reinhard library consisted of buffers which differed in pH and ionic strength and it is known that the ionic strength of a buffer solution has an effect on the stability of a protein. Therefore any changes in stability observed using the Reinhard buffer library could not be accounted for solely by the pH as the ionic strength variation introduced a contributing factor. A second library was therefore constructed with buffers of constant ionic strength to delineate between the influence of pH and ionic strength on protein stability. Evaluation of the AMT buffer library (Ellis and Morrison, 1982) yielded similar findings as the evaluation of the Reinhard library, whereby the cysteine proteases were found to be preferentially stabilised in a slightly acidic pH.

The effect of ionic strength on stability was clearly evident and impacted both cysteine proteases. This was observed, for instance, when comparing the  $T_M$  of papain when evaluated at pH 8.0 using the two different libraries. There, a significant difference in  $T_M$  of 2.09°C ( $p < 0.05$ ) was observed between the result obtained in AMT ( $T_M = 80.92^\circ\text{C}$ ) and that obtained in Bicine ( $T_M = 78.83^\circ\text{C}$ ). The difference in stability was accounted for by the higher ionic strength of the AMT buffer as compared to the Bicine buffer. This result correlates with findings by Yang *et al* (1994) and Dominy *et al* (2002) who tested the effect of ionic strength on protein stability. These authors observed that the overall net charge of the protein determined whether an increased

ionic strength favours protein stability or results in destabilisation of the protein (Yang and Honig, 1994; Dominy *et al*, 2002).

The  $T_M$  of papain was increased from 79.22°C under physiological conditions to 83.40°C using the AMT buffer at pH 4.5 which represented an increase in  $T_M$  of 4.18°C ( $p < 0.05$ ). The increase in  $T_M$  observed for  $TcoCATL_{cat}$  was even greater as the  $T_M$  increased from 47.12°C under physiological conditions to 65.80°C when using AMT buffer pH 4.5, which equates to a significant increase of 18.68°C ( $p < 0.05$ ). This significant shift in  $T_M$  possibly represents the transition of  $TcoCATL_{cat}$  in the monomeric form under physiological pH conditions to a dimeric form under acidic conditions. The lower melting point is indicative of the dissociation of a dimer and not the unfolding of the protein.

The  $T_M$  observed for  $TcoCATL_{cat}$  when tested under physiological conditions revealed a low  $T_M$  with a high  $dF/dT$  which indicated a higher raw fluorescence. The higher  $T_M$  and lower  $dF/dT$  (thus indicating a lower raw fluorescence obtained) was observed when using acidic pH conditions. This fluctuation in raw fluorescence was not initially anticipated as the concentration of the protein was kept constant. The conformational changes occurring with  $TcoCATL_{cat}$  under different pH conditions was not considered as a factor which contributed towards the decreased fluorescence as the same trend in fluorescence intensity was observed for papain for which there is no known conformational change due to pH changes. The trends observed in the raw fluorescence of both cysteine proteases indicated that the decrease in raw fluorescence of the proteins under acid pH conditions is not related to conformational changes of the protein occurring but may be due to the poor binding affinity of the SYPRO orange dye under acidic conditions. Alternatively, the proteins may be stabilised to a point and in such a manner that significantly fewer hydrophobic regions are exposed.

#### **4.4.2. Stabilisation of the selected proteases *via* the introduction of a ligand**

The thermal shift assay was used to detect changes in stability of the cysteine proteases when a ligand was introduced. For this study, E-64, the cysteine protease class-specific inhibitor, was considered for use in detecting a shift in the stability of the two cysteine proteases due to ligand binding as it is a strong inhibitor of cysteine proteases. Since E-64 was first isolated from a culture of *Aspergillus japonicus*, it has been used as a common and potent inhibitor for numerous cysteine proteases including those of the

papain family (Matsumoto *et al*, 1999). The inhibitor is classified as a covalent-type irreversible inhibitor and commonly binds to the  $S_n$  ( $n= 1\sim3$ ) subsites of cysteine proteases of the papain family due to the preserved tertiary structure of the C  $\alpha$ -backbone chains (Matsumoto *et al*, 1999). The N-terminal carboxyl group of E-64 forms three hydrogen bonds with the catalytic Gln-19, Cys-25, and His-159 of the cysteine protease. The O5, N10 and N6 atoms of E-64 participate in electrostatic short contacts with Gly-66 amide NH, Gly-66 amide O and Asp-158 amide O, respectively. These interactions contribute to the stabilisation of the E-64-papain complex and is also responsible for the orientation of the inhibitor into a direction suitable for the nucleophilic attack by the active cysteine. In the E-64-papain complex, the  $S_2$  subsite consists of three residues (Val-133, Val-157, and Ala-160). The leucyl side chains of E-64 are located near the entry of the  $S_2$  hydrophobic pocket and is in van der Waals contact with the side chain of Val-133 and also with Ala-160 (Kim *et al*, 1992; Matsumoto *et al*, 1999).

The described interaction of E-64 with papain and papain-like cysteine proteases was expected to cause an increased stability of the cysteine proteases used in this study, however, this was not observed using the TSA. E-64 did not generate a shift in the  $T_M$  of papain and *TcoCATL*<sub>cat</sub> regardless of the concentration tested or the pH of the buffer used in the assay. The inability of an irreversible inhibitor to affect a shift in  $T_M$  within the thermal shift assay has been observed previously (Rojas *et al*, 2015) and explained best within a review by Holdgate and Ward (2005). Within this review, the authors discussed the requirements for a  $T_M$  prediction assay to be valid. Specifically, they stated that ligand must not bind to the unfolded state of the protein, the ligand concentration must be greater than the equilibrium dissociation constant and that the complex transition between the unbound protein and the bound protein must be reversible (Holdgate and Ward, 2005). The reversible nature of the complex contributes towards the free energy of binding and the thermodynamic equilibrium within the system. As such, the use of an irreversible inhibitor prevents the accurate calculation of the free energy of binding as the thermodynamic equilibrium cannot be attained (Holdgate and Ward, 2005).

Due to the inability to detect a shift in the  $T_M$  of the relevant proteins by introducing E-64 as an inhibitor, the reversible inhibitor chymostatin was selected to serve as a positive control. Chymostatin was initially isolated from the filtrates of seven strains of

actinomycetes by Umezawa *et al* (1970) and was found to inhibit chymotrypsin and also papain to some extent. The isolated inhibitor appeared as three different forms which are chymostatin A, B and C with chymostatin A making up the majority of the mixture (Stein and Strimpler, 1987). Chymostatin is a covalent reversible protease inhibitor of mainly serine proteases but also binds weakly to cysteine proteases. The inhibitory activity of the protease is primarily due to the C-terminal phenylalanine aldehyde group which is responsible for the formation of a hemithiolacetal adduct with the nucleophilic thiol group of the cysteine protease (Umezawa, 1982; Tomkinson *et al*, 1992). There was also observed hydrogen bonding between the main chain of the enzyme and part of the main chain of the protease which was shown to stabilise the interaction (Umezawa, 1982; Tomkinson *et al*, 1992). Chymostatin has been identified as a non-toxic inhibitor and has been well tolerated in numerous intact cell systems (Place *et al*, 1987; Tomkinson *et al*, 1992). The inhibition of *TcoCATL<sub>cat</sub>* using chymostatin was demonstrated previously by Mendoza-Palomares *et al*, 2008, which provided evidence that chymostatin would be suitable for use in this study.

In this study, chymostatin was shown to effectively bind both papain and *TcoCATL<sub>cat</sub>* as evidenced by a shift in  $T_M$  was observed for both of the cysteine proteases in the thermal shift assay. The increase in  $T_M$  observed for the binding of a ligand to a protein has been observed to be proportionate to the concentration and affinity of the ligand being tested (Cimpmperman *et al*, 2008) and this was observed for both papain and *TcoCATL<sub>cat</sub>*. The increasing concentrations of chymostatin resulted in an increase in the  $T_M$  of papain and *TcoCATL<sub>cat</sub>* where it was observed that 50  $\mu$ M chymostatin resulted in the greatest shift in  $T_M$  of the cysteine proteases. Perhaps, most notably, the  $T_M$  of *TcoCATL<sub>cat</sub>* shifted significantly when in the presence of chymostatin from  $47.12 \pm 0.16$  to  $62.43 \pm 0.27$ . As it was established earlier in the study that the lower melting temperature is indicative of dimer dissociation while the higher melting temperature is indicative of monomer unfolding, it was concluded that chymostatin binds to and stabilises the monomeric form of *TcoCATL<sub>cat</sub>*. Thereafter, the ChemBridge library of compounds were tested at a concentration of 10  $\mu$ M in order to identify compounds capable of stabilising both cysteine proteases. However, none of the 50 compounds were able to stabilise papain or *TcoCATL<sub>cat</sub>* under the physiological conditions. This indicated that the compounds were incapable of binding to papain and *TcoCATL<sub>cat</sub>* or,

alternatively, the compound library may contain irreversible inhibitors which were not detected within the thermal shift assay.

#### **4.4.3. Increased protease stabilisation by pH manipulation and ligand binding**

The combination of chymostatin (at 50  $\mu\text{M}$ ) and the Reinhard buffer library was evaluated to determine the impact of combining pH manipulation and ligand binding on the stability of papain and *TcoCATL<sub>cat</sub>*. Here, it was observed that the combination of both factors resulted in the further stabilisation of both cysteine proteases and a maximum  $T_M$  was obtained for the cysteine proteases which surpassed the  $T_M$  of papain and *TcoCATL<sub>cat</sub>* when influenced by a single factor (pH or ligand). The affinity of chymostatin to bind to either cysteine proteases was, however, reduced under acidic conditions where the  $\text{pH} < 4.5$ . The melting points obtained under these low pH conditions revealed that the cysteine proteases appeared more stable in the absence of chymostatin. The ability for the thermal shift assay to detect changes in stability due to the combination of pH change and ligand binding may hold greater value when optimal activity of the protease is of importance. In this study, it was observed that the greatest increase in stability of *TcoCATL<sub>cat</sub>* was attributed to the combination of pH and ligand when sodium citrate pH 6.0 and 50  $\mu\text{M}$  chymostatin was used. Under these conditions, the  $T_M$  increased by  $3.17^\circ\text{C}$  from  $58.18 \pm 0.06^\circ\text{C}$  to  $61.35 \pm 0.09^\circ\text{C}$ , illustrating that the protease can be further stabilised by the addition of a ligand at a pH which favors optimal activity of the protease.

#### **4.5. The robustness and limitations of the thermal shift assay**

The thermal shift assay has become one of the primary methods utilised in the discovery of novel ligands which are capable of stabilising the target protein, the positive identification of stabilising ligands can thereafter be confirmed using other methods (Senisterra *et al*, 2010; Huynh and Partch, 2016). The assay has been regarded as a small-scale and high-throughput technique, however, this depends on the RT-PCR machine used, the availability of purified protein, pre-made buffer or additive screens availability as these contribute towards time delays (Senisterra *et al*, 2010; Huynh and Partch, 2016). An RT-PCR machine equipped with a 96-well plate reader is better suited for high-throughput as compared to a RT-PCR machine equipped with a rotor capable of running 36 PCR tubes in a single run.

According to Cimmerman *et al* (2008) the thermal shift assay has the ability to distinguish between weak and strong binding interactions when screening a library of compounds and this was described as one of the main advantages the assay holds over previously described methods used to determine the thermal stability of a protein. The assay was shown to determine the binding of weak ligands which was not detected by isothermal titration calorimetry. Another key advantage noted by Cimmerman *et al* (2008) was the low concentration of pure protein required for the assay. The assay was shown to provide valid results when the protein concentration used was 2  $\mu\text{M}$  for a protein with molecular mass of 35 kDa (Niesen *et al*, 2007; Cimmerman *et al*, 2008).

The requirements of the thermal shift assay may preclude the use of certain proteins. For instance, a protein with high surface hydrophobicity is not suitable for use within the thermal shift assay. This is due to the binding of SYPRO orange to these hydrophobic regions before the protein has unfolded, resulting in an incorrect melting temperature and an initial high fluorescence. This initial high fluorescence was also described as an event linked to protein aggregation. Therefore, the thermal shift assay would not be able to generate valid melt curves when dealing with a protein which has hydrophobic clusters present on the surface of the protein and when a protein has undergone aggregation. This phenomenon of protein aggregation, although not observed in this study, may be overcome by using the thermal shift assay to determine optimal conditions which prevent the protein from aggregating. This would be evident from the generation of a valid melt curve under optimal conditions.

A limiting factor of the thermal shift assay is that the melting point of a protein cannot exceed 95-100°C as this would cause the water present in samples to evaporate in an open plate-based method (Cimmerman *et al*, 2008). This temperature range is also a limitation due to the heating capability of the RT-PCR machine used. Another limiting factor explained by Kroeger *et al* (2017) was the result of quenched fluorescence of SYPRO orange due to the presence of EDTA. The quenched fluorescence was only observed when EDTA was used as a buffer component at  $\text{pH} > 9$ . Ethylenediaminetetraacetic acid is a chelating agent of metal ions and a common agent used in buffers which was used as a buffer component. It was observed that in the absence of a protein sample there are interactions between EDTA and SYPRO orange at  $\text{pH} > 9$  which resulted in the formation of supramolecular EDTA aggregates (Kroeger *et al*, 2017). These aggregates formed in a temperature-dependent manner and provided

a fluorescence signal similar to that obtained when a protein unfolds. The contributions of fluorescence resulted in incorrect melt plots for a protein. Therefore, EDTA cannot be used as a buffer component if a pH > 9 is required and SYPRO orange is the dye of choice. In this study, a pH range between 4 and 8 was used and therefore EDTA did not affect the fluorescence obtained.

The thermal shift assay is one of many techniques capable of measuring the stability of a protein, however, it is quickly becoming recognised as one of the simplest and least expensive methods (Niesen *et al*, 2007; Senisterra *et al*, 2010; Huynh and Partch, 2016). The ability of the assay to rapidly test many samples using a minimal protein concentration without the use of expensive equipment and the simple processing of data has led to the thermal shift assay becoming one of the leading assays in drug discovery studies focused on the detection of novel ligands (Andreotti *et al*, 2015). The assay has been described as robust and reproducible which is further supported by the low standard deviations calculated for each experiment run within this study.

#### **4.6. Conclusion**

The thermal shift assay was effective in determining the melting temperature for papain under physiological conditions of  $79.22 \pm 0.06^\circ\text{C}$ . The dissociation of the dimer for the catalytic domain of congoxin under physiological conditions was detected at a melting temperature of  $47.12 \pm 0.16^\circ\text{C}$  while the unfolding of the monomer was determined at temperatures exceeding  $60^\circ\text{C}$ . The thermal shift assay was able to identify pH and ligand binding as factors which contribute to protein stability and optimum conditions were determined for papain and *TcoCATL*. The most stable form of papain was observed at pH 5.0 with 50  $\mu\text{M}$  chymostatin which was more stable than papain under physiological conditions as evidenced by an increase in  $T_M$  of  $5.23^\circ\text{C}$ . Similarly, the most stable form of *TcoCATL*<sub>cat</sub> was also determined to be pH 5.0 with 50  $\mu\text{M}$  chymostatin while it was additionally shown that chymostatin binds to the monomeric and not the dimer form of the protein. Taken together, these findings hold value for further biochemical investigation of the two cysteine proteases and may provide useful guidelines to facilitate efforts to crystallise *TcoCATL*<sub>cat</sub>. This study also established the thermal shift assay to be a quick and efficient tool in detecting changes in the melting temperature of a protein when more than one factor is introduced.



#### **4.7. Future Studies**

The thermal stability of papain and the catalytic domain of congopain were both increased due to pH optimisation and the introduction of a known cysteine protease inhibitor. The 50 compounds obtained from ChemBridge library of compounds were unsuccessful in stabilising the cysteine proteases and thus we were unsuccessful in identifying new compounds which bound to papain or the catalytic domain of congopain. To rule out the possibility of the presence of irreversible inhibitors, the activity of the proteins in the presence of each compound should be evaluated by zymogram analysis. The thermal stability of both proteases should then be further evaluated by testing different sets of buffer libraries and testing larger compound libraries. A larger library of known cysteine protease inhibitors should also be screened to determine if any can further stabilise the cysteine proteases to a greater degree than was observed for chymostatin. Further evaluation of the thermal stability of the catalytic domain of congopain may provide favourable conditions. The optimised conditions may then be used in attempts to crystallise the highly relevant cysteine protease which is a potential drug target and diagnostic candidate.

## 5. REFERENCES

---

- Amri, E., Mamboya, F. (2012). Papain, a plant enzyme of biological importance: a review. *American Journal of Biochemistry and Biotechnology*, 8, 99-104.
- Andreotti, G., Monticelli, M., Cubellis, M. V. (2015). Looking for protein stabilizing drugs with thermal shift assay. *Drug testing and analysis*, 7, 831-834.
- Authié, E., Muteti, D. K., Mbawa, Z. R., Lonsdale-Eccles, J. D., Webster, P., Wells, C. W. (1992). Identification of a 33-kilodalton immunodominant antigen of *Trypanosoma congolense* as a cysteine protease. *Molecular and Biochemical Parasitology*, 56, 103-116.
- Bai, N., Roder, H., Dickson, A., Karanicolas, J. (2019). Isothermal analysis of thermofluor data can readily provide quantitative binding affinities. *Scientific Reports*, 9, 1-15.
- Balls, A. K., Lineweaver, H. (1939). Isolation and properties of crystalline papain. *Journal of Biological Chemistry*, 130, 669-686.
- Bloom, J. D., Labthavikul, S. T., Otey, C. R., Arnold, F. H. (2006). Protein stability promotes evolvability. *PNAS*, 103, 5869-5874.
- Boulangé, A., Serveau, C., Brillard, M., Minet, C., Gauthier, F., Diallo, A., Lalmanach, G., Authié, E. (2001). Functional expression of the catalytic domains of two cysteine proteinases from *Trypanosoma congolense*. *International Journal for Parasitology*, 31, 1435-1440.
- Boulangé, A. F., Khamadi, S. A., Pillay, D., Coetzer, T. H., Authié, E. (2011). Production of congopain, the major cysteine protease of *Trypanosoma (Nannomonas) congolense*, in *Pichia pastoris* reveals unexpected dimerisation at physiological pH. *Protein Expression and Purification*, 75, 95-103.
- Broom, A., Jacobi, Z., Trainor, K., Meiering, E. M. (2017). Computational tools help improve protein stability bit with a solubility tradeoff. *Journal of Biological Chemistry*, 292, 14349-14361.
- Bruce, D., Cardew, E., Freitag-Pohl, S., Pohl, E. (2019). How to stabilize protein: stability screens for thermal shift assays and nano differential scanning fluorimetry in the virus-X project. *Journal of Visualized Experiments*, 1, 1-11.
- Bruylants, G., Wouters, J., Michaux, C. (2005). Differential scanning calorimetry in life science: Thermodynamics, stability, molecular recognition and application in drug design. *Current Medicinal Chemistry*, 12, 2011-2020.

- Bunaciu, A. A., Aboul-Enein, H. Y., Hoang, V. D. (2015). Raman spectroscopy for protein analysis. *Applied spectroscopy reviews*, 50, 377-386.
- Celej, M. S., Montich, G. G., Fidelio, G. D. (2003). Protein stability induced by ligand binding correlates with changes in protein flexibility. *Protein Science*, 12, 1496-1506.
- Chagas, J. R., Authié, E., Serveau, C., Lalmanach, G., Juliano, L., Gauthier, F. (1997). A comparison of the enzymatic properties of the major cysteine proteinases from *Trypanosoma congolense* and *Trypanosoma cruzi*. *Molecular and Biochemical Parasitology*, 88, 85-94.
- Chen, T., Oakley, D. M. (1995). Thermal analysis of proteins of pharmaceutical interest. *Thermochimica Acta*, 248, 229-244.
- Chen, Y., Ding, F., Nie, H., Serohijos, A. W., Sharma, S., Wilcox, K. C., Yin, S., Dokholyan, N. V. (2008). Protein folding: then and now. *Archives of Biochemistry and Biophysics*, 469, 4-19.
- Cimpmperman, P., Baranauskiene, L., Jachimoviciute, S., Jachno, J., Torresan, J., Michailoviene, V., Matuilene, J., Sereikaite, J., Bumelis, V., Matulis, D. (2008). A quantitative model of thermal stabilization and destabilization of proteins by ligands. *Biophysical Journal*, 95, 3222-3231.
- Correa, D. H. A., Ramos, C. H. (2009). The use of circular dichroism spectroscopy to study protein folding, form and function. *African Journal of Biochemistry Research*, 3, 164-173.
- De La Cuesta, R. G., Goodacre, R., Ashton, L. (2014). Monitoring antibody aggregation in early drug development using Raman spectroscopy and perturbation-correlation moving windows. *Analytical Biochemistry*, 86, 11133-11140.
- Dennison, C., Lovrien, R. (1997). Three phase partitioning: concentration and purification of proteins. *Protein Expression and Purification*, 11, 149-161.
- Dill, K. A. (1990). Dominant forces in protein folding. *Biochemistry*, 29.
- Dominy, B. N., Perl, D., Schmid, F. X., Brooks, C. L. (2002). The effects of ionic strength on protein stability: the cold shock protein family. *Journal of Molecular Biology*, 319, 541-554.

- Dzwolak, W., Kato, M., Taniguchi, Y. (2002). Fourier transform infrared spectroscopy in high-pressure studies on proteins. *Biochimica et Biophysica Acta*, 1595, 131-144.
- Ellis, K. J., Morrison, T. F. (1982). Buffers of constant ionic strength for studying pH-dependent processes. *Methods in Enzymology*, 87, 405-426.
- Ericsson, U. B., Hallberg, B. M., Detitta, G. T., Dekker, N., Nordlund, P. (2006). Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Analytical Biochemistry*, 357, 289-298.
- Farady, C. J., Craik, C. S. (2010). Mechanisms of macromolecular protease inhibitors. *ChemBioChem*, 11, 2341-2346.
- Fernandez-Ballester, G., Castresana, J., Arrondo, J. R., Ferragut, J. A., Gonzalez-Ros, J. M. (1992). Protein stability and interaction of the nicotine acetylcholine receptor with cholinergic ligands studied by Fourier-transform infrared spectroscopy. *Biochemical Journal*, 288, 421-426.
- Flindt, M. L. H. (1979). Allergy to  $\alpha$ -amylase and papain. *The Lancet*, 313, 1407-1408.
- Fosado-Quiroz, R. E., Rojo-Dominguez, A. (2011). Metastability of papain and the molecular mechanism for its sequential acid-denaturation. *Protein Journal*, 30, 184-193.
- Fu, H., Grimsley, R. G., Scholtz, J. M., Pace, C. N. (2010). Increasing protein stability: importance of  $\Delta C_p$  and the denatured state. *Protein Science*, 19, 1044-1052.
- Garidel, P., Schott, H. (2006). Fourier-transform midinfrared spectroscopy for analysis and screening of liquid protein formulations. *Bioprocess Technical*, 48-55.
- Gill, P., Moghadam, T. T., Ranjbar, B. (2010). Differential scanning calorimetry techniques: Applications in biology and nanoscience. *Journal of Biomolecular Techniques*, 21, 167-193.
- Giordani, F., Morrison, L. J., Rowan, T. G., De Koning, H. P., Barrett, M. P. (2016). The animal trypanosomiases and their chemotherapy: a review. *Parasitology*, 143, 1862-1889.
- Gitonga, P. K., Ndungu, K., Murilla, G. A., Thande, P. C., Wamwiri, F. N., Auma, J. E., Ngae, G. N., Kibugu, J. K., Kurgat, R., Thuita, J. K. (2017). Differential virulence and tsetse fly transmissibility of *Trypanosoma congolense* and *Trypanosoma brucei* strains. *Onderstepoort Journal of Veterinary Research*, 84, 1-10.

- Greenfield, N. J. (2006). Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions. *Nature Protocols*, 1, 2527-2535.
- Gromiha, M. M. (2007). Prediction of protein stability upon point mutations. *Biochemical Society Transactions*, 35, 1569-1573.
- Gromiha, M. M., An, J., Kono, H., Oobatake, M., Uedaira, H., Sarai, A. (1999). ProTherm: thermodynamic database for proteins and mutants. *Nucleic Acids Research*, 27, 286-288.
- Grzonka, Z., Jankowsk, E., Kasprzykowski, F., Kasprzykowska, R., Lankiewicz, L., Wicz, W., Wiczczak, E., Ciarkowski, J., Drabik, P., Janowski, R., Kozak, M., Jaskólski, M., Grubb, A. (2001). Structural studies of cysteine proteases and their inhibitors. *Acta Biochimica Polonica*, 48, 1-20.
- Hanada, K., Tamai, M., Morimoto, S., Adachi, T., Ohmura, S., Sawada, J., Tanaka, I. (1978). Inhibitory activities of E-64 derivatives on papain. *Agricultural and Biological Chemistry*, 42, 537-541.
- Heussen, C., Dowdle, E. B. (1980). Electrophoretic analysis of plasminogen activators in polyacrylamide gels containing sodium dodecyl sulfate and copolymerized substrates. *Analytical Biochemistry*, 102, 196-202.
- Hinz, H. J., Steif, C., Vogl, T., Meyer, R., Renner, M., Ledermuller, R. (1993). Fundamentals of protein stability. *Pure and Applied Chemistry*, 55, 947-952.
- Hoffmann, B., Eichmuller, C., Steinhauser, O., Konrat, R. (2005). Rapid Assessment of protein structural stability and fold validation via NMR. *Proteomics*, 394, 142-175.
- Holdgate, G. A., Ward, W. H. J. (2005). Measurements of binding thermodynamics in drug discovery. *Drug Discovery Today*, 10, 1543-1550.
- Hooper, N. M. (2002). Proteases: a primer. *Essays in Biochemistry*, 38, 1-8.
- Huet, J., Looze, Y., Bartik, K., Raussens, V., Wintjens, R., Boussard, P. (2006). Structural characterization of the papaya cysteine proteinases at low pH. *Biochemical and Biophysical Research Communications*, 341, 620-626.
- Hummel, K. M., Penheiter, A. R., Gathman, A. C., Lilly, W. W. (1996). Anomalous estimation of protease molecular weights using gelatin-containing SDS-PAGE. *Analytical Biochemistry*, 233, 140-142.

- Huson, L. E. J., Authié, E., Boulangé, A. F., Goldring, J. P. D., Coetzer, T. H. (2009). Modulation of the immunogenicity of the *Trypanosoma congolense* cysteine protease, congopain, through complexation with  $\alpha$ 2-macroglobulin. *Veterinary Research*, 40, 1-12.
- Huynh, K., Partch, C. L. (2016). Current Protocols in protein science :Analysis of protein stability and ligand interactions by thermal shift assay. *Current Protocols in Protein Science*, 79, 1-19.
- Kamphuis, I. G., Kalk, K. H., Swarte, M. B., Drenth, J. (1984). Structure of papain refined at 1.65 Å resolution. *Journal of Molecular Biology*, 179, 233-256.
- Kateregga, J., Lubega, G. W., Lindblad, E. B., Authié, E., Coetzer, T. H., Boulange, A. F. (2012). Effect of adjuvants on the humoral immune response to congopain in mice and cattle. *BMC Veterinary Research*, 8, 1746-6148.
- Keeler, J. (2002). Understanding NMR spectroscopy. *Biochemistry*, 159-210.
- Kelly, S. M., Price, N. C. (2000). The use of circular dichroism in the investigation of protein structure and function. *Current Protein and Peptide Science*, 1, 349-384.
- Kim, M., Yamamoto, D., Matsumoto, K., Inoue, M., Ishida, T., Mizuno, H., Sumiya, S., Kitamura, K. (1992). Crystal structure of papain-E64-c complex. *Biochemical Journal*, 287, 797-803.
- Kimmel, J. R., Smith, E. R. (1954). Crystalline papain: I. preparation, specificity, and activation. *Journal of Biological Chemistry*, 207, 515-531.
- Klein, I. B., Kirsch, J. F. (1969). The activation of papain and the inhibition of the active enzyme by carbonyl reagents. *Journal of Biological Chemistry*, 244, 5928-5935.
- Kodre, K. V., Attarde, S. R., Yendhe, P. R., Patil, R. Y., Barge, V. U. (2014). Differential scanning calorimetry: A review. *Journal of Pharmaceutical Analysis*, 3, 11-22.
- Kong, J., Yu, S. (2007). Fourier transform infrared spectroscopic analysis of protein secondary structures. *Acta Biochimica et Biophysica Sinica*, 39, 549-559.

- Kroeger, T., Frieg, B., Zhang, T., Hansen, F. K., Marmann, A., Proksch, P., Nagel-Steger, L., Groth, G., Smits, S. H. J., Gohlke, H. (2017). EDTA aggregates induce SYPRO orange based fluorescence in thermal shift assay. *PLOS ONE*, 12, 1-21.
- Ku, T., Lu, P., Chan, C., Wang, T., Lai, S., Lyu, P., Hsiao, N. (2009). Predicting melting temperature directly from protein sequences. *Computational Biology and Chemistry*, 33, 445-450.
- Kwan, A. H., Mobli, M., Gooley, P. R., King, F. G., Mackay, J. P. (2011). Macromolecular NMR spectroscopy for the non- spectroscopist. *The FEBS Journal*, 278, 687-703.
- Kyte, J., Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology*, 157, 105-132.
- Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, 680-685.
- Lalmanach, G., Boulangé, A., Serveau, C., Lecaille, F., Scharfstein, J., Gauthier, F., Authié, E. (2002). Congopain from *Trypanosoma congolense*: Drug target and vaccine candidate. *Biological Chemistry*, 383, 739-749.
- Lalmanach, G., Lecaille, F., Chagas, J. R., Authié, E., Scharfstein, J., Juliano, M. A., Gauthier, F. (1998). Inhibition of trypanosomal cysteine proteinases by their propeptides. *Journal of Biological Chemistry*, 273, 25112-25116.
- Lambri, M., Roda, A., Dordoni, R., Fumi, M. D., De Faveri, D. M. (2014). Mild process for dehydrated food-grade crude papain powder from papaya fresh pulp: lab-scale and pilot plant experiments. *Chemical Engineering Transactions*, 38, 7-12.
- Laskar, A., Chatterjee, A. (2009). Protease-Revisiting the types and potential. *Online Journal of Biotechnology research*, 1, 55-61.
- Lavinder, J. J., Hari, S. B., Sullivan, B. J., Magliery, T. J. (2009). High-throughput thermal scanning: a general, rapid dye-binding thermal shift screen for protein engineering. *Journal of the American Chemical Society*, 131, 3794-3795.
- Lecaille, F., Authié, E., Moreau, T., Serveau, C., Gauthier, F., Lalmanach, G. (2001). Subsite specificity of trypanosomal cathepsin L-like cysteine proteases. *European Journal of Biochemistry*, 268, 2733-2741.

- Lecaille, F., Kaleta, J., Bromme, D. (2002). Human and parasitic papain-like cysteine proteases: their role in physiology and pathology and recent developments in inhibitor design. *Chemical Reviews*, 102, 4459-4488.
- Li, C. H., Nguyen, X., Narhi, L., Chemmalil, L., Towers, E., Muzammil, S., Gabrielson, J., Jiang, Y. (2011). Applications of circular dichroism (CD) for structural analysis of proteins: qualification of near-and far UV CD for protein higher order structural analysis. *Journal of Pharmaceutical Sciences*, 100, 4642-4654.
- Li, Y., Kang, C. (2017). Solution NMR spectroscopy in target-based drug discovery. *Molecules*, 22, 1-21.
- Liu, H., Hu, M., Wang, Q., Cheng, L., Zhang, Z. (2018). Role of papain-like cysteine proteases in plant development. *Frontiers in Plant Science*, 9, 1-10.
- Magliery, T. J. (2015). Protein stability: computation, sequence statistics, and new experimental methods. *Current Opinion in Structural Biology*, 33, 161-168.
- Malek, K., Norazan, M., Ramaness, P., Othman, N. Z., Malek, R., Aziz, R., Aladdin, A., Enshasy, H. E. (2016). Cysteine proteases from *Carica papaya*: An important enzyme group of many industrial applications. *Journal of Pharmacy and Biological Sciences*, 11, 11-16.
- Matsumoto, K., Mizoue, K., Kitamurei, K., Tse, W., Huber, C. P., Ishida, T. (1999). Structural basis of inhibition of cysteine proteases by E-64 and its derivatives. *Biopolymers (Peptide Science)*, 51.
- Miyazaki, Y., Doi, N., Koma, T., Adachi, A., Nomaguchi, M. (2017). Novel in vitro screening system based on differential scanning fluorimetry to search for small molecules against the disassembly or assembly of HIV-1 capsid protein. *Frontiers in Microbiology*, 8, 1-5.
- Monti, R., Basilio, C. A., Trevisan, H. C., Contiero, J. (2000). Purification of papain from fresh latex of *Carica papaya*. *Brazilian Archives of Biology and Technology*, 43, 501-507.
- Moosavi-Movahedi, A. A., Saboury, A. A., Hosseinkham, S., Lohrasbi-Nejad, A., Habibi-Rezaei, M., Maghami, P., Atri, M., Fotouhi, L. (2016). Protein stability, folding, disaggregation and etiology of conformational malfunctions. *Biomacromolecular Journal*, 2, 8-20.
- Mótyán, J. A., Tóth, F., Tözsér, J. (2013). Research applications of proteolytic enzymes in molecular biology. *Biomolecules*, 3, 923-942.



- New, S. Y., Aung, K. M., Lim, G. L., Hong, S., Tan, S. K., Lu, Y., Cheung, E., Su, X. (2014). Fast screening of ligand-protein Interactions based on ligand-induced protein stabilization of gold nanoparticles. *Analytical Chemistry*, 86, 2361-2370.
- Niesen, F. H., Berglund, H., Vedadi, M. (2007). The use of differential fluorimetry to detect ligand interactions that promote protein stability. *Nature Protocols*, 2, 2212-2221.
- Otto, H., Schirmeister, T. C. (1997). Cysteine proteases and their inhibitors. *Chemical Reviews*, 97, 133-171.
- Pace, C. N. (1990). Conformational stability of globular proteins. *Trends in Biochemical Sciences*, 15, 14-17.
- Pace, C. N. (1992). Contribution of the hydrophobic effect to globular protein stability. *Journal of Molecular Biology*, 226, 29-35.
- Pace, C. N., Grimsley, R. G., Scholtz, J. M. (2009). Protein ionizable groups: pK values and their cotribution to protein stability and solubility. *The Journal of Biological Chemistry*, 284, 13285-13289.
- Pan, D., Hill, A. P., Kashou, A., Wilson, K. A., Tan-Wilson, A. (2011). Electrophoretic transfer protein zymography. *Analytical Biochemistry*, 411, 277-283.
- Pantoliano, M. W., Petrella, E. C., Kwasnoski, J. D., Lobanov, V. S., Myslik, J., Graf, E., Carver, T., Asel, E., Springer, B. A., Lane, P., Salemme, F. R. (2001). High-density miniaturized thermal shift assays as a general strategy for drug discovery. *Journal of Biomolecular Screening*, 6, 429-440.
- Paolini, S., Tanfani, F., Fini, C., Bertoli, E., Pelosi, P. (1999). Porcine odorant-binding protein: structural stability and ligand affinities measured by Fourier-transform infrared spectroscopy and fluorescence spectroscopy. *Biochimica et Biophysica Acta*, 1431, 179-188.
- Patel, H., Bhoi, M. N., Borad, M. A., Dalvadi, A. D., Dalsania, K. V. (2012). Extraction and application of papain enzyme on degradation of drug. *International Journal of Pharmacy and Biological Science*, 2, 113-115.
- Pillay, D. 2010. Identification and characterisation of novel pathogenic factors of *Trypanosoma congolense*. Pietermaritzburg: University of KwaZulu-Natal.

- Place, G. A., Chetland, J., Galpin, I. J., Beynon, R. J. (1987). The effect of analogues of chymostatin on lysosomal and non-lysosomal components of protein degradation in isolated hepatocytes. *Biochimica et Biophysica Acta*, 925, 185-193.
- Qiagen 2010. Rapid, high-throughput assessment of protein stability on the Rotor-Gene Q cyclor. *QIAGEN - Literature: News 2010 e7*. QIAGEN.
- Ragoonanan, V., Aksan, A. (2012). Protein Stabilization. *Transfusion Medicine and Hemotherapy*, 34, 246-252.
- Rantanji, K. D., Derrick, J. P., Dearman, R. J., Kimber, I. (2014). Immunogenicity of therapeutic proteins: influence of aggregation. *Journal of Immunotoxicology*, 11, 99-109.
- Rawlings, N. D., Barrett, A. J. (1993). Evolutionary families of peptidases *Biochemistry Journal*, 290, 205-218.
- Rawlings, N. D., Barrett, A. J., Bateman, A. (2010). MEROPS: the peptidase database. *Nucleic Acids Research*, 38, D227-D233.
- Rawlings, N. D., Barrett, A. J., Thomas, P. D., Huang, X., Bateman, A., Finn, R. D. (2018). The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Research*, 46, D624-D632.
- Reinhard, L., Mayerhofer, H., Geerlof, A., Mueller-Dieckmann, J., Weiss, M. S. (2012). Optimization of protein buffer cocktails using thermofluor. *Acta Crystallographica F69*, 209-214.
- Rojas, R. J., Edmondson, D. E., Almos, T., Scott, R., Massari, M. E. (2015). Reversible and irreversible small molecule inhibitors of monoamine oxidase B (MAO-B) investigated by biophysical techniques. *Bioorganic and Medicinal Chemistry*, 23, 770-778.
- Sahin, A., Ascencio, C., Izotte, J., Pillay, D., Coustou, V., Karembe, H., Baltz, T. (2014). The Susceptibility of *Trypanosoma congolense* and *Trypanosoma brucei* to isometamidium chloride and its synthetic impurities. *Veterinary Parasitology*, 203, 270-275.
- Schechter, I., Berger, A. (1967). On the size of the active site in proteases I papain. *Biochemical and biophysical research communications*, 27, 1-6.

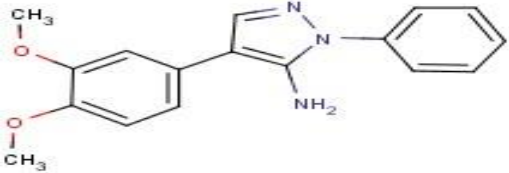
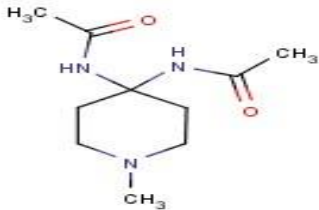
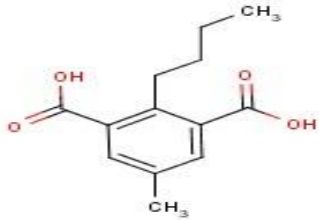
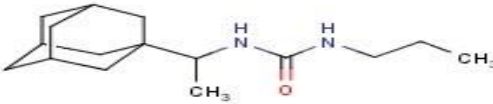
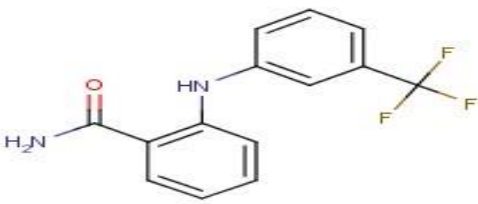
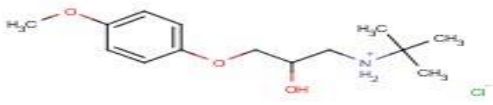
- Scheiblhofer, S., Laimer, J., Machado, Y., Weiss, R., Thalhamer, J. (2017). Influence of protein fold stability on immunogenicity and its implications for vaccine design. *Expert Review of Vaccines*, 16, 479-489.
- Scott, D. J., Kummer, L., Tremmel, D., Pluckthun, A. (2013). Stabilizing membrane proteins through protein engineering. *Current Opinion in Chemical Biology*, 17, 427-435.
- Senisterra, G. A., Ghanei, H., Khutoreskaya, G., Dobrovetsky, E., Edwards, A. M., Privé, G. G., Vedadi, M. (2010). Assessing the stability of membrane proteins to detect ligand binding using differential static light scattering. *Journal of Biomolecular Screening*, 314-320.
- Shulman, B. A., Kim, P. S., Dobson, C. M., Redfield, C. (1997). A residue-specific NMR view of the non-cooperative unfolding of a molten globule. *Nature structural biology*, 4, 630-634.
- Song, J., Tan, H., Perry, A. J., Akutsu, T., Webb, G. I., Whisstock, J. C., Pike, R. N. (2012). PROSPER: an integrated feature-based tool for predicting protease substrate cleavage sites. *PLOS One*, 7, 1-23.
- Sorrentino, S., Yakolev, G. I., Libonati, M. (1981). Dimerization of deoxyribonuclease I, lysozyme and papain. *European Journal of Biochemistry*, 124, 183-189.
- Stein, R. L., Strimpler, A. M. (1987). Slow-binding inhibition of chymotrypsin and cathepsin G by the peptide aldehyde chymostatin. *Biochemistry*, 26, 2611-2615.
- Talley, K., Alexov, E. (2010). On the pH-optimum of activity and stability of proteins. *Proteins*, 78, 2699-2706.
- Thomas, A. S., Elcock, A. H. (2007). Molecular dynamics simulations of hydrophobic associations in aqueous salt solutions indicate a connection between water hydrogen bonding and the Hofmeister effect. *Journal of the American Chemical Society*, 129, 14887-14898.
- Tomkinson, N. P., Galpin, I. J., Beynon, R. J. (1992). Synthetic analogues of chymostatin. *Biochemical Journal*, 286, 475-480.
- Tuma, R. (2005). Raman spectroscopy of proteins: from peptides to large assemblies. *Journal of Raman Spectroscopy*, 36, 307-319.
- Turk, D., Podobnik, M., Turk, B. (1998). Revised definition of substrate binding sites of papain-like cysteine proteases. *Biological Chemistry*, 379, 137-147.

- Turk, V., Stoka, V., Vasiljeva, O., Renko, M., Sun, T., Turk, B., Turk, D. (2012). Cysteine cathepsins: From structure, function and regulation to new frontiers. *Biochimica et Biophysica Acta*, 1824, 68-88.
- Umezawa, H. (1982). Low-molecular weight enzyme inhibitors of microbial origin. *Annual Review of Microbiology*, 36, 75-99.
- Umezawa, H., Aoyagi, T., Morishima, H., Kunimoto, S., Matsuzaki, M., Hamada, M., Takeuchi, T. (1970). Chymostatin, a new chymotrypsin inhibitor produced by Actinomycetes. *The Journal of Antibiotics*, 23, 425-427.
- Uson, I., Sheldrick, G. M. (1999). Advances in direct methods for protein crystallography. *Current Opinion in Structural Biology*, 9, 643-648.
- Van Den Burg, B., Eijssink, V. G. (2002). Selection of mutations for increased protein stability. *Current Opinion in Biotechnology*, 13, 333-337.
- Van Mierlo, C. P. M., Steensma, E. (2000). Protein folding and stability investigated by fluorescence, circular dichroism (CD), and nuclear magnetic resonance (NMR) spectroscopy: the flavodoxin story. *Journal of Biotechnology*, 79, 281-298.
- Varughese, K. I., Ahmed, F. R., Carey, P. R., Hasnain, S., Huber, C. P., Storer, A. C. (1989). Crystal structure of a papain-E-64 complex. *Biochemistry*, 28, 1330-1332.
- Vedadi, M., Niesen, F. H., Allali-Hassani, A., Fedorov, O. Y., Finerty, P. J., Wasney, G. A., Yeung, R., Arrowsmith, C., Ball, L. J., Berglund, H., Hui, R., Marsden, B. D., Nordlund, P., Sundstrom, M., Weigelt, J., Edwards, A. M. (2006). Chemical screening methods to identify ligands that promote protein stability, protein crystallization, and structure determination. *PNAS*, 103, 15835-15840.
- Verma, S., Dixit, R., Pandey, K. C. (2016). Cysteine proteases: Modes of activation and future prospects as pharmacological targets *Frontiers in Pharmacology*, 7, 1-12.
- Vivoli, M., Novak, H. R., Littlechild, J. A., Harmer, N. J. (2014). Determination of protein-ligand interactions using differential scanning fluorimetry. *Journal of Visualized Experiments*, 91, 1-13.
- Waldron, T. T., Murphy, K. P. (2003). Stabilization of proteins by ligand binding: application to drug screening and determination of unfolding energetics. *Biochemistry*, 42, 5058-5064.

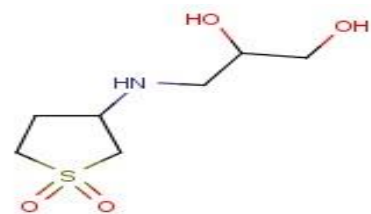
- Wen, Z. (2007). Raman spectroscopy of protein pharmaceuticals. *Journal of Pharmaceutical Sciences*, 96, 2861-2878.
- Wlodawer, A., Minor, W., Dauter, Z., Jaskolski, M. (2007). Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures. *The FEBS Journal*, 275, 1-21.
- Yang, A., Honig, B. (1994). Structural origins of pH and ionic strength effects on protein stability. *Journal of Molecular Biology*, 237, 602-614.
- Yang, A. S., Honig, B. (1993). On the pH dependence of protein stability. *Journal of Molecular Biology*, 231, 459-474.
- Yeh, A. P., Mcmillan, A., Stowell, M. H. B. (2006). Rapid and simple protein-stability screens: Application to membrane proteins. *Biological Crystallography*, D62, 451-457.

## APPENDIX A

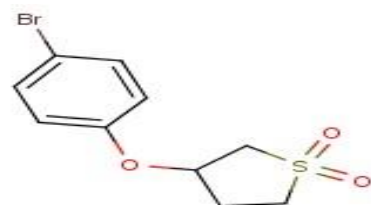
**Table A.1: The chemical structures of the 50 compounds obtained from the ChemBridge library of compounds**

Compound ID	Compound name	Compound structure
5118317	4-(3,4-dimethoxyphenyl)-1-phenyl-1H-pyrazol-5-amine	
5118841	N, N'-(1-methyl-4,4-piperidinediyl) diacetamide	
5119061	2-butyl-5-methylisophthalic acid	
5142981	N-[1-(1-adamantyl) ethyl]-N'-propylurea	
5144439	2-{[3-(trifluoromethyl)phenyl] amino} benzamide	
5144520	N-(tert-butyl)-2-hydroxy-3-(4-methoxyphenoxy)-1-propanaminium chloride	

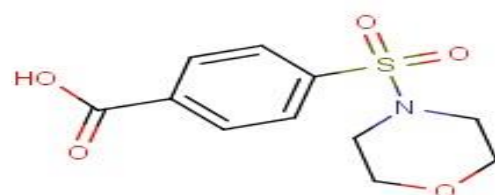
5155819 3-[(1,1-dioxidotetrahydro-3-thienyl) amino]-1,2-propanediol



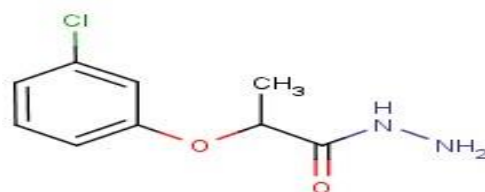
5155858 3-(4-bromophenoxy) tetrahydrothiophene 1,1-dioxide



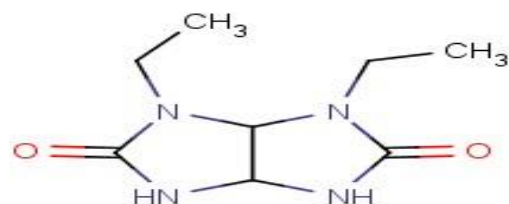
5156995 4-(4-morpholinylsulfonyl) benzoic acid



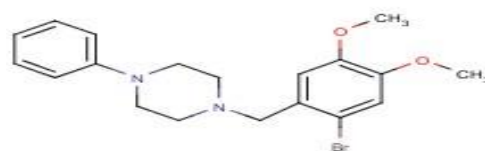
5189169 2-(3-chlorophenoxy) propanohydrazide



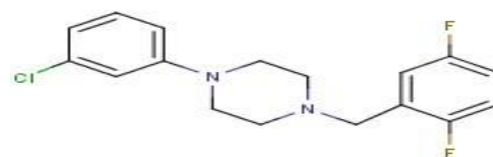
5233951 1,6-diethyltetrahydroimidazo[4,5-d] imidazole-2,5(1H,3H)-dione



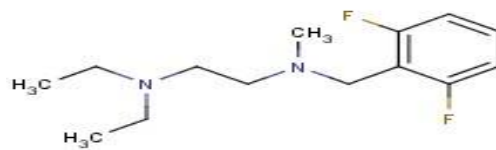
5411732 1-(2-bromo-4,5-dimethoxybenzyl)-4-phenylpiperazine



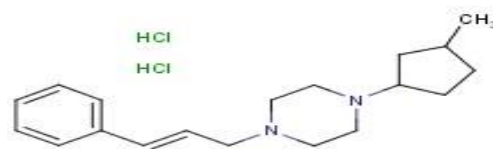
5414621 1-(3-chlorophenyl)-4-(2,5-difluorobenzyl) piperazine



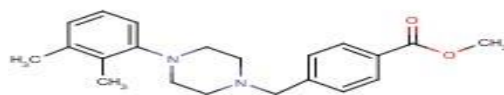
5415800 N-(2,6-difluorobenzyl)-N',  
N'-diethyl-N-methyl-1,2-  
ethanediamine



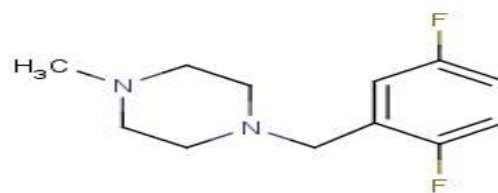
5417271 1-(3-methylcyclopentyl)-4-  
(3-phenylprop-2-en-1-yl)  
piperazine dihydrochloride



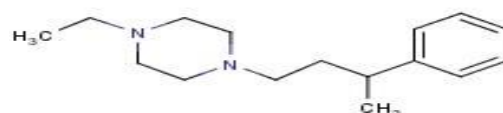
5418608 methyl 4-{[4-(2,3-  
dimethylphenyl)-1-  
piperazinyl] methyl}  
benzoate



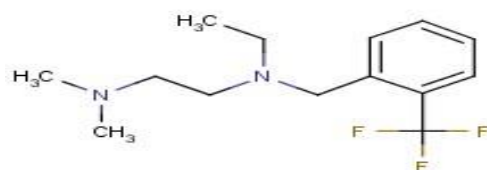
5419034 1-(2,5-difluorobenzyl)-4-  
methylpiperazine



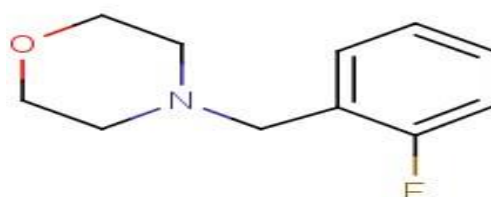
5430819 1-ethyl-4-(3-phenylbutyl)  
piperazine



5430906 N-ethyl-N', N'-dimethyl-N-  
[2-(trifluoromethyl) benzyl]-  
1,2-ethanediamine

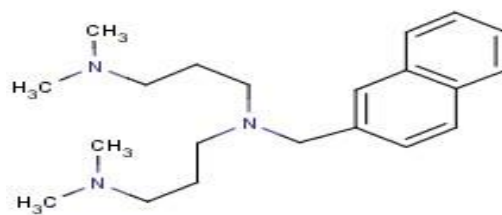


5431331 4-(2-fluorobenzyl)  
morpholine

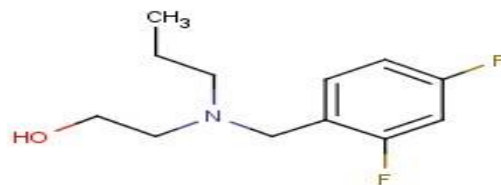




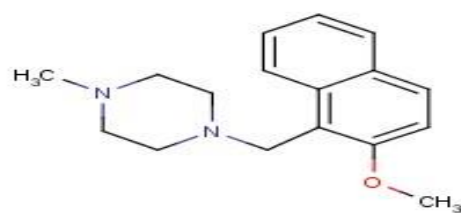
5431334 N-[3-(dimethylamino)propyl]-N', N'-dimethyl-N-(2-naphthylmethyl)-1,3-propanediamine



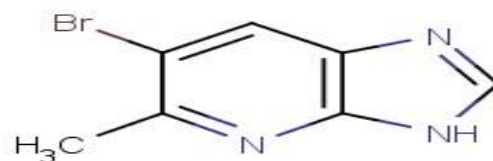
5431798 2-[(2,4-difluorobenzyl)(propyl)amino] ethanol



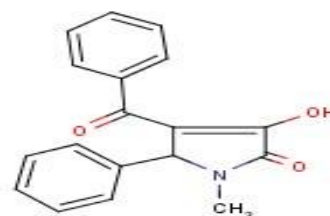
5431935 1-[(2-methoxy-1-naphthyl)methyl]-4-methylpiperazine



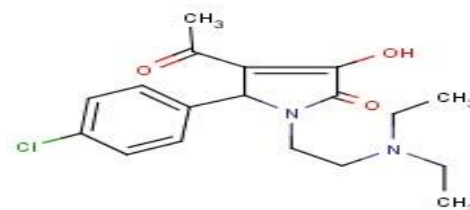
5722012 6-bromo-5-methyl-3H-imidazo[4,5-b]pyridine

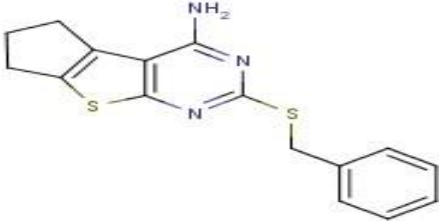
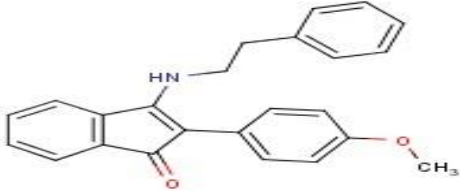
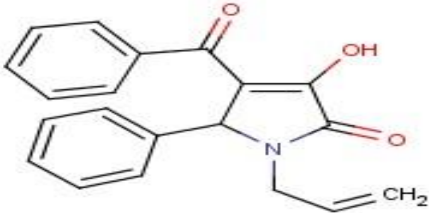
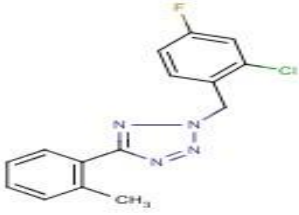
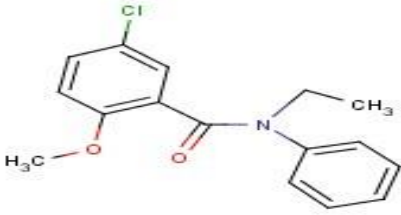
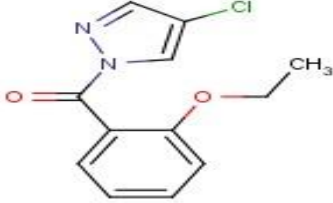
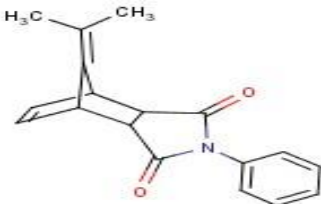


5787093 4-benzoyl-3-hydroxy-1-methyl-5-phenyl-1,5-dihydro-2H-pyrrol-2-one

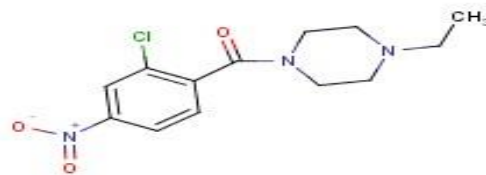


5791634 4-acetyl-5-(4-chlorophenyl)-1-[2-(diethylamino) ethyl]-3-hydroxy-1,5-dihydro-2H-pyrrol-2-one

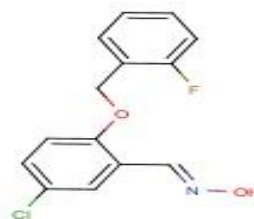


6136720	2-(benzylthio)-6,7-dihydro-5H-cyclopenta [4,5] thieno[2,3-d] pyrimidin-4-amine	
6194607	2-(4-methoxyphenyl)-3-[(2-phenylethyl) amino]-1H-inden-1-one	
6197483	1-allyl-4-benzoyl-3-hydroxy-5-phenyl-1,5-dihydro-2H-pyrrol-2-one	
6271753	2-(2-chloro-4-fluorobenzyl)-5-(2-methylphenyl)-2H-tetrazole	
6612023	5-chloro-N-ethyl-2-methoxy-N-phenylbenzamide	
6660132	4-chloro-1-(2-ethoxybenzoyl)-1H-pyrazole	
6714292	10-(1-methylethylidene)-4-phenyl-4-azatricyclo [5.2.1.0~2,6~] dec-8-ene-3,5-dione	

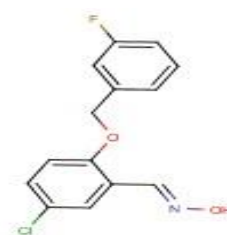
6800944

1-(2-chloro-4-nitrobenzoyl)-  
4-ethylpiperazine

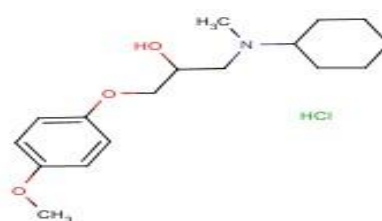
6832681

5-chloro-2-[(2-fluorobenzyl)  
oxy] benzaldehyde oxime

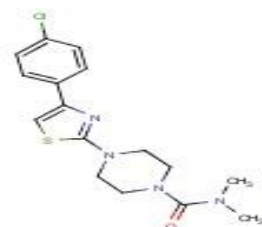
6843219

5-chloro-2-[(3-fluorobenzyl)  
oxy] benzaldehyde oxime

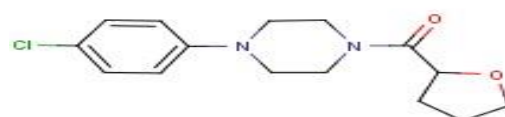
6943696

1-  
[cyclohexyl(methyl)amino]-  
3-(4-methoxyphenoxy)-2-  
propanol hydrochloride

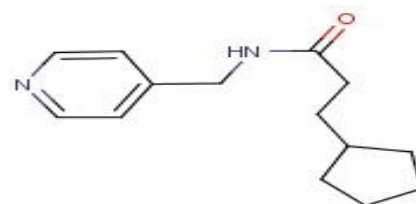
6950620

4-[4-(4-chlorophenyl)-1,3-  
thiazol-2-yl]-N, N-dimethyl-  
1-piperazinecarboxamide

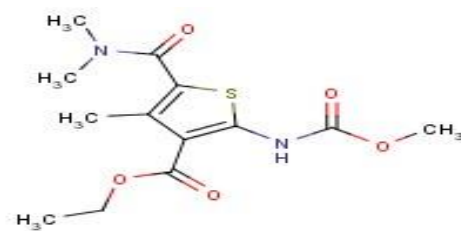
7275637

1-(4-chlorophenyl)-4-  
(tetrahydro-2-  
furanylcarbonyl) piperazine

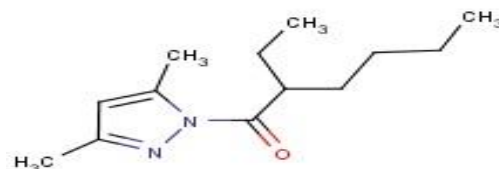
7293667

3-cyclopentyl-N-(4-  
pyridinylmethyl)  
propanamide

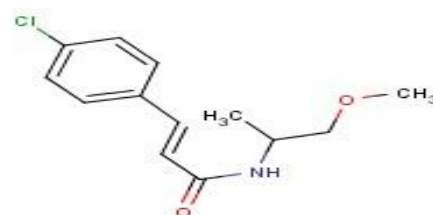
7294926 ethyl 5-  
[(dimethylamino)carbonyl]-  
2-[(methoxycarbonyl)amino]-  
4-methyl-3-  
thiophenecarboxylate



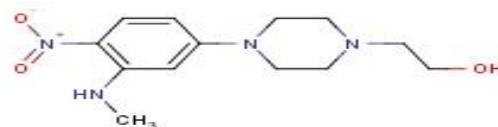
7299305 1-(2-ethylhexanoyl)-3,5-  
dimethyl-1H-pyrazole



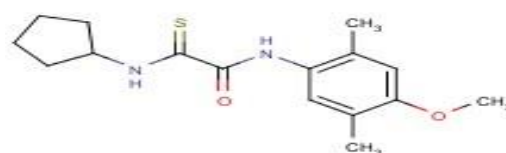
7364709 3-(4-chlorophenyl)-N-(2-  
methoxy-1-methylethyl)  
acrylamide



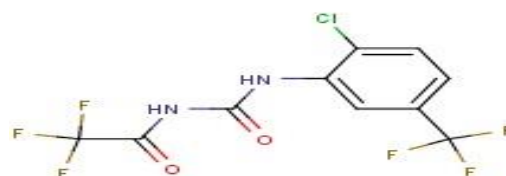
7951028 2-{4-[3-(methylamino)-4-  
nitrophenyl]-1-piperazinyl}  
ethanol



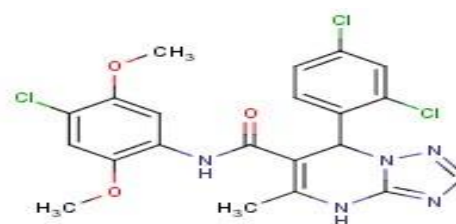
7952968 2-(cyclopentylamino)-N-(4-  
methoxy-2,5-  
dimethylphenyl)-2-  
thioxoacetamide



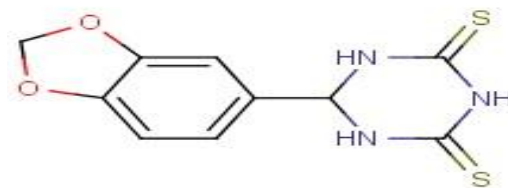
7953011 N-({[2-chloro-5-  
(trifluoromethyl) phenyl]  
amino} carbonyl)-2,2,2-  
trifluoroacetamide



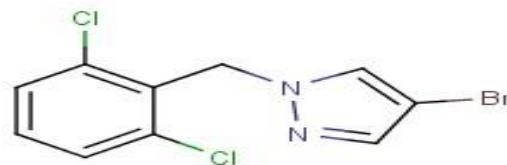
7953379 N-(4-chloro-2,5-  
dimethoxyphenyl)-7-(2,4-  
dichlorophenyl)-5-methyl-  
4,7-dihydro [1,2,4]  
triazolo[1,5-a] pyrimidine-6-  
carboxamide



7953623

6-(1,3-benzodioxol-5-yl)-  
1,3,5-triazinane-2,4-dithione

7955822

4-bromo-1-(2,6-  
dichlorobenzyl)-1H-pyrazole

7959564

4,5-dimethyl-2-(4-  
nitrophenyl)-1H-imidazol-1-  
ol 3-oxide