The Synthesis and Evaluation of Novel Calpain-I Inhibitors

by

Sarah Elizabeth Adams

A thesis submitted to
Cardiff University
for the degree of
DOCTOR OF PHILOSOPHY

School of Chemistry
Cardiff University
June 2013

Declaration

This work has not previously been submitted in substance for any degree or award at this or any other university or place of learning, nor is being submitted concurrently in canditure for any degree or other award.

Signed ……………………………………… Date ………………………………………

Statement 1

This thesis is being submitted in partial fulfilment of the requirements for the degree of PhD.

Signed ……………………………………… Date ………………………………………

Statement 2

This thesis is the result of my own independent work/investigation, except where otherwise stated. Other sources are acknowledged by explicit references. The views expressed are my own.

Signed ……………………………………… Date ………………………………………

Statement 3

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan and for the title and summary to be made available to outside organisations.

Signed ……………………………………… Date ………………………………………

Abstract

The calcium activated cysteine protease calpain-I has a pivotal role in a variety of physiological processes within the human body. In particular calpain-I enables the cell spreading and subsequent chemotaxis behaviour of neutrophils in response to tissue damage. Neutrophils are linked to the pathological condition rheumatoid arthritis and so calpain-I is considered be a valuable therapeutic target. Many inhibitors of calpain-I are highly non-selective with the exception of two small molecule synthetic inhibitors. A phenyl and an indole-based α-mercaptoacrylic acid have shown a slight selectivity towards calpain-I over other cysteine proteases.

In this work 24 novel monohalogenated α-mercaptoacrylic acid inhibitors were prepared based on these lead structures using Vilsmeier-Haack chemistry followed by Knoevenagel condensation of the resulting aromatic aldehydes as key steps. The thiols within the α-mercaptoacrylic acid moiety demonstrated a tendency to form disulfide bridges in solution. Analysis of this disulfide formation through ^{1}H NMR spectroscopy, UV-Vis spectrophotometry and HPLC showed that the monomeric form was active under the reducing conditions used in subsequent assays.

The analogues were tested as inhibitors of calpain-I revealing that bromoindole based inhibitors were the most potent. Selected compounds showed ~10 fold selectivity towards calpain-I versus calpain-II. In live neutrophils they were capable of slowing the cell spreading process by up to 70%. When live neutrophils containing the inhibitor were irradiated with 410 nm light, the cells completely lost the ability to spread.

To show that these compounds were allosteric inhibitors the calcium binding domain PEF(S) was expressed in *E. coli* and purified using anion exchange chromatography and size exclusion chromatography. Solution of X-ray co-crystal structures of the calpain PEF(S) domain with two different inhibitors revealed that they bind to the protein in a similar fashion as an α-helical domain of calpastatin, the endogenous inhibitor of calpain.

Acknowledgements

Table of Contents

List of Figures

List of Schemes

List of Tables

List of abbreviations

A	Adenosine
Ac$_2$O	Acetic anhydride
AcOH	Acetic acid
Acyl-CoA	Acyl-coenzyme A
AMC	7-amido-4-coumarin
APCI	Atmospheric pressure chemical ionisation
APS	Ammonium persulfate
ATP	Adenosine triphosphate
Bak	BCL-2 homologous antagonist killer
Bax	BCL-2 associated X protein
βME	β-mercaptoethanol
Bid	BH3 interacting domain
BSA	Bovine serum albumin
C	Cytosine
c	Concentration
Ca^{2+}	Calcium ion
CCP4	Collaborative computational project number 4
CD	Cicular Dichroism
Co(II)salen	*N,N′*-bis(salicylidene)ethylenediaminocobalt(II)
COSY	Homonuclear correlation spectroscopy
CV	Column volume
DABCYL	4-(Dimathylaminophenyl)diazenylbenzoic acid
DEAE	Diethylaminoethanol
DEPT	Distortionless enhancement by polarisation transfer
dH$_2$O	Deionised water
DNA	Deoxyribose nucleic acid
DMF	*N,N*-Dimethylformamide
DMSO	Dimethylsulfoxide
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EFH	EF-hand

EI	Electron ionisation
ES	Electrospray ionisation
FAM	Fluorescein amidite
FMLP	Formyl-methionyl-leucyl-phenylalanine
FPLC	Fast protein liquid chromatography
FRET	Förster/fluorescence resonance energy transfer
G	Guanosine
HEPES	2-[4-(2-Hydroxyethyl)piperazin-1-yl]ethanesulfonic acid
HMBC	Heteronuclear multiple-bond correlation spectroscopy
HPLC	High pressure/performance liquid chromatography
HRMS	High resolution mass spectrometry
HSQC	Heteronuclear single-quantum correlation spectroscopy
IC_{50}	Half maximal inhibitory concentration
IP_3	Inositol triphosphate
IPTG	Isopropyl β-$_D$-1-thiogalactopyranoside
kb	Kilobasepairs
K_iPO_4	Potassium phosphate
l	Pathlength
LB	Luria-Bertani growth media
MALDI-TOF	Matrix assisted laser desorption ionisation – time of flight
MeCN	Acetonitrile
MeOD	Deuterated methanol
MeOH	Methanol
mp	Melting point
MW	Molecular weight
MWCO	Molecular weight cut-off
n	Number of peptide bonds
NaOAc	Sodium acetate
NMR	Nuclear magnetic resonance
NSI	Nanospray ionisation
OD	Optical density
p53	Protein 53
PEG	Polyethylene glycol

RA	Rheumatoid arthritis
RCF	Relative centrifugal force
r.m.s.d	Root mean square deviation
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate – polyacrylamide gel electrophoresis
T	Thymidine
TEMED	*N,N,N',N'*-Tetramethylenediamine
TFA	Trifluoroacetic acid
TNS	6-(*p*-toluidino)-2-naphthanlenesulfonic acid
Tris	Tris(hydroxymethyl)aminomethane
UV-Vis	Ultraviolet-visible spectroscopy
vdW	van der Waals force
ε	Extinction coefficient
θ	CD signal (millidegrees)
λ_{max}	Maximum absorbtion wavelength
υ	Wavenumber

In this work, amino acids are represented as either their one or three letter codes;

Amino Acid	3 Letter code	1 Letter code
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamic acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

1 Introduction

1.1 Rheumatoid Arthritis

Rheumatoid arthritis (RA) is one of the most common forms of inflammatory arthritis and one of the most prevalent autoimmune diseases.[1] It is an autoimmune condition where the body's natural defence mechanism becomes over stimulated and attacks the body itself.[2] Though the exact causes of RA are unknown, there are links to both genetics[3] and environmental factors such as smoking.[4] The damage that occurs through RA causes severe pain and disability to the patient and through development of secondary conditions, such as cardiovascular disease and pulmonary fibrosis, leads to an increased mortality rate.[5, 6] The condition is known to affect approximately 1% of the world's population and in the UK alone costs the economy almost £8 billion each year (2010 figures).[6, 7]

In RA, the joints are attacked at the synovial membrane by a number of leukocytes and lymphocytes (white blood cells). This process is thought to be initiated *via* a specific subset of lymphocytes known as T cells, more specifically T-helper cells.[8] T-helper cells are activated in response to antigen presenting cells found within the synovium.[8] The presence of these cells causes an influx of macrophages,[9] a form of leukocyte, into the synovial lining. As a consequence the release of cytokines is stimulated.[8, 10, 11] These pro-inflammatory cytokines, such as TNFα and interleukin-1, act as signalling molecules for the attraction and activation of other leukocytes and lymphocytes to the area, such as neutrophils, chondrocytes and fibroblasts[12] The activated chondrocytes and fibroblasts secrete metalloproteases and reactive oxygen species that damage the cartilage and tissue within the surrounding area.[8] The constant bombardment of white blood cells eventually causes the formation of granular pannus tissue, which grows over the articular cartilage.[6, 13] This pannus tissue is composed of synovial fibroblasts and mononuclear cells and leads to further destruction of the cartilage and bone due to the high concentration of metalloproteases (Figure 1.1).[13-15] In addition the synovial fluid within the joint increases in volume and decreases in viscosity leading to the loss of lubrication within the joint and further damage (Figure 1.1). Synovial fibroblasts within the arthritic joint can also spread to other previously unaffected joints.[16]

The role of the neutrophil, a polymorphonuclear granulocyte, is generally forgotten within rheumatoid arthritis.[17] Neutrophils, the most abundant form of white blood cells, are found within the joint, with the highest populations found within the synovial fluid.[15, 18, 19] Neutrophils contain granules that release their toxic contents into the surrounding area and therefore can cause severe damage to the cartilage and bone (Section 1.2).[17, 20, 21]
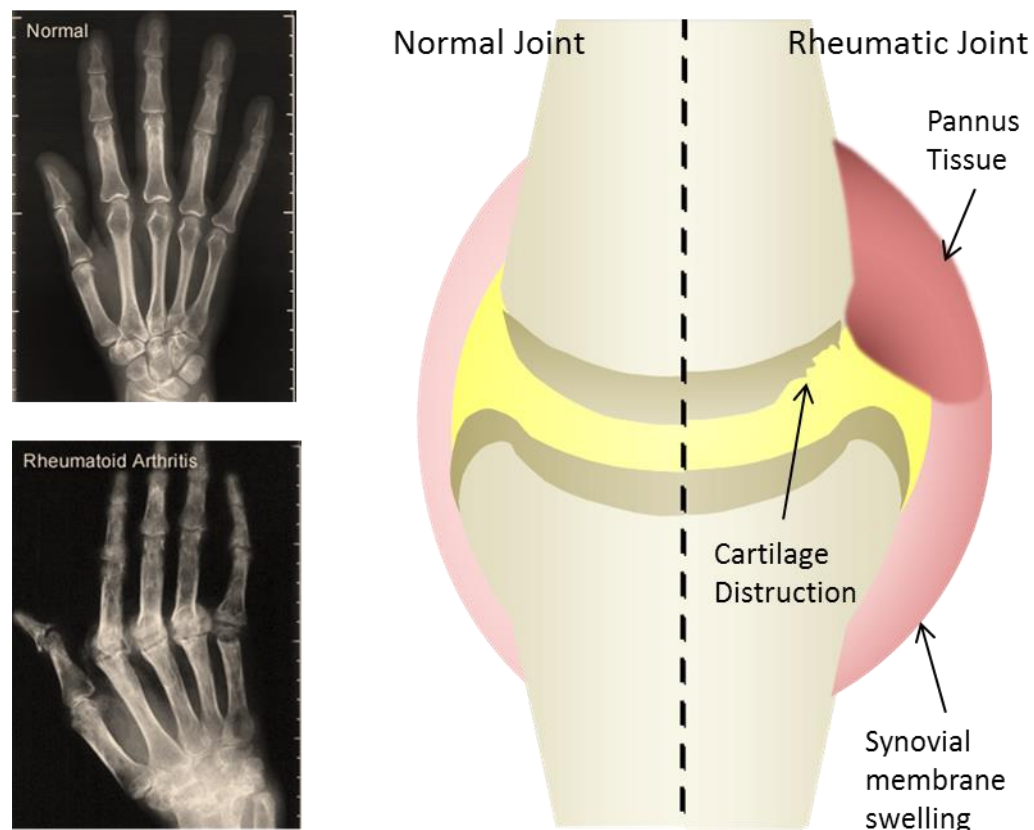
Figure 1.1: *Image depicting the damage caused to a hand afflicted with rheumatoid arthritis (top) in comparison to a normal hand (bottom). A diagram representing a joint with one side afflicted with rheumatoid arthritis (right) and the normal joint (left).[1]*

1.2 Neutrophils

Neutrophils are a granular leukocyte and are the most abundant group in the immune system making up 40-65% of all white blood cells.[22] They are typically the first cells to reach the site of infection or inflammation.[22] Granular leukocytes are so called as they contain cytoplasmic granules, the contents of which are highly toxic towards foreign bodies (*e.g.* microbes).[22] These granules contain a wide variety of proteinases and hydrolases as well as proteins that have the capability to generate reactive oxygen species (ROS).[21] Upon normal

activation at the site of action, the cells are phagocytic and encapsulate the foreign bodies in a vacuole. Neutrophils are 12-15 μm in circumference and are capable of encapsulating up to eight separate zymosan particles, which are 1-2 μm in circumference.[23] The cell then releases the contents of the granules into the vacuole formed; it is these contents that destroy the foreign body (Figure 1.2).[21, 24, 25] Neutrophils have a relatively short lifetime of 5-7 days within the body.[22] Once a cell has reached the end of its life cycle it undergoes apoptosis and the remnants of the neutrophil are taken up by macrophages.[25]
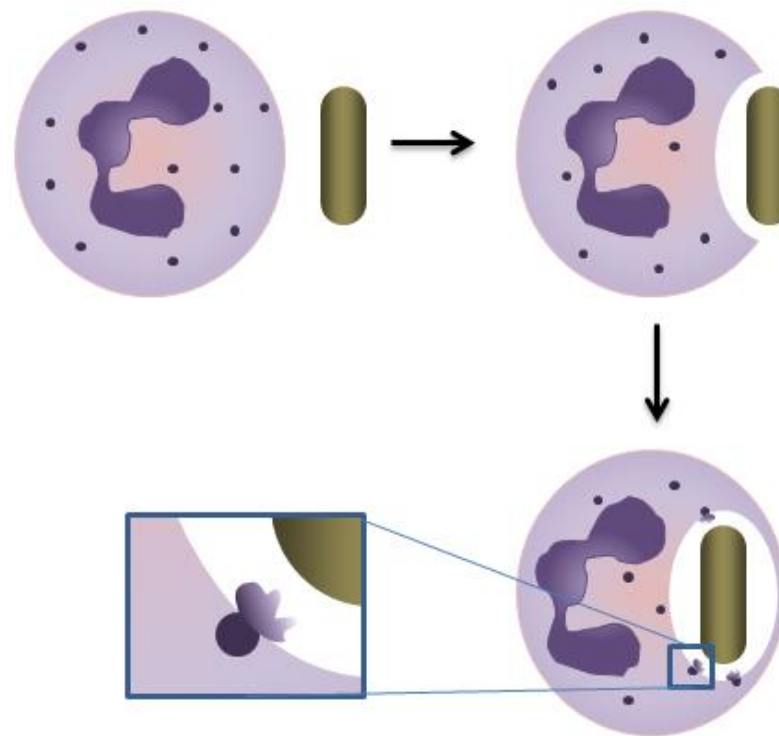
Figure 1.2: A neutrophil undergoing phagocytosis and upon consuming the microbe releasing granules into the vacuole containing the microbe, these granules contain proteases, hydrolases and reactive oxygen species (ROS).[21, 22]

To reach a site of damage or infection neutrophils must migrate across several tissue barriers beginning with blood vessel walls. Initially neutrophils detect small molecule chemoattractants, for example *N*-formyl peptides and interleukin-8, which are released by cells in the damaged area.[22, 26, 27] Through detection of these chemoattractants the neutrophils are triggered to migrate towards the site and stimulate the activation of integrins, cell adhesion molecules.[27] Migration of neutrophils from the bloodstream occurs after the cells have undergone a rapid shape change to adhere to the endothelium of the

blood vessel and extrude themselves across the tissue barrier.[23, 28] In the bloodstream neutrophils are roughly spherical in shape with the cell surface membrane comprised of numerous wrinkles all along the surface, these wrinkles are held together through two different protein scaffolds.[23] L-selectin and β_2-integrin are bound in the membrane and are localised upon the peaks and valleys of the wrinkles in the cell membrane, respectively (Figure 1.3).[28] It is postulated that the glycoprotein L-selectin is bound to the intracellular protein actin through a protein 'bridge' ezrin.[23, 29] Actin is bound to the membrane bound protein β_2-integrin through talin. This model is based upon what has been reported for the wrinkles found within lymphocytes (Figure 1.3).[23, 29] In response to the chemoattractants calcium (Ca^{2+}) is released. The release of calcium leads to the activation of the cystolic cysteine protease calpain-I (Section 1.5) at the plasma membrane.[30, 31] The activated protease then cleaves the proteins holding the wrinkles in place (talin and ezrin) allowing the cell membrane to spread and flatten (Figure 1.3).[23, 32-34] Once the neutrophil has spread the L-selectins and β-integrins are capable of interacting and adhering to membrane bound proteins on the endothelial wall.[22] The neutrophil then undergoes chemotaxis, whereby the cell migrates towards an ever increasing concentration of chemoattractants.[35] In order to move the neutrophil forward the actin filaments below the cell membrane are used as protrusions that push the membrane forward.[22, 36] Detachment at the rear of the neutrophil occurs *via* the recycling of β-integrins towards the 'front' of the cell with the aid of calcium and calcineurin.[37] The process of chemotaxis occurs until the site of damage is reached.[35]
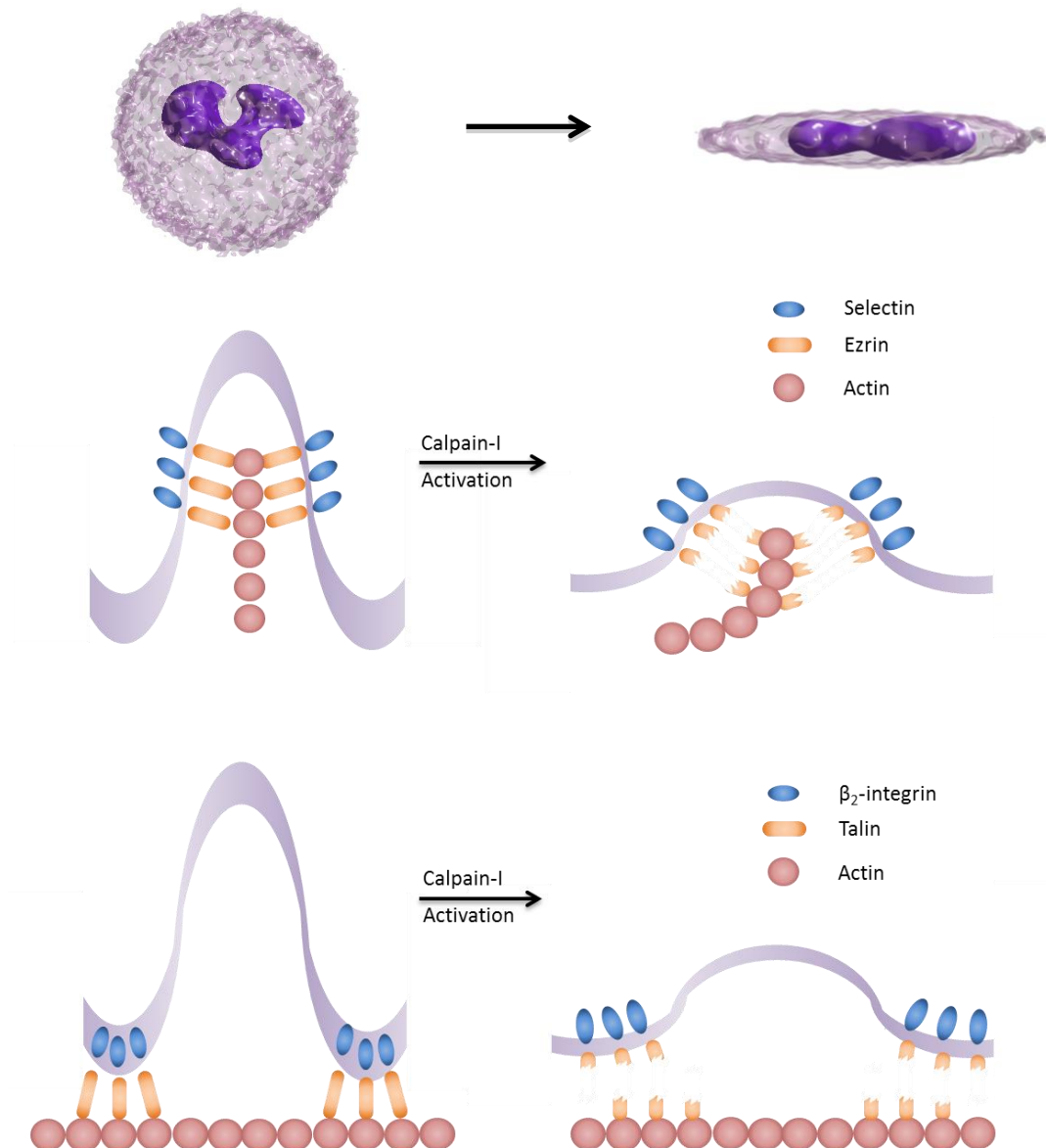
***Figure 1.3:** The structure of a neutrophil in the bloodstream with wrinkles on the cell surface membrane. Detailed structures of the individual wrinkles with L-selectin/β_2-integrin bound to ezrin/talin which is bound to the intracellular protein actin (peak of the wrinkle/'valley' of the wrinkle).[23]*

1.3 The Calpain Family

The calpain family is a group of calcium activated cysteine proteases within the papain super-family.[38, 39] As well as calpains the papain superfamily consists of bleomycin hydrolases and papains.[38, 40] The calpain family has 15 known isoforms which all contain the catalytic active site known as the CysPc domain (Section 1.6.2).[41] Some of these isoforms are ubiquitously expressed throughout the human body, such as the two most studied members calpain-I and calpain-

II;[42] others are tissue specific, such as calpain-III (p94), which is found in the skeletal muscle.[43] This family comprises two subsets; the classical calpains, of which there are nine known isoforms that are solely activated by calcium and are structurally similar. The second subset are the non-classical calpains, these isoforms possess calcium binding properties but also feature other activation methods such as zinc finger domains.[44] Calpains are modulator proteases meaning that instead of hydrolysing a protein substrate to destroy it, such as caspases and proteasomes, they modify and repurpose proteins to alternate functions within the body.[41]

1.4 Cysteine Protease – The mechanism of action
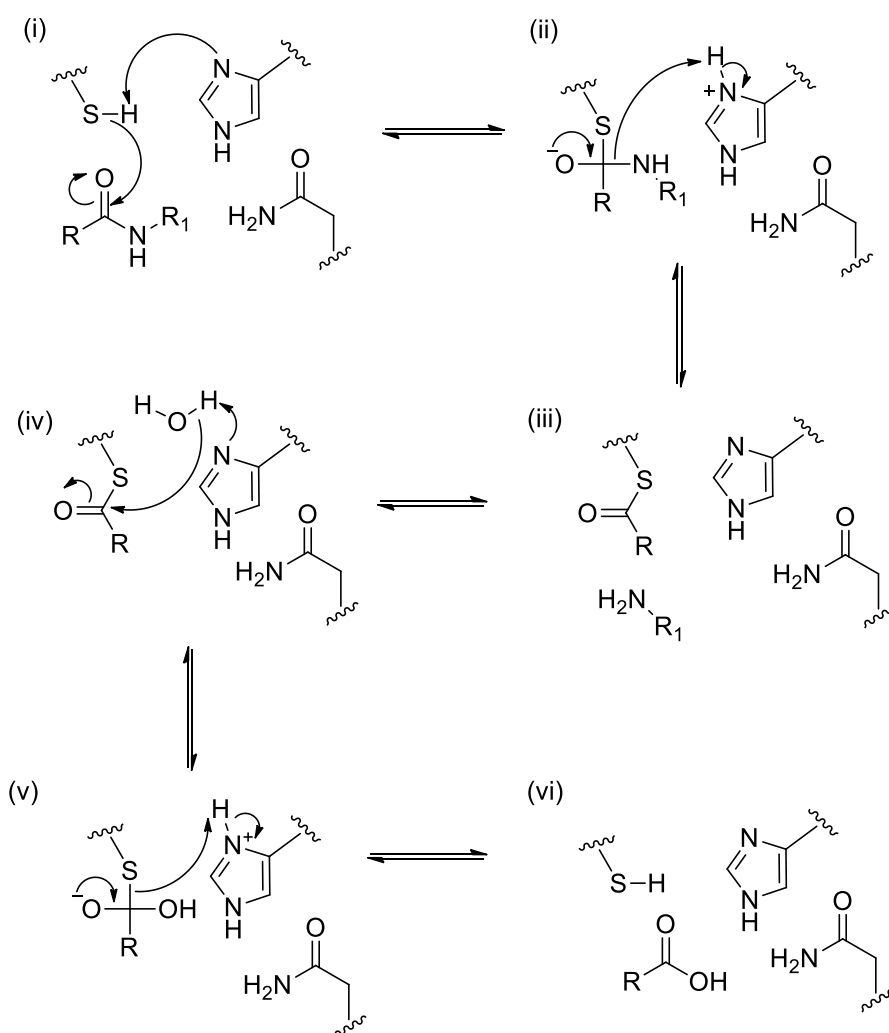
Figure 1.4: *General mechanism of the hydrolysis of a peptide bond within the active site of a cysteine protease.[45]*

To catalyse the hydrolysis of the target proteins, cysteine proteases use a cysteine residue within a catalytic triad that is completed with a histidine and an asparagine residue.[46] The asparagine is utilised to manoeuvre the histidine residue into the correct position to act as a base for deprotonation of the sulfhydryl of the cysteine residue (Figure 1.4).[45] The nucleophilic thiolate then attacks the carbonyl group of the target peptide bond (Figure 1.4).[45] The incorporation of water then completes the hydrolysis and the peptide bond is cleaved and the cysteine residue is regenerated to carry out further hydrolysis (Figure 1.4).[45]

1.5 Calpain-I and calpain-II

Calpain-I and calpain-II are heterodimeric proteins composed of a large subunit, known as CAPN1 and CAPN2 respectively, and a small subunit known as CAPNS1.[44] CAPN1 comprises 714 amino acids (MW 81,889)[47] which is 14 more than CAPN2 (MW 80,005).[48] The sequence similarity between the large subunits of these two isoforms of calpain is 62%.[48] CAPNS1, also known as the regulatory subunit, is composed of 268 amino acids (MW 28,316) and is identical in both calpain-I and calpain-II.[49] The large subunit comprises four domains and the small subunit two (Figure 1.5).[50] The N-terminal domain is the anchor α-helix (Section 1.6.1), this is followed by the proteolytic core of the enzyme known as the CysPc domain. The CysPc domain is divided into two halves, known as PC1 and PC2 (Section 1.6.2).[41, 51] The third domain is the C2L domain and is similar to the C2 phospholipid binding domains found in phospholipases (Section 1.6.3).[41, 51, 52] The fourth domain within the large subunit is a calcium binding domain that is composed of five EF hand motifs four of which bind calcium ions; it is known as the PEF(L) domain (Section 1.6.4).[41, 51] The two domains that are found in the small subunit (CAPNS1) are the N-terminal glycine rich (GR) domain (Section 1.6.5) and the PEF(S) calcium binding domain, which is also composed of five EF hand motifs (Section 1.6.4).[41, 51]

(A)

(B)
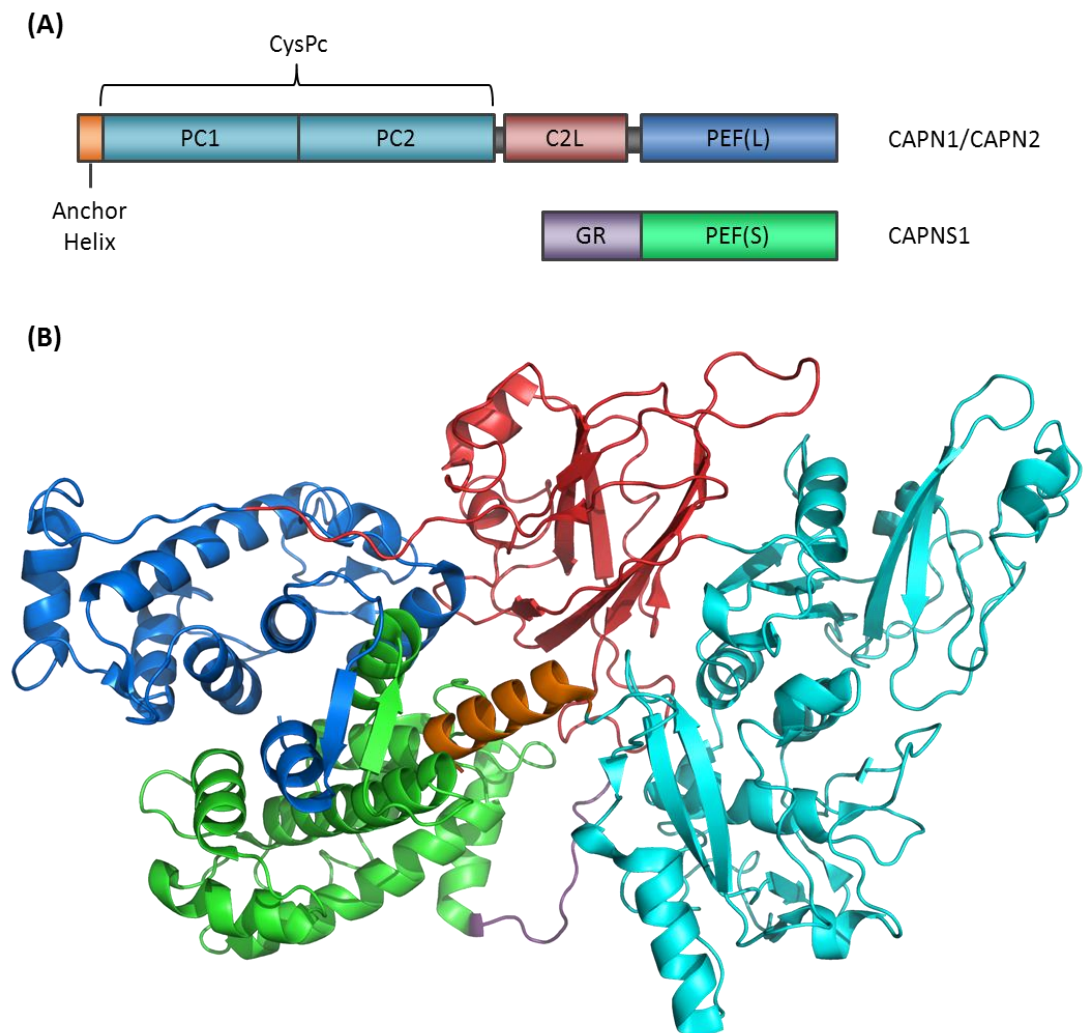
Figure 1.5: *(A) Domain structure of CAPN1 and CAPN2 along with the CAPNS1. (B) A secondary structure representation of inactive CAPN2 (PDB:1KFU).[53] The large subunit; the anchor helix (orange), CysPc (cyan), C2L (red) and PEF(L) domain (blue). Small subunit; GR domain (purple) and PEF(S) domain (green). Image rendered in PyMOL as are all subsequent protein images.[54]*

The most significant functional difference between calpain-I and calpain-II is the concentration of calcium ions that is required for activation.[50] ~50 µM of calcium is required for the activation of calpain-I where as ~0.35 mM of calcium is required for the activation of calpain-II.[50] These required concentrations of calcium give rise to the two other names for these isoforms; µ-calpain and m-calpain. The concentration of calcium required to activate these two isoforms was measured *in vitro* and both are higher than the levels of calcium ions that are found *in vivo*. Hence there must be other factors that come into play for the activation of these proteases within the body, such as the binding of activator proteins (Section 1.8).[55] Both of these isoforms of calpain are inhibited by the

same regulatory endogenous inhibitor, calpastatin (Section 1.7), a highly specific inhibitor for conventional calpains.[56]

1.6 Domain structure

To date there are only full length X-ray structures published for calpain-II. These have enabled the classification of the domains within the structure (Section 1.5). There is only a hybrid crystal structure of calpain-I currently available, which contains the N-terminal helix and PEF(L) domains of calpain-II and the the CysPc and C2L domains of calpain-I.[57] The calpain-I-like structure was very similar to that of calpain-II so the structural analysis of the individual domains (Sections 1.6.1 to 1.6.5) will be based on the structures of calpain-II.[53,55,57]

1.6.1 The N-terminal anchor α-helix

Situated at the N terminus of the protein, the anchor α-helix domain is a small domain of 19 residues in CAPN2 (Figure 1.6) and leads into the CysPc domain, the proteolytic core (Section 1.6.2).[53] The sequence alignment of CAPN1 with CAPN2 indicates that this anchor α-helix is ten residues longer in CAPN1 than CAPN2, though there is no published structure demonstrating the purpose of this.[58] The increase in the length of this domain, in calpain-I, has been shown to account for ~20% of the difference in the concentrations of calcium that is required for the activation of the two isoforms.[59] The first 40 residues of calpain-II were substituted with those of calpain-I and the concentration of calcium that was required to activate the protease was reduced to 250 μM.

Prior to calpain-II activation, the α-helix interacts with the PEF(S) domain (Section 1.6.4) (Figure 1.6), which enables clamping of the two subunits together.[53] Numerous hydrophobic residues from the PEF(L) domain and the anchor α-helix interact with one another to stabilise the inactive conformation of the protease.[60] Upon activation of calpain-II, the α-helix is released relieving some of the restraints that hold the two halves of the CysPc domain apart. The release of the α-helix is proposed to be followed by autolysis. Autolysis enables activation of calpain-II with a lower concentration of calcium ions.[61] However

this autolysis also causes the heterodimer to become unstable and dissociation occurs, which in turn may lead to a loss of substrate specificity for calpain-II.[62]
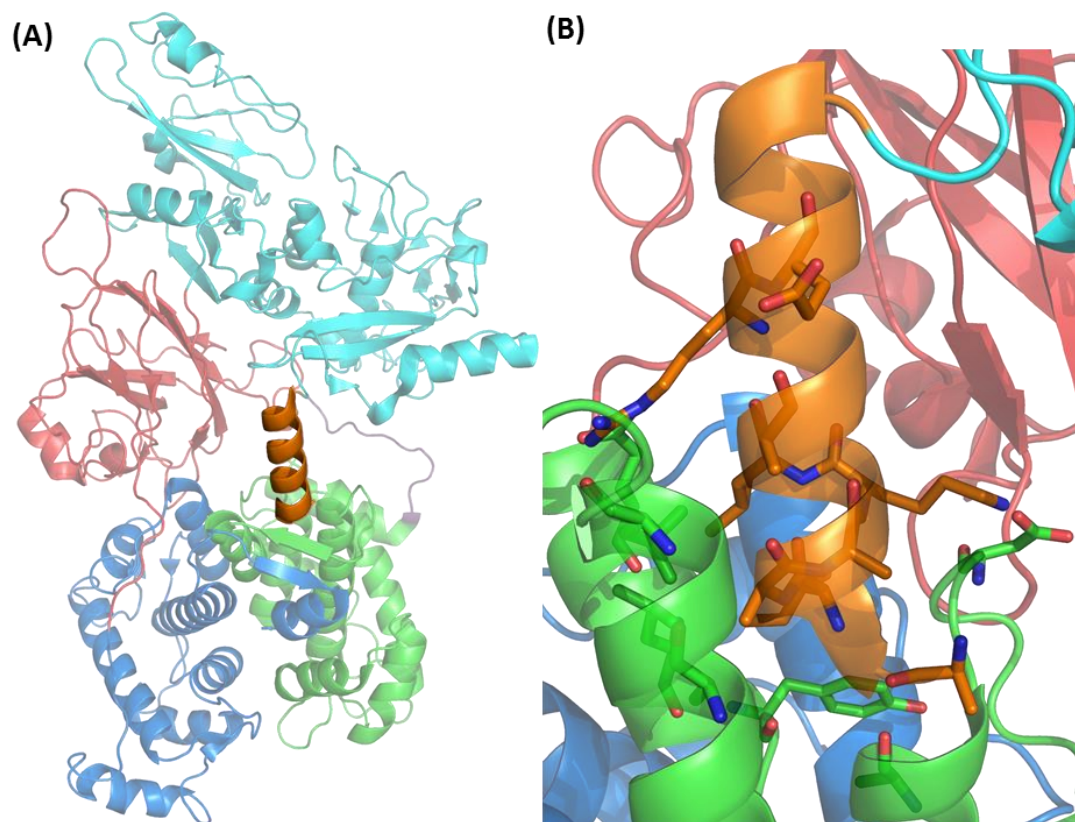
Figure 1.6: *(A) A secondary structure representation of the 19 residue N-terminal α-helix of anchor in orange. (B) The interaction between the anchor helix and PEF(S) that, amongst other interactions, holds the catalytic active site open.[53]*

1.6.2   The CysPc proteolytic domain

This proteolytic domain is where the catalytic triad, cysteine (Cys), histidine (His) and asparagine (Asn), is situated. This domain is composed of two subdomains that are known as PC1 and PC2.[44] PC1 contains the catalytically active Cys residue and PC2 completes the catalytic triad with the His and the Asn.[53, 60] These two subdomains are found in both calpain-I and calpain-II. The catalytic triad is composed of Cys 115, His 272 and Asn 296 in calpain-I,[50] whereas in CAPN2 the catalytic triad is at positions Cys 105, His 262 and Asn 286.[63] The secondary structure of this domain shares significant topology with papain, hence its inclusion in the papain superfamily, where the residues that are found within the immediate vicinity of the active site of calpain are very similar to those that are found in papain.[53, 60]

The secondary structure of the PC1 subdomain is composed of two anti-parallel β-sheets and a series of α-helices.[60] In the structure of calpain-II the Cys is situated in a short α-helix.[60] PC2 is also composed of a series of inter-connected α-helices and two anti-parallel β-sheets.[60] The residues of the catalytic triad are situated on the interface between PC1 and PC2.
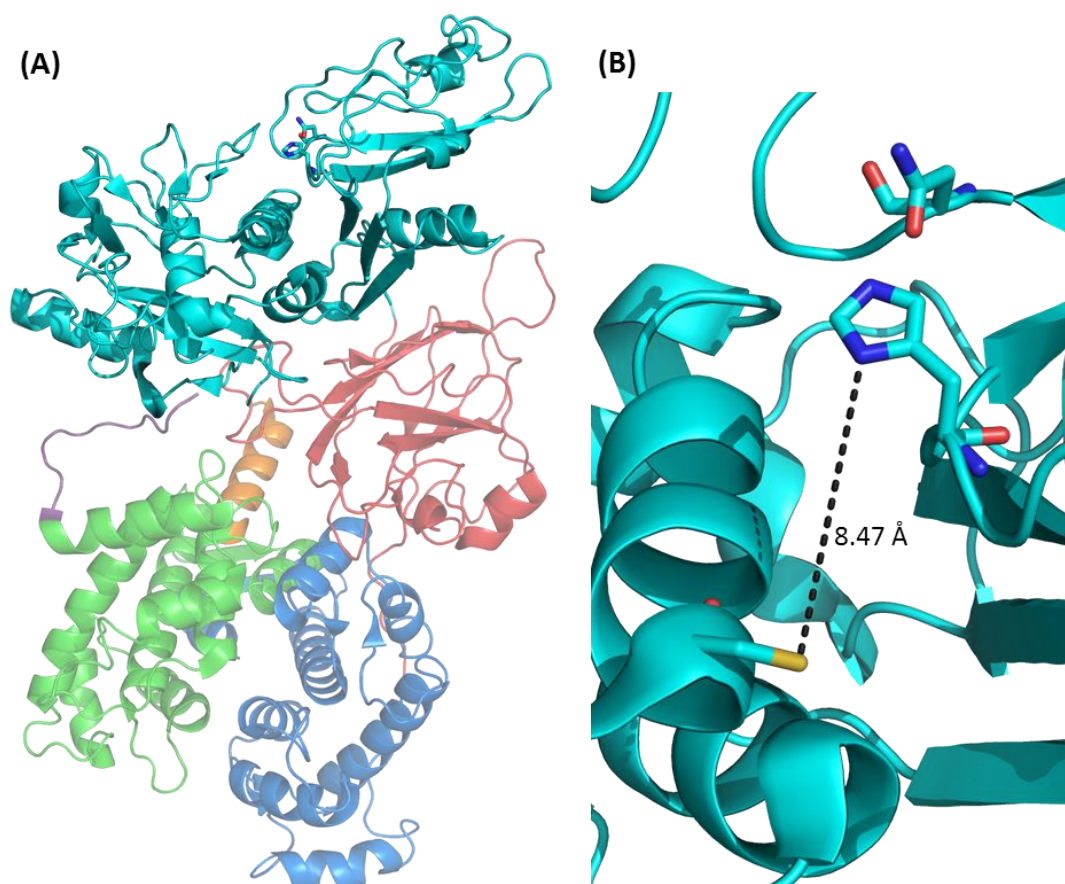
Figure 1.7: *(A) A secondary structure representation of the CysPc domain of calpain-II and its position within the protease, highlighted in cyan. (B) The catalytic triad in the inactive conformation of the CysPc domain of calpain-II, with the distance between the C105 and H286 at 8.47 Å.[53]*

When calpain-II is inactive the catalytic triad is held apart through structural restraints, with Cys105 and the His262 are held 8.5 Å apart from one another in calpain-II (Figure 1.7).[53] One of these structural restraints is the anchoring helix bound to the PEF(S) domain. The binding of calcium ions to PEF(S) causes structural movement of this domain, which in turn releases the anchor α-helix releaving the structural restraints.[60] Another set of structural restraints is the negatively charged loop of the C2L domain that interacts with a series of positively charged residues in the PC2 subdomain. Upon binding calcium these

electrostatic interactions are released to allow for PC1 and PC2 to move towards one another and form the active form of the protease (see Section 1.6.3) (Figure 1.8).[64] The CysPc domain binds two calcium ions which contributes to the movement of the separated portions of the catalytic triad towards one another for catalysis to occur (see Section 1.8 for further discussion).[65]

1.6.3   The C2L domain

This 130 residue domain has a β-sheet based secondary structure and it acts as a bridge between the CysPc domain (Section 1.6.2) and the PEF(L) domain (Section 1.6.4). The C2L domain folds into a pair of four stranded parallel β-sheets and is broadly similar to the structure of C2 binding domains.[60] C2 binding domains are known to bind to phospholipids but for this to occur there is a precursor step where the domain binds calcium.[66] The C2L domain contains a negatively charged loop (Figure 1.8) that comprises residues 392 to 402 in calpain-II and residues 403 to 414 in calpain-I. When the whole of the C2L domain is expressed separately from the rest of the enzyme it does bind calcium with an affinity similar to PEF(L) (Section 1.6.4).[67] The C2L domain has a higher affinity for calcium when phospholipids are also bound to the domain.[67] The phospholipid binding nature of this domain gives evidence of how the protease is capable of migrating from the cytosol, to the site of action at the membrane where activation of the enzyme occurs.[68] The calcium binding nature of the C2L domain has been contradicted with structural evidence where the acidic loop appears not to bind calcium in a crystal structure of calpain II bound with its endogenous inhibitor, calpastatin (Section 1.7) despite calcium ions being present elsewhere in the structure.[69]
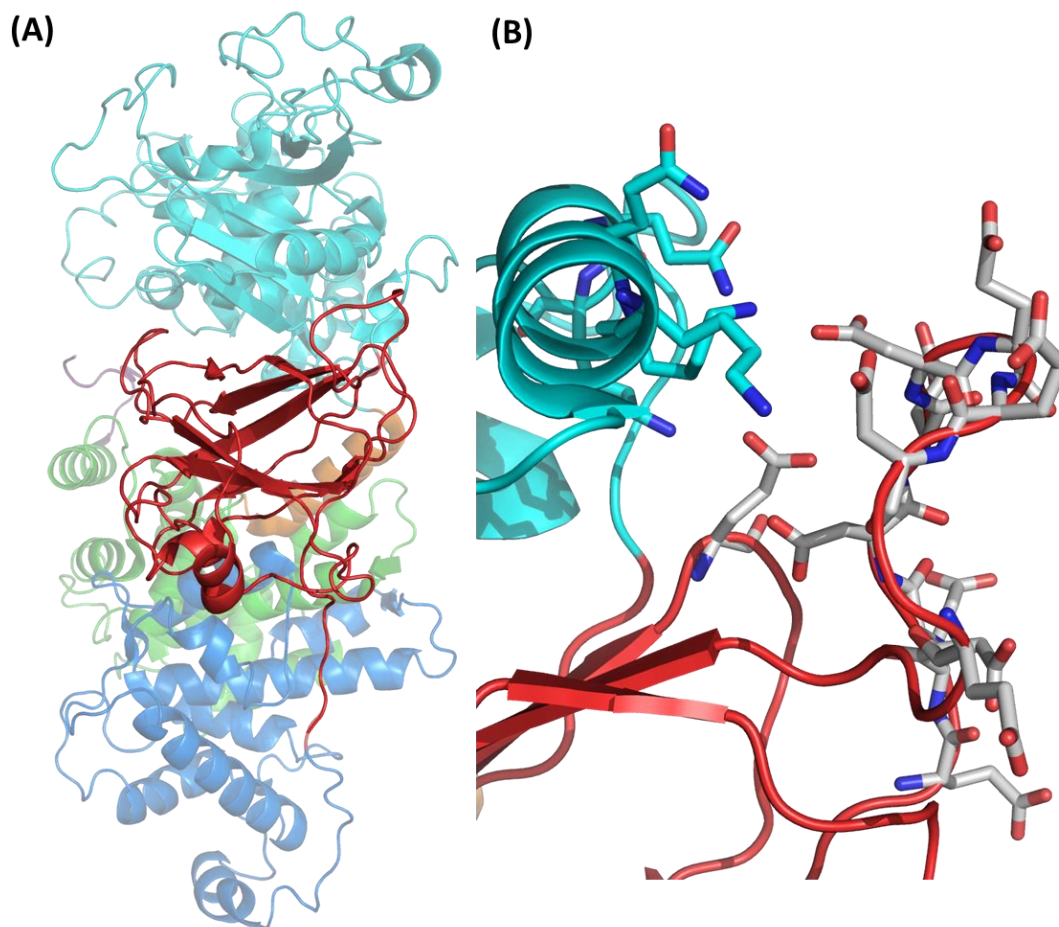
Figure 1.8: *(A) A secondary structure representation of calpain-II C2L domain, in red, (B) a close up of the negatively charged loop (residues 392 to 402) and its interactions with the positively charged residues in CysPc that keep the protease in an inactive conformation (PDB:1KFU).[53]*

The negatively charged loop, of the C2L domain, interacts with a series of positively charged lysines found in CysPc domain (Figure 1.8). These interactions hold the cleft between PC1 and PC2 open and therefore keep the protease in an inactive conformation.[64] Another negatively charged residue which is situated on the linker between the CL2 domain and the PEF(L) domain, Glu504, interacts with Lys230 in PC2 subdomain.[70] Mutations to the negatively charged loop and Glu504 cause lowering of the concentration of calcium that is required for activation *in vitro*, demonstrating that these electrostatic interactions are integral for keeping the protease in its inactive form.[70] Upon activation, disruption of the electrostatic interactions occurs allowing the active site to 'close' and proteolysis to occur.

1.6.4 PEF(L) and PEF(S) domains

These two domains are both calcium binding domains. PEF(L) is at the C-terminus of the large subunit and PEF(S) is one half of the small subunit.[44] Both of these domains are composed of five EF-hands.[53, 60]
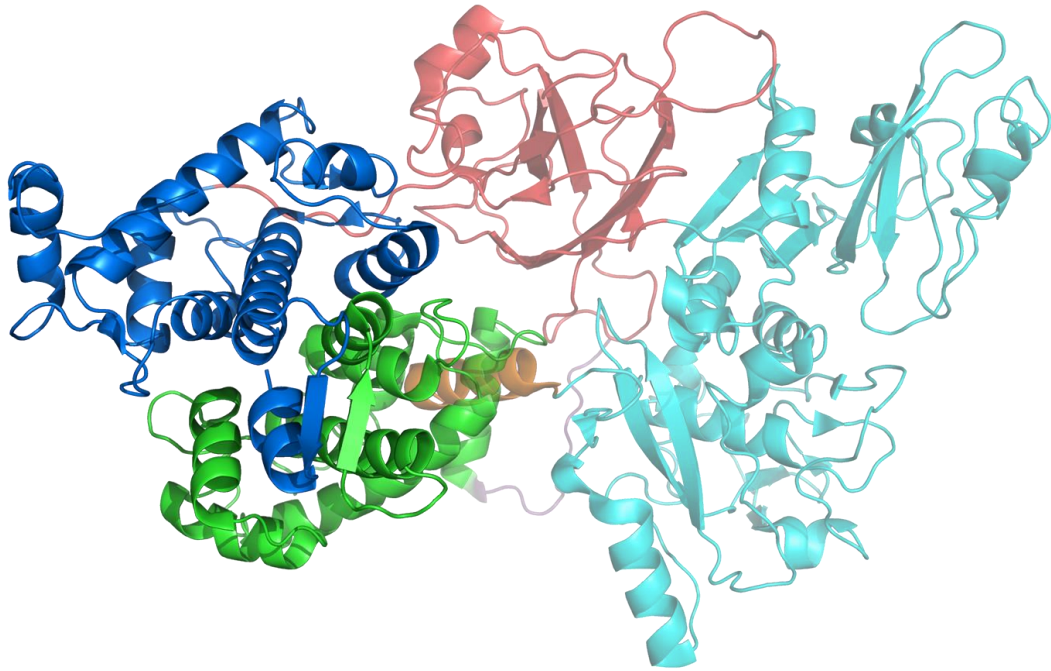
Figure 1.9: *A secondary structure representation showing the position of calpain PEF(L) in blue, and PEF(S) in green, within the whole protein structure (PDB:1KFU).[53]*

EF-hands

EF hands are composed of two perpendicular α-helices that are linked together through an interhelical loop, the loop contains residues that coordinate a single calcium ion (Figure 1.10).[71] EF-hands (EFH) contain the X • Y • Z • -Y • –X • • –Z motif that is indicative of a calcium binding domain. Where X, Y, Z, -X, -Y and – Z are coordination ligands of calcium and the dots represent the intervening residues in the EF hand sequence.[72] Typical coordinating functional groups are the carbonyls of the peptide bond, or negatively charged side chains such as the carboxylates of aspartate or glutamate, or hydroxyl groups such as in threonine.[73] To complete the coordination of the calcium ion other molecules, such as water, may also be used.[73] There are two forms of the EF-hand loop; the canonical and the non-canonical loop, both of which are found within PEF(L) and PEF(S) of calpain-I and calpain-II (Table 1.1).[74]

EF-hand Motif	X	*	Y	*	Z	*	-Y	*	-X	*	*	-Z
EFH-2 domain VI (Canonical)	**D**	S	**D**	T	**T**	G	**K**	L	G	F	E	**E**
EFH-1 domain VI (Non-Canonical)	L	**A**	G	D	**D**	M	**E**	V	S	A	T	**E**

Table 1.1: Sequence alignment of a canonical (EFH-2 of PEF(S)) and a non-canonical (EFH-1 of PEF(S)) against the EFH binding motif, these sequences are from rat calpain-II.[74]

The canonical EF-loop uses the calcium binding motif noted above, where five residues of the six bind to the calcium ion. The pentagonal bipyramidal structure is completed with the presence of a water molecule.[75] EFH-2 of domain VI clearly demonstrates canonical binding; with X, Y, Z, -Y and –Z represented by residues within the peptide sequence; Asp154, Asp156, Thr158, Lys160 and Glu165 (Figure 1.10).[74] In the case of Lys160 the carbonyl from the peptide backbone is used to coordinate the calcium ion. –X is not represented by a residue but is instead replaced with a water molecule for the completion of the pentagonal bipyramid coordination.[74]

With non-canonical binding of a calcium ion the length of the peptide chain is generally shorter than 12 residues.[73] Therefore fewer side chains and carbonyls are used for binding and extra water molecules are required to complete the coordination of calcium.[73] An example of a non-canonical binding motif is EFH-1 of PEF(S), with 11 residues within the coordinating sequence as opposed to the 12 residues found in the canonical sequence (Table 1.1).[74] The coordinating residues are Ala111, Asp114, Glu116 and Glu121, with the carbonyls of the peptide backbone of Ala111 and Glu116 coordinating the Ca^{2+} ion (Figure 1.10 (B)).[74] The coordination sphere is completed with two water molecules (Figure 1.10 (B)).[74]
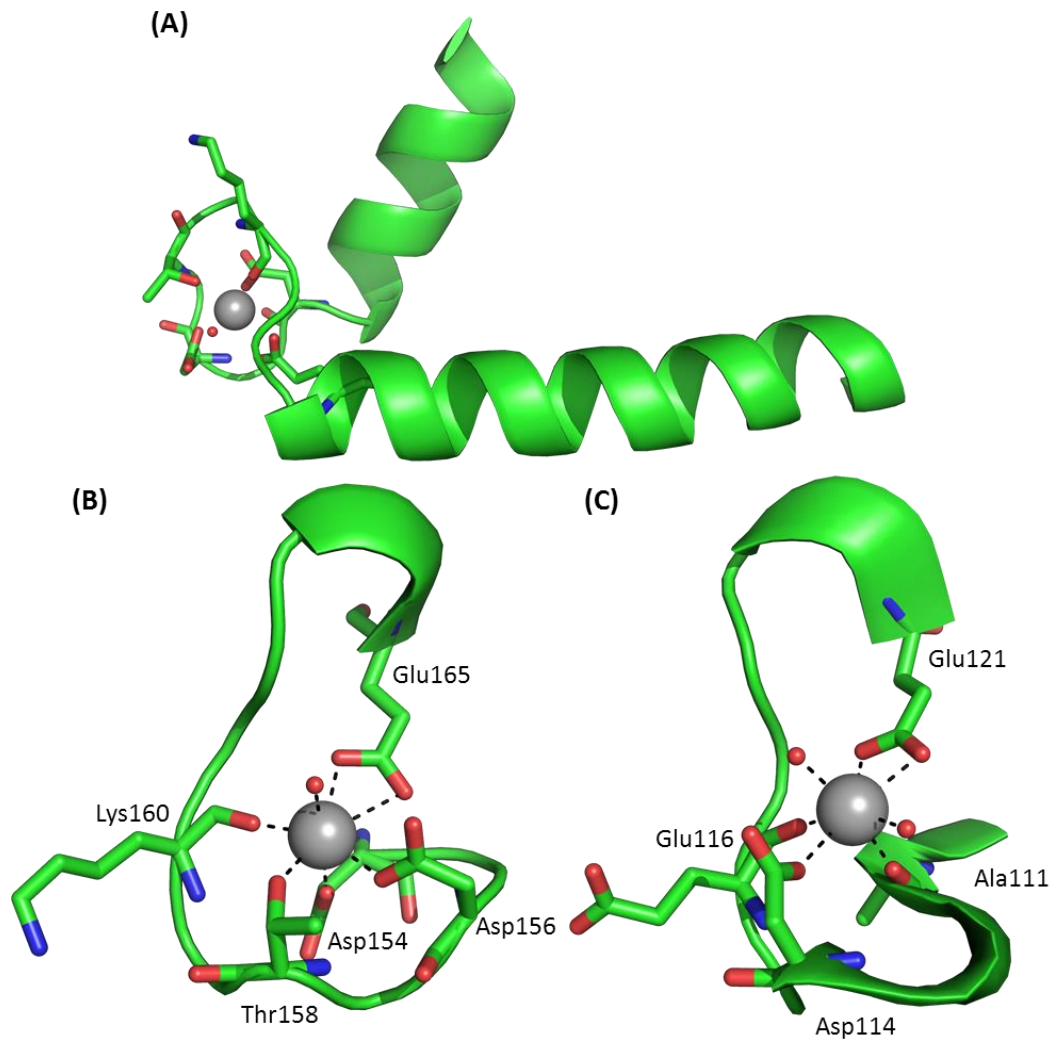
Figure 1.10: *(A) A secondary structure representation of the full structure of an EF-hand with the helices perpendicular and interconnected through the calcium binding loop. (B) A stick representation EFH-2 of calpain PEF(S), demonstrating canonical binding and (C) EFH-1 of calpain PEF(S) demonstrating non-canonical calcium binding. All the binding residues are labelled [(B) and (C)], the water molecules are shown as a red spheres and the Ca^{2+} ions are shown as a grey sphere. (PDB: 1DVI)[74]*

PEF(L) domain

This calcium binding domain varies between the two isoforms of calpain. As noted earlier the concentration of calcium required to activate calpain-I is 7-fold lower than that required for the activation of calpain-II.[50] The sequence similarity between the PEF(L) domains found in human calpains-I and II is 48%.[48] Transformation of PEF(L) residues of calpain-I into the residues of those found in calpain-II and *vice versa* demonstrated that PEF(L) is important for the difference in the concentration of calcium that are required for the activation of each isoform.[59] There is no full crystal structure of calpain-I and therefore there

is no structure of PEF(L), so the discrepancy in the concentration required to activate both isoforms cannot yet be explained through structural analysis of PEF(L). The calcium binding residues in the PEF(L) domains of both isoforms are highly conserved, so the differing calcium concentrations required for activation of the isoforms is unlikely to be due to the ligands that coordinate the individual calcium ions (Table 1.2).[76]

EF-hand Motif	X	*	Y	*	Z	*	-Y	*	-X	*	*	-Z
EFH-1 (CAPN1)	L	**A**	G	D	**D**	M	**E**	I	S	V	F	**E**
EFH-1 (CAPN2)	L	**A**	G	E	**D**	A	**E**	I	S	A	K	**E**
EFH-2 (CAPN1)	**D**	R	**D**	G	**N**	G	**K**	L	G	L	V	**E**
EFH-2 (CAPN2)	**D**	E	**D**	G	**S**	G	**K**	L	G	L	K	**E**
EFH-3 (CAPN1)	**D**	L	**D**	K	**S**	G	**S**	M	S	A	Y	**E**
EFH-3 (CAPN2)	**D**	V	**D**	R	**S**	G	**T**	M	N	S	Y	**E**
EFH-4 (CAPN1)	Y	S	E	P	D	L	A	V	**D**	F	**D**	N
EFH-4 (CAPN2)	F	A	D	D	E	L	I	I	**D**	F	**D**	N

Table 1.2: *A table comparing all of the Ca²⁺ ion binding residues for each of the calcium binding EF-hands for both CAPN1 (proposed since there is no crystal structure) and CAPN2 (rattus novigenius).[74] The calcium binding residues are highlighted in bold.*

The PEF(L) domain of calpain-II comprises eight α-helices that form five individual EF-hands (Figure 1.9). Four of these loops bind calcium ions. Two bind calcium ions in the canonical EF-hand binding motif (EFH-2 and EFH-3) and two in the non-canonical formation (*vide supra*). Mutations to the individual EF hands of the large subunit of calpain-II found that the EFH-3 was vital for determining the concentration of calcium that is required for activation of the protease.[77] The order of affinity for calcium to the EFH motifs is as follows; EFH-3>EFH-2>EFH-1>>EFH-4.[77] Instead of binding calcium, EFH-5 is involved in formation of the heterodimeric structure of calpain-I and calpain-II. This is acheived through a so called EF-handshake with its respective partner in PEF(S) (*vide infra*).

There is only one published calcium-bound structure of full calpain-II and therefore PEF(L); this was achieved through the binding of the endogenous inhibitor calpastatin (see Section 1.7).[69, 78] A structure of heterodimeric calpain-II with calcium bound without calpastatin has yet to be obtained, as the protein has a tendency to aggregate in the presence of high concentrations of calcium.[78]

The EF handshake

The fifth EF-hand of both PEF(L) and PEF(S) forms an integral part of the heterodimeric interface.[79] The interacting residues that form this so called EF handshake are aromatic and hydrophobic residues (Figure 1.11).[53, 60, 76]
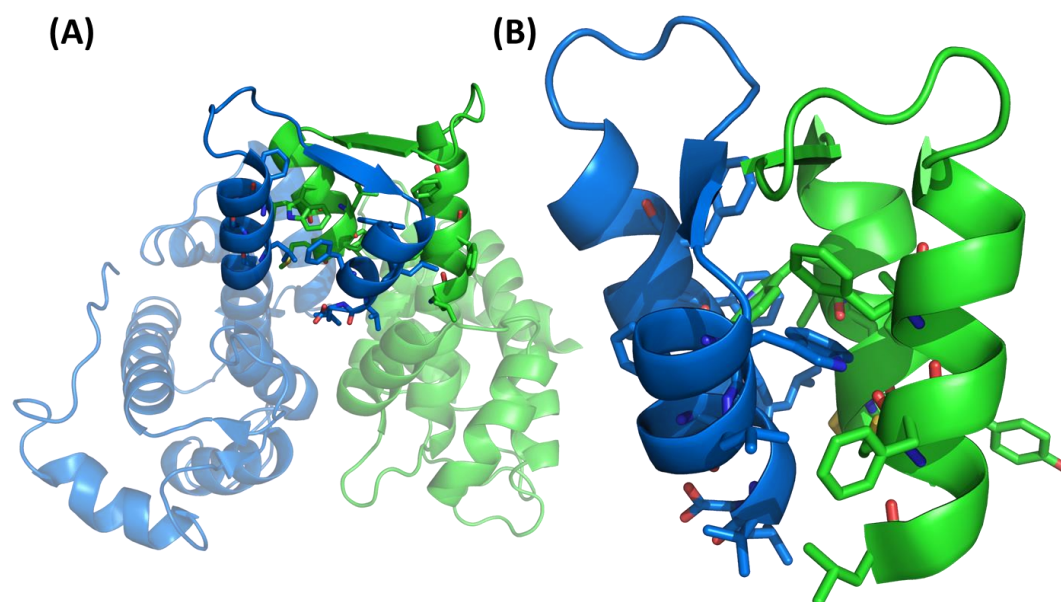
EFH-5 PEF(L) CAPN1	EFH-5 PEF(L) CAPN2	EFH-5 PEF(L) CAPNS1
Leu682	Leu669	Leu236
Met685	Leu672	Met239
Phe686	Phe673	Phe240
Phe689	Phe676	Phe243
Trp707	Trp694	Trp261
Leu708	Leu695	Leu262
Leu710	Phe697	Leu264
Met712	Val699	Met266
Phe713	Leu700	Tyr267

Figure 1.11: (A) A secondary structure representation of PEF(L) in blue, and PEF(S) in green, with EFH-5 highlighted, (B) a close-up of EFH-5 of both PEF(L) and PEF(S) with the hydrophobic residues represented as sticks (PDB:1KFU). A table showing the interacting residues on EFH-5 in both the PEF(L) domain and the PEF(S) domain which form the heterodimer of rat calpain.[50]

PEF(S) domain

PEF(S) is part of the small subunit, which is also known as the regulatory subunit and is identical in both calpain-I and calpain-II.[50] The regulatory subunit is so-called due to the fact that it forms a secondary layer of protection to prevent over-activation of the protease.[50] PEF(S) and PEF(L) are structurally

very similar. PEF(S) comprises five EF-hands four of which bind calcium and the fifth binds to the corresponding EFH-5 in PEF(L) (Figure 1.11).[53, 60] Of the four EF hands, that bind calcium ions, two coordinate calcium in a canonical fashion and the other two bind calcium in a non-canonical fashion (Table 1.3).[74] PEF(S) is also capable of forming a homodimer, where two molecules of PEF(S) interact with one another through EFH-5 (Figure 1.12).[80, 81]

EF-hand Motif	X	*	Y	*	Z	*	-Y	*	-X	*	*	-Z
EFH-1 (CAPNS1)	L	**A**	G	D	**D**	M	**E**	V	S	A	T	**E**
EFH-2 (CAPNS1)	**D**	S	**D**	T	**T**	G	**K**	L	G	F	E	**E**
EFH-3 (CAPNS1)	**D**	T	**D**	R	**S**	G	**T**	I	G	S	N	**E**
EFH-4 (CAPNS1)	Y	S	D	E	T	G	N	M	**D**	F	**D**	N

Table 1.3: *A table showing the residues in each of the EF hands of the calcium binding domain PEF(S).[74]*
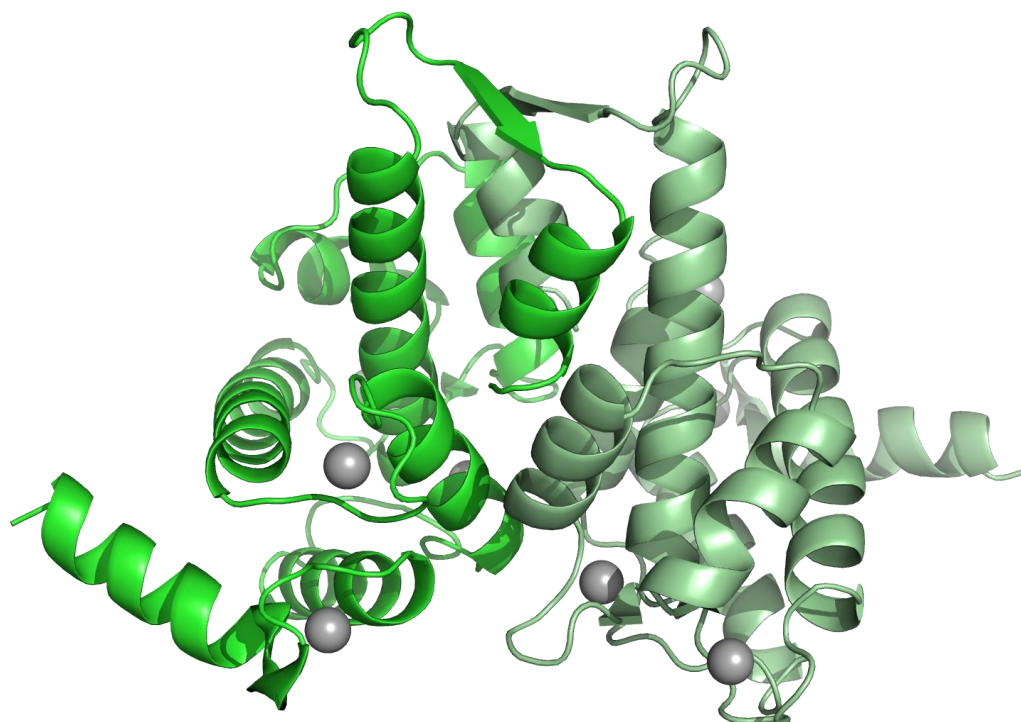
Figure 1.12: *A secondary structure representation of a crystal structure of PEF(S) demonstrating the formation of a homodimer. One monomer is shown in green and the other in pale green (PDB: 1DVI).*

The structure of PEF(S) crystals grown in the presence of varying concentrations of calcium ions revealed EFH-4 required 20 mM calcium chloride before a chelated ion was visible; 20 times more than EFH-1 to EFH-3.[74, 77] The higher concentration of calcium chloride that was required for

coordination to EFH-4, relative to other EF-hands, suggests that it is less important for activation of calpain-I and calpain-II.

It has been postulated that binding of Ca^{2+} to the heterodimeric complex causes dissociation of the two subunits, with subsequent homodimerisation of the small subunit.[78] Small r.m.s.d. values calculated between calcium bound PEF(S) and calcium-free PEF(S) crystal structures implied that calcium binding does not directly induce subunit separation.[74, 82]

1.6.5 The glycine rich domain

This 95 residue domain contains 40 glycines.[51] There is no structural data on this domain due to its flexible nature.[53] However there are two sequences that are postulated to form structured regions. The first is the sequence GTAMRILGGVI, which is postulated to form an oblique α-helix that interacts with phospholipids. The second is a series of five prolines; prolines are structurally constrained and therefore this sequence is highly rigid.[50, 83, 84] The flexibility of the domain in combination with the phospholipid binding α-helix suggests that this domain is used for the binding of the protease to the cell membrane.[50]

1.7 Calpastatin

Calpastatin is the endogenous inhibitor of both calpain-I and calpain-II and is composed of five domains totalling a molecular weight of 76,000.[85] The N-terminal domain is known as the L domain and is separate from the four individual inhibitory domains, each of which binds a single molecule of the heterodimeric calpain (Figure 1.13).[56, 85, 86] The inhibitory domains are numbered I-IV from the L-domain.[87] Each domain differs in affinity for calpain with the order being: I>IV>III>II, the K_d values of the domains range from 4.5 pM to 4.0 nM.[87]
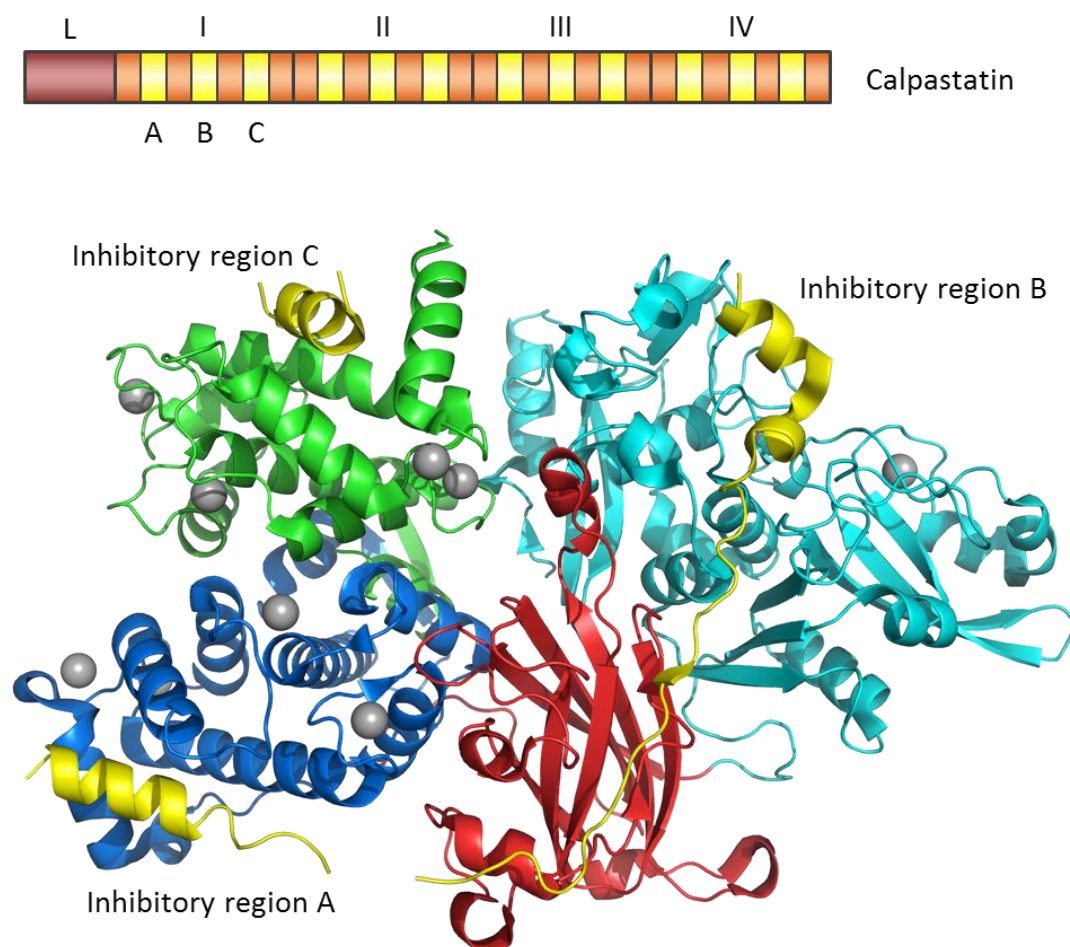
Figure 1.13: *(A) A diagram representing the inhibitory domains of calpastatin; domain L and domains I to IV. The inhibitory regions, A, B and C are highlighted in yellow within each of those domains. (B) A secondary structure representation of inhibitory regions A, B and C bound to the heterodimeric structure of calpain-II; with CysPc in cyan, C2L in red, PEF(L) in blue and PEF(S) in green (PDB:3BOW).[69]*

The inhibitory domains are all approximately 140 amino acids in length and are constructed of three inhibitory regions that bind to different domains of calpain. The inhibitory regions are known as regions A, B and C (Figure 1.13).[86] Regions A and C are α-helices that bind to hydrophobic grooves in PEF(L) and PEF(S), respectively.[69] The hydrophobic grooves that bind regions A and C are formed between α-helix 1 and 2 of EFH-1 and α-helix 4 of EFH-2 in both PEF(L) and PEF(S).[69] Inhibitory region B is inherently unstructured and initially interacts with the C2L domain, the poly peptide then runs through the narrow active site in the CysPc domain, from PC2 to PC1. Region B terminates in an α-helix that interacts with a hydrophobic groove of PC1 (Figure 1.13).[69] The structure of calpastatin domain IV bound to calpain-II shows several coordinated calcium ions, implying that the endogenous inhibitor is used to sequester and inactivate

calpain following activation of the protease, preventing any further proteolysis.[69]

Many peptide inhibitors have been produced based on the sequence of inhibitory region B of calpastatin and were found to be capable of inhibiting both calpain-I and calpain-II.[88] In stark contrast peptides derived from regions A and C, without inhibitory region B, have been found to lower the concentration of calcium required for the activation of calpain-I and calpain-II.[89]

1.8 Activation of calpain-I and calpain-II

Calcium binds to calpain-I and calpain-II *in vitro* with concentrations of ~50 µM and ~0.35 mM respectively. This leads to structural alterations within the protease that enable the enzyme to reach its active form.[50] The only calcium bound X-ray crystal structure of calpain is of calpain-II bound to one domain of calpastatin.[69, 78] Hence a variety of alternative techniques have been used to indirectly study the conformational changes that occur during calcium mediated activation. These include mutation of residues within the enzyme and the solution of the structure of isolated calcium binding domains to determine the structural effect upon binding calcium.[65, 70]

The penta-EF hand domains, PEF(L) and PEF(S), are thought to have more of a regulatory role than playing a major part in the activation of calpain upon binding calcium ions. It has been shown that there is very little movement in the overall structure of these domains when calcium is present.[74, 81] However, both of these domains are important for the determination of the concentration of calcium that is required in the activation of both calpain-I and calpain-II.[59, 77] The greatest movement that was visualised in these domains was demonstrated at EFH-1 of PEF(S). This same movement may therefore occur with EFH-1 of PEF(L). Upon binding calcium to the PEF(L) domain, the polypeptide linker between the first α-helix of EFH-1 and the C2L domain is thought to extend and enable the disruption of the electrostatic interactions of the C2L domain and the PC2 domain.[74, 90, 91] There is an important salt bridge interaction between the residues Glu504 and Lys234 and mutations at this point dramatically decrease the concentration of calcium that is required for activation.[64, 70] Other

interactions between the domain interface of PEF(L) and C2L are also thought to be affected through calcium binding in the PEF(L) domain.[51] Upon the binding of calcium to EFH2 of the PEF(S) domain repulsive interactions with the N-terminal anchor α-helix are presumed to occur releasing the constraints that hold the active site subdomains apart.[62] The negatively charged loop of the C2L domain that interacts with the PC2 subdomain is thought to bind calcium, like other known C2 domains, disrupting the electrostatic interactions between these domains.[66, 67] However the crystal structure of the calcium bound calpain/calpastatin system is in conflict with this due to the absence of Ca^{2+} observed within this domain.[69]

The disruption of the electrostatic interactions between the proteolytic core and the auxiliary domains frees the restraints holding the PC1 and PC2 subdomains apart, alignment of the catalytic triad occurs after PC1 and PC2 bind calcium themselves.[92] The two subdomains contain a non-EF hand calcium binding site each, this was observed through the crystallisation of the proteolytic core of calpain-I without any other domains in the presence of calcium ions (Figure 1.14).[65] Negatively charged residues in the two loops and water molecules, one in PC1 and the other in PC2, are used to bind the calcium.[65, 92] The calcium binding to the two loops in the subdomains allows for the two halves of the CysPc domain to rotate approximately 25° into the active conformation.

In calpain-I the CysPc domain alone is also a calcium activated cysteine protease with a similar activation concentration as the parent molecule at ~40 μM .[65] In contrast the CysPc domain for calpain-II alone was found to be either only weakly active[93] or completely inactive.[94] In the former case the specific activity of the proteolytic core of calpain-I alone only reached 10% of the parent molecule, demonstrating the proteolytic core requires the other domains for the maximum activity of the enzyme.[65]
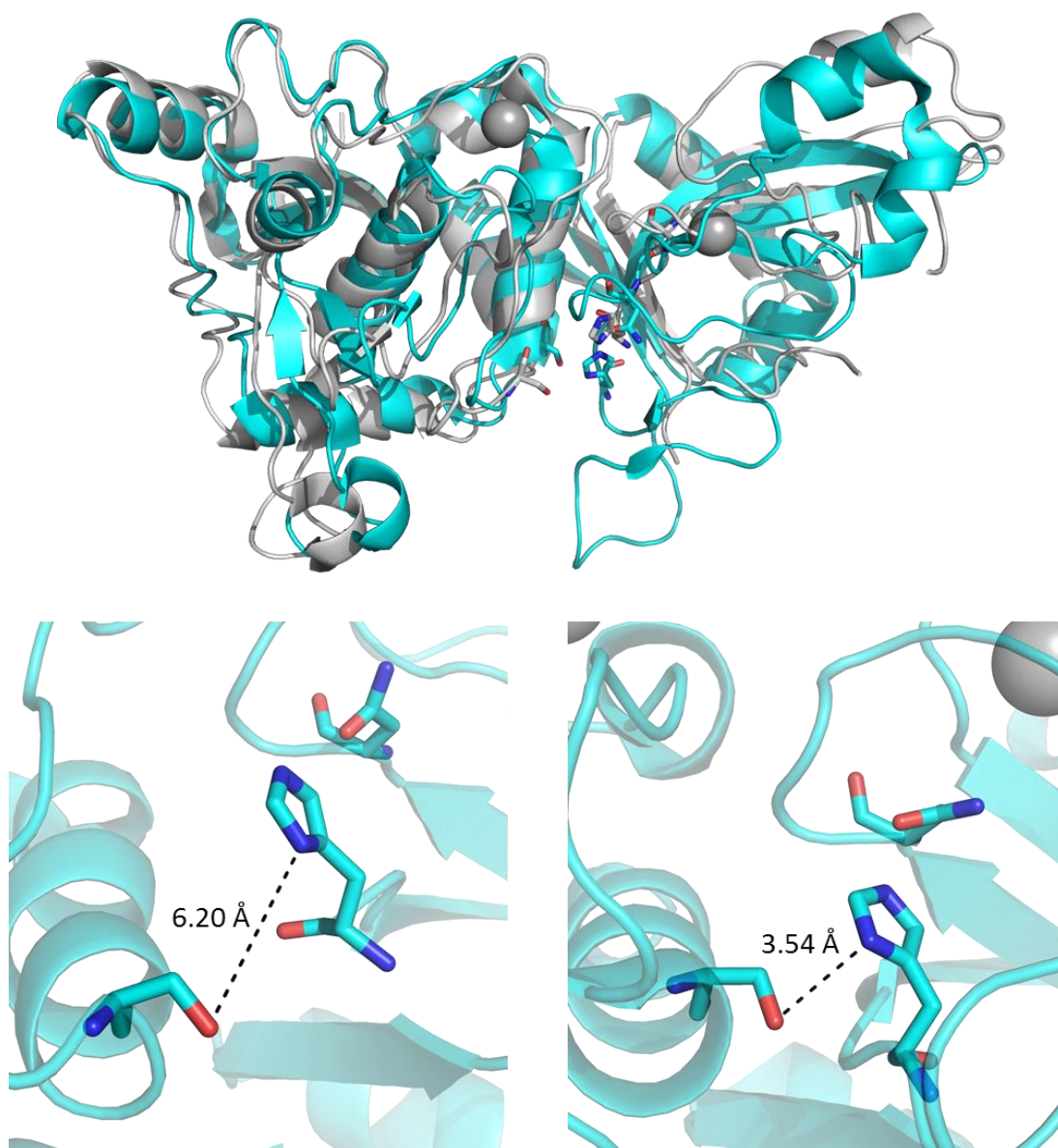
Figure 1.14: *A secondary structure representation the inactive CysPc domain of calpain-I, derived from a hybrid structure of calpain-I and calpain-II (grey),[57] overlaid with the calcium bound CysPc domain which is in the active conformation (cyan).[65] A close up of the catalytic triad, with the cysteine mutated to a serine, the inactive protease (left) and the active protease (right).*

Therefore in-order to activate calpain with calcium a two-step activation route has been proposed. The initial step requires a release of the restraints that hold the proteolytic core in an inactive conformation through the binding of calcium to the PEF domains on both subunits. Once calcium is bound the electrostatic interactions between the C2L domain and the CysPc domain are disrupted, as well as the interactions between the anchor helix and PEF(S). The subsequent step is the binding of calcium to the proteolytic core to allow PC1 and PC2 to

rotate towards one another forming the catalytic triad and proteolysis may then occur (Figure 1.15).
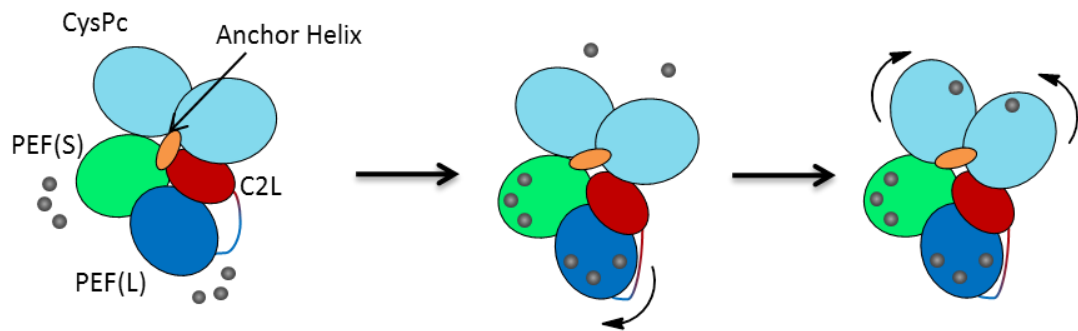
Figure 1.15: *A schematic representation of how calcium binds to calpain through a postulated two-step process; calcium initially binds to the PEF domains releasing the anchor helix and the electrostatic interactions that hold PC1 and PC2 in an inactive conformation, calcium then binds to PC1 and PC2 causing an alignment of the proteolytic core.*

1.8.1   Factors that lead to activation *in vivo*

The *in vitro* concentrations of calcium required for the activation of the two isoforms is much greater than the cellular concentrations which reach a maximum of 1 μM.[50, 95] Though there are certain local high concentrations of calcium, such as just beneath the membrane in neutrophils, where the concentration can reach up to ∼30 μM, this remains too low for the activation of either calpain isoform.[31] Therefore, to lower the concentration required for the activation of calpain-I and calpain-II other factors must be involved. These include the binding of phospholipids,[96, 97] autolysis, the binding of activator proteins,[55, 98, 99] and phosphorylation.[100]

Phospholipids have been observed to bind to the C2L and GR domains of calpain-I or calpain-II.[67, 83, 84] This decreases the concentration of calcium required for activation implying that the binding to phospholipids may enable activation *in vivo*.[96, 97] The ability of phospholipids to activate calpain-I and calpain-II suggests that these proteins are mainly active at the cell membrane.[101, 102] The binding of phospholipids has also been found to lower the concentration of calcium required for autolysis *via* an intramolecular reaction.[103] The autolysis in turn allows for a lowering in the required concentration for activation *in vitro* and therefore potentially *in vivo* (Figure 1.16).[96, 103] Autolysis of the large subunit of calpain-II occurs at the N-terminus

leaving a molecular weight of 78,000, from 80,000, this happens under low concentrations of calcium, leaving a less stable but readily active protease.[104] Further autolysis of calpain-II leads to an even less stable protease where the large subunit has a molecular weight of 76,000 which can also be activated at lower concentrations of calcium. Due to the instability of the structure of autolysed calpain-II, the fragments of the protease can be difficult to detect *in vivo*.[104] Therefore it is difficult to establish whether autolysis is a precursor step to the full activation of calpain.[104] Calpain-I has also been found to undergo autolysis which results in a reduction of the concentration of calcium required for the activation.[105] Phospholipid binding is contradicted by Cong *et al.* however, who state that neither calpain-I or calpain-II undergo autolysis through the binding of phosphatidylinositol-1,3-bisphosphate.[106] The concentration of calcium required for autolysis to occur, both with and without phosphatidylinositol-1,3-bisphosphate, was higher than the known *in vitro* activation concentrations for both isoforms.[106] Therefore they state that there must be another mechanism utilised for activation.[106]
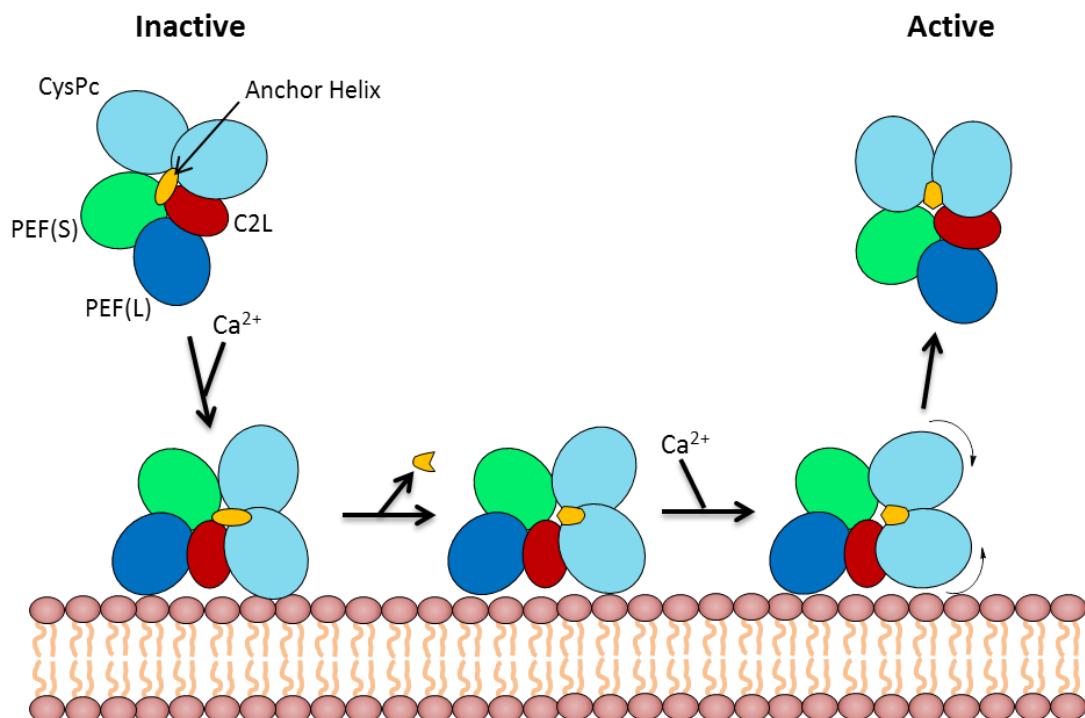
Figure 1.16: *A proposed activation mechanism of calpain-I/II via membrane association (phospholipid binding) and the subsequent autolysis.[101, 103] Initially the inactive protease binds to the cell membrane, followed by the autolysis of the anchor helix. This lowers the calcium concentration required for activation and following this the active site moves together and the protease is release from the membrane in its activated form.*

Another postulate is that association of activator proteins may lower the concentration of calcium that is required for the activation of both calpain-I and calpain-II (Figure 1.17). Acyl-CoA binding protein was found to decrease the concentration of Ca^{2+} ions that are required by over 50 fold.[55] Another protein activator that was proposed by the same group is UK114, a tumour antigen from goat liver, that decreased the activation concentration required by over 10 fold in the case of calpain-I.[98] A second activator protein, of otherwise unknown function, for calpain-I was also found in human neutrophils.[99] UK114 from *Drosophila* was used as an activator of calpain by Farkas A. *et al.* but they found that it did not act as an activator protein at all and was found to act merely as a molecular chaperone for other proteins such as citrate synthase.[107]
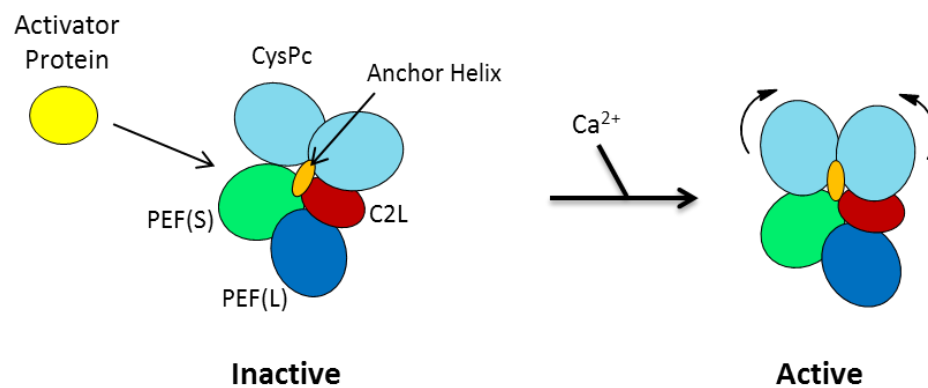
Figure 1.17: Another theory; an activator protein is associated with calpain to lower the concentration of Ca^{2+} ions required for activation.[55, 98, 108] The activator protein binds to the protease, lowering the concentration of calcium that is required for activation. Calcium then binds to the protease allowing the active site to close into the active form of the protease.

1.9 Physiological function and known substrates

Calpain-I and calpain-II are ubiquitously expressed throughout the human body and have been linked to many physiological roles. Determining the exact physiological role of these proteases has been difficult due to the way that they are analysed, mainly through use of inhibitors.[109] Caution must be used with conclusions drawn through the use of inhibitors however, due to the lack of specificity of the inhibitors produced to date. The only calpain-selective inhibitor, calpastatin, is not cell permeable.[109] There are many substrates known to be cleaved by calpain-I and calpain-II *in vitro*, these can be used to predict the function of the protease *in vivo*.

1.9.1   Known substrates of calpain-I and calpain-II

Calpain-I and calpain-II have highly similar substrate specificity and over 100 substrates have been discovered to date through *in vitro* studies.[50, 109] These substrates include cytoskeletal proteins, several kinases and phosphatases, some membrane bound proteins and a few cytosolic proteins.[50, 109, 110] Several of the substrates are cytoskeletal proteins that are mainly found in the region of the cellular membrane, examples of these include ezrin, talin, α-fodrin.[111-113] Calmodulin dependent proteins either membrane bound or in the cytosol are susceptible to calpain mediated hydrolysis, for example Ca^{2+}-ATPase and the phosphatase calcineurin.[110, 114, 115] Other cytosolic enzymes that get hydrolysed by calpain-I include caspase 12 during apoptosis.[116]

Calpain demonstrates a preference for certain amino acids at specific positions within the active site.[109] Through sequencing analysis of numerous substrates of calpain the most common of the amino acids that occur at position P1 and P1', either side of the cleavage site, are tyrosine, lysine or an arginine at P1 and serine at P1' (Table 1.4).[117] However synthetic peptide library screening found that the optimum amino acids were either leucine or phenylalanine at P1 and at P1' a methionine, alanine or an arginine (Table 1.4).[118] This does not match any protein sequences of the known substrates of calpain to date.[119] Proline residues are preferred at certain points within the sequence of the substrate (e.g. at P3, P2' and P3'), this implies that there is a lack of secondary structure at the cleavage site.[117] The lack of secondary structure may be important for the recognition of the substrate by the protease.[117] This was corroborated with allignment of the structures of papain and calpain-I active sites with the inhibitor leupeptin bound (Section 1.11.2).[120] This revealed that the inhibitor resides deeper within the active site of calpain-I than papain, with the calcium binding loops surrounding the active site.[120] The narrow active site of calpain suggests that only a disordered secondary structure of the substrate is capable of accessing the active site.[120]
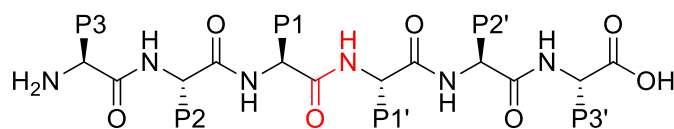
Sequence	P3	P2	P1	P1'	P2'	P3'
A	Trp>Pro	Leu>Thr >Val	Lys>Tyr >Arg	Ser	Pro	Pro>Trp
B	Phe>Leu >Pro	Leu>Val	Leu=Phe	Met>Ala >Arg	Glu	Arg>Lys

Table 1.4: *The optimum sequences of amino acids at the cleavage site for calpain-I the cleavage site is shown in red, sequence A was derived from the sequences of 49 known substrates,[117] sequence B was derived from a peptide library screen.[118]*

1.9.2   Physiological function

Through the specific amino acids that are required for substrate activity, numerous substrates have been identified *in vitro* and can be used to imply roles for the protease *in vivo*. However, caution must be used as the environment within cells greatly differs to the conditions used *in vitro*, for example the concentration of calcium that is used for the activation. Both calpain-I and calpain-II have a variety of postulated roles within the human body including involvement in signal transduction,[121] the cell cycle,[109, 122] platelet function,[123] prevention of erythrocyte deformability,[124] apoptosis[125] and cell spreading and migration.[126]

The cell cycle

Both isoforms of calpain have been shown to affect progression through the cell cycle.[122] Microinjections of calpain-II at different phases of the cell cycle and especially during mitosis revealed that the cell progresses rapidly through the phases of mitosis.[127] When calpain-II was added to cells whilst in the metaphase the cells proceeded to the anaphase in a reduced time, and similarly it could accelerate the progression from the prophase to the interphase.[127] Inhibition of a calpain-like protease in fibroblasts has been shown to stop the growth of these cells at the G_1 stage of growth and therefore there was no progression to the S phase.[128]

Platelet function

Platelet agonists cause an increase in the concentration of intracellular calcium, leading to cytoskeletal remodelling and the secretion of α-granules for membrane fusion.[129, 130] The increase in calcium levels in platelets leads to the activation of calpains for the hydrolysis of cytoskeletal proteins such as fodrin.[131] The inhibition of calpain in platelets, with a cell permeable calpastatin derived peptide, demonstrated that the protease has an essential role with secretion, aggregation and the spreading of platelets.[132] Knockout of the gene encoding for calpain-I reveal that this isoform is important for platelet secretion and aggregation.[133]

There is a disagreement as to the expression levels of both these proteases in platelets. Studies have found that calpain-II makes up approximately 2% of all the protein within platelets, distributed evenly within the cytosol and at the membrane of platelets.[134] This is disputed, however, with claims that calpain-I is the dominant protease and not calpain-II.[135, 136]

Apoptosis

Programmed cell death involves both calpain-I and calpain-II. In the earlier stages of apoptosis calpain-I is activated through the release of calcium from the endoplasmic reticulum.[137, 138] The activated protease hydrolyses Bid, a proapoptotic protein.[139] This truncated Bid (tBid) is now active and translocates to the mitochondria allowing for oligomerisation with the fellow pro-apoptotic proteins, Bak and Bax.[138] These oligomers form pores in the mitochondrial membrane releasing cytochrome c thereby continuing the mitochondrial mediated apoptosis cascade.[138]

Caspases are a well-known group of cysteine proteases that are intrinsically linked to the apoptosis pathway. Calpains are linked to the cleavage and activation of these enzymes during apoptosis, caspase 12 is just one example.[116] Calpain-I and calpain-II have also been found within mitochondria and have been associated with the latter stages of apoptosis, through the truncation and transportation of apoptosis inducing factor from the mitochondria.[140-142] Joshi

et al. have, however, contradicted this proposed physiological role of calpain-I as they did not observe this phenomenon.[143]

Cell spreading and migration

Calpain-I appears to play an integral role in the ability of cell membranes to spread and hence enable the migration of neutrophils (Section 1.2).[23] The two isoforms of calpain are important for the motility of other cell types also.[144, 145] Many of the putative substrates of calpain are located at the cell membrane and form focal adhesion complexes and are cleaved *in vivo*.[146] These proteins include ezrin,[34] talin[33] and β$_2$-integrins.[147, 148] Integrin complexes comprise many proteins that are membrane bound proteins that span the plasma membrane and protrude out into the extracellular space. These integrins allow for the adhesion of the cell to the endothelial wall, beginning the process of migration to the target site of action.

Calpains are implicated in the cell spreading capability of T-cells and fibroblasts. Calpain-II is a vital component in the ability of the cell to form these integrin interaction complexes and therefore in the capability of the cell to spread.[145, 149] Caution must be used in this interpretation as inhibitors were used to observe this process and there is the poor specificity of most inhibitors towards individual calpain isoforms.[149] A decrease in the activity of calpain-II through the over expression of calpastatin suppressed the cell spreading capability of fibroblasts, as the cells were not capable of forming protrusions of the cell membrane.[150, 151]

1.10 Calpain related diseases

Calpain-I and calpain-II are both linked to many diseases, through either the disruption of the calpain-calpastatin system or the over-activation of calpain. As well as rheumatoid arthritis (Section 1.1) both isoforms are linked to Alzheimer's disease,[152, 153] ischemic cell death,[154] cancer,[155] cataracts,[156] multiple sclerosis[157] and malaria.[158]

1.10.1 Alzheimer's disease

Alzheimer's disease is the death of neurons within the brain due to aggregation of β-amyloid particles and tau protein to form β-amyloid plaques and neuronal filaments respectively. [159, 160] Cellular extracts from neurons of Alzheimer's patients when compared to healthy brain neurons, post-mortem, indicated an unusually high activity of calpain-I.[152] This increase in calpain activity may lead to damage of the cellular membranes since calpain-I is known to cleave cytoskeletal proteins, and induction of neuronal apoptosis through the cleavage of proapoptotic protein p53.[101] However, in contrast to this, a decrease in calpain-I activity is linked to the formation of β amyloid plaques, due to the lack of proteolysis of the precursor protein α-secretase.[161] Also the over-expression of calpain-II has been detected in the brains of patients suffering from Alzheimer's disease, through the use of a calpain-II specific antigen.[162, 163]

1.10.2 Ischemic cell death

During ischemic associated cell death which generally occurs as a consequence of strokes, cardiac arrest, or through trauma, there is a rapid increase in calcium concentration within the cells.[164] The increase in calcium levels leads to the over activation of calpain and in turn, to necrotic cell death. The rapid increase in calcium levels is generally thought to be caused, within the brain, by the excessive influx of the major excitatory neurotransmitter glutamate.[165, 166] When glutamate reaches these excessive levels it is considered a neurotoxin. It causes the depolarisation of the membrane, causing sodium and calcium ions to flood into the cytosol and consequently activating a variety of proteins, such as calpain.[165] The activation of calpains within the cytosol then leads to neuronal apoptosis and the degradation of cytoskeletal proteins.[159, 165] With cardiac ischemia the over-activation of calpain-I has been linked to the release of the apoptosis inducing factor from the mitochondria within these cells, leading to cell death.[167] The degradation of cytoskeletal proteins may also enable cell death.[165]

1.10.3 Calpain and cancer

A link between cancer and calpain has been established through the detection of an over expression of both calpain-I and calpain-II in a variety of cancers.[168-170] Also a decrease in the expression of the endogenous inhibitor calpastatin occurs in cancer cells found within endometrial cancer.[155] The over activation of calpain is considered to play a major role in the metastasis of cancer.[155, 168] The down regulation of calpain-II, as well as the use of inhibitors, has demonstrated that the invasiveness of prostate cancer cells can be significantly reduced.[169]

Due to the numerous diseases that both of these isoforms of calpain have been linked to, these are attractive targets for the development of new inhibitors. The knockout of calpain-I within cells, and therefore the loss of calpain activity, demonstrated a deformity within erythrocytes[124] and platelets,[133] whereas the knockout of calpain-II is embryonically lethal.[171] Therefore the development of inhibitors that are selective for calpain-I is desirable.

1.11 Inhibition of calpain

The optimum selective inhibitor for calpain-I and calpain-II is the endogenous inhibitor calpastatin (Section 1.7). This inhibitor is both potent and selective towards calpain-I and calpain-II. The main problem in the use of this inhibitor *in vivo* is that due to its size it is not cell permeable. Additionally it binds to both of the isoforms and the total inhibition of calpain-II is detrimental to life.[171] Therefore efforts have been made to develop inhibitors that are potent, cell permeable and selective between these two calpain isoforms.[172, 173]

1.11.1 Calpastatin based inhibitors

Inhibitors have been developed based on the sequence of calpastatin inhibitory region B, which binds to the active site of calpain-I and calpain-II.[173] CP1B is a 27 residue peptide based on inhibitory region B of calpastatin and inhibited both isoforms of calpain with a K_i of ~0.2 nM, though the protein was incapable of penetrating the cell membrane (Table 1.5).[88] The cell penetrating peptide penetratin was attached to CP1B, to make PCP1B, this did not affect the

apparent potency of the inhibitor towards calpain (K_i = ~0.5 nM) (Table 1.5).[174] Truncation of penetratin to an 8 residue peptide (from a 17 residue peptide) did not affect the cell penetrating ability of CP1B but increased the potency with a K_i of 0.18 nM (Table 1.5).[175] Since these inhibitors were based on calpastatin, they were not selective between the isoforms.

Inhibitor	Sequence	K_i (nM)
CP1B[88]	Ac-DPMSSTYIEELGKREVTIPPKYRELLA-NH$_2$	0.20
PCP1B[174]	Ac-CDPMSSTYIEELGKREVTIPPKYRELLA-NH$_2$ | Ac-CRQIKIWFQNRRMKWKK-NH$_2$	0.50
8-mer-PCP1B[175]	Ac-CDPMSSTYIEELGKREVTIPPKYRELLA-NH$_2$ | Ac-CRRMKWKK-NH$_2$	0.18

Table 1.5: *The sequences of inhibitors derived from calpastatin inhibitory region B with their reported K_i values.*

1.11.2 Inhibitors containing warheads

Inhibitors that contain so-called warheads attack the cysteine within the active site meaning that the inhibitor is likely to bind to many other cysteine proteases since they have similar acitive sites. Inhibiting several different cysteine proteases would be to the detriment of normal body function. Functional groups commonly found in these warheads are epoxides, aldehydes and hemiacetals.[172] The most well known inhibitors are E-64 (**1**),[176] leupeptin (**3**) , calpeptin (**4**) and ALLN (**5**).

E-64 (**1**) contains an epoxide based warhead and is known to irreversibly alkylate the catalytic cysteine residue (Figure 1.19).[120] The peptide is non-selective towards calpains (Figure 1.18).[177] E64 inhibits calpain-I and calpain-II with IC$_{50}$ values of 1.53 μM and 1.09 μM respectively.[178] A derivative of this inhibitor has been developed, WRH(R,R) (**2**), which has greater selectivity towards calpains over other cysteine proteases, such as cathepsins (Figure 1.18). Compound **2** inhibits calpain-I and calpain-II with IC$_{50}$ values of 0.46 μM and 0.10 μM respectively.[179]
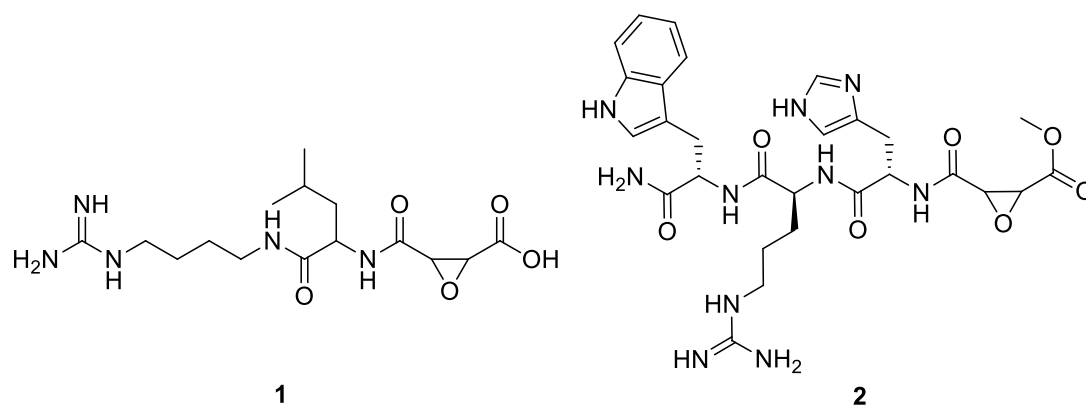
Mechanism
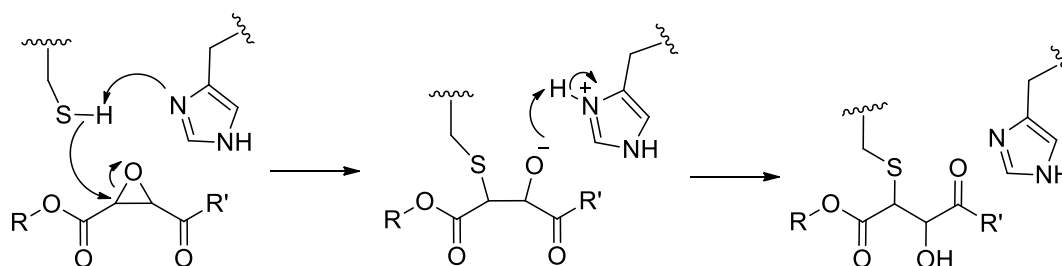
Figure 1.18: *Peptide based warhead inhibitors with epoxide functional groups that alkylate the cysteine. (**1**) E64; IC$_{50}$ = 1.53 µM (calpain-I) and 1.09 µM (calpain-II), (**2**) WRH(R,R); IC$_{50}$ = 0.46 µM (calpain-I) and 0.10 µM (calpain-II).[178, 179] The general mechanism of epoxide based inhibition of the cysteine protease.*
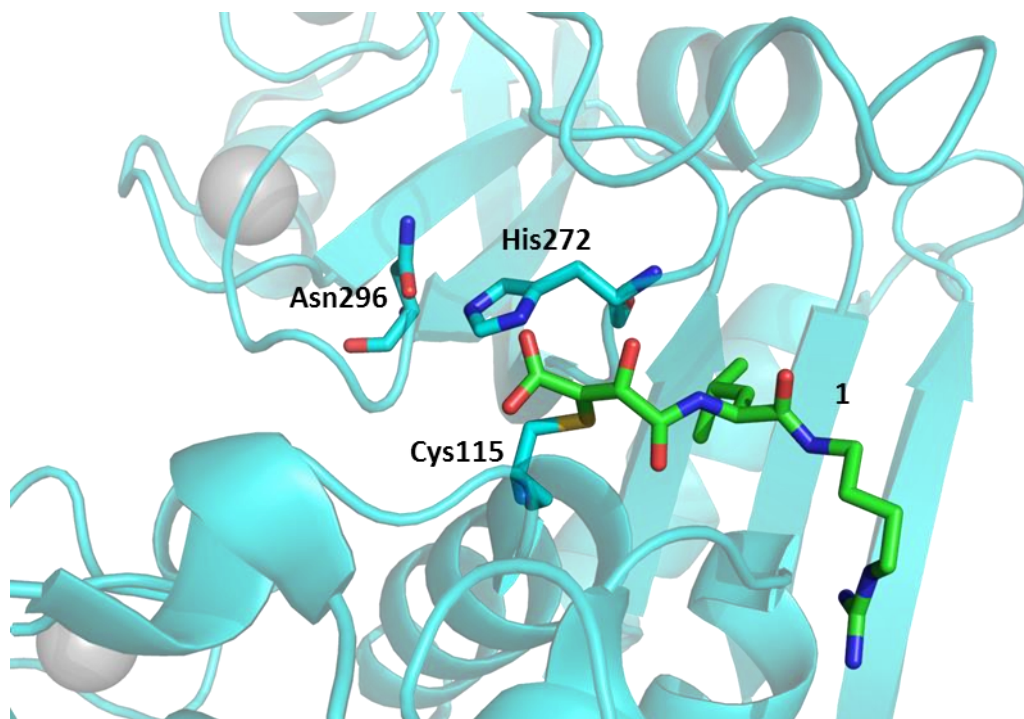
Figure 1.19: *A secondary structure representation of the CysPc domain of calpain-I, the catalytic triad is represented as sticks with the cys115 having been alkylated with **1** (PDB:1TLO).[120]*

Inhibitors that contain an aldehyde warhead attack the catalytic cysteine residue but are fully reversible inhibitors. Leupeptin (**3**) (Figure 1.21), an example of this family, had been found in various *Streptomyces* species and is highly potent towards calpain-I and calpain-II with IC_{50} values of 0.27 µM and 0.38 µM respectively (Figure 1.21).[172, 178] Leupeptin is unspecific, attacking other cysteine proteases as well as some serine proteases (Figure 1.20).[180] Derivatives of **3** such as calpeptin (**4**) have been produced, that possess the ability to pass through the cell membrane, as well as being highly potent towards calpain-I and calpain-II (Figure 1.21).[178, 181, 182] Calpeptin inhibits calpain-I and calpain-II with IC_{50} values of 0.010 µM and 0.014 µM respectively.[178] ALLN (**5**) and ALLM (**6**), or calpain inhibitor I and calpain inhibitor II as they are otherwise known, are two other peptide inhibitors that contain aldehyde warheads (Figure 1.21). These are synthetic inhibitors and both are cell permeable; they inhibit calpain-I with IC_{50} values of 0.023 µM and 0.13 µM for and calpain-II with IC_{50} values of 0.022 µM and 0.23 µM respectively. However both were found to be more potent inhibitors of cathepsins than of calpains.[178]
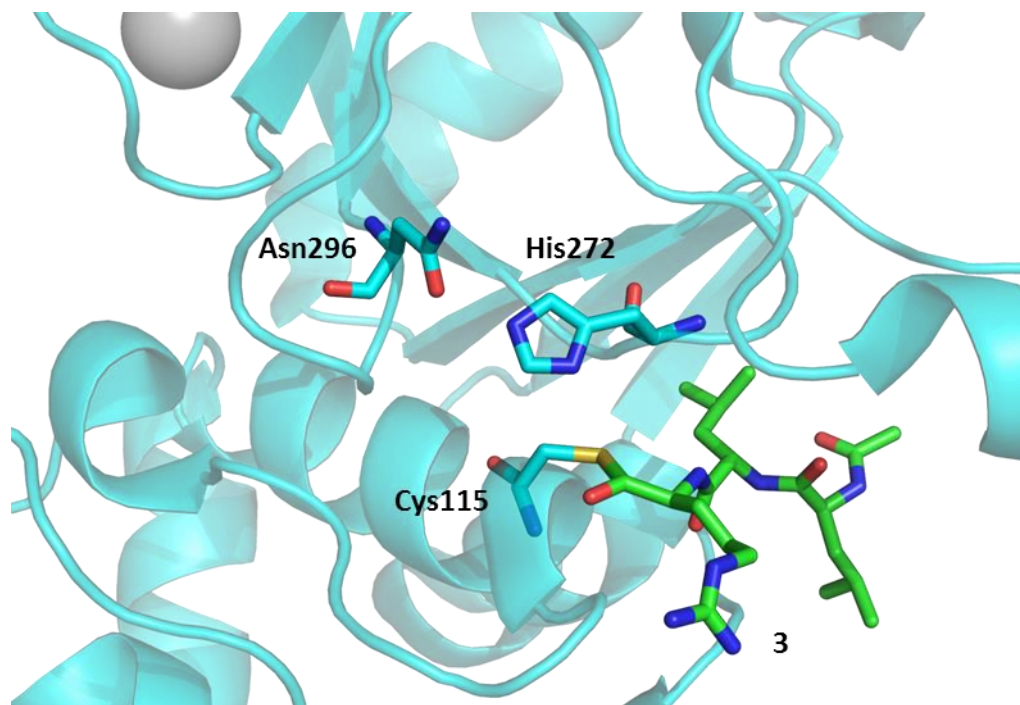
Figure 1.20: *A secondary structure representation of the CysPc domain of calpain-I, the catalytic triad is represented as sticks with the aldehyde based inhibitor **3** alkylating the proteolytic cysteine (PDB:1TLO).[120]*
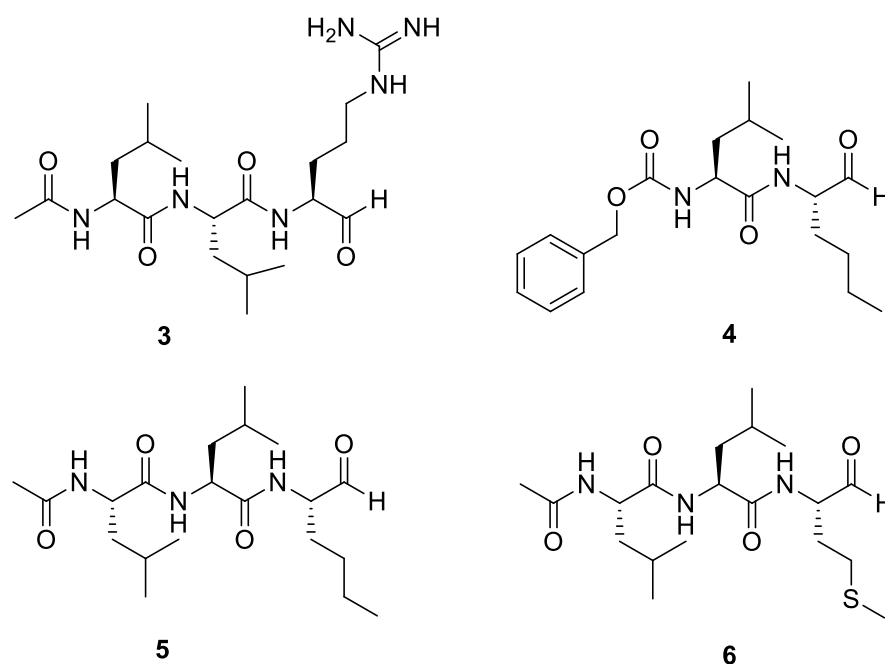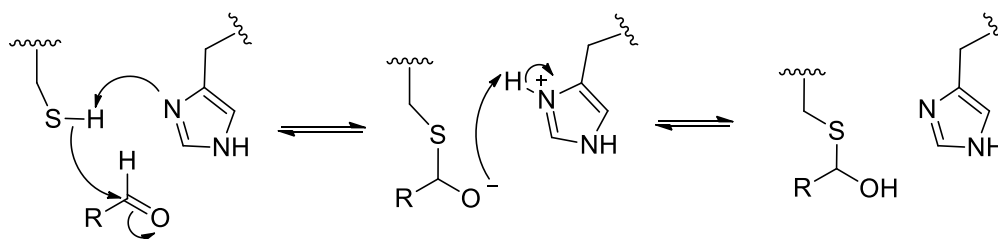
Figure 1.21: *The structures of peptide based inhibitors that contain aldehyde warheads; (3) leupeptin; IC_{50} = 0.27 μM (calpain-I) and 0.38 μM (calpain-II),[178] (4) calpeptin; IC_{50} = 0.010 μM (calpain-I) and 0.014 μM (calpain-II),[178] (5) ALLN; IC_{50} = 0.023 μM (calpain-I) and 0.022 μM (calpain-II),[172] and (6) ALLM; IC_{50} = 0.13 μM (calpain-I) and 0.23 μM (calpain-II),.[172] The mechanism of action.*

1.11.3 Inhibitors without warheads

Ideally, calpain-I selective inhibitors should not contain warheads which alkylate the active site cysteine residue, due to the similar substrate specificities of both isoforms. Numerous non-warhead based inhibitors have been discovered from natural sources as well as through rational design. A synthetic peptide based compound **7** has been reported to be the most potent inhibitor to date, with an IC_{50} value for calpain-I at 87 pM (Figure 1.22).[183] Compound **7** does not interact with other cysteine proteases such as papain or cathepsins, although no data has been recorded for its interaction with calpain-II.[183] This

compound contains a biphenyl linker between two peptide chains, it is thought that it prevents calpain activation through the chelation of calcium ions.[183]
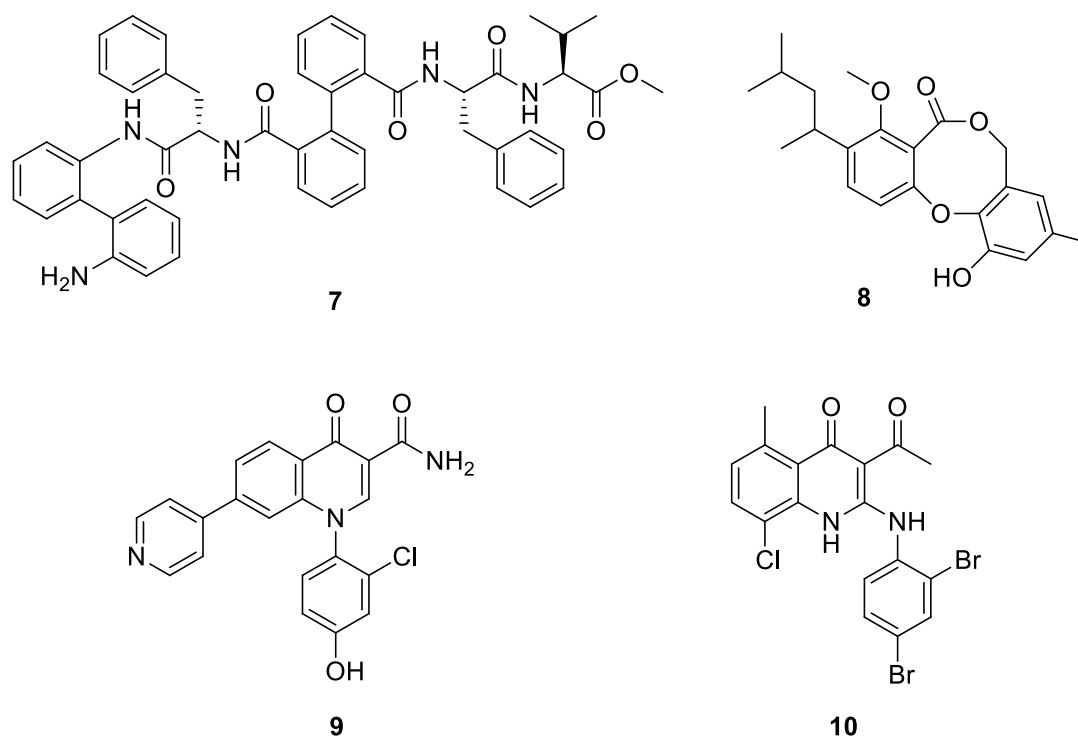
Figure 1.22: *The structures of non-warhead based inhibitors; (**7**) a biphenyl derivative that is one of the most potent inhibitor of calpain-I to date (IC$_{50}$ = 87 pM)[183], (**8**) penicillide (IC$_{50}$ = 7.1 μM) (calpain-II), (**9**) quinolinecorboxamide derivative (IC$_{50}$ = 0.50 μM) and (**10**) another quinolinecarboxamide derivative (IC$_{50}$ = 0.28 μM).*

Other non-peptide based inhibitors have also been found to inhibit calpain such as penicillide (**8**), a polyketide that was found in *Penicillium sp.* It has activity against calpain-II with an IC$_{50}$ of 7.1 μM (Figure 1.22).[172] Quinoline derivatives have also shown inhibitory properties towards calpain-I, with derivatives of 3-quinolinecarboxamide **9** and **10** with IC$_{50}$ values of 0.5 μM and 0.28 μM respectively.[184] Docking studies suggested that the molecule binds within the active site of calpain-I.[185] These compounds inhibit calpain-I, but no data on inhibition of calpain-II has been published. However there are two small molecules that have shown a modest selectivity towards calpain-I over calpain-II (Section 1.11.4).

1.11.4 Small molecule inhibitors
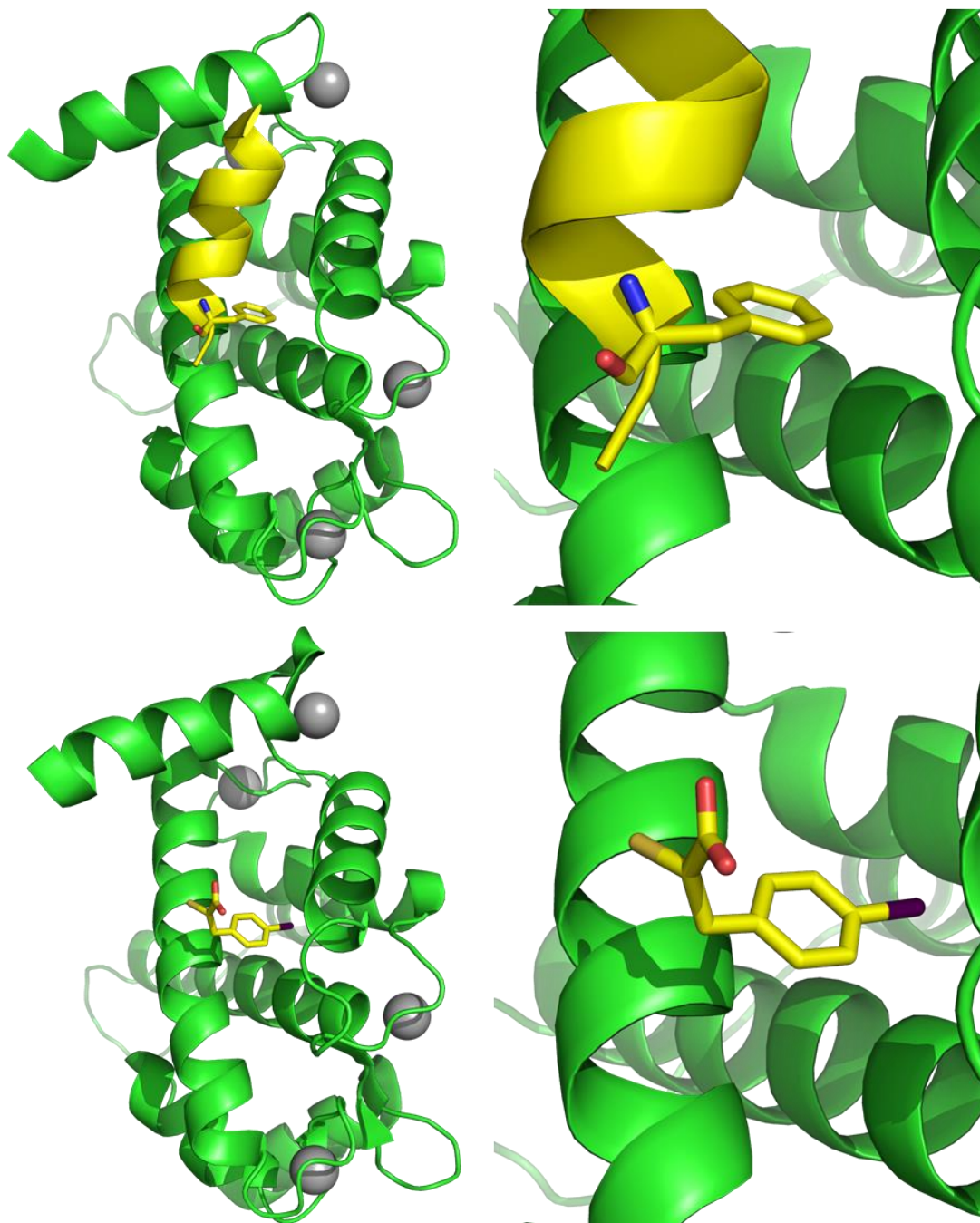
Figure 1.23: *A secondary structure representation of calpastatin inhibitory region C bound to porcine PEF(S) (PDB:1NX1). A secondary structure representation of PD150606 bound to the same region of porcine PEF(S) (PDB:1NX3).[186]*

Two small molecule inhibitors were developed in 1996 that displayed selectivity towards calpain-I over calpain-II.[187] Known as PD150606 (**11**) and PD151746 (**12**) these small molecule mercaptoacrylic acid inhibitors are potent and do not permanently bind to the enzyme (Figure 1.24). PD150606 is a

phenyl based mercaptoacrylic acid with an iodine in the *para* position (Figure 1.24). The other inhibitor, PD151746 is an indole based mercaptoacrylic acid, with the mercaptoacrylic acid and fluorine situated at positions 3 and 5, respectively (Figure 1.24).

These inhibitors are allosteric inhibitors targetting the PEF(L) and the PEF(S) domains of the heterodimeric structure. A crystal structure of PD150606 bound to PEF(S) demonstrated that it binds to the same region of PEF(S) as calpastatin region C (Figure 1.23).[81, 186] PD151746 displayed 20 fold selectivity towards calpain-I over calpain-II. The phenyl derivative was less selective for calpain-I, though it was found to be a slightly more potent inhibitor.[187] The discrepancy in the potency of these derivatives may imply that these inhibitors also interact with the PEF(L) domain as well as with the small subunit, since PEF(S) is identical in both isoforms.

11. K_I (μM)
calpain-I = 0.21
calpain-II = 0.37

12. K_I (μM)
calpain-I = 0.26
calpain-II = 5.33
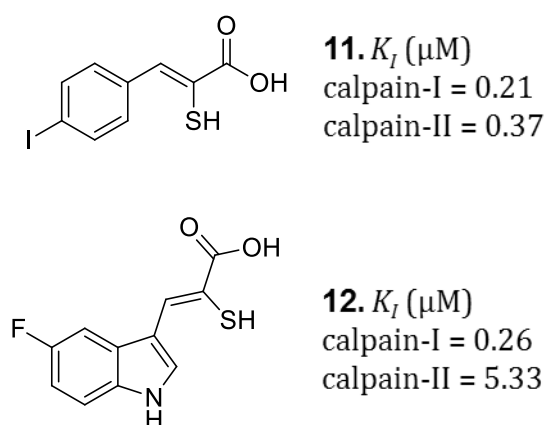
Figure 1.24: The structures of PD150606 and PD151746.[187]

1.12 Aims

The exact physiological roles of these proteases are poorly understood and hence also their roles in disease processes. One reason for the lack of knowledge is the paucity of specific inhibitors for each isoform. Therefore it is important to develop inhibitors that are selective towards calpain-I over calpain-II and *vice versa*.

The aim of this work was to develop a series of monohalogenated mercaptoacrylic acids based on the structures of PD150606 and PD151746.

These inhibitors were to contain the same skeletal structures with the mercaptoacrylic acid present at the same position of both the indole and phenyl ring. These compounds were then to be tested as inhibitors of calpain-I and calpain-II. X-ray crystallography could then be used to determine how they bind to the PEF(S) domain. The most potent and specific compounds could then be employed for the study of the physiological and pathophysiological effects of calpain-I inhibition.

2 Synthesis of α-mercaptoacrylic acids for inhibition of calpain-I

2.1 α-Mercaptoacrylic acid derivatives

The aim of the work discussed in this chapter was to produce a series of α-mercaptoacrylic acid derivatives based upon the two structures of PD150606 and PD151746. The target compounds were both phenyl and indole based derivatives, with a single halogen substituent (fluorine, chlorine, bromine or iodine) at all possible positions of the six membered ring in each case (Figure 2.1). Once prepared, they were to be tested as inhibitors of both calpain-I and calpain-II.
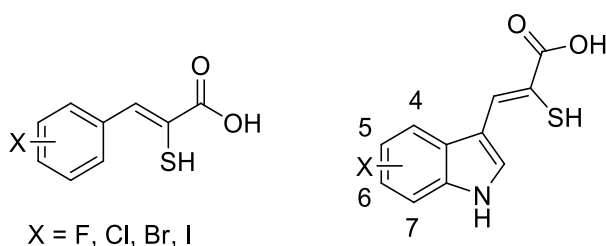
Figure 2.1: *The structure of the target compounds.*

2.1.1 Synthetic routes to α-mercaptoacrylic acid derivatives

The synthetic routes used for preparing these compounds are summarised in Scheme 2.1. Phenyl and indole derivatives were prepared from substituted benzaldehydes or indole-3-carboxaldehydes, respectively, *via* a Knoevenagel condensation with 2-thioxothiazolidin-4-one followed by hydrolysis to the appropriate α-mercaptoacrylic acid compounds (reactions (v) and (vi) in Scheme 2.1).[188] In some cases a Vilsmeier-Haack reaction [(iv) in Scheme 2.1] was used to obtain the required aldehyde from the parent arene.[189]

Moreover, several of the halogenated indoles required synthesis prior to introduction of the aldehyde functionality. For example 4-iodoindole was prepared using a two-step synthesis with the use of thallium(III)trifluoroacetate [(iii) of Scheme 2.1 and Section 2.2].[190] Preparation of 6-iodoindole was achieved using a five step synthesis. The initial starting material 6-nitroindoline was transformed using a Sandmeyer reaction as the key step to introduce the halogen atom [(ii) in Scheme 2.1], see Section 2.3.[191] Finally, two 7-substituted indole derivatives required synthesis, 7-fluoroindole and 7-bromoindole. These were prepared *via* a Bartoli indole synthesis from 1-

fluoro-2-nitrobenzene and 1-bromo-2-nitrobenzene, respectively [(i) of Scheme 2.1] (see Section 2.4 for more details).[192]
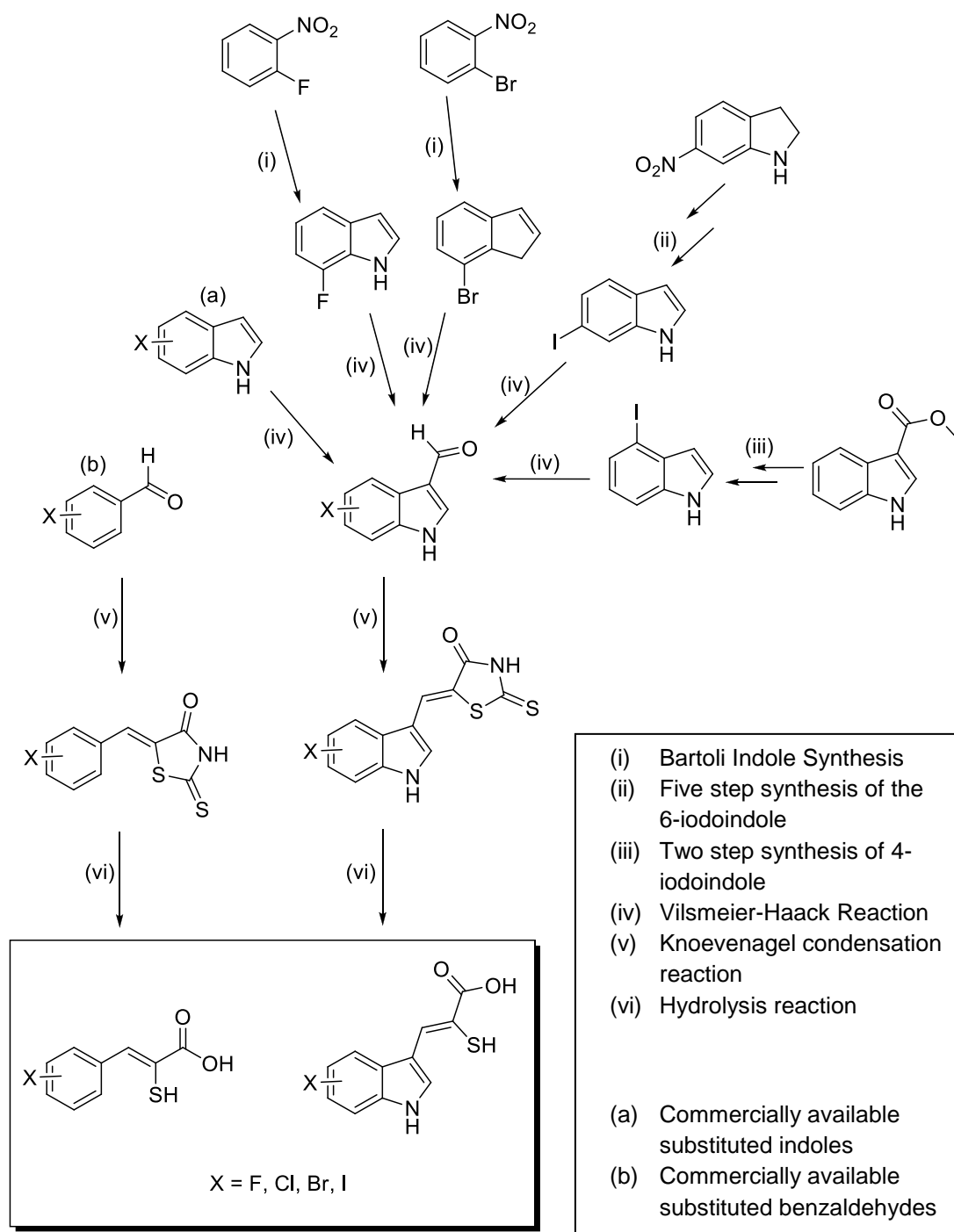
Scheme 2.1: *The synthetic routes for preparation of the target α-mercaptoacrylic acids.*

2.2 Synthesis of 4-iodoindole
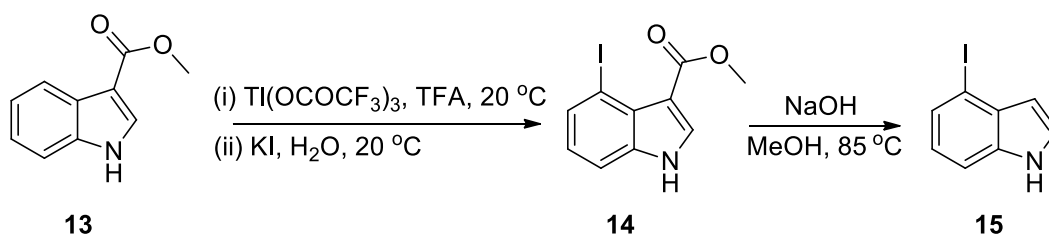
Scheme 2.2: *The formation of 4-iodoindole (15) from methyl indole-3-carboxylate (13).*[190, 193]

Here the treatment of **13** with thallium(III)trifluoroacetate resulted in substitution at position 4 of the indole ring. Treatment of the intermediate with potassium iodide (KI) replaced the thallium with iodide.[193] The resulting compound **14** was then hydrolysed using sodium hydroxide to obtain 4-iodoindole (**15**) in an overall yield of 33% (Scheme 2.2).
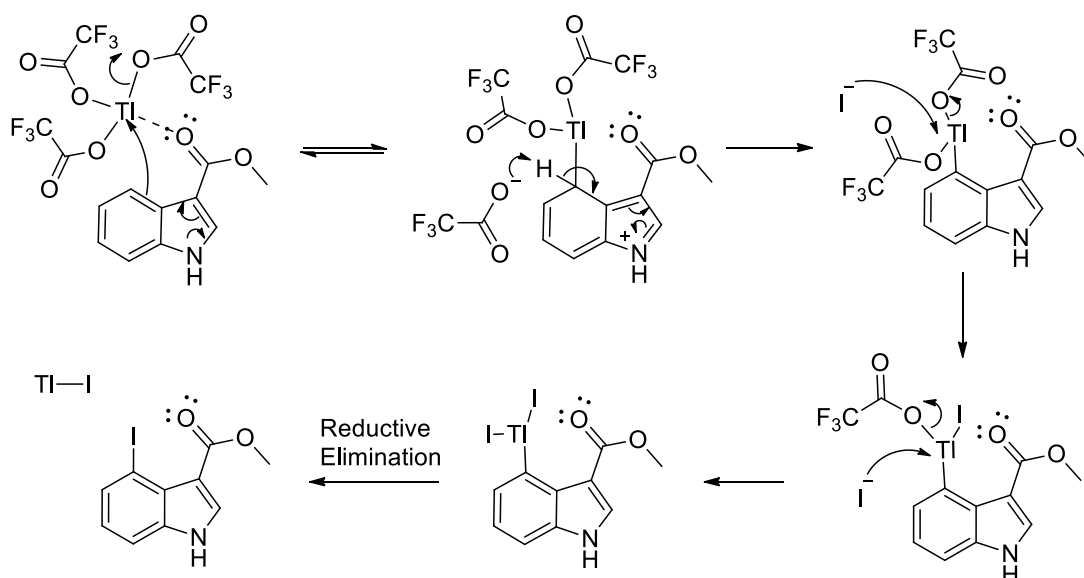
Scheme 2.3: *The coordination of thallium(III)trifluoroacetate to position 4 of methyl indole-3-carboxylate and the subsequent replacement of thallium with iodide.*[194]

In the first step thallium(III)trifluoroacetate coordinates to the oxygen in the carbonyl group of the methyl ester at position 3 and it is this that directs it to position 4 (Scheme 2.3). [190, 195] Deprotonation at the *ipso* position generates a thallium substituted aromatic ring (Scheme 2.3).[195] This intermediate then reacts with KI in water to form methyl 4-iodoindole-3-carboxylate (Scheme 2.3).[190] The addition of KI causes the relatively weak coordination bond of the thallium ion to the indole to break and be replaced with iodine (Scheme 2.3). Hydrolysis of the methyl ester with 40% sodium hydroxide solution for 1.5

hours created a carboxylate that then spontaneoutsly lost CO_2 to give the desired product.

2.3 Synthesis of 6-iodoindole

To prepare 6-iodoindole a five step synthetic route was used. 6-Nitroindoline was acetylated to protect the NH group (Scheme 2.4) (Section 2.3.1).[191] The resulting compound was reduced using a palladium catalysed hydrogenation reaction to form *N*-acetyl-6-aminoindoline (Scheme 2.4) (Section 2.3.2).[191] This compound then underwent a Sandmeyer reaction, to generate a diazonium salt. Reaction with potassium iodide then resulted in *N*-acetyl-6-iodoindoline (Scheme 2.4) (Section 2.3.3).[191, 196] The acetyl group was then removed by base catalysed hydrolysis with 40% sodium hydroxide in water (Scheme 2.4) (Section 2.3.4).[191] The final step was air oxidation of the indoline ring in the presence of a Co(II)salen catalyst (Scheme 2.4) (Section 2.3.5).[191, 197]
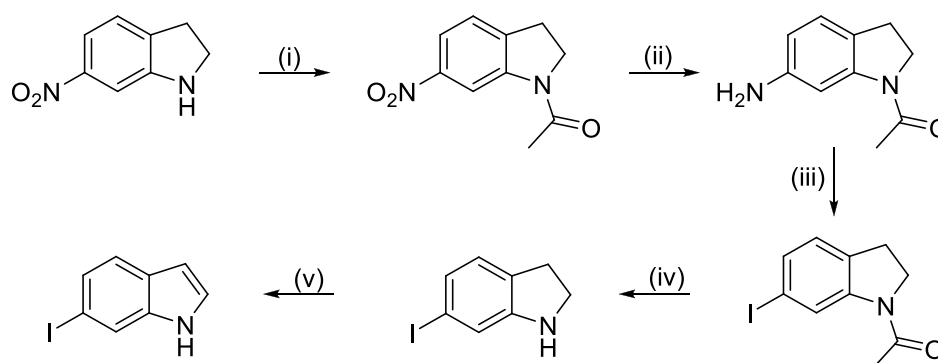
Scheme 2.4: *The synthetic route for the preparation of 6-iodoindole. (i) Ac₂O, 130 °C, (ii) H₂, 10%-Pd/C, 25 °C, (iii) NaNO₂, KI, AcOH, 0 °C, (iv) NaOH, MeOH, 75 °C, (v) Co(II)salen, air, MeOH, 25 °C*[191]

2.3.1 Acetylation of 6-nitroindoline
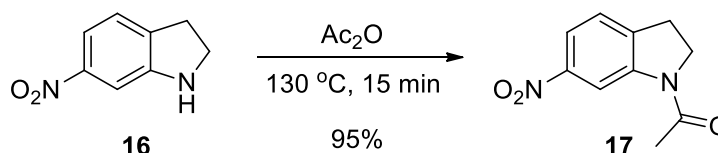
Scheme 2.5: *The acetylation reaction of 6-nitroindoline to form N-acetyl-6-nitroindoline.*

To protect the amino group of **16**, it was heated with acetic anhydride under reflux for 15 minutes (Scheme 2.5). The reaction was then allowed to cool to room temperature and water was added to precipitate the yellow compound **17** in 95% yield after collection by filtration and drying under reduced pressure.

2.3.2 Hydrogenation of *N*-acetyl-6-nitroindoline
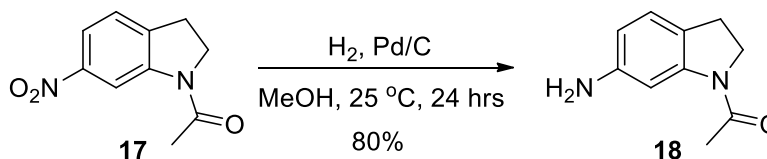
Scheme 2.6: *Reduction of N-acetyl-6-nitroindoline to form N-acetyl-6-aminoindoline.*

After reduction of **17** with hydrogen gas (under 1 atmosphere's pressure) over 10% palladium on activated charcoal **18** was isolated as a white solid in 80% yield (Scheme 2.6).

2.3.3 The Sandmeyer reaction
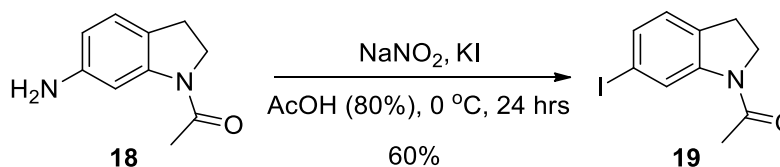
Scheme 2.7: *The Sandmeyer reaction.[196]*

Compound **18** was reacted with nitrosonium ions formed through the reaction of 80% aqueous acetic acid with sodium nitrite (Scheme 2.8). The nitrosonium ions then react with the amino substituent of the starting material to form a diazonium salt and a molecule of water. The diazonium functionality is a good leaving group so upon reaction with the KI via a radical nucleophilic reaction, nitrogen gas leaves and is replaced with iodide (Scheme 2.8).[198] After purification by flash chromatography on silica gel **19** was isolated in 60% yield.

(A)

(B)
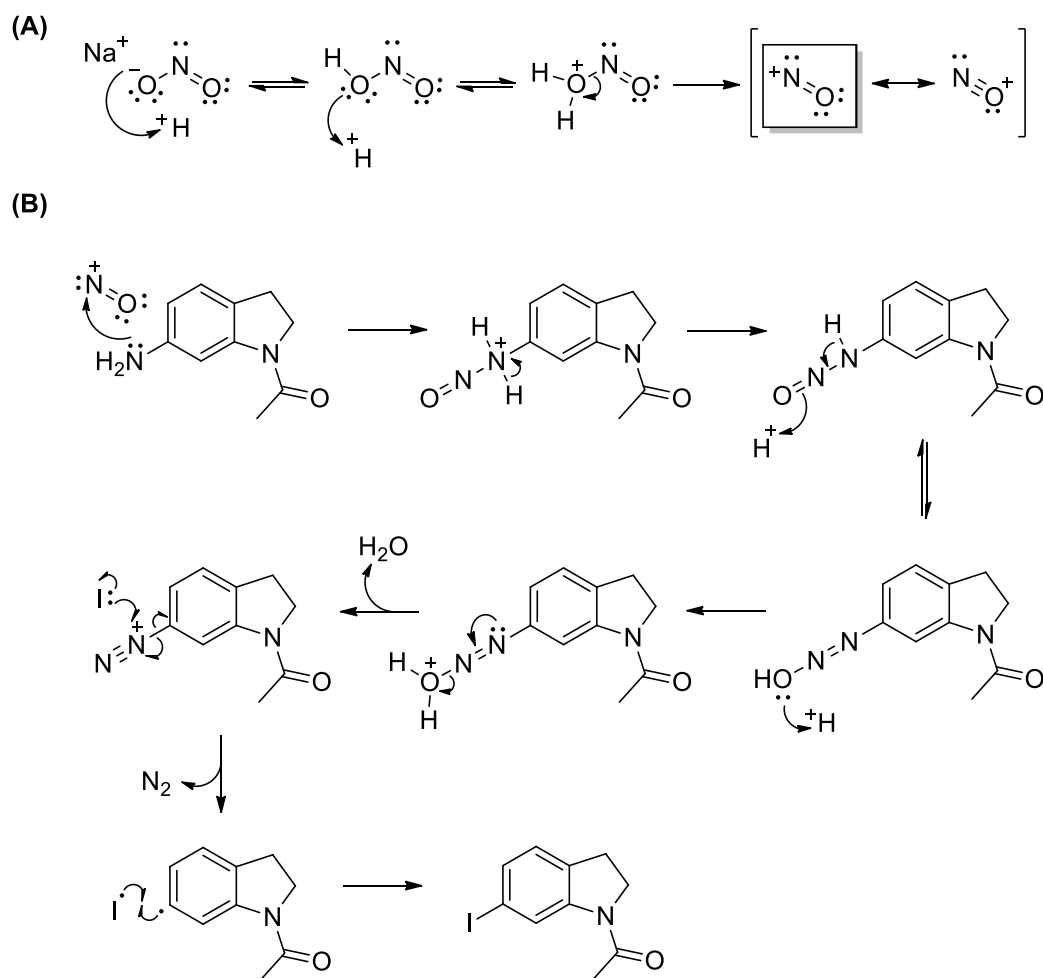
Scheme 2.8: *(A) The mechanism of formation of the nitrosonium ion. (B) The mechanism for the formation of* **19**.[198, 199]

2.3.4 Hydrolysis of the protecting group
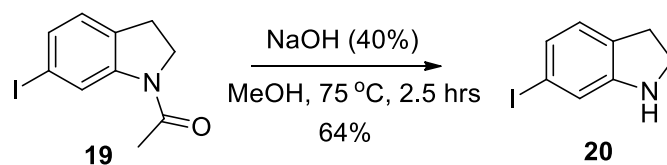
Scheme 2.9: *Hydrolysis of the N-acetyl protecting group.*

The acyl protecting group of **19** was removed through treatment with 40% NaOH solution in methanol for 2.5 hours (Scheme 2.9). The desired product **20** was isolated in 64% yield after purification by flash chromatography on silica gel.

2.3.5 Oxidation of indoline
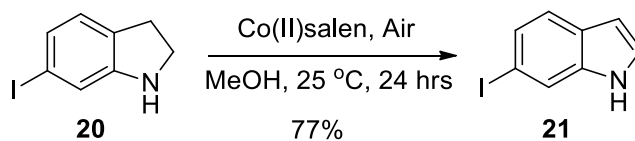
Scheme 2.10: *Conversion of indoline to indole.[197]*

N,N′-Bis(salicylidene)ethylenediaminocobalt(II) was used as a catalyst to carry out the oxidation. This catalyst uses oxygen to oxidise the indoline to form indole.[197] Air was used as the source of oxygen for the reaction and was bubbled through the stirred reaction mixture for 24 hr. The desired product was isolated as pale brown crystals in a yield of 77%.

2.4 Bartoli indole synthesis
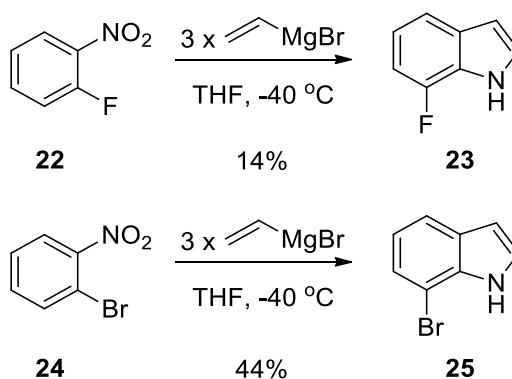
Scheme 2.11: *The Bartoli indole synthesis.[200]*

The Bartoli indole synthesis is used to form 7-substituted indoles through the reaction of three equivalents of vinyl magnesium bromide with 2-substituted nitrobenzenes (Scheme 2.11).[192] This reaction was used for the formation of 7-fluoroindole (**23**) and 7-bromoindole (**25**) in yields of 14% and 44% respectively.
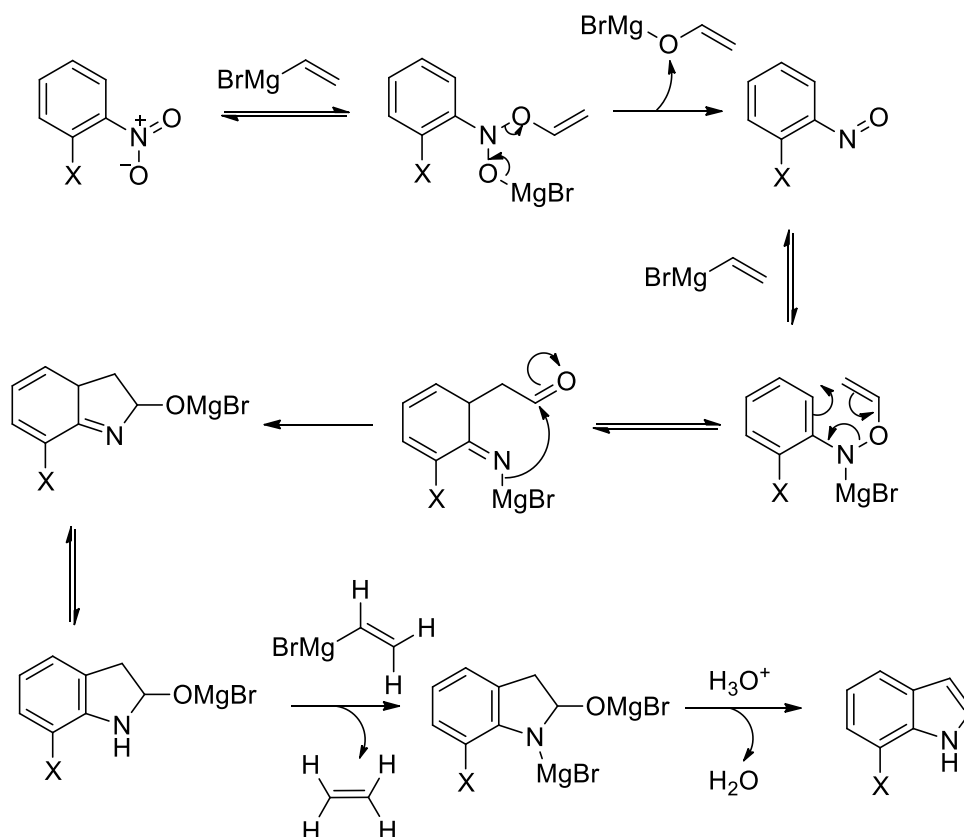
Scheme 2.12: *The reaction mechanism for the Bartoli indole synthesis.*[200]

2.5 Vilsmeier-Haack reaction
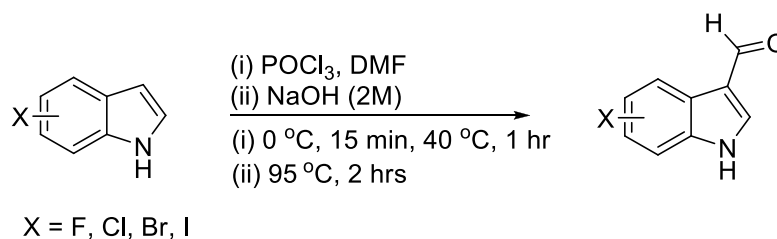
Scheme 2.13: *A reaction scheme for the Vilsmeier-Haack reaction.*[189]

To form several of the required indole-3-carboxaldehydes a Vilsmeier-Haack reaction was used.[201] In this reaction *N,N*-dimethylformamide (DMF) was reacted with phosphorus oxychloride (POCl$_3$) to form a *N*-(chloromethylene)-*N*-methylmethanimminium cation (Scheme 2.14). The intermediate was then attacked by the nucleophilic aromatic ring; treatment with water hydrolysed the resulting imminum ion to form an aldehyde (Scheme 2.14).[201]
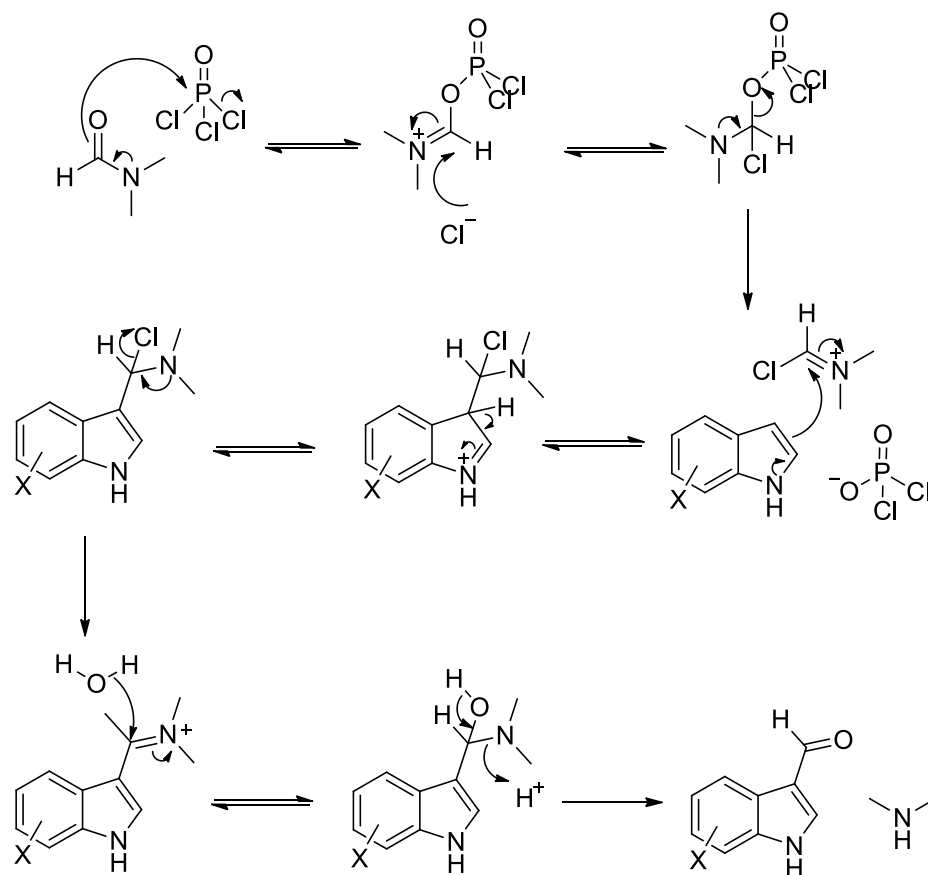
Scheme 2.14: *The mechanism of the Vilsmeier- Haack reaction.*[201]

This reaction was carried out on the following substituted indoles; 4-fluoroindole, 4-chloroindole, 4-bromoindole, 4-iodoindole, 5-iodoindole, 6-fluoroindole, 6-chloroindole, 6-iodoindole, 7-fluoroindole, 7-chloroindole, 7-bromoindole and 7-iodoindole. Yields were generally excellent ranging from 54% to 98% (Figure 2.2).

26: 98% **27:** 80% **28:** 94% **29:** 92%

30: 95% **31:** 77% **32:** 96% **33:** 95%

34: 54% **35:** 89% **36:** 82% **37:** 92%
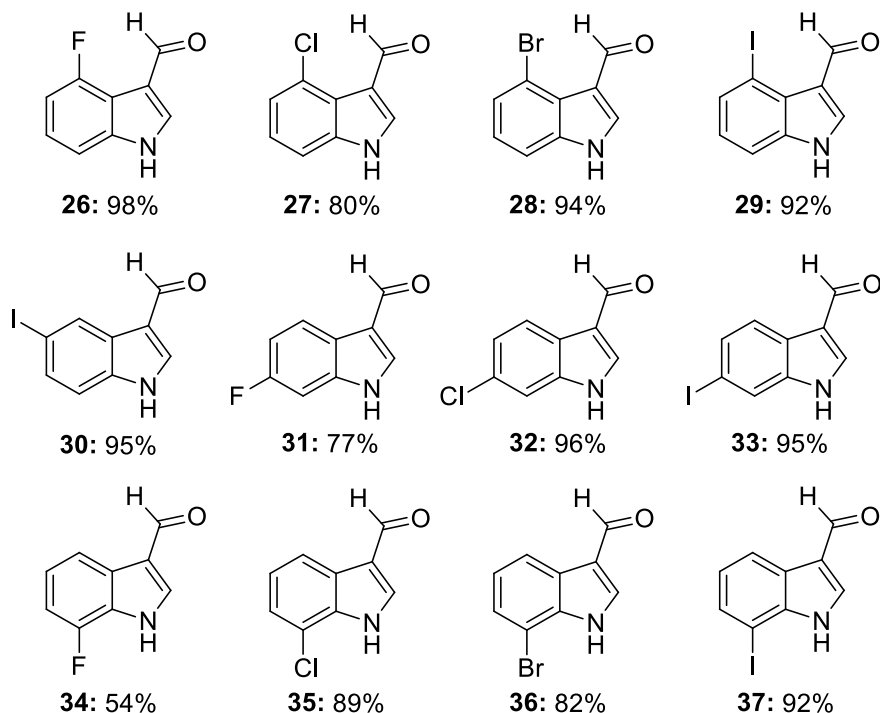
Figure 2.2: *The compounds produced using the Vilsmeier-Haack reaction.*

2.6 Synthesis of α-mercaptoacrylic acid

The desired α-mercaptoacrylic acid functionality was formed *via* a Knoevenagel condensation and was carried out between a substituted benzaldehyde or indole-3-carboxaldehyde and 2-thioxothazolidin-4-one followed by base catalysed hydrolysis of the resulting condensation product (Section 2.6.1 and Section 2.6.2).

2.6.1 Knovenagel condensation reaction
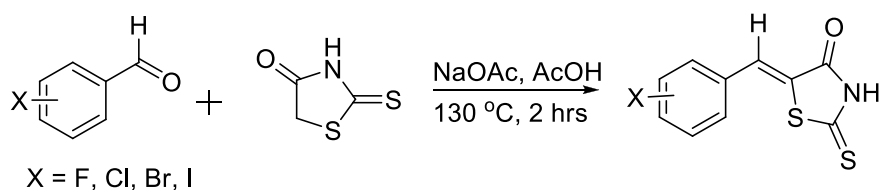
Scheme 2.15: *The Knoevenagel condensation reaction of 2-thioxothiazolidin-4-one with a halo substituted benzaldehyde.*[188]

A halo-substituted benzaldehyde or indole-3-carboxaldehyde was reacted with 2-thioxothiazolidin-4-one using sodium acetate in acetic acid as the catalyst (Scheme 2.15).
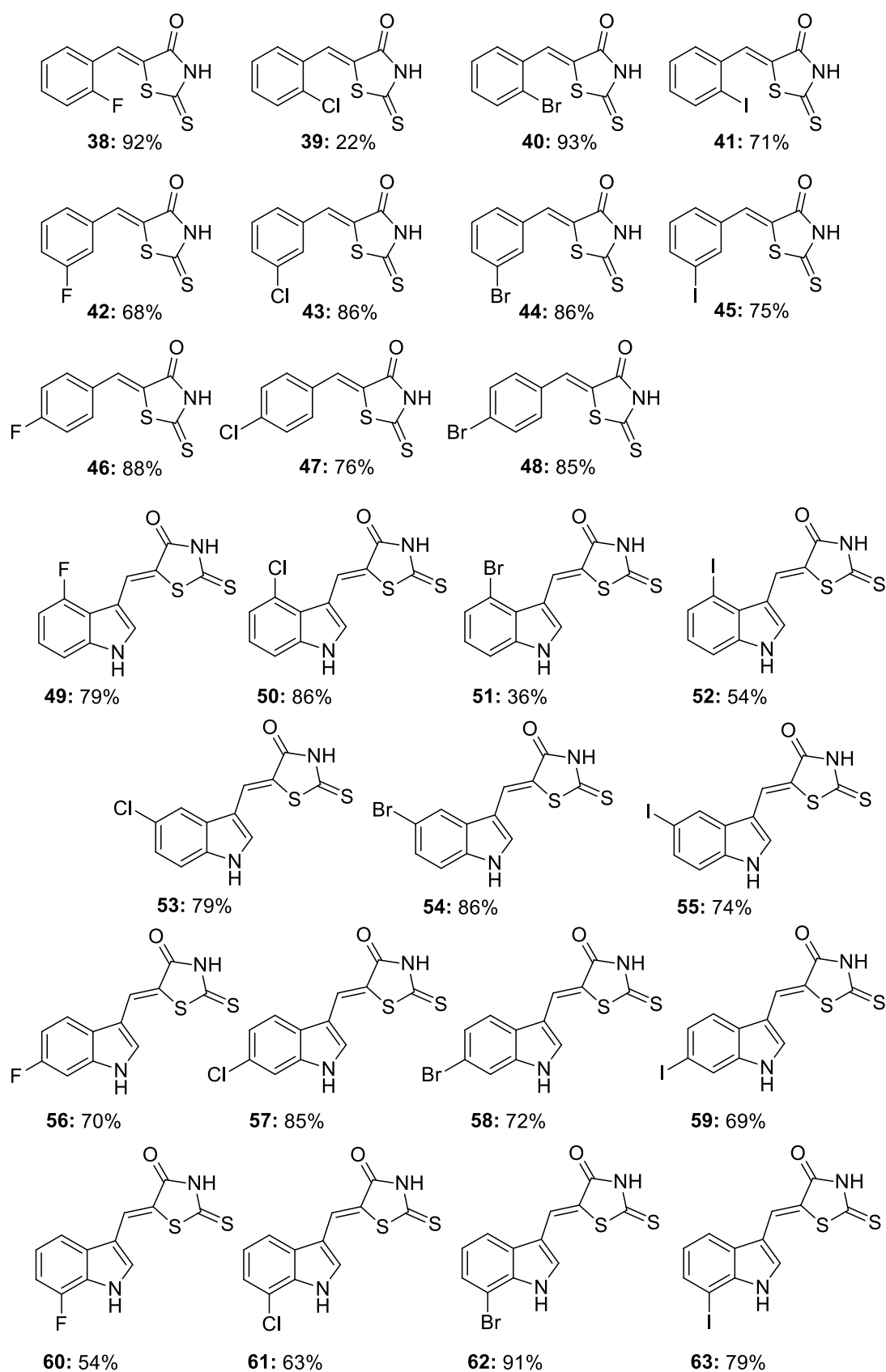
Figure 2.3: *The yields of the phenyl and indole based derivatives after the Knoevenagel condensation.*

The yields achieved for the majority of the compounds were above 60%, with the exception of compounds **39**, **51**, **52** and **60** (Figure 2.3). The lower yields achieved for compounds **39**, **51** and **52** may be down to the proximity of the halogen towards the 2-thioxothiazolidin-4-one ring.

2.6.2   Base-catalysed hydrolysis
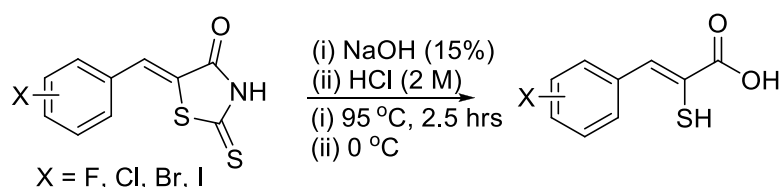
Scheme 2.16: *The base catalysed hydrolysis of the 2-thioxothiazolidin-4-one ring to form the α-mercaptoacrylic acid substituent.*[188]

The compounds produced by the Knoevenagel condensation reaction (Section 2.6.1) were subjected to a base-catalysed hydrolysis to form the α-mercaptoacrylic acid (Scheme 2.16). After refluxing in 15% aqueous sodium hydroxide solution the reaction was cooled to 5 °C and quenched with 2M hydrochloric acid. The precipitates that formed were filtered and air-dried under reduced pressure. Through mass spectrometry, ^{1}H NMR and ^{13}C NMR spectroscopy 24 compounds were identified as the desired α-mercaptoacrylic acids. The thiol of the indole based mercaptoacrylic acids was visible in the ^{1}H NMR spectra as a broad peak at ~5 ppm, when carried out in DMSO-d_6. The proton of the NH in the 2-thioxothiazolidin-4-one ring was no longer visible at ~14 ppm in the ^{1}H NMR spectra.

The hydrolysis of compounds **41** and **52** yielded compounds **67** and **78**, which were not the desired α-mercaptoacrylic acids. The thiol group had replaced the halogen through an *ipso* substitution to give a benzothiophene, **67**, in the case of the phenyl derivative and a thiopyranoindole-2-carboxylic acid in the case of the indole, **78** (Figure 2.4) (Section 2.6.3). All of the compounds, apart from **81**, were isolated in yields in excess of 50 % (Figure 2.4).

64: 91%   **65:** 63%   **66:** 84%   **67:** 72%

68: 68%   **69:** 79%   **70:** 75%   **71:** 68%

72: 88%   **73:** 58%   **74:** 82%

75: 87%   **76:** 81%   **77:** 71%   **78:** 68%

79: 97%   **80:** 89%   **81:** 29%

82: 56%   **83:** 52%   **84:** 76%   **85:** 67%

86: 78%   **87:** 81%   **88:** 86%   **89:** 81%
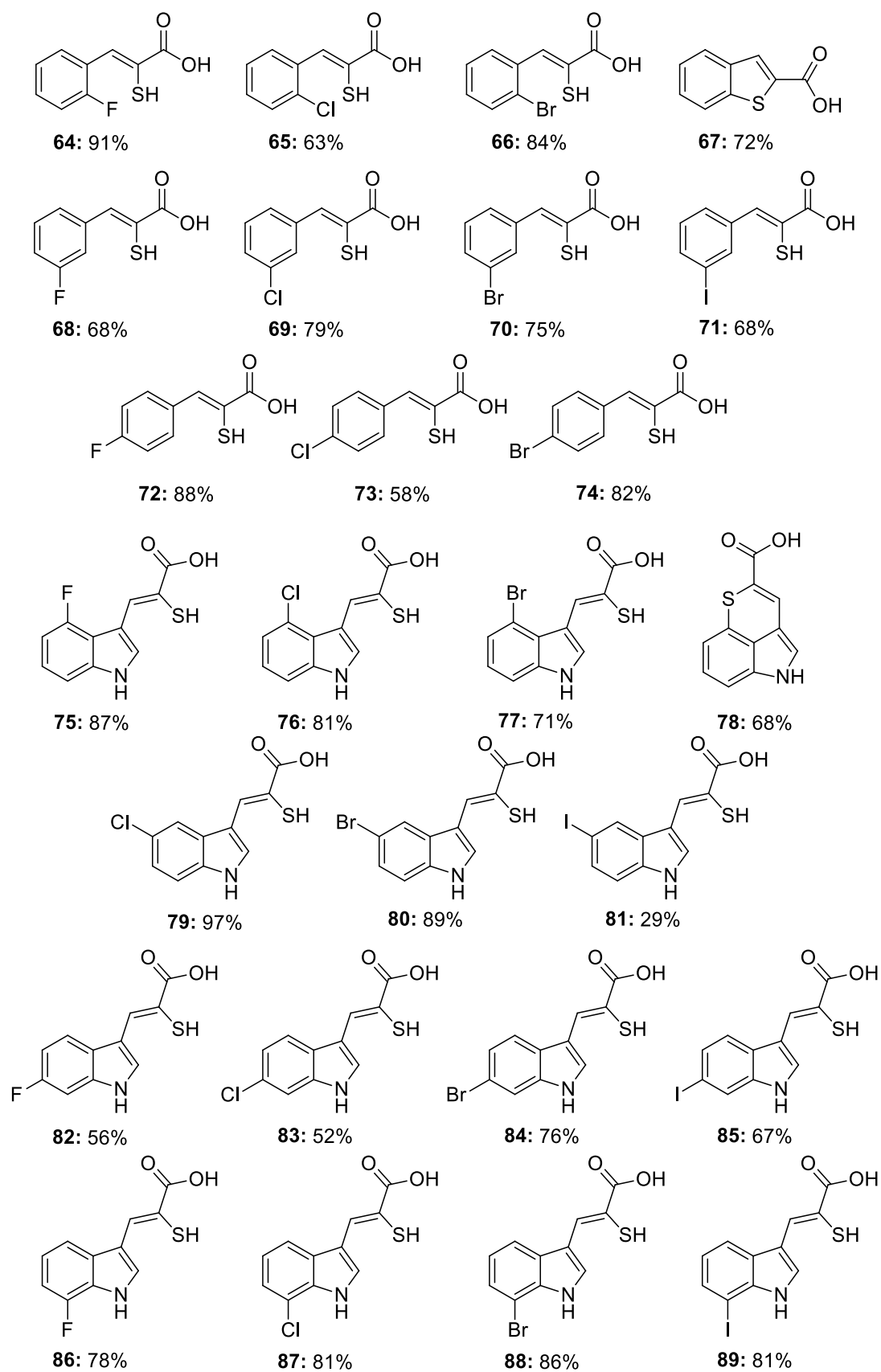
Figure 2.4: *The final phenyl and indole α-mercaptoacrylic acid derivatives.*

2.6.3   Intra-molecular cyclisation

The target compounds (*Z*)-3-(2-iodophenyl)-2-mercaptoacrylic acid and (*Z*)-3-(4-iodoindol-3-yl)-2-mercaptoacrylic acid could not be prepared using this chemistry since during the hydrolysis reaction the compound underwent a cyclisation reaction (Scheme 2.17).
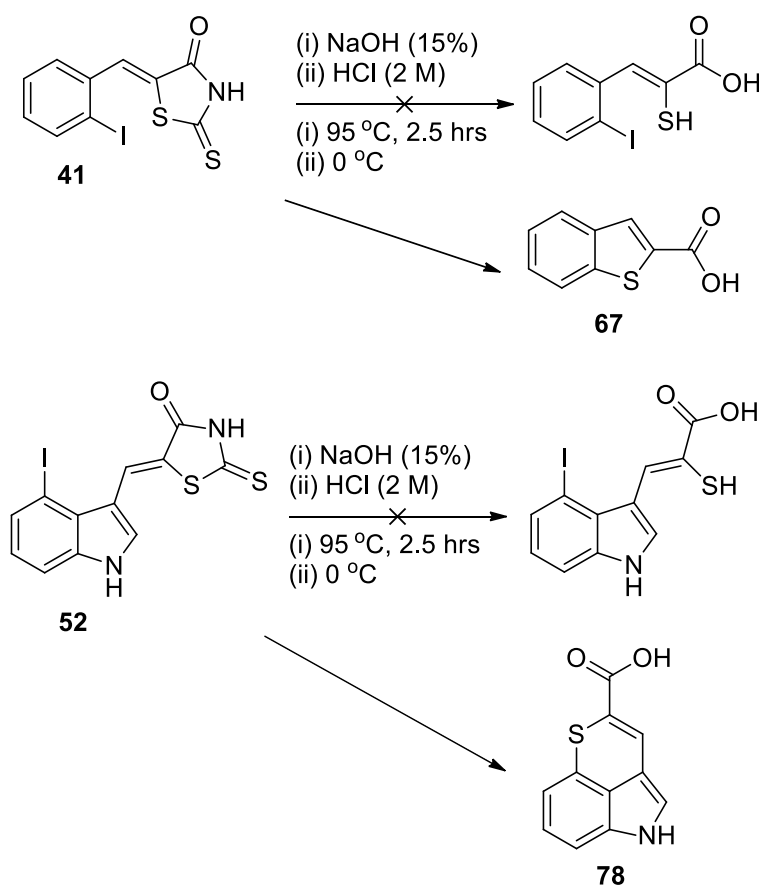
Scheme 2.17: *The base catalysed hydrolysis of **41** and **52** resulting in undesired products.*

This intra-molecular *ipso*-substitution reaction presumably occurs *via* the nucleophilic thiolate that forms during the hydrolysis. The thiolate attacks at the iodo-substituted carbon and the iodide is then lost to form the benzothiophene product (Scheme 2.18).
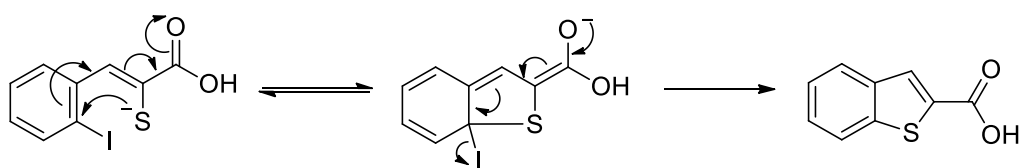
Scheme 2.18: *Proposed mechanism for the thiophene formation.*

The intra-molecular reaction only occurs when the iodine is attached to the *ortho* position relative to the mercaptoacrylic acid substituent. I⁻ is a much better leaving group than fluoride, chloride and bromide, therefore it appears that cyclisation is faster than the formation of the α-mercaptoacrylic acid substituent.

2.7 Disulfide formation

The final α-mercaptoacrylic acid compounds contain a thiol group. Therefore it was important to determine whether a disulfide bond forms between two thiols (Scheme 2.19).
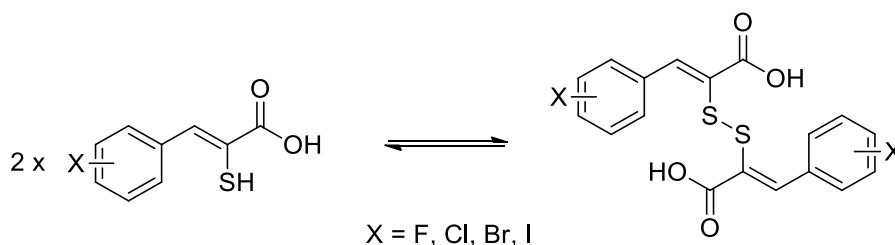
Scheme 2.19: *Formation of the disulfide bond between two molecules of α-mercaptoacrylic acid.*

2.7.1   Disulfide formation with indole based α-mercaptoacrylic acids

To determine whether the disulfide formation occurred with the indole based α-mercaptoacrylic acids, compound **88** was used as a model compound (Scheme 2.20). Disulfide formation occurs spontaneously in air or by the addition of oxidising agents and may be reversed by the addition of reducing agents such as dithiothreitol (DTT). Observation of any disulfide formation was carried out using three techniques; ^{1}H NMR spectroscopy, UV-Vis spectrophotometry and HPLC analysis. ^{1}H NMR spectroscopy was used to observe the change in the chemical shift as dimerization occurs. This was carried out in DMSO-d_6 and as this is an oxidising solvent the addition of DTT would have little effect on the dimeric compound. UV-Vis spectrophotometric analysis was used to follow disulfide formation and reduction through the absorbance of the reduced form at 324 nm. Reversed phase HPLC analysis was also used to observe the formation of the disulfide and the reduction back to the monomeric compound through the change in retention time.
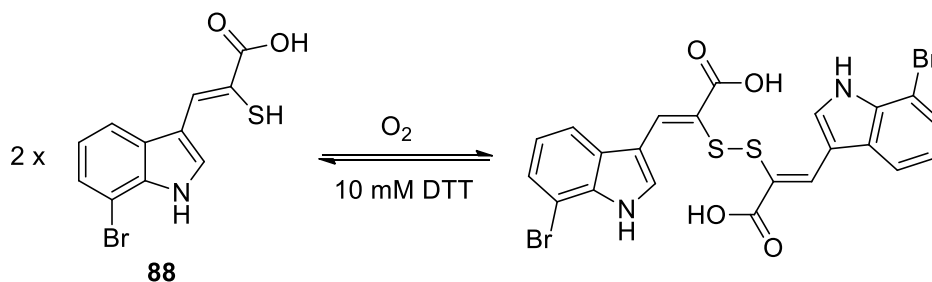
Scheme 2.20: *Oxidation of **88** to form a disulfide.*

UV-Vis spectroscopy
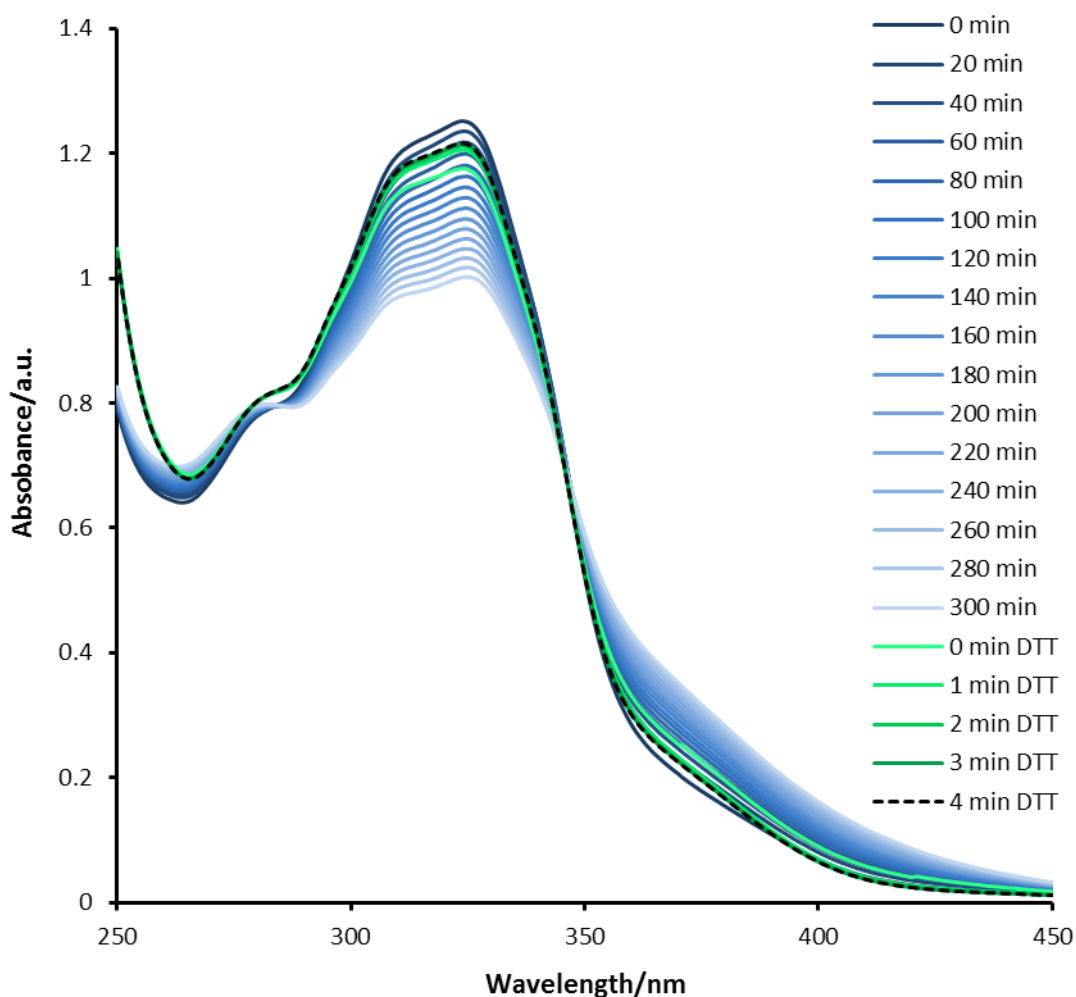
Figure 2.5: *UV-Vis spectra of **88** in a 1:1 mixture of acetonitrile and 100 mM potassium phosphate buffer at pH 7.4, recorded every 20 minutes for 300 minutes, where the process of gradual oxidation was observed. This was followed by the addition of DTT (10 mM) to reduce the compound back to its monomeric form.*

The reaction was carried out in a 1 ml cuvette using a 1:1 solution of acetonitrile and 100 mM potassium phosphate buffer at pH 7.4. The solvent mixture was used to ensure the compound remained in solution over the course

of the experiment. 2 μl of **88** (50 mM in DMSO) was added to the solution to make a total concentration of 100 μM. Spectra were recorded from 250 – 450 nm every 20 minutes for 300 minutes. The λ_{max} at 324 nm diminished in intensity over the course of the experiment, resulting from structural alterations to the compound in solution (Figure 2.5).

10 μl of 1 M dithiothreitol (DTT) was then added to the solution giving a total concentration of 10 mM. UV-Vis spectra were then recorded every 60 seconds until the compound appeared to have been reduced back to the monomeric form (Figure 2.5). The addition of DTT revealed an increase in the absorbance at 324nm and that the reaction was reversible (Figure 2.5). The absorbance observed, at 324 nm, 4 minutes after the addition of DTT to the solution did not reach the original absorbance at the start of the experiment. This may be due to the gradual degradation of the compound in aqueous solutions, evidence for this degradation was also seen in the HPLC analysis (*vide infra*).

HPLC analysis
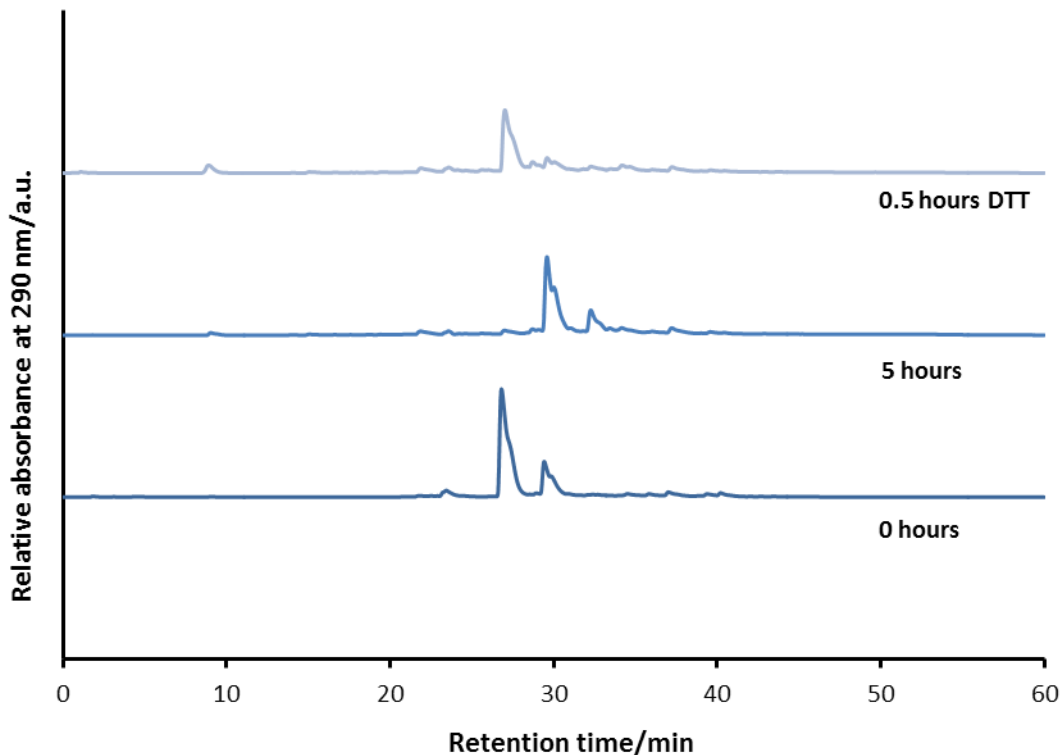
Figure 2.6: *HPLC traces demonstrating that the monomeric **88** (retention of 26.8 minutes bottom trace) changes to give another compound of retention time of 29.6 minutes (middle trace). 10 mM DTT was added to the solution and the compound rapidly returns to is original form (top trace).*

Compound **88** (6.7 mM) was dissolved in 0.5 ml of a 1:1 mixture of acetonitrile and 100 mM phosphate buffer at pH 7.4. 20 μl of this sample was analysed by revesed phase HPLC. Elution was carried out a reversed phase analytical column (Acclaim® 120, C18, 3 μm, 120 Å, 150 × 4.6 mm) with the wavelength of the detector set to 290 nm. The solvent gradient ranged from 100% water (0.1% trifluoroacetic acid (TFA)) to 100% acetonitrile (0.1% TFA) over the course of 60 minutes, with the flow rate set to 1 ml/min.

The retention time of **88** at 0 hours was 26.8 minutes (Figure 2.6). After 5 hours in solution (at 20 °C) a second 20 μl sample was analysed. The peak at 26.8 min had disappeared and a second peak was clearly visible, at 29.6 minutes (Figure 2.6). 5 μl of a 1M DTT solution (in 100 mM phosphate buffer) was added to the solution of **88** to give a total concentration of 10 mM DTT, after 0.5 hours at 20 °C a 20μl sample was analysed to reveal that the majority of the compound had been reduced to the monomeric form. Again over the duration of the experiment it appears that some of the compound had been lost due to degredation in aqueous solution. The HPLC traces indicate that the loss of the compound is likely to be due to decomposition, as the number of peaks observed in the chromatogram increased over time (Figure 2.6).

¹H NMR spectroscopic analysis

¹H NMR spectroscopic analysis of **88** revealed that in DMSO the compound alters structurally, over the course of 22 hours (Figure 2.7). The resonances arising from the protons in the vicinity of the mercaptoacrylic acid substituent showed the most dramatic changes in chemical shift (Figure 2.7). The signals for the C3 proton, the alkene proton and the NH proton moved by 0.51, 0.21 and 0.20 ppm in chemical shift, respectively, during this time (Figure 2.7). This is consistent with disulfide formation since the greatest changes in chemical shift are found in the protons localised around the mercaptoacrylic acid. After 22 hours in DMSO-d_6 other peaks are visible, that are not related to the disulfide compound, this is yet another indication that these compounds are not stable in solution.

Proton	δ_H (Monomer)	δ_H (Dimer)
1	12.07 ppm	12.27 ppm
2	8.00 ppm	8.21 ppm
3	7.95 ppm	8.46 ppm
4	7.80 ppm	7.78 ppm
5	7.10 ppm	7.13 ppm
6	7.46 ppm	7.47 ppm
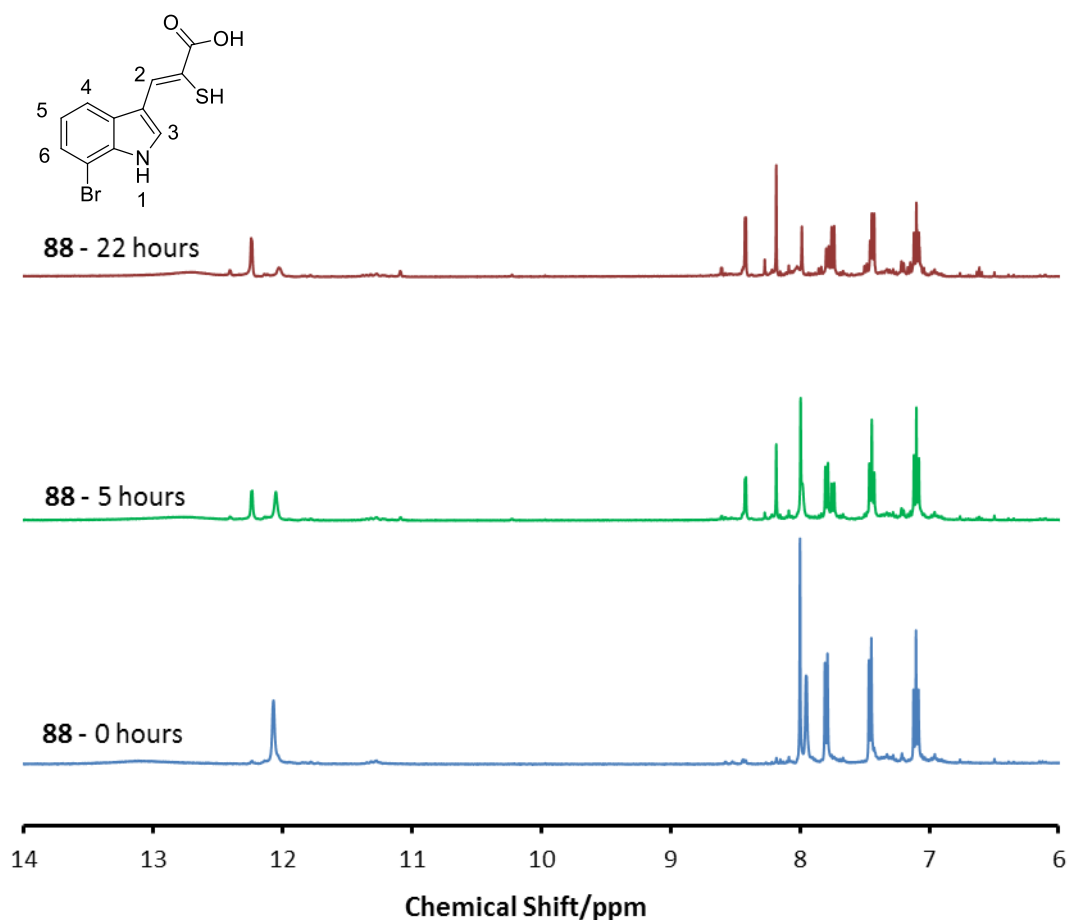
Figure 2.7: *¹H NMR spectra of **88** in DMSO-d₆ at 0 hours (blue), 5 hours (green) and 22 hours (red) and table listing the chemical shifts and assignment of the resonances due to each proton.*

2.7.2 Phenyl based α-mercaptoacrylic acids

Two techniques were used for the detection of the dimerisation of phenyl based inhibitors, UV-Vis spectroscopy and HPLC analysis. The analyses were carried out using **74** (Scheme 2.21).
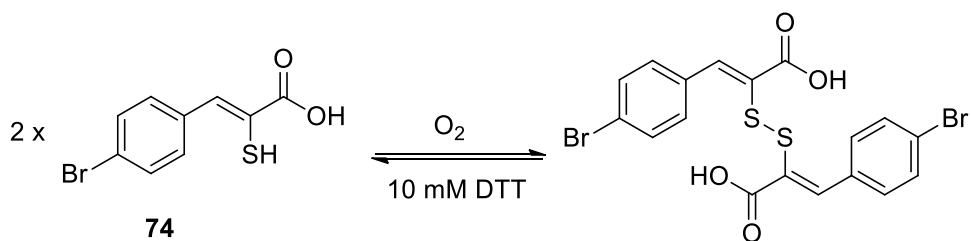
Scheme 2.21: *Oxidation of **74** to form a disulfide.*

UV-Vis spectroscopy
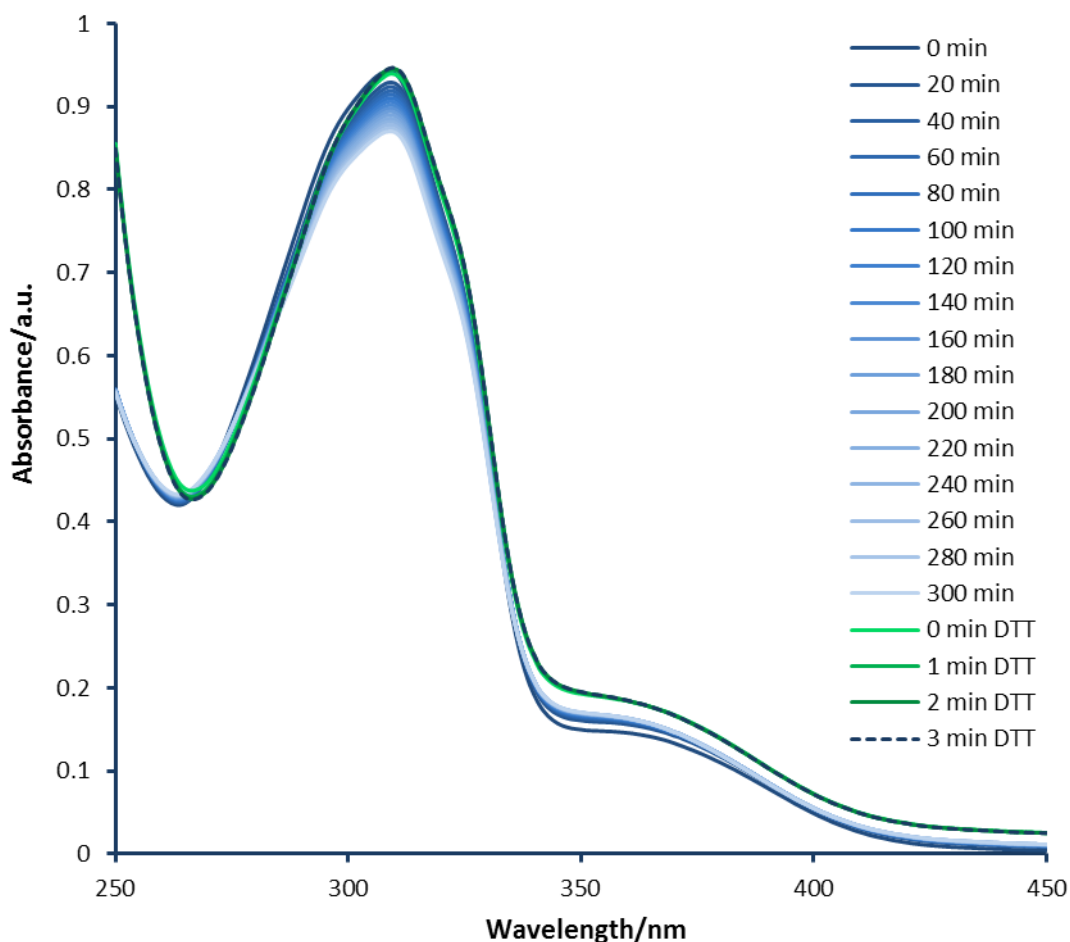
Figure 2.8: *UV-Vis spectra of **74** in a 1:1 mixture of acetonitrile and 100 mM potassium phosphate buffer at pH 7.4, recorded every 20 minutes for 300 minutes, where the process of gradual oxidation was observed. This was followed by the addition of DTT (10 mM) to reduce the compound back to its monomeric form.*

Compound **74** (100 μM) was dissolved in 1 ml of a 1:1 mixture of acetonitrile and 100 mM potassium phosphate buffer at pH 7.4, UV-Vis spectra were then recorded from 450 – 250 nm every 20 minutes for 300 minutes. The spectra revealed that the λ_{max} of 309 nm decreases steadily over time, but at a slower rate than **88** (*vide supra*). DTT was added (final concentration 10 mM) after 300

minutes and spectra were recorded every 60 seconds until absorbance at 309 nm stopped increasing. The compound was reduced to the monomeric form within 3 minutes (Figure 2.8). The slower oxidation of **74** implies that the compound is more stable in the monomeric form than compound **88** is in solution.

HPLC analysis

This technique was used to establish whether the disulfide that forms is fully reduced by the addition of 10 mM DTT. **74** was dissolved in 1 ml of a 1:1 mixture of acetonitrile and potassium phosphate buffer at pH 7.4, to make a total concentration of 5.6 mM. 20 µl of this sample was analysed by reversed phase HPLC in an identical manner to that used for the indole derivative **88** (*vide supra*).

The retention time of **74** at 0 hours was 30.0 minutes (Figure 2.9). The solution containing the inhibitor was left for 5 hours at 20 °C and a 20 µl sample was analysed revealing the formation of a new peak at 33.3 min. The solution was left for a further 19 hours before the sample was analysed again revealing that the peak at 30.0 min had disappeared (Figure 2.9). DTT was added to a total concentration of 10 mM, after 0.5 hours the HPLC analysis was repeated to determine whether the compound had returned the original monomeric form. The compound had fully reduced after 0.5 hours (Figure 2.9).
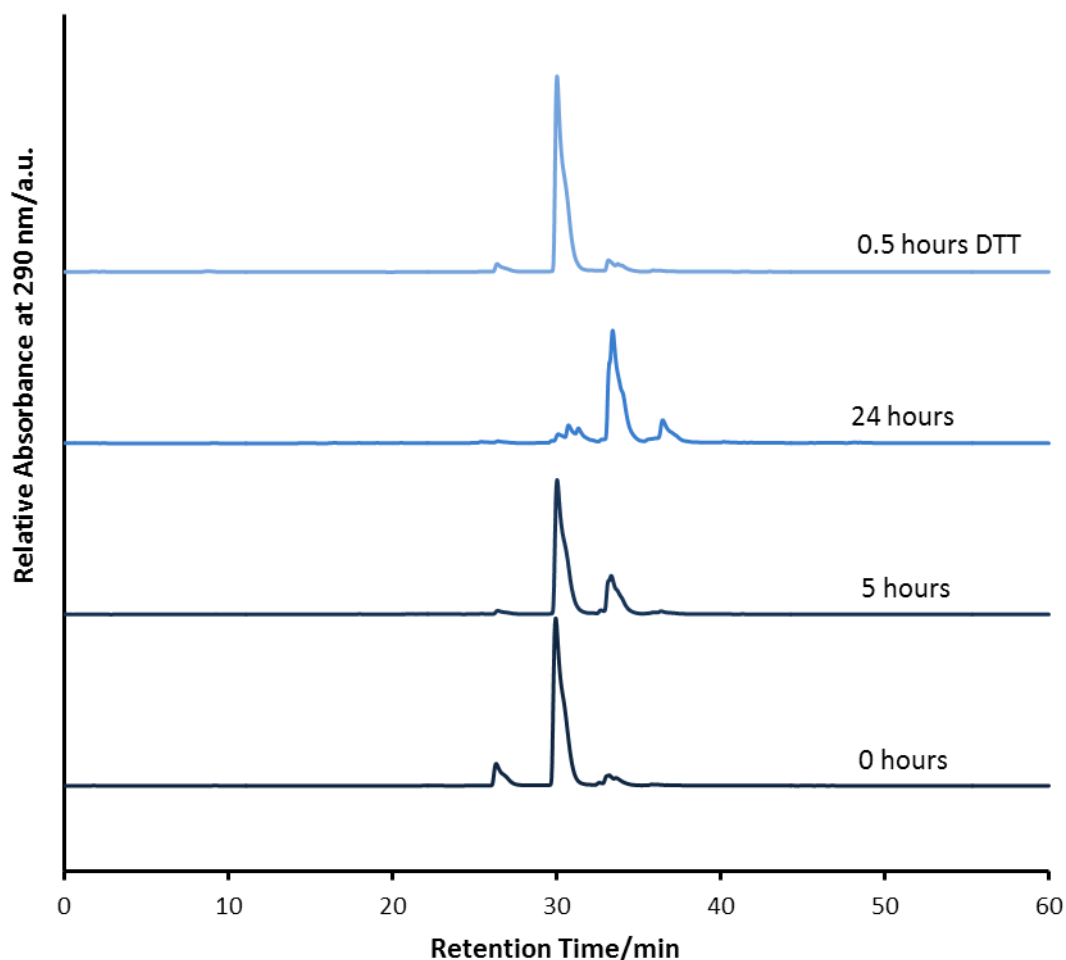
Figure 2.9: *Chromatograms demonstrating that the monomeric **74** (30.0 minutes bottom trace) forms a disulfide with a retention time of 33.3 minutes (middle trace). 10 mM DTT was added to the solution and the compound rapidly underwent a reduction to the monomeric form (top trace).*

2.7.3 Discussion

Both compound **88** and **74** readily react to form disulfides leading to dimers. The combination of the UV-Vis spectra and the HPLC chromatograms revealed that **88** has a greater tendency to dimerise than its phenyl counterpart. Through the addition of 10 mM DTT the dimers are swiftly reduced back to the monomeric form. This reducing agent is representative of the reducing conditions that are found within the cell, as the cytosol contains reducing agents such as glutathione.[202] Therefore when these inhibitors were tested with calpain-I and calpain-II in the fluorogenic assays and within the neutrophil itself (Section 2.8), the compound that was active was the very likely monomeric mercaptoacrylic acid and not the disulfide form of the compound.

2.8 Testing for calpain-I inhibition

Each α-mercaptoacrylic acid was tested for its inhibitory properties toward both calpain-I and calpain-II. FRET based assays were used to determine the IC_{50} values for each inhibitor. The two assays used different substrates, the first substrate was calpain-I specific (Section 2.8.1) and the second could be hydrolysed by both isoforms (Section 2.8.2). Calpain-I inhibition was preferred over inhibition of calpain-II as inhibition of calpain-II is embryonically lethal.[171]

All inhibition assays were performed by Dr. C. Parr, Dr E. Robinson and Prof. M. B. Hallett (School of Medicine, Cardiff University, Cardiff).

2.8.1 Calpain-I specific FRET based assay

This FRET based assay employed a peptide linker derived from the known substrate α-fodrin between the donor chromophore and the acceptor chromophore.[203, 204] α-Fodrin, also known as non-erythroid α-spectrin, is a major structural protein found in the hippocampus. It is used in the cytoskeletal binding of cell plasma membrane to actin filaments within the cell.[205, 206] Calpain-I is known to cleave this protein and the cleavage site is therefore used as substrate for a FRET based inhibition assay.[203, 204, 206]

The donor chromophore was fluorescein amidite (FAM) attached to the N-terminal lysine side chain (Figure 2.10). The acceptor chromophore was 4-(dimethylaminophenyl)diazenylbenzoic acid (DABCYL) covalently linked to the side chain of a C-terminal lysine (Figure 2.10). The linker between the two chromophores was the hexapeptide Glu-Val-Tyr-Gly-Met-Met, therefore giving an octapeptide in total - Lys(FAM)-Glu-Val-Tyr-Gly-Met-Met-Lys(DABCYL).[206] The peptide bond that lies between Tyr and Gly is hydrolysed by calpain-I allowing fluorescence to be detected.
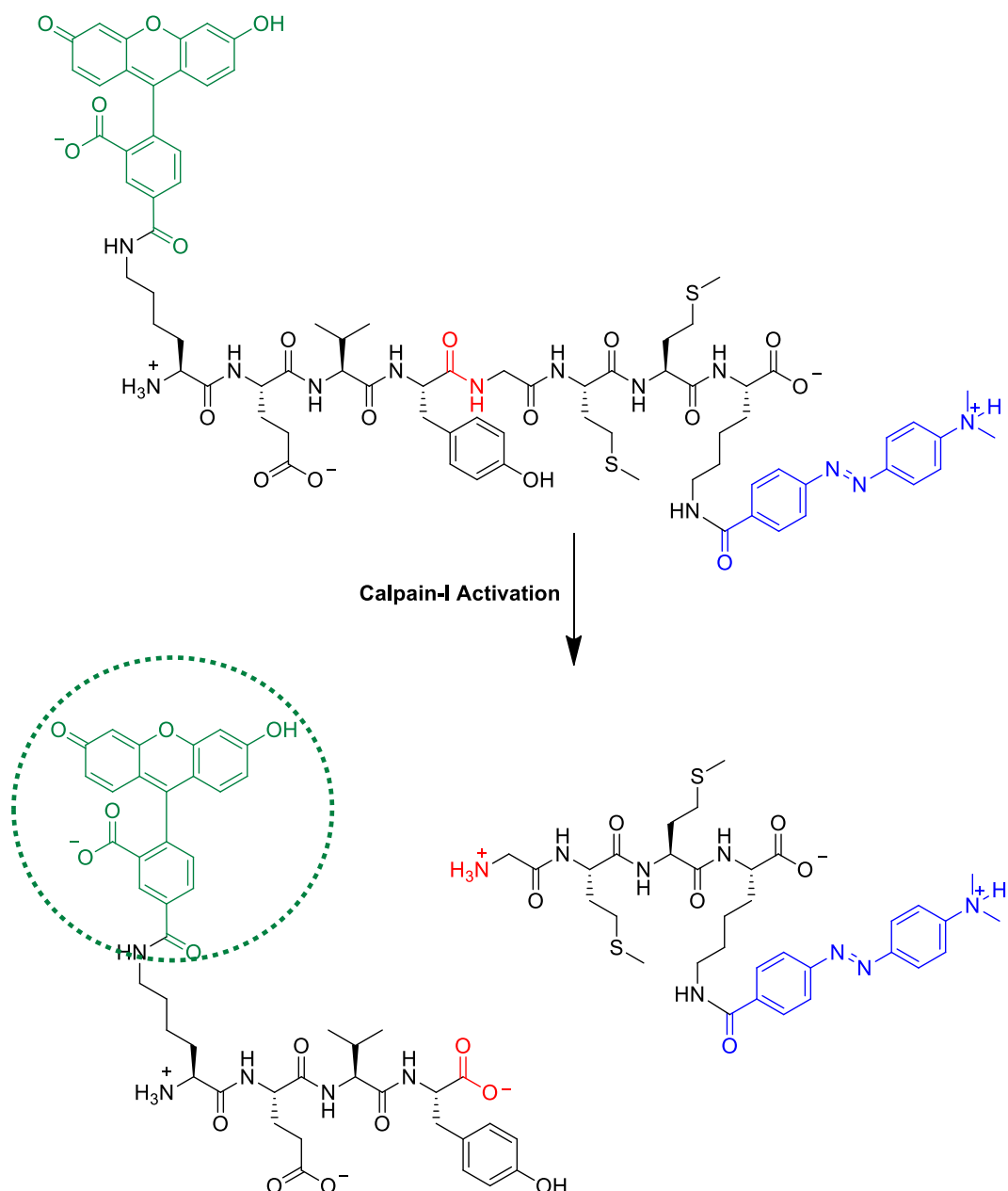
Figure 2.10: *The calpain-I specific fluorogenic substrate derived from α-fodrin. Fluorescein is bound to the side chain of the N-terminal lysine and DABCYL is attached to the C-terminal lysine side chain. The cleavage site is highlighted in red.*

The fluorogenic assay was set-up in a 96-well plate reader with different concentrations of the final compounds in each well (0, 0.005, 0.05, 0.5, 5, 50, 500 µM). Each well also contained buffer (10 mM HEPES, 10 mM DTT, 0.5 mM EDTA, 0.1% BSA at pH 7.2), 25 nM human calpain-I and 1 µM of the fluorogenic substrate. The excitation wavelength was 490 nm and the emission wavelength of 520 nm was detected by the plate reader. The fluorescence of the assay solutions in each well was monitored prior to the activation of calpain-I to

obtain a baseline. Due to the internal fluorescence quenching of the substrate no signal was detected. 5 mM $CaCl_2$ was then added to initiate the calpain activity. Upon activation calpain-I cleaves the peptide bond between the tyrosine (Tyr) and the glycine (Gly) of the substrate. The cleaved substrate allowed the fluorescence of FAM to be detected (Figure 2.10). The rate of increase of fluorescence intensity was directly proportional to the activity of the enzyme. Thus increasing the concentration of inhibitor suppressed the rate of increase of fluorescence intensity. The concentration of inhibitor that was required to suppress the rate of increase of fluorescence intensity by half was recorded as the IC_{50} value of each inhibitor.

Compound	IC_{50} (µM)	Compound	IC_{50} (µM)
64	21.9 ± 8.4	77	0.006 ± 0.003
65	11.9 ± 8.9	79	0.25 ± 0.21
66	21.0 ± 5.2	80	0.006 ± 0.003
68	19.0 ± 6.2	81	0.10 ± 0.09
69	4.2 ± 1.8	82	0.01 ± 0.007
70	4.0 ± 1.2	83	0.21 ± 0.10
71	4.3 ± 2.1	84	0.007 ± 0.002
72	6.9 ± 1.4	85	0.17 ± 0.09
73	2.5 ± 1.4	86	0.12 ± 0.009
74	1.5 ± 0.9	87	0.01 ± 0.01
75	0.01 ± 0.009	88	0.002 ± 0.001
76	0.02 ± 0.01	89	0.09 ± 0.09

Table 2.1: *The IC_{50} values (µM) of the α-mercaptoacrylic acids for the inhibition of calpain-I. Assays were performed with the calpain-I specific FRET substrate (1 mM), calpain-I (25 nM) and an assay buffer at pH 7.2 consisting of HEPES (10 mM), DTT (10 mM), EDTA (0.5 Mm) and BSA (0.1%). Reactions were initiated by addition of $CaCl_2$ (5 mM). Each figure quoted is the average of four readings.*

The data obtained revealed that the phenyl derivatives were all potent inhibitors with IC_{50} values in the micromolar region with the most potent derivatives to be 4-substituted phenyl mercapotacrylates. The 2-substituted compounds proved the least potent implying that the favoured position for substitution in this series is as far away as possible from the mercaptoacrylate

group. Bromine substituted compounds proved to be the most potent inhibitors, with **74** ((*Z*)-3-(4-bromophenyl)-2-mercaptoacrylic acid) having an IC$_{50}$ of 1.5 μM. Therefore in order of potency of the final compounds the position of the substituent favoured is *2<3<4* and the nature of the halogen substituent varies as; F<Cl<Br≈I.

The data for the indole derivatives indicates that the potency of the derivative significantly increases as the aromatic ring system changes from a phenyl ring to a 3-substituted indole system, as these compounds all gave sub-micromolar IC$_{50}$ values. In contrast to the phenyl derivatives, the most potent derivatives had the halogen in position 4 and 7 on the indole ring, position 4 being closer to the mercaptoacrylic acid substituent than positions 5 and 6. The compounds that contain halogens in position 4 all gave IC$_{50}$ values of 20 nM or lower. When the halogen was substituted into position 7 the inhibitors gave IC$_{50}$ values of 120 nM or lower. The compounds containing halogens at positions 5 and 6 on the indole ring are less potent towards calpain-I than those with substituents at positions 4 and 7, though the IC$_{50}$ values obtained were still in the nanomolar region. The reason for the reduced potency of the compounds may be the size of the pocket that the inhibitor resides in, which may be too small for the bulky iodine. Fluorine and chlorine substituents at positions 5 and 6 may be disfavoured due to unfavourable electronic interactions with the halogens within the hydrophobic pocket. The most potent inhibitors contained a bromine substituent regardless of the position of the halogen, with the IC$_{50}$ values of the bromine substituted indole based inhibitors all in the low nanomolar region, ranging from 2 nM to 7 nM.

Therefore the most potent inhibitors of calpain-I in these two series contained bromine substituents. The indole based mercaptoacrylic acids were more potent inhibitors of calpain I than the phenyl based inhibitors.

2.8.2 Calpain-I and calpain-II assay

The substrate that was used for this assay was not specific for calpain-I and could be used to determine the inhibition of calpain-II. Therefore the results obtained for this assay with both isoforms were directly comparable. This assay

was also a FRET based assay (Section 2.8.1) and was carried out in a 96-well plate reader. The substrate used in this assay contained 7-amido-4-coumarin (AMC), which fluoresces prior to cleavage but not as intensely as it does once hydrolysed. The peptide substrate was Leu-Leu-Val-Tyr, with a succinic acid attached at the N terminus and AMC attached at the C terminus (Figure 2.11).[207] Upon activation of calpain through addition of $CaCl_2$ the protease hydrolyses the peptide bond between tyrosine and AMC. The hydrolysis of the bond relieves the quenching effect of the peptide and fluorescence was detected at the emission wavelength 460 nm.[207]
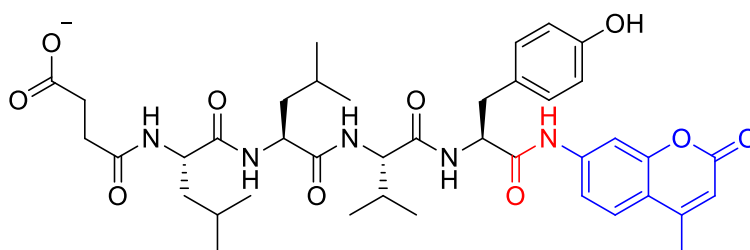
Figure 2.11: *The structure of Suc-Leu-Leu-Val-Tyr-AMC. The cleavage site is highlighted in red.*

Testing with calpain-I

The buffer and the method used for this assay were the same as that used for the calpain-I specific assay, with $CaCl_2$ used as the source of calcium for the initiation of the assay. The excitation wavelength of the fluorophore was 360 nm and the emission wavelength 460 nm. The concentrations of the inhibitors used in the assays were 0, 0.005, 0.05, 0.5, 5, 50, 500 µM. The IC_{50} values were defined as the concentration of inhibitor required to suppress the rate of the increase of fluorescence by 50% (Table 2.2).

The IC_{50} values that were obtained for the non-specific assay were much higher than those measured for the calpain-I specific assay (Table 2.2). The poor specificity of the substrate may be the cause for the discrepancy in the IC_{50} data that was obtained from the two assays (Table 2.1 and Table 2.2). The same trends were noted however for both the position and nature of the halogen substrate in the most potent phenyl based substrate. Most of the 4-substituted analogues gave low micromolar IC_{50} values in this assay.

Compound	IC$_{50}$ (µM)	Compound	IC$_{50}$ (µM)
64	23.2	77	0.10
65	20.7	79	0.05
66	9.1	80	0.02
68	31.5	81	0.09
69	13.1	82	0.30
70	8.4	83	0.10
71	23.9	84	0.09
72	5.6	85	0.30
73	15.0	86	0.11
74	1.5	87	0.09
75	0.30	88	0.07
76	0.14	89	0.22

Table 2.2: *The IC$_{50}$ values of the α-mercaptoacrylic acid derivatives determined from the AMC based FRET assay.*

The indole based mercaptoacrylic acid derivatives, when tested with this assay, gave IC$_{50}$ values that were in the nanomolar region. Therefore the indole derivatives have a higher potency than the phenyl derivatives, corroborating the data obtained from the calpain-I specific assay (Section 2.8.1). Bromoindole inhibitors displayed a higher potency towards the enzyme than any other halo-substituted inhibitors, with the IC$_{50}$ values ranging from 20 nM to 100 nM. The data obtained from this assay does not appear to show a definitive trend towards where the halogen should be positioned on the indole ring. However there is an indication that position 5 and position 7 of the indole ring are important for potency. This contradicts the data obtained from the calpain-I specific assay (Section 2.8.1), where position 4 not position 5 was important for a higher potency.

Testing with calpain-II

The AMC-based assay was also used to test the inhibitors against calpain-II. The same assay conditions that were utilised for testing the AMC fluorogenic substrate with calpain-I were employed (Section 2.8.2). The results indicated

that these compounds are selective towards calpain-I over calpain-II, but remain capable of inhibiting calpain-II (Figure 2.12).
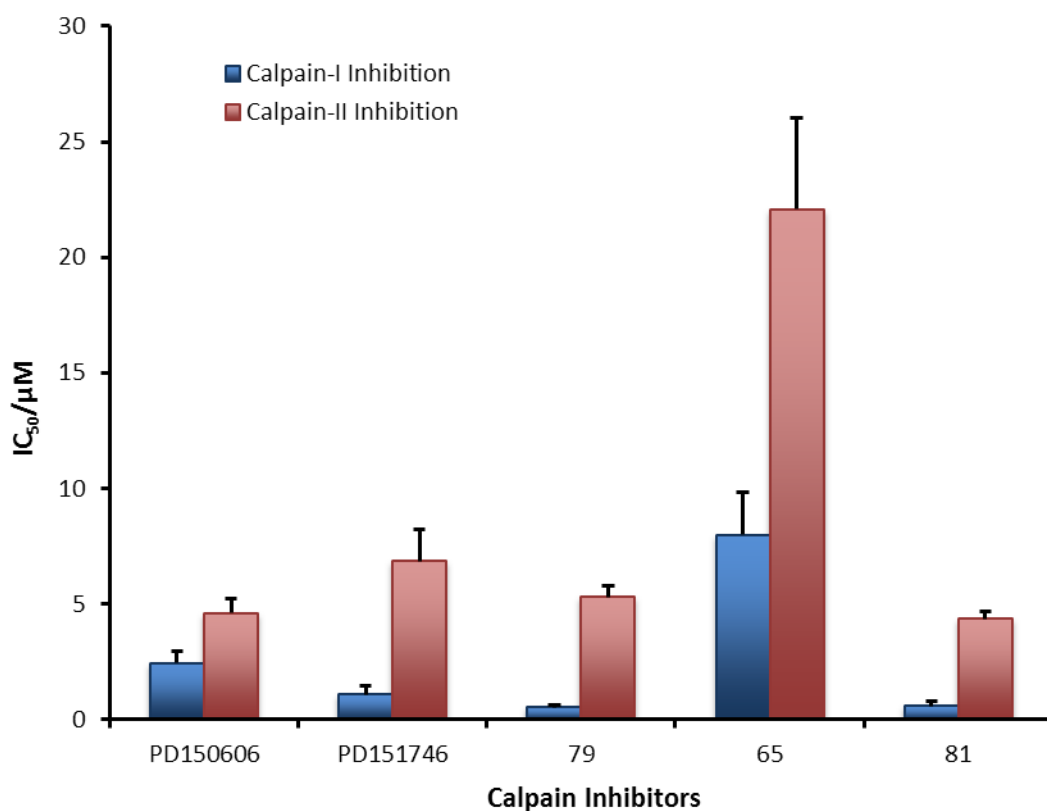
Figure 2.12: *A graph representing the difference in potency of selected mercaptoacrylic acid compound towards calpain-I and calpain-II. This data was measured with the fluorogenic AMC substrate.*

The results demonstrate that the inhibitors are moderately selective for calpain-I over calpain-II, no matter whether their structure is a phenyl based or indole based derivative (Figure 2.12). The indole based inhibitors were found to have a slightly higher selectivity towards calpain-I than the phenyl based derivatives however.

2.8.3 In cell testing of inhibitors

Live neutrophils purified from human blood were used to determine whether these inhibitors could pass through cell membranes. The cells had to be used immediately due to the short lifetime of neutrophils (5-7 days). The cells were suspended in Krebs medium (120 mM NaCl, 4.9 mM KH_2PO_4, 1.2 mM $MgSO_4$, 1.2 mM $CaCl_2$, 1.3 mM KCl, 25 mM HEPES and 0.1% BSA at pH 7.4) along with 10 μM of each of the inhibitors (added as a 50 mM solution in DMSO). 100 μl of the

cell solution containing the inhibitor was loaded onto a coverslip containing 100 μl of the Krebs medium prewarmed to 37 °C. Once the cells make contact with the glass from the coverslip they spontaneously start to spread. The initial cell area was monitored with a Leica microscope and images were collected over a 3 minute period whilst the cells spread. Approximately 50 cells were monitored so an average surface area could be determined. 0.1% DMSO was used as a control to determine whether this solvent has an effect upon the neutrophils since this was the solvent that was used to create the stock solutions of the mercaptoacrylic acid compounds.
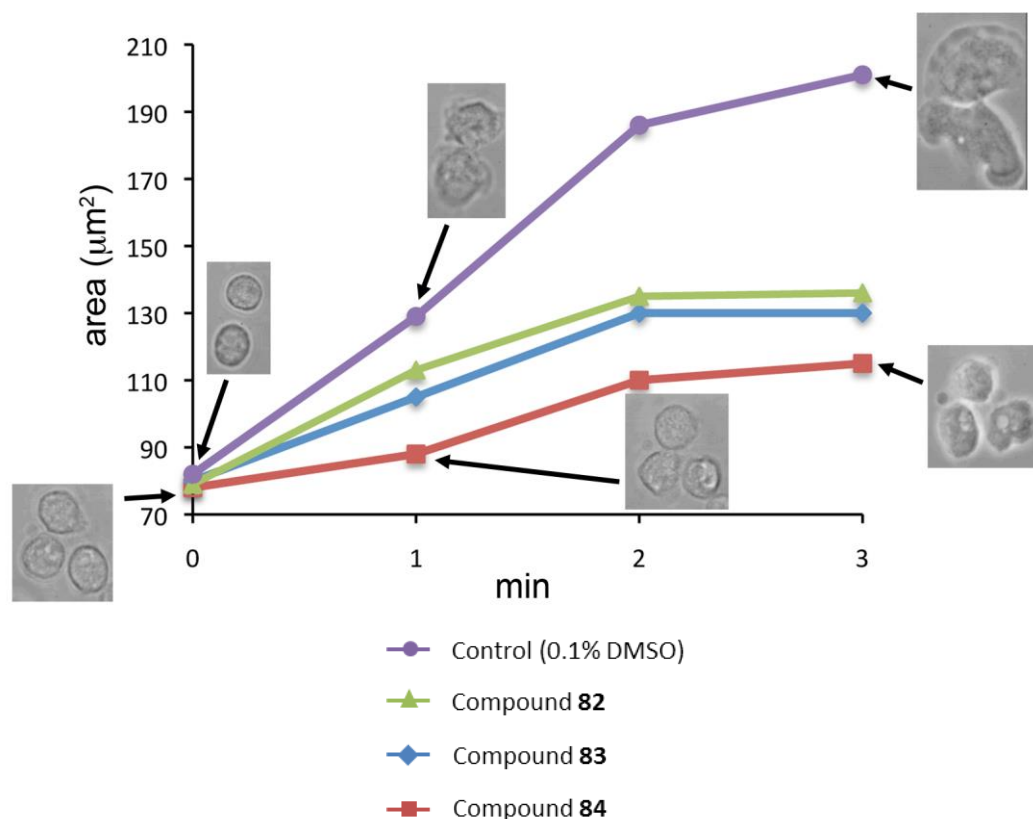
Figure 2.13: *A graph demonstrating that the addition of 10 μM of 6 substituted indole inhibitors slows down the rate of neutrophil spreading. The control was performed with 0.1% DMSO and showed a150% increase in cell surface area during the assay, the inhibitors reduced the cell spreading area by 50 – 70%.*

The cells in the control spread ~2.5× the original area of the cell. The control assay demonstrated the DMSO has little effect on the cell spreading capability of the neutrophil (Figure 2.13). The addition of the 6-substituted indole mercaptoacrylic acids inhibited the capability of the cells' spreading process. Compound **84** appeared to be the most effective at slowing the cell spreading process with a reduction of ~70% in comparison to the control. Both **83** and **82**

were less effective but curtailed the neutrophil spreading with a reduction of ~60% and ~50%, respectively (Figure 2.13). This shows clearly that calpain-I is instrumental in accelerating the cell spreading process as its inhibition within a live cell causes the cell spreading process to slow down. These inhibitors do not completely stop the capability of the cells to spread - this could be disastrous for the immune system. They do however slow the cell spreading process down, which may be beneficial as a treatment for inflammatory diseases.

2.9 Photochemical effect

When the inhibitors were tested in neutrophils, an intriguing phenomenon was observed. If the cells containing the inhibitor were irradiated with 410 nm light the cells would lose their ability to spread. 410 nm light is normally used to release caged inositol triphosphate (IP$_3$) within the cell, as the release of this messenger stimulates release of calcium ions from intracellular stores.[208-210] Caged IP$_3$ is microinjected into individual neutrophils, and upon irradiation the cells begin to spread through the release of calcium and the activation of calpain-I.[211]

The neutrophils were observed with a Leica microscope and were suspended in Krebs medium (Section 2.8.3) and 10 μM of **81** ((*Z*)-3-(5-iodoindol-3-yl)-2-mercaptoacrylic acid). The cells were irradiated, with a 410 nm laser, in a select area in the field of view of the microscope, then for the activation of the cell spreading process the small peptide formyl-methionyl-leucyl-phenylalanine (FMLP) was added. FMLP is a known chemoattractant that influences the activation of neutrophils and the addition of this peptide causes the cells to spread and flatten and to migrate towards the source of the chemoattractant.[212, 213] After 5 minutes incubation with FMLP, the cells contained within the irradiated area remained spherical, whereas all the cells that were outside of the area had spread and formed a flatter shape (Figure 2.14).

In section 2.8.3 it was demonstrated that the addition of the α-mercaptoacrylic acid derivatives slowed down the rate of neutrophil cell spreading. The fact that upon irradiation the ability for the cells to spread is completely lost implies that

calpain is permanently inhibited, as it has been shown that this protease has an integral role in the cell spreading process. The apparent irrevesible inhibition of calpain-I hints that irradiation might affect the structure of **81** and cause it to permanently bind to the enzyme.
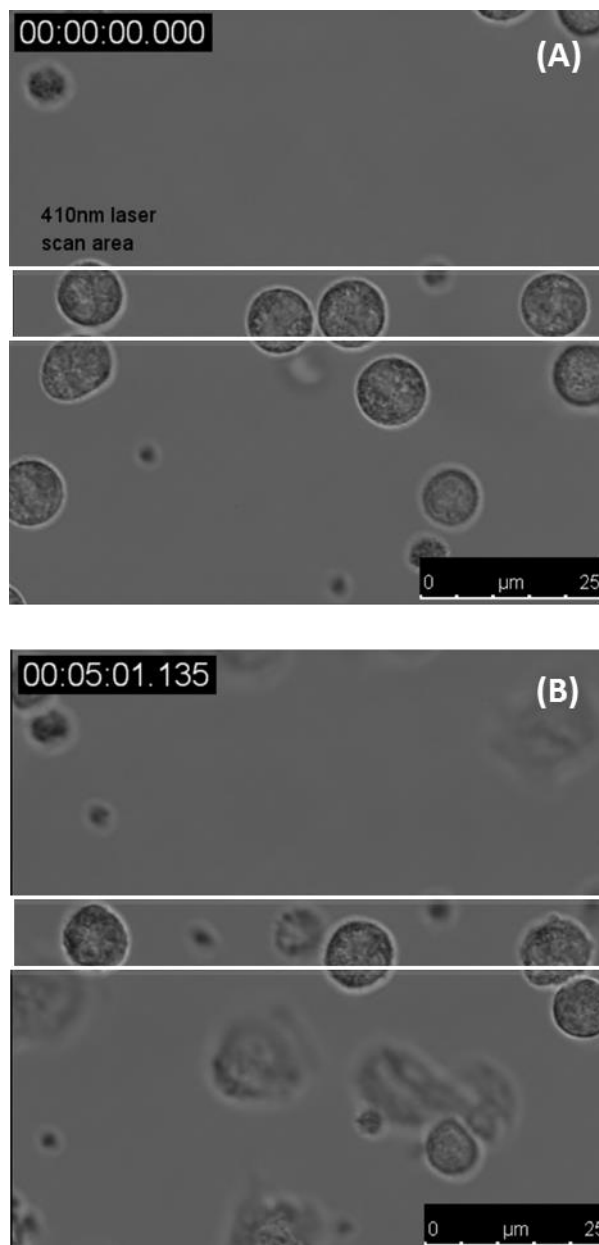
Figure 2.14: *(A) A microscope image prior to the addition of the peptide FMLP, the area with the cells that were irradiated with 405 nm light is highlighted in blue, (B) An image collected 5 minutes after FMLP addition, the cells within the irradiation area are held in a spherical shape.*

2.9.1   Photo-inhibition of calpain-I

To determine whether the photo-inhibition reaction occurs with calpain-I the AMC assay (Section 2.8.2) was used. 20 nM of calpain-I was used under four

different assay conditions in the presence or absence of 1 µM of **81** and with or without irradiation (Figure 2.15). Samples were irradiated for 5 minutes with a 100 W light source with a 360 nm band pass filter. 100 µl aliquots were then placed in a 96 well plate and 5 mM $CaCl_2$ was added to activate the enzyme. The cleavage of the substrate was monitored continuously for an hour with the excitation wavelength set to 360 nm and the emission wavelength set to 460 nm.

Conditions	Calpain-I	Inhibitor	Irradiation
1	20 nm	N/A (DMSO control)	N/A
2	20 nm	N/A (DMSO control	360 nm (5 min)
3	20 nm	**81**	N/A
4	20 nm	**81**	360 nm (5 min)
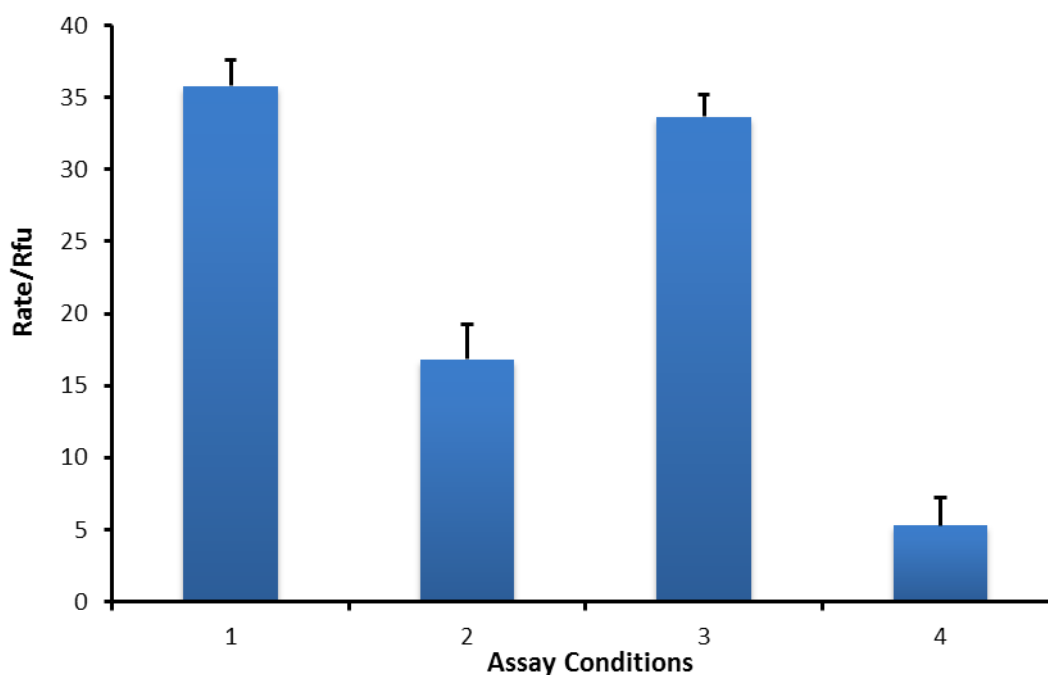
Figure 2.15: *The conditions used for the irradiation experiment on calpain-I using the AMC assay and the relative rates of the increase in fluorescence in the AMC assay, with the different irradiation conditions in calpain-I.*

The results of the AMC fluorescence assay revealed that the irradiation of the protein alone has an effect on the protease itself with a 50% drop in the activity of the protease (Figure 2.15). The addition of 1 µM of **81** showed that without light there was a minimal drop in the overall rate of the increase in fluorescence, ~6%. Irradiated inhibitor however suppressed the activity of the enzyme by approximately 85%. The irradiated inhibitor sample shows a

decrease in the enzyme activity by 31%, when compared to the irradiated control sample. Taking all these results into account, irradiation of the inhibitor dramatically decreases the overall activity of calpain-I. Comparison of this with the activity of the enzyme in the presence of inhibitor without light indicates that the light has a dramatic effect on the potency of the inhibitor, increasing it greatly.

The data from the AMC assay and the cell spreading assay might demonstrate that the light has a dramatic effect on the potency of the inhibitor, however it appears that damage may also occur to the enzyme itself as the AMC assay showed a decrease in the activity of the enzyme by 50% upon irradiation.

2.10 Conclusions

The synthetic routes that were employed for the production of these α-mercaptoacrylic acids were generally high yielding and generated a library of potent inhibitors of calpain-I. The two step synthesis to form the mercaptoacrylic acids from substituted benzaldehydes or indole-3-carboxaldehydes gave overall yields varying from 50% to 94% with a few minor exceptions. For the synthesis of the indole-3-carboxaldehydes a Vilsmeier-Haack procedure was employed to transform suitably substituted indoles, with yields of the desired products ranging from 54-98%.[189] The Bartoli indole synthesis was required for the production of 7-substituted indoles; 14% and 43% yields were recorded for 7-fluoroindole and 7-bromoindole respectively.[192] For the five step synthesis that yielded 6-iodoindole, each step recorded a yield that was at least 50% or above, this resulted in an overall yield of 22%. The thallium trifluoroacetate based chemistry required for the synthesis of 4-iodoindole gave an overall yield of 43% for the three steps, although unfortunately the 4-iodoindole final product could not be prepared with this methodology due to an undesired *ipso*-substitution reaction in the final hydrolysis step.

The analogues proved to be potent and moderately selective towards calpain-I, with the size and position of the halogen important for the potency of inhibition. The bromide analogues proved to be most potent analogues of this class

prepared to date. The assays demonstrated that the larger indole ring system, gave inhibitors that were more potent than the phenyl based inhibitors.

The testing of inhibitors within live neutrophils revealed that these mercaptoacrylic acid analogues were effective at slowing down the rate of cell spreading, providing further evidence that calpain-I is instrumental in accelerating the cell spreading process. Additional experiments corroborated this when irradiation of neutrophils with 410 nm light completely ceased the cells spreading process. Testing calpain-I with the AMC assay and with incident light revealed that the potency of the inhibitor dramatically increases when light is shone upon the enzyme-inhibitor complex.

This work demonstrates that the mercaptoacrylic analogues prepared here may lead to a viable small molecule based treatment of the autoimmune disease rheumatoid arthritis.

3 PEF(S); Expression, Purification and Characterisation

3.1 α-Mercaptoacrylic acids and PEF(S)

X-ray crystal structures demonstrated that the phenyl based α-mercaptoacrylic acid inhibitor, PD150606, binds to a hydrophobic pocket of PEF(S) (Figure 3.1).[81, 186] This is also the pocket that binds inhibitory region C of calpastatin.[81] The aromatic ring on PD150606 interacts with leucine and phenylalanine residues within this pocket and the α-mercaptoacrylic acid moiety interacts with hydrophilic residues on the surface of the protein.
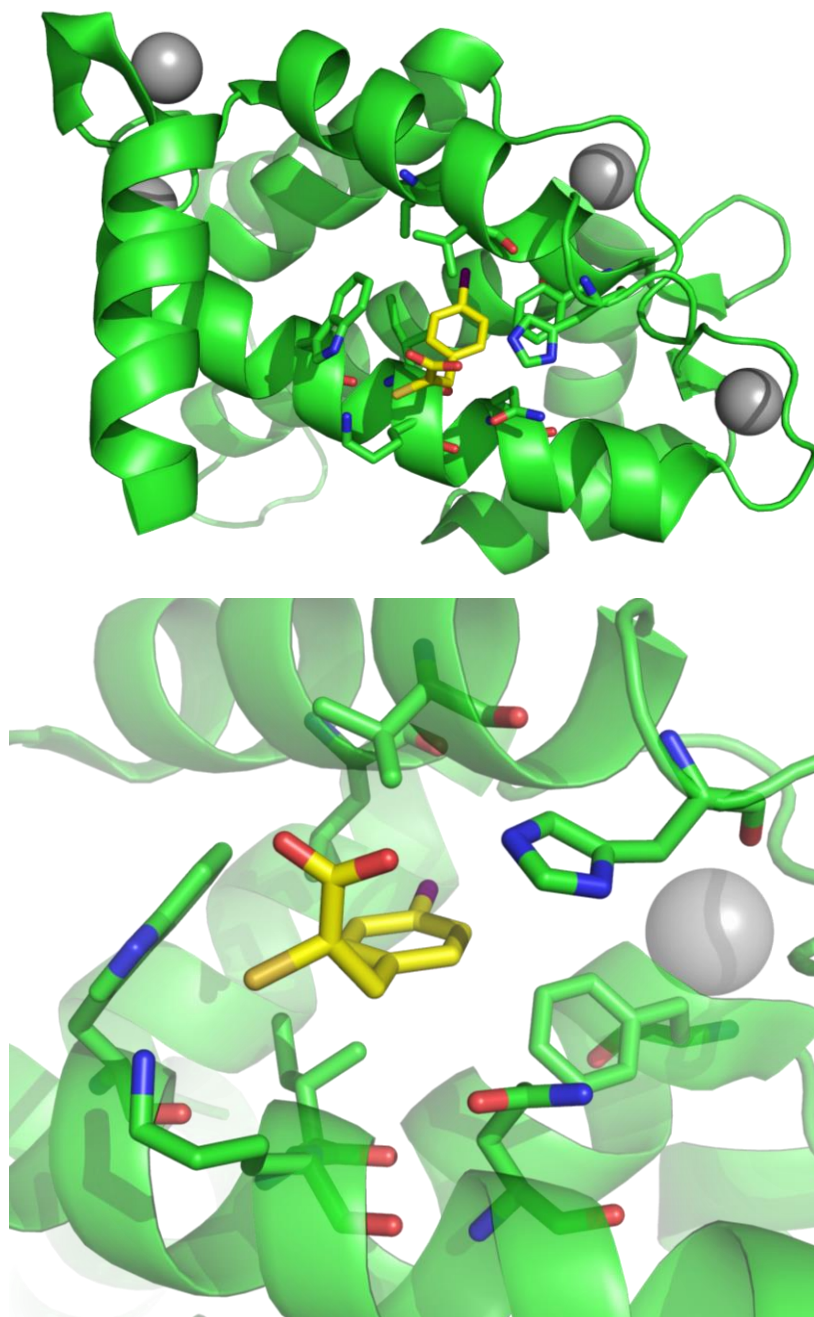
Figure 3.1: *A cartoon representation of PD150606, bound to PEF(S) of calpain, the aromatic ring of the inhibitor binds in a hydrophobic pocket (PDB:1NX3).[186]*

To determine whether the α-mercaptoacrylic acid derivatives described in Chapter 2 bind to calpain-I in a similar fashion, a gene encoding for the porcine PEF(S) (MW 19,992) was expressed in *E. coli* and the resulting protein purified and characterised.

3.2 Sub-cloning *PEF(S)* into an expression vector
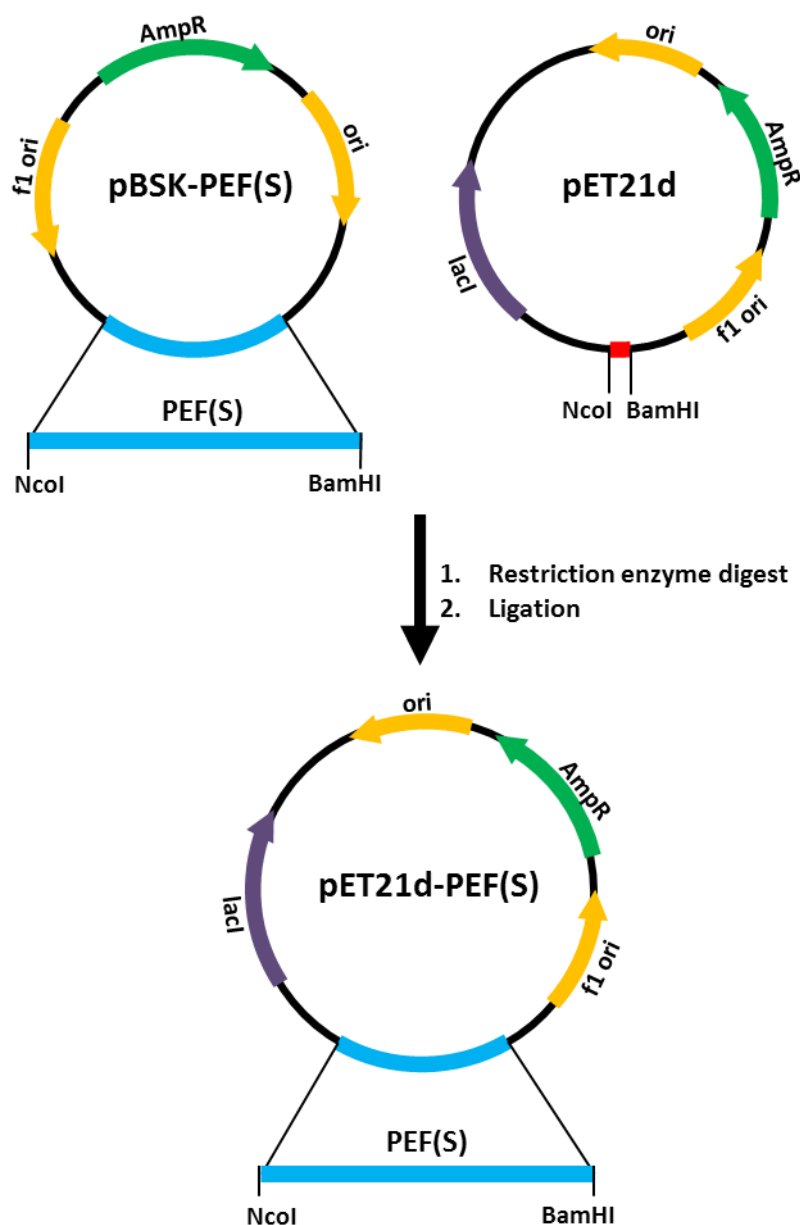
Figure 3.2: *Plasmid maps of pBSK-PEF(S) with PEF(S) gene highlighted in blue and the expression vector pET21d, the unique restriction sites are noted on the maps. Both vectors were digested with NcoI and BamHI, the PEF(S) gene was ligated into the pET21d vector.*

A cDNA for porcine PEF(S), codon optimised for *E. coli*, was purchased from Epoch in a pBSK vector (pBSK-*PEF(S)*) (Epoch Biolabs, Texas, USA). The gene was located between an *Nco*I and a *Bam*HI restriction site (Figure 3.2). *PEF(S)* was sub-cloned into pET21d vector (Figure 3.2) for expression.

3.2.1 Restriction endonuclease digestion
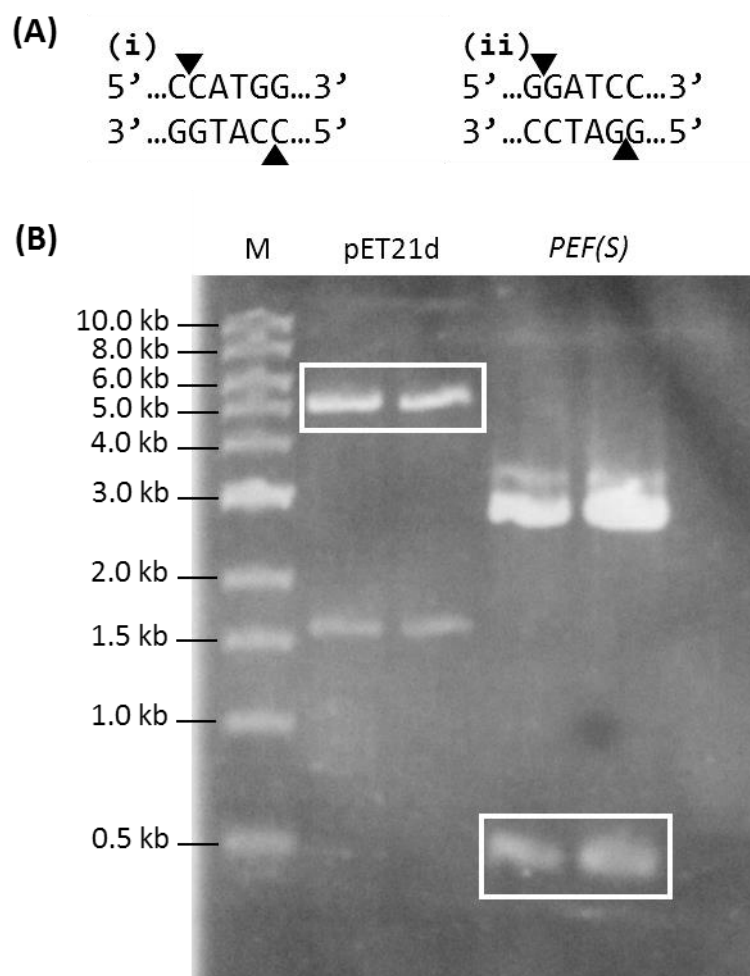
Figure 3.3: *(A) The sequences of recognition by the restriction endonucleases NcoI (i) and BamHI (ii). The black triangles represent the site of cleavage. (B) An image of the agarose gel of the products resulting from digestion of the pBSK vector containing the PEF(S) gene, 530 base pairs, (highlighted in a white box) and the pET21-d expression vector (highlighted in a white box), M = marker.*

The *PEF(S)* gene was excised from the pBSK vector using restriction endonucleases NcoI-HF (high fidelity) and BamHI. Both pET21d and pBSK-*PEF(S)* were incubated with the restriction endonucleases sequentially. The DNA was incubated at 37 °C for 3 hours with NcoI-HF followed by BamHI for 4 hours (Section 6.3.6). The same sequence of digestions was carried out on the

pET21d vector. Agarose gel electrophoresis separated the fragments (Figure 3.3). The bands of DNA corresponding to the correct molecular weight for the *PEF(S)* gene and digested pET21d expression vector were retrieved from the gel and used for the ligation of the pET21d-*PEF(S)* construct.

3.2.2 DNA Ligation

For the ligation, pET21d and PEF(S) were incubated with T4 ligase in ratios of 1:3 and 1:6. The resulting ligated DNA was then incubated and was used to transform *E. coli* XL1-Blue super competent cells. Following a 10 ml overnight culture the DNA was purified using QIAprep spin miniprep kit according to the manufacturers instructions. The DNA was sequenced to reveal that the *PEF(S)* had been successfully ligated into pET21d (Figure 3.4).

```
ATTTTGTTTACTTTAAGAAGGAGATATACCATGGAGGAAGTCCGCCAATTTCNCCGTCTG
 I   L   F   T   L   R   R   R   Y   T   M   E   E   V   R   Q   F   X   R   L
TTTGCACAACTGGCAGGTGATGATATGGAAGTTAGTGCGACAGAACTGATGAACATCCTG
 F   A   Q   L   A   G   D   D   M   E   V   S   A   T   E   L   M   N   I   L
AATAAAGTTGTTACACGTCATCCGGATCTGAAAACTGATGGGTTTGGGATCGACACGTGT
 N   K   V   V   T   R   H   P   D   L   K   T   D   G   F   G   I   D   T   C
CGTAGCATGGTGGCCGTGATGGATAGCGACACGACTGGTAAACTGGGCTTCGAAGAATTC
 R   S   M   V   A   V   M   D   S   D   T   T   G   K   L   G   F   E   E   F
AAATACCTGTGGAATAACATCAAGAAATGGCAGGCGATCTACAAACAATTTGATGTGGAT
 K   Y   L   W   N   N   I   K   K   W   Q   A   I   Y   K   Q   F   D   V   D
CGCAGCGGTACCATTGGTAGTAGCGAACTGCCGGGTGCTTTTGAAGCGGCTGGCTTTCAC
 R   S   G   T   I   G   S   S   E   L   P   G   A   F   E   A   A   G   F   H
CTGAACGAGCATCTGTATTCTATGATTATCCGTCGCTATAGTGATGAGGGCGGCAATATG
 L   N   E   H   L   Y   S   M   I   I   R   R   Y   S   D   E   G   G   N   M
GACTTCGATAACTTCATTTCTTGCCTGGTACGTCTGGACGCCATGTTCCGCGCCTTCAAG
 D   F   D   N   F   I   S   C   L   V   R   L   D   A   M   F   R   A   F   K
TCTCTGGACAAAGACGGCACCGGCCAGATTCAGGTCAATATTCAGGAGTGGCTGCAACTG
 S   L   D   K   D   G   T   G   Q   I   Q   V   N   I   Q   E   W   L   Q   L
ACCATGTATAGCTAATGAGGATCCGAATTCGAGCTCCGTCGACAAGCTTGCGGCCGCACT
 T   M   Y   S   -   -   G   S   E   F   E   L   R   R   Q   A   C   G   R   T
CGAGCACCACCACCACCACCACTGAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGA
 R   A   P   P   P   P   P   L   R   S   G   C   -   Q   S   P   K   G   S   -
```

Figure 3.4: *DNA sequence of the* PEF(S) *gene inserted (red) into the pET21-d expression vector (black), with the* NcoI *and* BamHI *restriction sites highlighted in blue. The amino acid sequence that the gene encodes for is shown below the DNA sequence, as a one letter code.*

3.3 Expression of the *PEF(S)* gene

Expression tests were carried out on the pET21d-*PEF(S)* construct using the *E. coli* expression strains BL21-CodonPlus(DE3)-RP [214] and BL21-CodonPlus(DE3)-RIL.[214] The cells were grown in 500 ml of ampicillin (1 μg/ml) selective LB medium at 37 °C (shaking at 150 rpm), until an optical density (OD) of 0.6 a.u. at 600 nm was reached. IPTG (63 mg) was then added to give a total concentration of 0.5 mM and the expression was carried out at 37 °C for 4 hours. A 1 ml sample of the culture was collected every hour. The samples that were collected were analysed by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) (Section 6.3.7) to determine whether the expression of the gene had been a success (Figure 3.5). The SDS-PAGE indicated that the optimum time for expression of the gene was 3 hours at 37 °C using BL21-CodonPlus(DE3)-RP *E. coli* cells.
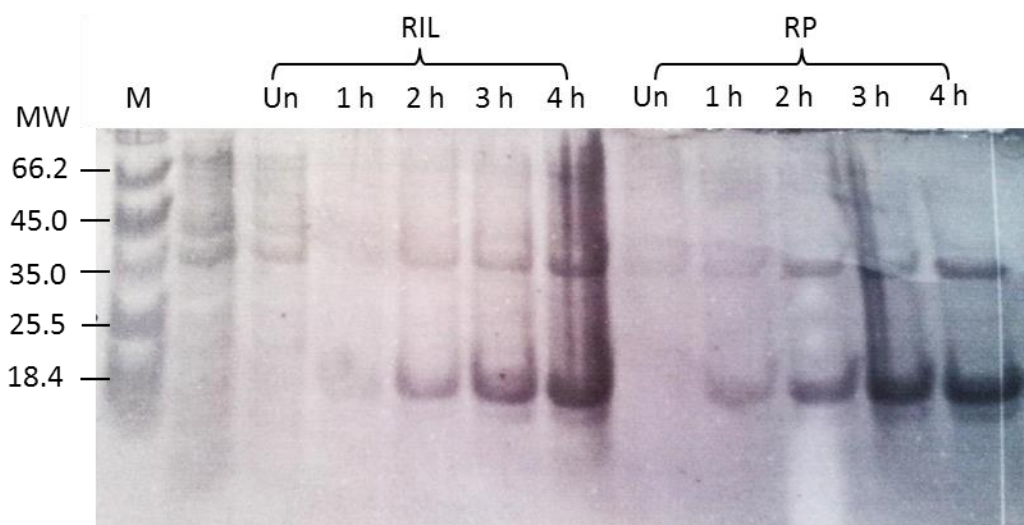
Figure 3.5: *The SDS polyacrylamide gel that was used to assess the performance of the small scale expression experiments. M = molecular weight marker, Un = uninduced cells, 1 h = 1 hour after induction, 2 h = 2 hours after induction, 3 h = 3 hours after induction, 4 h = 4 hours after induction, RIL = BL21-CodonPlus(DE3)-RIL, RP = BL21-CodonPlus(DE3)-RP and MW = molecular weight (× 1000).*

3.3.1 Large scale expression and purification of *PEF(S)*

Transformed BL21-CodonPlus(DE3)-RP *E. coli* cells were grown in 6 x 500 ml of ampicillin (1 μg/ml) selective LB media at 37 °C (shaking at 150 rpm). Once the cells reached an OD of 0.6 a.u. at 600 nm, the cells were induced with IPTG. After further incubation for 3 hours the cells were harvested by high-speed

centrifugation at 4 °C (6080 RCF, 15 min). The pellet was re-suspended in cell lysis buffer (Tris-Base (20 mM), βME (5 mM) and EDTA (2 mM) at pH 8.0). After sonication and clarification by centrifugation (30310 RCF, 30 min), the pellet and the crude cell extract were analysed by SDS-PAGE revealing that the over-expressed protein was soluble and was clearly visible as a band at a MW ~20,000.

3.4 Purification of PEF(S)

A two-step purification process was required to gain protein of sufficient purity for protein crystallography. Anion exchange chromatography using diethylaminoethyl-sepharose (DEAE) was used to purify the protein from the crude cell extract. The column was attached to a fast protein liquid chromatography instrument (FPLC). The FPLC instrument allowed for an increasing NaCl concentration to be applied to the column whilst the UV absorption of the eluent was monitored at 280 nm. The protein was applied to the column at a salt concentration of 0 M NaCl (Purification buffer A, section 6.3.4). A linear salt gradient was then applied to the column from 0 M to 1 M NaCl (20 mM Tris-Base, 2 mM EDTA and 5 mM βME at pH 8.0) over 6 CV, the eluent was collected in 6 ml fractions (Figure 3.6).

A 20 µl sample was taken from each of the fractions that contained eluted proteins (with an UV absorbance above 500 mAU) and analysed using SDS-PAGE, to determine which fractions contained the desired protein (Figure 3.6). The resulting gel revealed that the protein eluted from the DEAE anion exchange column at a salt concentration of ~0.3 M NaCl. Even though the anion exchange column had purified PEF(S) from many of the other proteins that were found within the cell extract, the protein is required to be at a higher level of purity than that obtained from this purification to be suitable for protein X-ray crystallography. Therefore the fractions that contained the protein were pooled and concentrated for further purification. The solution (~42 ml) was concentrated to 15 ml, using an Amicon ultrafiltration membrane with a molecular weight cut off of 10,000.
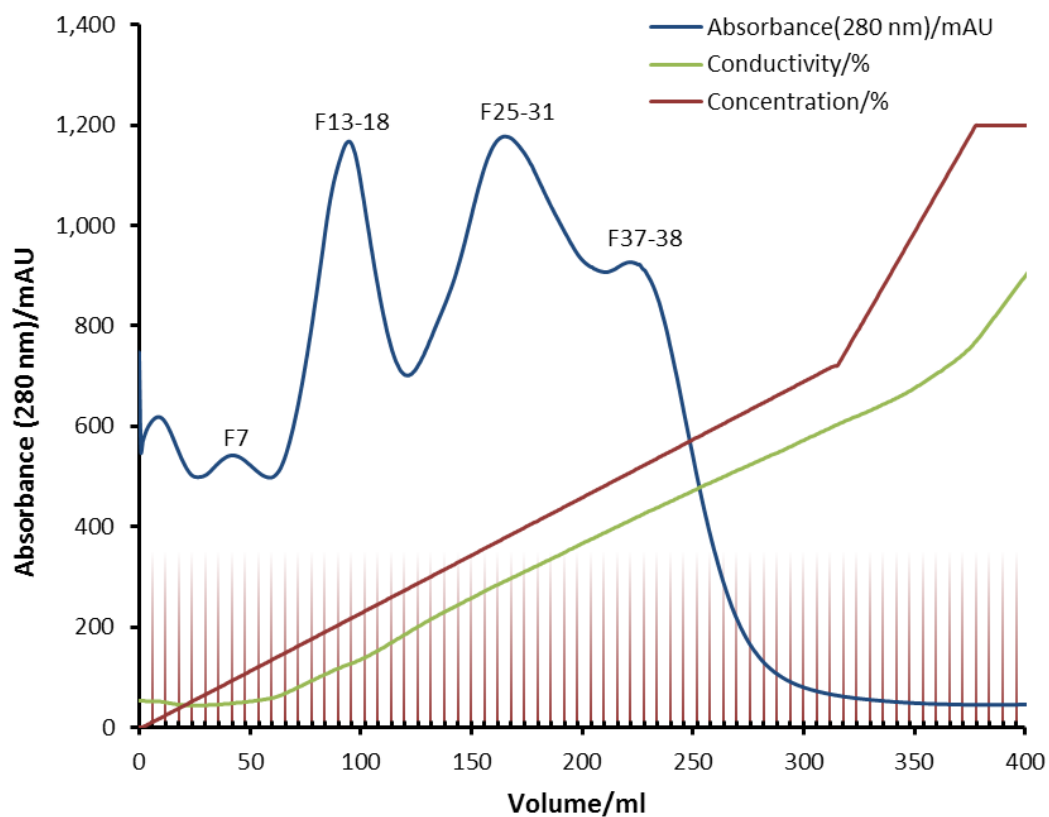
Fraction	Volume/ml	NaCl Concentration/%	NaCl Concentration/M
7	37-43	6.6-7.7	0.07-0.08
13-18	73-109	13.2-18.7	0.13-0.19
25-31	145-187	26.2-34.2	0.26-0.34
37-38	217-229	39.8-42.1	0.40-0.42
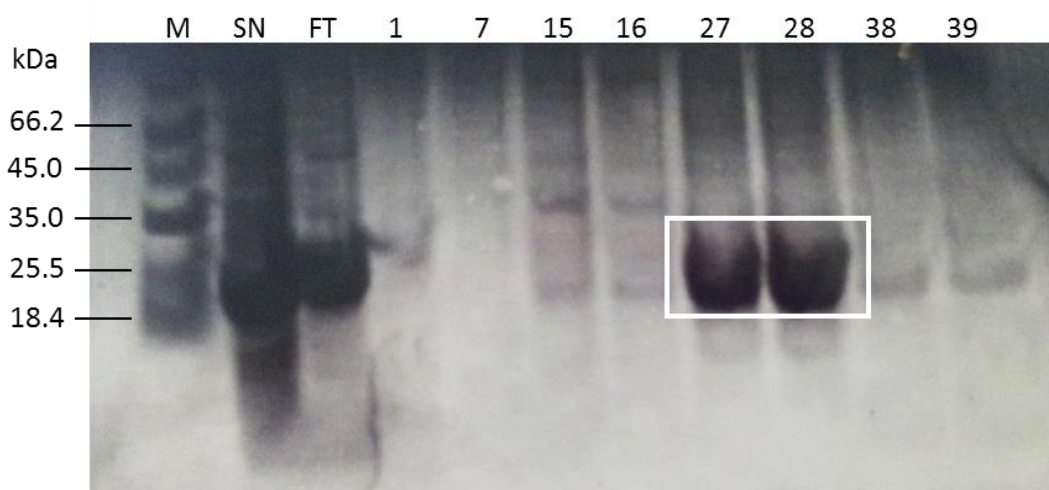
Figure 3.6: *The chromatogram resulting from the purification of the crude cell extract from the expression of porcine* PEF(S) *gene, with the use of a DEAE anion exchange column, the red lines represent the fractions collected. The table shows the fractions collected and their relative column and salt concentration. The SDS-PAGE of the eluted fractions from the DEAE anion exchange column, where M = marker, SN = supernatant solution, FT = flow through and the numbers correspond to the fractions collected with a level of UV absorbance above 500 mAU, detected through the FPLC. The white box represents the desired PEF(S) protein (MW 20,000).*

Size exclusion chromatography

This method of purification, also known as gel filtration, separates proteins according to their molecular weight.[215] Larger molecules elute with a shorter retention time than smaller molecules.[216] The size exclusion column that was used for purification of PEF(S) was a SuperDex™ 75 (GE Healthcare Life Sciences, Buckinghamshire, UK), this column resolves of proteins with a molecular weights ranging from 3,000 to 70,000.[215] The size exclusion column (CV 320 ml) was connected to the FPLC instrument and equilibrated in a buffer containing Tris-Base (20 mM), EDTA (2 mM), βME (5 mM) and NaCl (0.15 M) at a pH of 8.0 (Purification buffer C, Section 6.3.4). The 15 ml solution of PEF(S) protein was applied to the size exclusion column and eluted using 1.5 CV of protein purification buffer C. The eluent from the column was monitored at 280 nm, 10 ml fractions were collected (Figure 3.7).
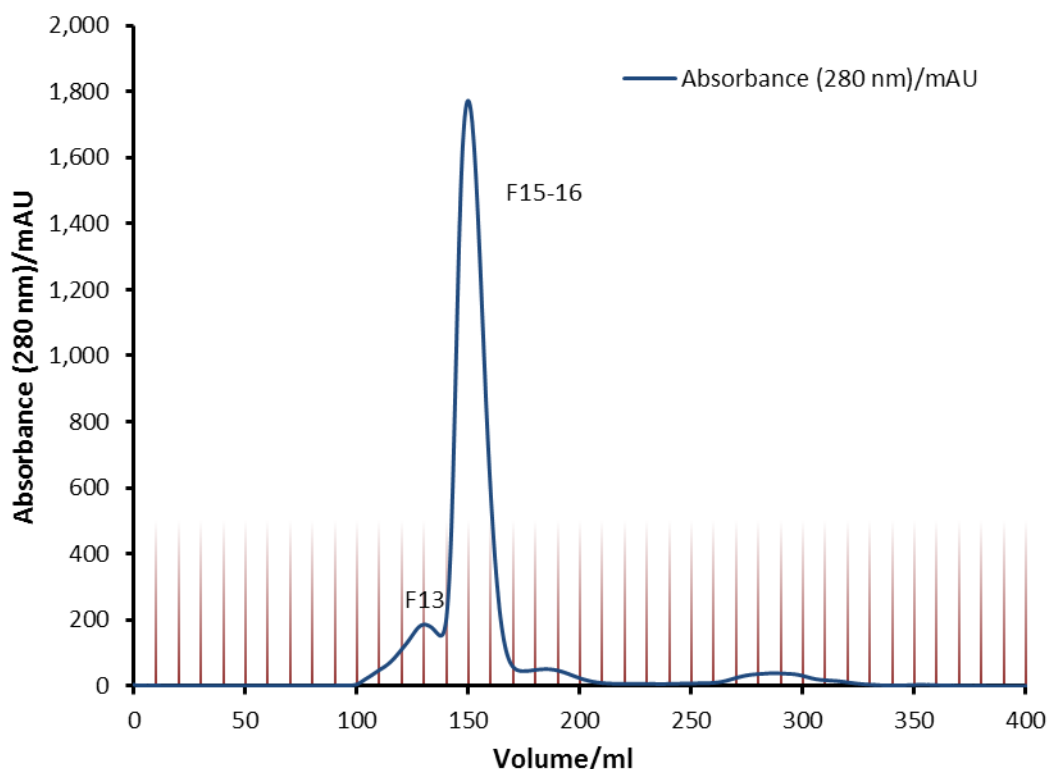
Figure 3.7: *The chromatogram resolution from the purification of PEF(S) with the Superdex 75 size exclusion column, fractions with UV absorbance >100 mAU (280 nm) are labelled. The red lines represent the fractions collected.*

The three fractions with the highest absorbance at 280 nm were analysed with SDS-PAGE (Figure 3.8), to reveal which of them contained PEF(S). The required

protein PEF(S) corresponded to the large peak that absorbed at 1800 mAU (fractions 15 and 16) and it was judged to be ~95 % pure.
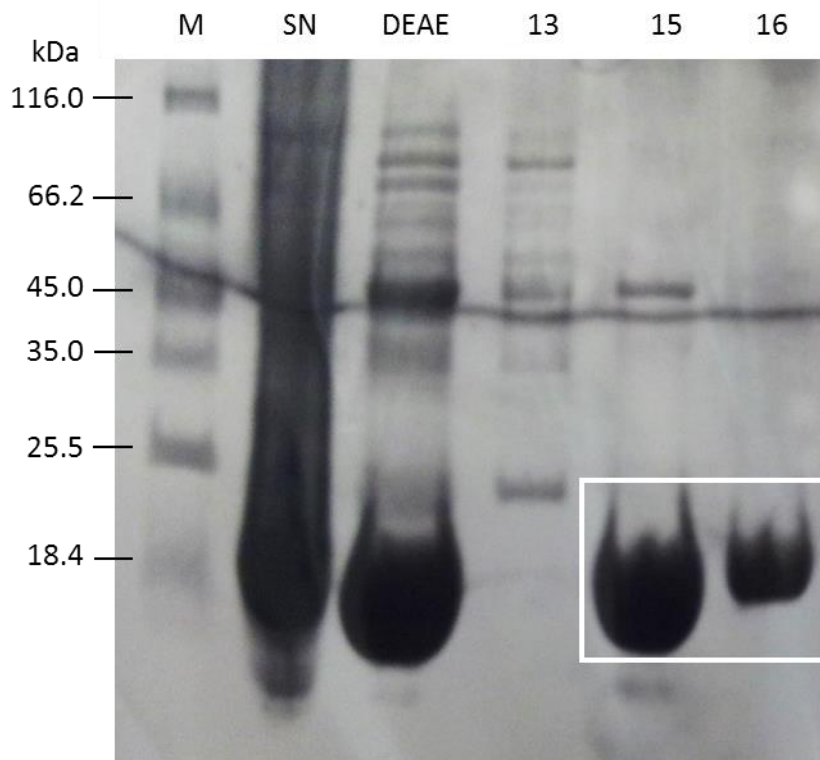
Figure 3.8: *The SDS-PAGE of the fractions that were eluted from the size exclusion column, with M = marker, SN = crude sample before the DEAE column and DEAE = fractions from the DEAE column. The white box highlights the PEF(S) protein (MW 20,000).*

The pooled fractions 15 and 16 were dialysed against a buffer containing K_iPO_4 (5 mM), EDTA (2 mM) and βME (5 mM) at pH 7.0 (Dialysis buffer, Section 6.3.4). The 20 ml solution containing PEF(S) was concentrated to 2 ml by Amicon ultrafiltration as above. The concentration of the resulting protein solution was measured with UV-Vis spectroscopy, at different wavelengths (210, 215 and 220 nm) (Section 6.3.7).[217] The result of the concentration measurement revealed that the protein was capable of reaching high concentrations in solution without precipitating. A final concentration of 2.2 mM (43 mg/ml) was measured. The overall yield obtained from 3 l culture was approximately 86 mg of protein.

3.5 Characterisation of PEF(S)

In addition to SDS-PAGE, the purified protein was characterised using mass spectrometry, circular dichroism spectroscopy and analytical size exclusion chromatography.

3.5.1 Mass spectrometry

The mass of the protein was determined using matrix-assisted laser desorption/ionisation - time of flight mass spectrometry (MALDI-TOF). A 1:1 mixture of the protein solution at 2 µM (in phosphate buffer) and matrix solution (sinapinic acid in acetonitrile/water solution), was placed on a MALDI plate and analysed in a MALDI-TOF Micro MX spectrophotometer. The mass spectrum obtained contained a major peak at *m/z* 19988.81 which is 4 mass units off the predicted mass of the protein (*m/z* 19992.69) (Figure 3.9). A second peak at *m/z* 999.75 is consistent with a protein ion of +2 charge. The molecular ion is consistent with the predicted mass of the PEF(S) domain.
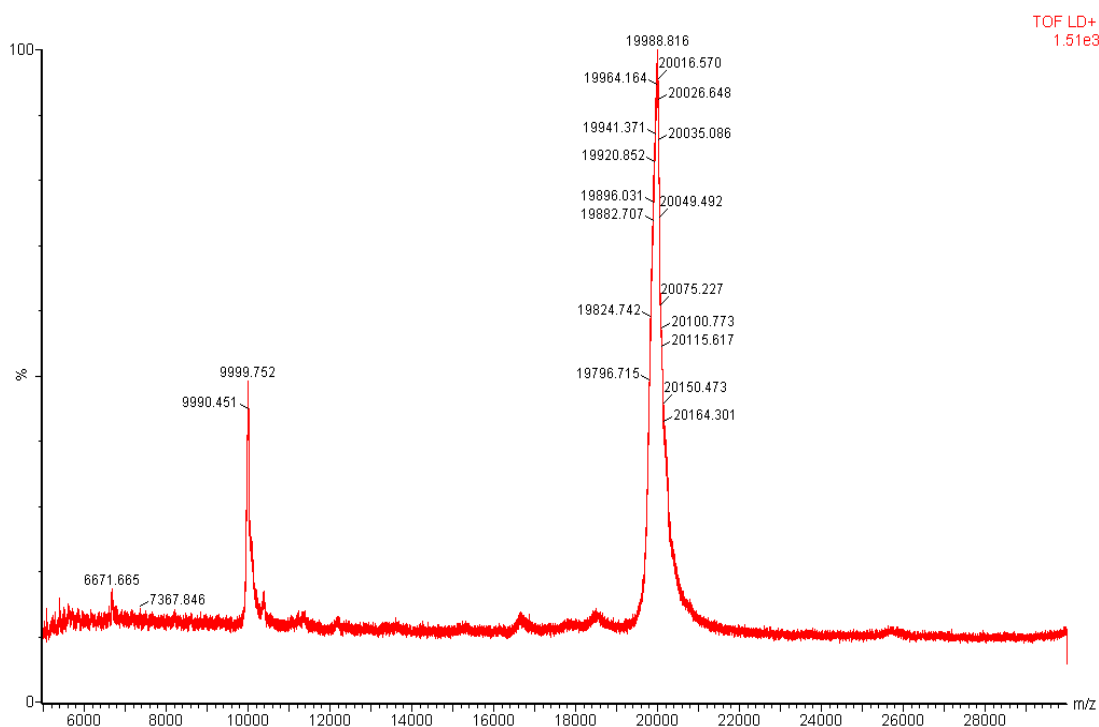
Figure 3.9: *The mass spectrum of the purified protein obtained through MALDI-TOF mass spectrometry, where the peak found at 9999.75 is the molecular ion with a +2 charge and 19988.82 is the PEF(S) molecular ion of +1 charge.*

3.5.2   Analytical size exclusion chromatography

To determine whether the protein is a monomer, dimer or takes any other form in solution, the sample was passed through an analytical size exclusion column. 500 μl of the dilute sample of purified protein from the SuperDex 75 (CV 320 ml) purification step (Section 3.3.1), was injected onto an analytical SuperDex 75 (CV 26 ml), which was connected to an FPLC instrument. The column was equilibrated in purification buffer C (Section 6.3.4).
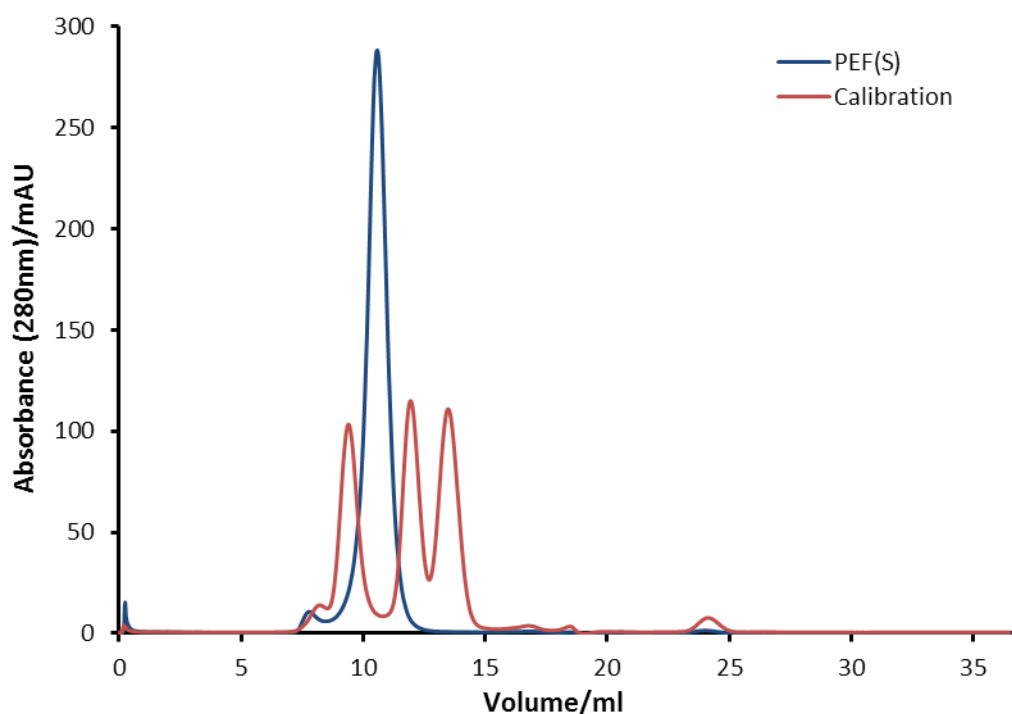
Figure 3.10: *Analytical size exclusion chromatogram of porcine PEF(S), shown in blue, overlaid on the calibration chromatogram shown in red where the first protein to elute is BSA (MW 66,000), the second trypsin inhibitor (MW 20,090) and the final protein is RNAase A (MW 13,700).*

The chromatogram obtained (Figure 3.10) was compared to a calibration chromatogram using the same analytical column. The proteins that were used to calibrate the size exclusion column were: bovine serum albumin (BSA molecular weight of 66,000), trypsin inhibitor (from soya beans, 20,090) and ribonuclease A (from bovine pancreas, 13,700). When compared to the calibration chromatogram the protein eluted at 10.7 ml which was in between the elution times of BSA and the trypsin inhibitor, implying that the mass of the protein was approximately 40,000. This means that the protein formed a dimer in solution, as the molecular weight of monomeric protein is 20,000 (Section

3.5.2). This corroborates previous reports of the PEF(S) domain forming a dimer in solution.[80]

3.5.3 Circular Dichroism spectroscopy

Proteins can fold into three main types of secondary structure; they can be α-helical, β-sheets or a random loop. The secondary structure of the protein can be analysed using circular polarised light. When the angle of the light emitted from the circular polarised light source hits a chiral molecule within a sample it changes in orientation. The change in orientation of the polarised light allows for the change in signal at a given wavelength, it is this signal alteration that allows for the determination of the overall secondary structure of a protein. Negative peaks at 208nm and 220 nm and a positive peak at 192 in a CD spectrum are indicative of α-helices.[218] If the protein structure contains a lot of β-sheets a negative peak is observed at 216 nm with two positive peaks at 200 nm.[218] When the protein structure is mainly composed of random coils the CD spectrum that is obtained has a negative peak at 195 nm (Figure 3.11).[219]
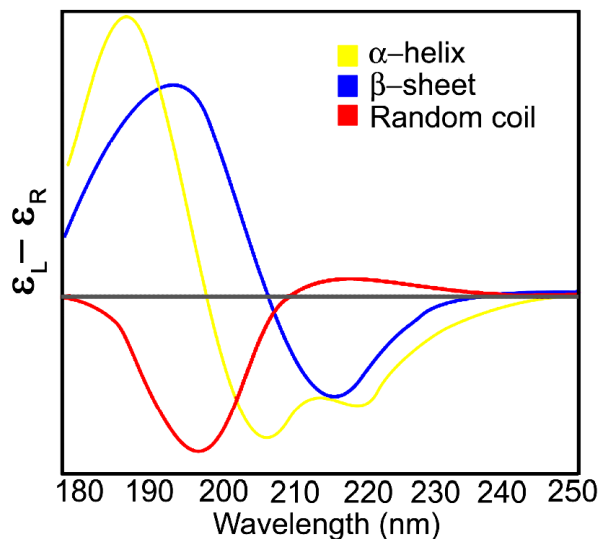
Figure 3.11: *Examples of the CD spectra obtained from ideal protein structure consisting of an α-helical structure (yellow), β-sheet (blue) and random coil (red).[219]*

The protein solution was diluted to 20 μM in a CD cuvette containing 350 μl of degassed K_iPO_4 (10 mM) at pH 7.0. A CD spectrum was then recorded between 350 nm to 190 nm, with 1 nm steps at 20 °C (Section 6.3.7). The data obtained contained two minima at 220 nm and 208 nm indicating that the majority of the secondary structure was α-helical (Figure 3.12).
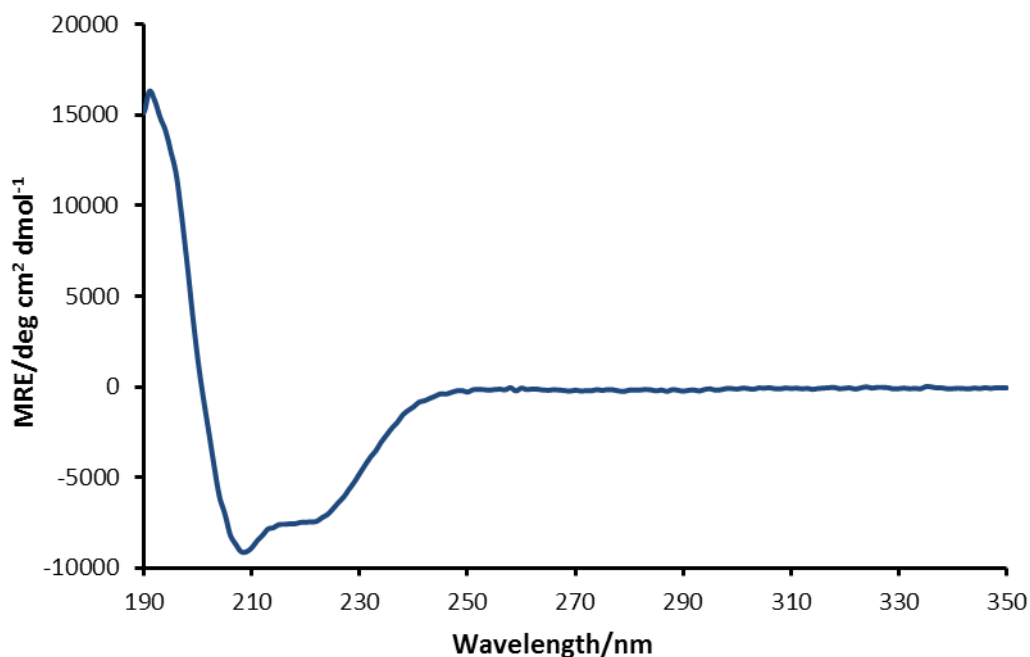
Figure 3.12: *The CD spectrum of PEF(S) showing that it forms a mainly α-helical secondary structure.*

The raw data was then de-convoluted to determine the percentage of the different forms of secondary structure found within PEF(S).[220-222] This revealed that the secondary structure of the protein was 62% α-helical (Table 3.1). The dominance of α-helices within the secondary structure correlates with the published X-ray crystal structure of PEF(S).[74]

Secondary structure	PEF(S)
α-Helix	62%
β-sheets	6%
β-turns	11%
Unordered	21%
Total	100%

Table 3.1: *The relative percentages of the secondary structure found within PEF(S).[220, 221]*

3.6 Conclusion

The expression and purification of PEF(S) yielded a useful quantity of pure protein. The characterisation of the over-expressed protein indicated that it resembled the desired PEF(S) domain. With the aid of mass spectrometry it was shown that the mass of the expressed protein was identical, within experimental error, to the calculated mass of PEF(S). Circular dichroism

spectroscopy indicated that the secondary structure of the protein was composed mostly of α-helices, consistent with the published X-ray crystal structure.[74] Finally analytical size exclusion chromatography demonstrated that the protein forms a homodimer in solution. The formation of the dimer is consistent with the EFH-5 within each monomer binding to its respective partner forming an EF-handshake, as previously observed in the crystal structure of PEF(S) and the structure of full calpain-II.[74, 76]

The fully characterised protein was then used to determine whether the α-mercaptoacrylic acid derivatives prepared as described in chapter 2 bound to this part of calpain-I.

4 The Interaction of α-Mercaptoacrylic acids with PEF(S)

4.1 Interaction of α-mercaptoacrylic acids with PEF(S)

The synthesis of α-mercaptoacrylic acid compounds was discussed in Chapter 2 where it was shown that they are inhibitors of calpain. Previous X-ray crystallography work on the parent mercaptoacrylic acid compound PD150606 indicated that it binds to the PEF(S) domain.[81, 186] To determine whether the inhibitors prepared in this project interact with this calcium binding domain, a fluorescence displacement assay[187] was carried out using the PEF(S) protein expressed and purified as described in Chapter 3. The α-mercaptoacrylic acid derivatives do not fluoresce, but can be used to displace a fluorescent compound with a lower binding affinity for the target protein domain. The probe used was 6-(*p*-toluidino)-2-naphthalenesulfonic acid (TNS). The complex of TNS with PEF(S) was titrated with compound **74** (Figure 4.1).[187]
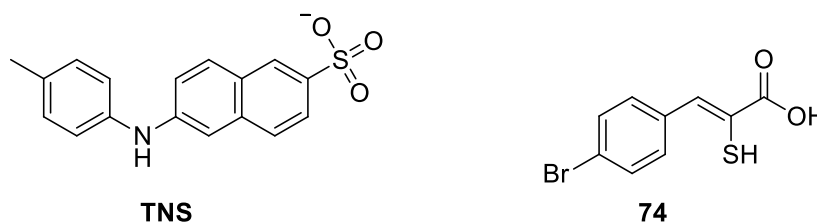
Figure 4.1: *The structure of the fluorescent probe **TNS** and the inhibitor **74**.*

TNS has an absorption maximum at 340 nm and an emission maximum at 445 nm. A 20 μM solution of the PEF(S) in 20 mM Tris-HCl, 0.1 mM EDTA was incubated with 20 μM TNS for five minutes and initial fluorescence emission spectra were obtained from 350 nm to 600 nm in both the presence and absence of 1.1 mM $CaCl_2$. The inhibitor was then added and new fluorescence spectra were measured. Reduction in the intensity of the fluorescent signal from TNS at 447 nm indicated its displacement from a binding site of the protein into the solution where its fluorescence was quenched by water. The spectra obtained revealed that when calcium was present the fluorescence of TNS-PEF(S) complex increased in intensity by 100%. Upon addition of **74** to the solution to a final concentration of 83 μM the fluorescence intensity decreased in both the calcium containing and the calcium free samples by 55% and 14%, respectively (Figure 4.2). Even though the decrease of the fluorescent signal was 4 times greater in the presence of calcium the signal still fell to the same

final intensity, implying that calcium does not affect inhibitor binding or TNS displacement, but instead affects the fluorescence intensity of TNS itself. Most importantly, the inhibitor seemed able to bind to PEF(S) in the presence and absence of calcium.
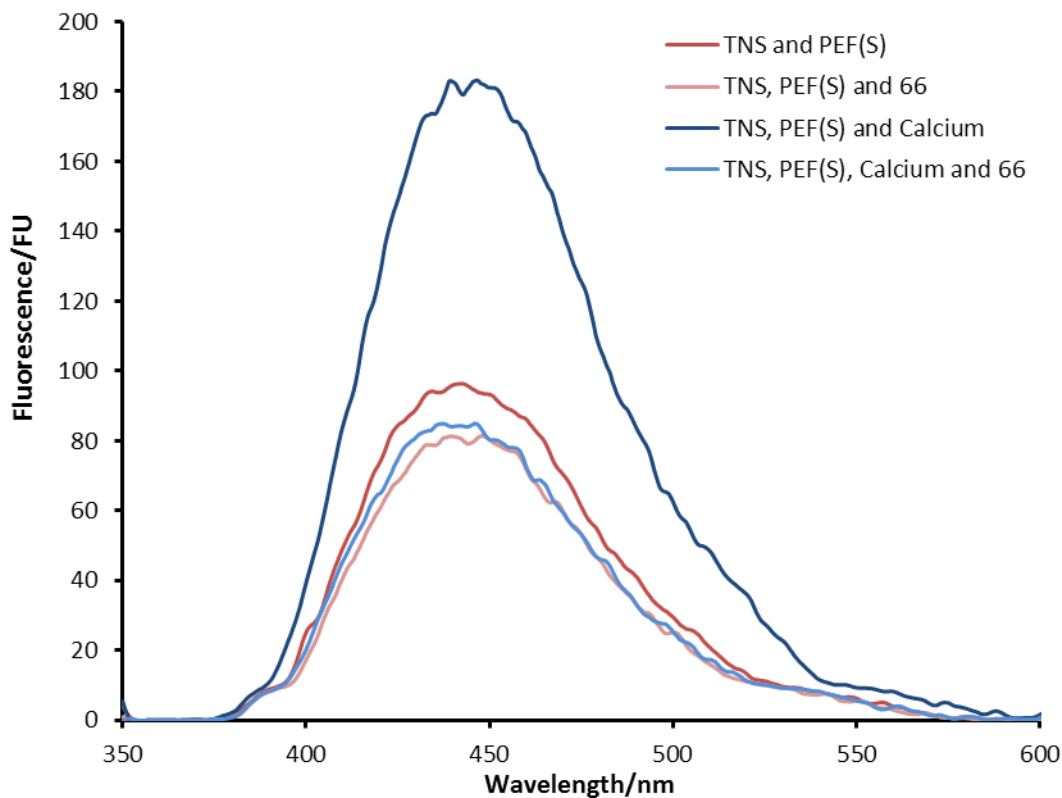
Figure 4.2: *The fluorescence spectra of the fluorescent probe, **TNS**, bound to PEF(S) and the displacement of **TNS** through the addition of **74**. Initial spectra were obtained from 3 ml solutions containing 20 μM PEF(S), 20 μM TNS, 0.1 mM EDTA and 20 mM Tris-HCl at pH 7.4, both with or without 1.1 mM CaCl$_2$. Compound **74** (83 μM) was added reducing intensity of the fluorescence at 447 nm. Excitation = 340 nm and emission = 445 nm.*

4.2 Single crystal X-ray diffraction

Having established the PEF(S) domain as a binding site for α-mercaptoacrylic acid derivatives single crystal X-ray diffraction was used to elucidate the mode of binding.

4.2.1 Screening

To obtain the optimal conditions for the formation of crystals, commercial wide screens were carried out. These screens contain a variety of precipitating solutions based upon the conditions that are known to crystallise many proteins. Three different screens were used: JCSG-*plus*™, PACT-*premier*™ and a

ProPlex screen (Molecular Dimensions, Suffolk, UK). Each screen contains 96 different precipitating conditions.[223, 224] The JCSG-*plus*™ screen is an optimised sparse matrix screen, specifically designed to incorporate a wide range of precipitants[223, 225] and is generally used in tandem with the PEG based PACT-*premier*™ screen.[225] The PACT-*premier*™ screen contains precipitants that vary in pH, cationic and anionic concentrations, whilst maintaining constant PEG concentrations and molecular weights.[226] The *ProPlex* screen is optimised for the crystallisation of protein complexes and is a targeted sparse matrix screen based on solutions used to successfully crystallise protein complexes.[224]

4.2.2 Plate set-up

The most common crystallisation method applied to proteins is vapour diffusion, either by sitting drop or hanging drop methods (Figure 4.3). With the sitting drop approach the protein solution is contained in a small depression beside a deeper well containing the precipitating solution. The protein solution is generally a mixture of 1:1 purified protein at 10 mg/ml and the precipitating solution. In the hanging drop scenario the protein solution hangs from a surface above the precipitant. In both cases the volume of the precipitant, in the deep well, is normally in excess of 150 × the volume of the protein drop. The precipitant in solution is much more concentrated than the drop containing the protein. Hence to achieve equilibrium, water vapour diffuses into the deep well containing the concentrated precipitant solution. The diffusion of the water from the protein-containing droplet gradually increases the concentration of the protein within the drop, allowing the protein to aggregate and form crystals.
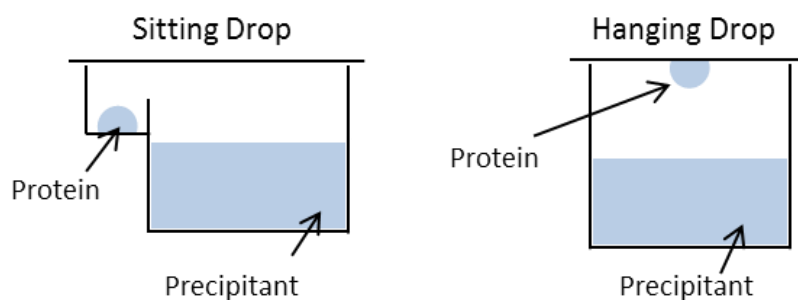
Figure 4.3: *Diagram of the sitting drop and the hanging drop techniques used in protein crystallography.*

The sitting drop technique was used to screen and crystallise PEF(S). 96 Well plates were used with each of the JCSG-*plus*™, PACT-*premier*™ and *ProPlex* screening solutions. The plates were set-up using an Art-Robbins Phoenix dispensing robot (Alpha Biotech, Glasgow, UK). 70 μl of the crystallisation solution was dispensed into the deep well followed by 0.2 μl of the same solution dispensed in the small well. 0.2 μl of the PEF(S) solution (5 mg/ml) was then dispensed into the small well of the crystallisation plate. The plates were sealed and were stored at 20 °C until crystals formed.

A number of crystals formed over a short period of time (~1 day). The crystals were collected using a cryoloop and flash frozen in liquid nitrogen, as the crystal would not survive X-ray irradiation at room temperature. The samples were taken to the synchrotron (Diamond Light Source, Oxfordshire, UK) and crystals were sequentially mounted onto a goniometer under a stream of gaseous nitrogen at 100 K in order to maintain the temperature and stability of the crystal. The crystal was then irradiated with X-rays of wavelength ~0.976 Å. The diffraction patterns were recorded using an ADSC Q315 charge-coupled device detector. Xia2 software reduced and analysed the data set with the CCP4 package (Collaborative Computational Project number 4).[227] A model of a similar protein was then used to fit to the density map (PDB:1NX2) as a starting point for structure solution.[186]

4.3 PEF(S) structure – initial crystallisation conditions

A diffraction pattern was obtained at a resolution of 2.09 Å in C2 space group from a crystal grown using a precipitation solution that contained 1 M sodium citrate and 0.1 M HEPES at pH 7.0, which was found in the *ProPlex* screen. Initially the data set was run through the molecular replacement programme Phaser-MR from the CCP4 program suite[228] using the PDB file 1NX2, which is a previously published structure of *holo*-PEF(S).[186] The data was then carried through numerous cycles of refinement with COOT and REFMAC5 until the model and the density map agreed with one another and no further refinements could be made. [229, 230]

Phaser-MR is used for the molecular replacement of a known protein structure (model) into the asymmetric unit of the unknown crystal in order to phase the data.[228] Rigid body refinement is used to optimise the positioning of the model into the density map, while avoiding deviations of the model when positioned in the data.[228] A log likelihood score is given, based on Bayesian statistics, which is representative of whether the model is a good fit to the data. The program then outputs a PDB file that best fits the data, and the resulting phases are used to calculate an observed density map and a difference density map. These maps are used for model adjustment using COOT for a better fit of the model. The improved model is then subjected to refinement with REFMAC to make even finer adjustments without manual intervention, while minimising the difference between the observed and calculated structure factors.[230] The process is repeated iteratively until no further improvement can be made. The quality of fit is monitored by the residual factor, or R-factor, which needs to be minimised to below 0.25 (25%) on average. This metric is considered to be sub-optimal, so a subset of the observations (reflections or data points) are set aside from the refinement calculations, but their R-factor, the so-called free R-factor or R_{free}, ought to be reduced if the improvement upon refinement is genuine.

COOT (Computational Object-Oriented Toolkit) is used to manually adjust the model against both the observed density and difference maps obtained from PHASER-MR or REFMAC. A number of parameters are used for the validation of the data, these include the Ramachandran plot, geometrical parameters, density fit analysis and B-factor analysis.[229]

Data Reduction Statistics	*Holo*-PEF(S)
Unit cell dimensions/Å	a = 183.16, b = 41.05, c = 48.23
Unit cell angles/°	a = 90.00, b = 101.54, c = 90.00
Space Group	C 1 2 1
Resolution range/Å	47.26 – 2.09 Å
Total Reflections	62904
Unique Reflections	18402
Completeness/% (last shell)	87.1 (71.7)
I/σ (last shell)	9.4 (2.4)
R(merge)/% (last shell)	12.7 (90.4)
B(iso) from Wilson/Å^2	38.61
Refinement Statistics	
Protein atoms excluding H	2790
Calcium atoms	2
Solvent molecules	72
R factor/%	25.08
R_{free}/%	32.97
R.m.s. bond length/Å	0.0122
R.m.s. bond angle/°	1.4738

Table 4.1: *The data statistics from structure of* holo-*PEF(S), with the final refined data statistics.*

The statistics from the data demonstrate that the unit cell is different from the previously reported data, having a C2 space group as opposed to $P3_121$.[186] The R factor and the R_{free} for the refined data-set are 24.5% and 31.3%, respectively (Table 4.1). The values obtained for the R factor gives an indication as to how well the refined model fits the data obtained and R_{free} is based on reflections that were never used in refinement and therefore cannot bias the model.[231, 232] The difference between the R factor and the R_{free} gives an overall indication as to how well the model fits to the data.[231] The refined structure demonstrated that the protein is a dimer, which is the form the protein takes in solution as shown by analytical size exclusion chromatography (Chapter 3). The density map shows little density at the N-termini of both monomers but does show a calcium ion bound to each monomer within the structure.

4.3.1 PEF(S) soak with α-mercaptoacrylic acid derivatives

With crystallisation conditions in hand a 48-well sitting drop plate was set-up with the precipitating conditions obtained from the *ProPlex* screen (*vide supra*).

The precipitant used was 1.4 M sodium citrate, 0.1 M HEPES at pH 7.0. 300 μl of this solution was placed into the deep well and 1 μl of the precipitant was placed in the sitting drop with the equivalent quantity of PEF(S) (5 mg/ml). The plates were sealed with an adhesive film and left at 20 °C until crystals formed (~1 week).

24 hours prior to crystal harvesting, they were soaked with different inhibitors at a variety of concentrations (Table 4.2). All inhibitors contained a bromine atom, as these α-mercaptoacrylic acid derivatives had shown the highest activity as calpain-I inhibitors (Chapter 2).

Inhibitor	Concentration /mM	Ratio/inhibitor to protein	Resolution /Å
77	1.0	4:1	6.3
80	0.5	2:1	2.6
	0.75	3:1	2.5
	0.75	3:1	2.6
84	0.5	2:1	2.5
	0.5	2:1	2.7
	0.75	3:1	2.9
	1.0	4:1	3.3
88	0.5	2:1	2.4
	0.75	3:1	2.9
66	0.75	3:1	2.4
	1.0	4:1	3.0
70	0.75	3:1	2.9
	1.0	4:1	No data
74	0.75	3:1	No data
	1.0	4:1	No data

Table 4.2: A table collating all of the data that was recorded from the soaks that were carried out of PEF(S) for 24 hours at 20 °C. Soaking solutions contained 10 mM DTT within the solution.

The data sets were uniformly of lower resolution than the data set obtained for the *holo* structure, an effect that is often observed in soaking experiments. Though all the data remained in the C2 space group, the lower resolution implied that the soaks had caused some damage to the lattice of the *holo*-PEF(S) protein. The damage to the crystal lattice suggested uptake of the ligands into the structure of the protein.

Molecular replacement was carried out on the data sets from the soaked crystals using the *holo* structure previously refined. Unfortunately, processing and refinement revealed little or no extra density that could indicate another molecule within the structure.

4.4 PEF(S) structure – new crystallisation conditions

Due to these results, alternative conditions for forming crystals of PEF(S) were used. The new precipitating solution contained 50 mM sodium cacodylate, 12.5% (w/v) PEG 6000 and 20 mM calcium chloride at pH 7.0 (Section 6.4.3).[233] The crystallisation plates were set up using the hanging drop vapour diffusion technique, where 1 ml of the precipitant was placed in the well and 3 μl of precipitant was mixed with 1 μl of deionised water and 3 μl of protein solution (10 mg/ml in 5 mM sodium cacodylate, 2 mM EDTA, 5 mM βME, 1 mM $CaCl_2$ at pH 7.0) in the drop. Crystals formed after 3 days at 20 °C (Figure 4.4).
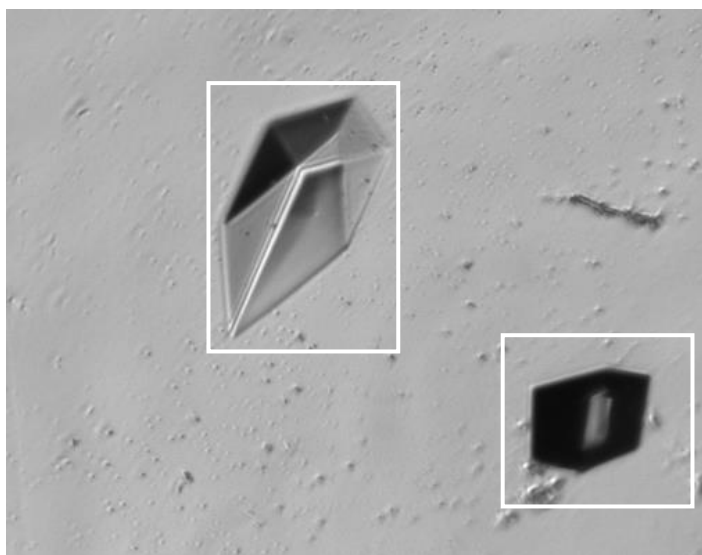
Figure 4.4: *Image taken of a crystals formed (highlighted in the white boxes) in 50 mM sodium cacodylate, 12.5% PEG 6000 and 20 mM $CaCl_2$ at pH 7.0, PEF(S) concentration was 10 mg/ml.*

Data was again collected at the Diamond Light Source and scaled and reduced using xia2 to yield a data set with a resolution of 1.79 Å in the $P2_1$ space group. The reduced data was fitted with a model based on *holo* porcine PEF(S) (PDB 1ALV); this was used due to the correspondence of the space groups.[81] Many iterations of refinement with COOT and REFMAC5 were carried out to obtain

the final structure. The statistics obtained from the refinement process revealed a final R factor of 18.6% and R_{free} of 22.4% (Table 4.3).

Data Reduction Statistics	*Holo*-PEF(S)
Unit cell dimensions/Å	a = 49.53, b = 70.56, c = 58.67
Unit cell angles/°	a = 90.00, b = 108.79, c = 90.00
Space Group	P 1 2_1 1
Resolution/Å	39.05 – 1.79 Å
Total Reflections	133085
Unique Reflections	35505
Completeness/% (last shell)	98.4 (99.9)
I/σ (last shell)	16.6 (2.1)
R-merge/% (last shell)	5.0 (7.1)
B(iso) from Wilson/Å^2	24.49
Refinement Statistics	
Protein atoms excluding H	2790
Calcium atoms	8
Solvent molecules	296
R factor/%	18.62
R_{free}/%	22.45
R.m.s. bond length/Å	0.019
R.m.s. bond angle/°	1.904

Table 4.3: *The data statistics from structure of* holo-*PEF(S), with the final refined data statistics.*

The refined structure of PEF(S) is highly similar to other reported crystallographic structures.[81, 186] Superposition of the *holo*-PEF(S) structure with the PDB entry 1ALV revealed that it is identical to the published structure of the protein (Figure 4.5).[81] Comparison of the residues in chain A of the refined structure and the previously published structure was carried out with Superpose from CCP4.[234] Superpose aligned the PDB entry 1ALV to the *holo*-PEF(S) structure and calculated the root mean square deviation (r.m.s.d) of each amino acid between the structures. The r.m.s.d. values obtained were plotted on a graph and revealed that the residues that show the greatest positional root mean square deviation (r.m.s.d.) were localised found the loop regions of the structure (Figure 4.5).

(A)

(B)
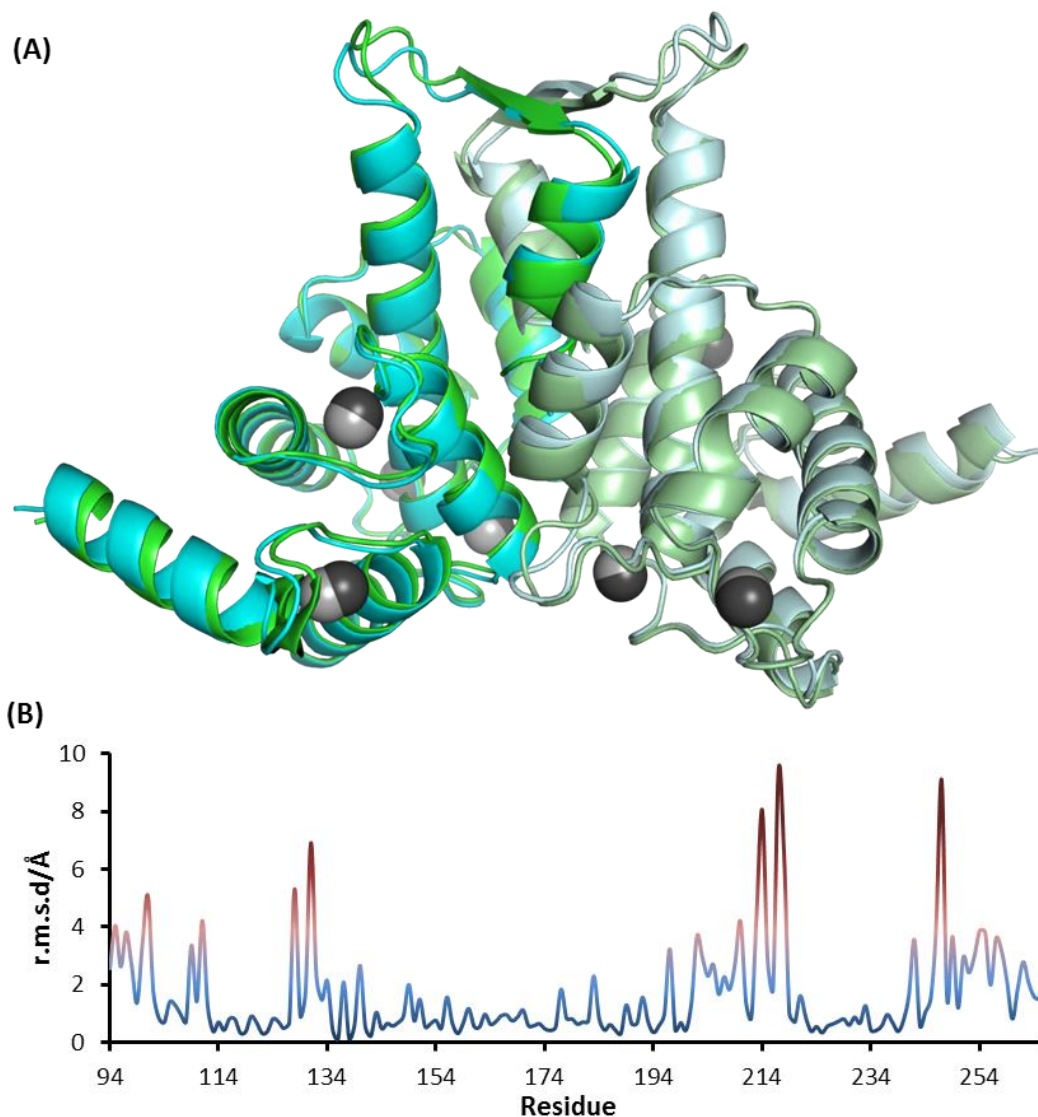
Figure 4.5: *(A) Secondary structure alignment of* holo-*PEF(S) (cyan and pale cyan) derived from the density map and a published structure of PEF(S) (green and dark green) (PDB 1ALV) using PyMOL.[54] The calcium ions represented as grey and dark grey spheres for* holo-*PEF(S) and the published structure.[81] (B) A graph representing the r.m.s.d. of the residues as calculated by Superpose in CCP4, the comparison was carried out between chain A of* holo-*PEF(S) and the PDB 1ALV.[81, 234]*

PEF(S) forms a homodimer in solution as shown by size exclusion chromatography (Chapter 3). In the crystalline form the protein remains as a homodimer with EFH-5 of each monomer in a classical EF-handshake (Figure 4.5). The mainly hydrophobic residues of the EFH-5 bind to one another to form the dimer interface and comprise Val254, Ile252, Leu262 and Trp259 (Figure 4.6).
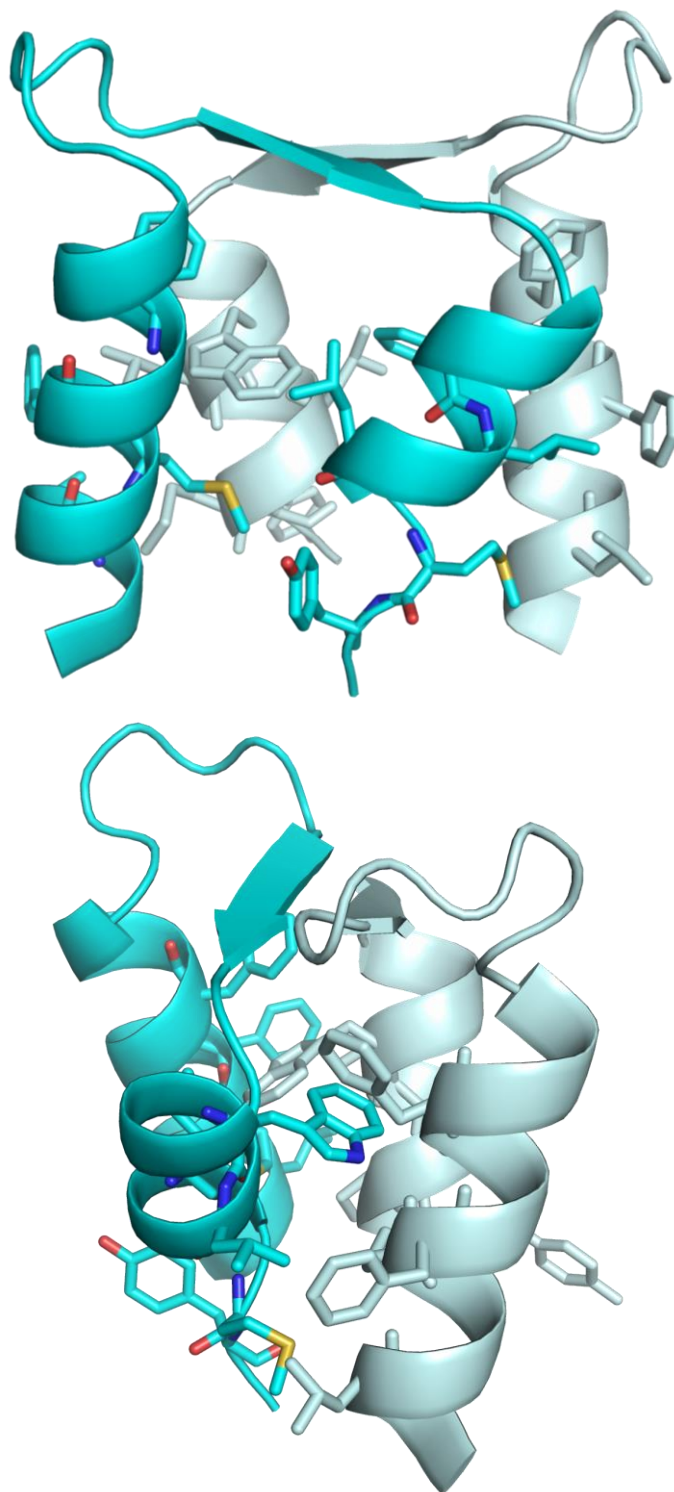
Figure 4.6: *Secondary structure representations of the hydrophobic interactions between the residues of EFH-5 at the interface between two monomers of PEF(S).*

The other four EF-hands within the structure all bind calcium ions, in the same manner as previously published structures of porcine PEF(S) (Figure 4.7). With EF-hands 2 and 3 the bound calcium ion interacts with numerous amino acids, and a single water molecule completes the coordination shell around Ca^{2+}. EF-

hand 1 required two water molecules to complete the binding of calcium. The fourth EF-hand also does not demonstrate the canonical EF-hand binding motif (Section 1.6.4), as the helix-loop-helix motif is not used to bind the Ca^{2+}, the four residues used are Asp135, Asp223, Asp225 and Asn226 (Figure 4.7).[72]
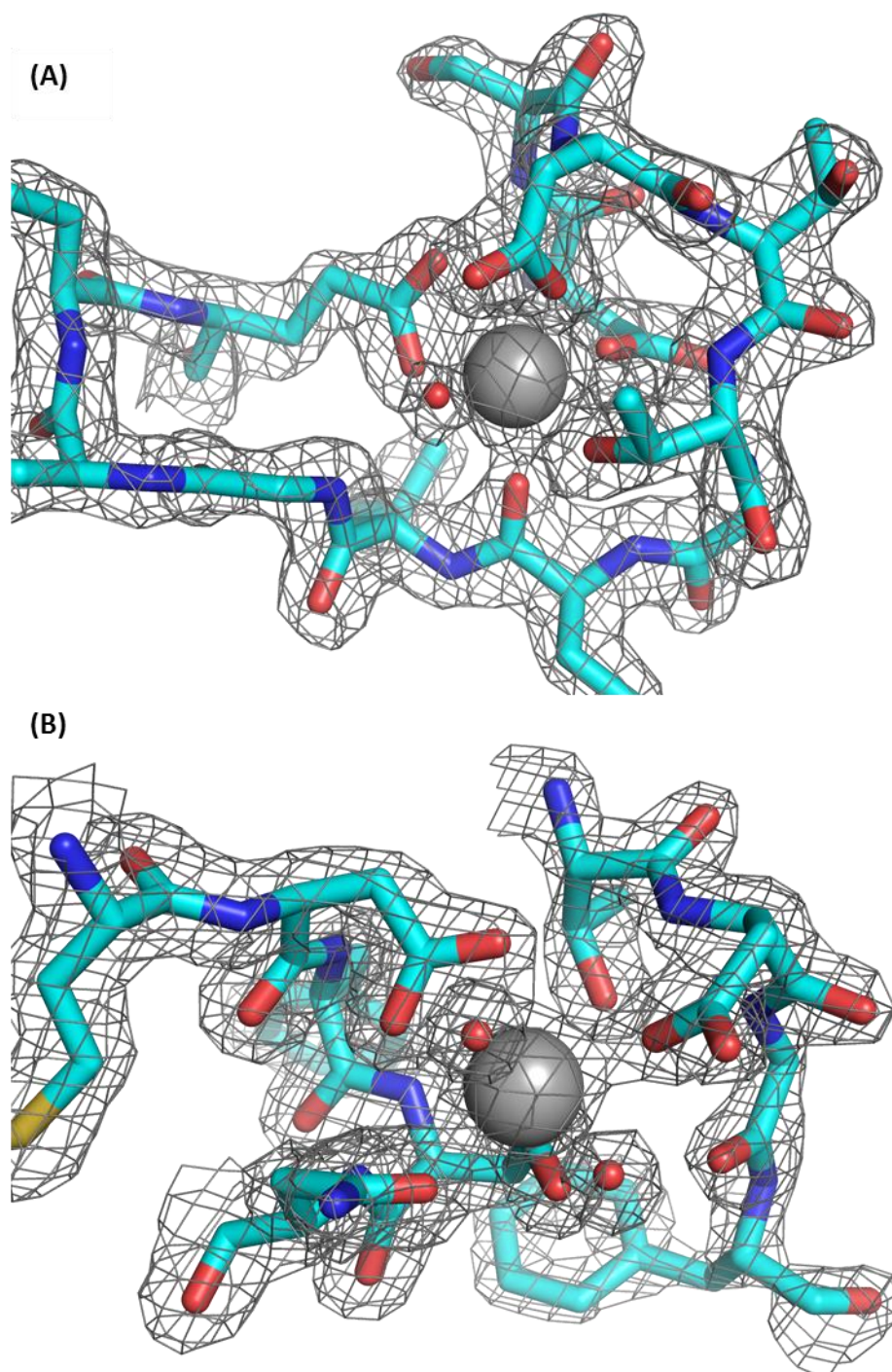
Figure 4.7: *(A) A stick representation of EFH-2 which binding calcium in a canonical fashion and (B) EFH-4 where the binding site is non-canonical. The density map is contoured to 1.5σ. Calcium ion is represented as a sphere.*

4.5 Inhibitor interaction with PEF(S)

To obtain a complex of the inhibitor, the compounds were soaked into crystals preformed under the new conditions. Each inhibitor was dissolved in DMSO to create a 50 mM stock solution, these stock solutions were then diluted into the precipitation buffer used to form the crystals, 50 mM sodium cacodylate, 12.5% (w/v) PEG 6000 and 20 mM $CaCl_2$ at pH 7.0 with the addition of 10 mM DTT. The overall concentration of the inhibitor in the solution was 8 mM. 1 µl of the inhibitor solution was added to the 7 µl drop that contained the crystals, giving an overall final concentration of 1 mM of the inhibitor within the drop. The hanging drops were resealed and the inhibitor was left to soak into the crystals for 24 hours. (*Z*)-3-(4-bromophenyl)-2-mercaptoacrylic acid (**74**) and (*Z*)-3-(6-bromoindol-3-yl)-2-mercaptoacrylic acid (**84**) were chosen as representative example inhibitors (Figure 4.8).
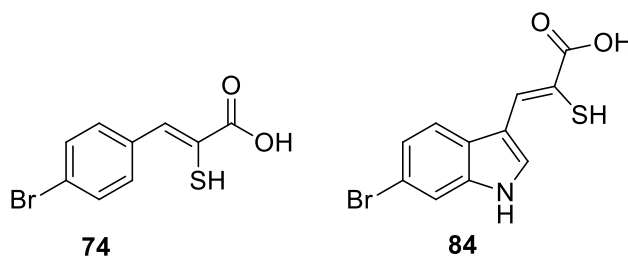
Figure 4.8: *The structures of (Z)-3-(4-bromophenyl)-2-mercaptoacrylic acid (**74**) and (Z)-3-(6-bromoindol-3-yl)-2-mercaptoacrylic acid (**84**).*

Molecular replacement was carried out on the reduced mtz file, with Phaser from the CCP4 package using the refined *holo* structure model (Section 4.4).[228] The data was then carried through sequential rounds of refinement with COOT and REFMAC5 as before.[229, 230]

4.5.1 Structure of phenyl derivative 74 bound to PEF(S)

Compound **74** was the most potent inhibitor of the phenyl based α-mercaptoacrylic acid based compounds with an IC_{50} value of 1.5 µM (Chapter 2). The data set extended to a resolution of 2.10 Å. The statistics obtained from refinement reveal that the model and the density map fit well, with the R factor and R_{free} with values of 18.7% and 26.0% (Table 4.4).

Crystallographic Data	PEF(S) with 74
Unit cell dimensions/Å	a = 49.74, b = 70.29, c = 57.92
Unit cell angles/°	a = 90.00, b = 108.42, c = 90.00
Space Group	P 1 2₁ 1
Resolution/Å	39.18 – 2.10 Å
Total Reflections	79602
Unique Reflections	20053
Completeness/% (last shell)	90.4 (91.1)
1/σ (last shell)	8.4 (2.2)
R-merge/% (last shell)	9.6 (57.6)
B(iso) from Wilson/Å²	28.86
Refinement Statisitics	
Protein atoms excluding H	2790
Calcium atoms	8
Ligand atoms	26
Solvent atoms	194
R factor/%	18.72
R_{free}/%	25.96
R.m.s. bond length/Å	0.015
R.m.s. bond angle/°	1.716

Table 4.4: *The data statistics from the structure of PEF(S) soaked with **74**, with the final refined data statistics.*

The refinement of the model revealed a volume of unoccupied positive difference density, where **74** was situated. After further refinement, the density accounted for 50% of that expected for the inhibitor molecule. This area of density was present in both copies of PEF(S) within the asymmetric unit (Figure 4.9).
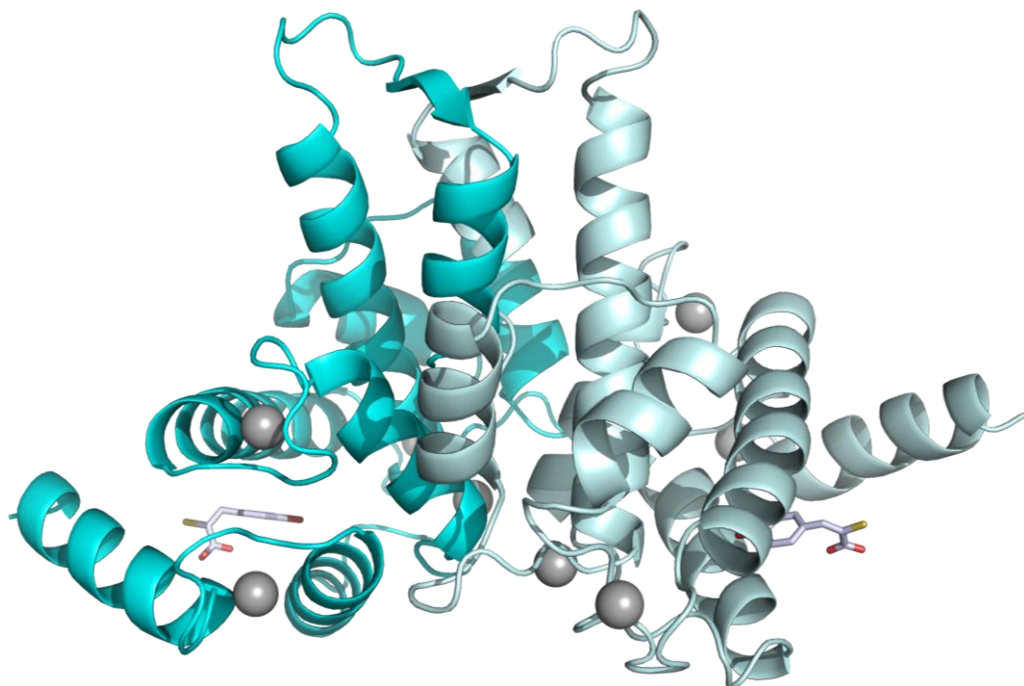
Figure 4.9: *A cartoon representation of the dimer of PEF(S) with two molecules of **74** bound, represented as white sticks.*

The residues at the interface between **74** and the hydrophobic pocket were calculated with PDBePISA and all the individual interactions below 4.0 Å were calculated using Contact.[234, 235] The analysis of the interface revealed that extensive interactions occur between the protein and the aromatic portion of the inhibitor. These hydrophobic interactions occurred with both aromatic residues and alkyl residues (Table 4.5). The hydrophilic residues that interact with the α-mercaptoacrylic acid are Arg128 and Gln173 (Table 4.5). Lys170 appeared to interact primarily through the alkyl chain rather than with the hydrophilic ammonium end group. Measurement of the contacts between the ligand and its binding pocket indicates that there are no significant hydrogen bonds that enable the ligand to bind tightly to the pocket (Table 4.5) (Figure 4.10). Both Arg128 and Lys170 contain highly flexible side chains of course, so in reality these are capable of moving in solution to form hydrogen bonds with the α-mercaptoacrylic acid.
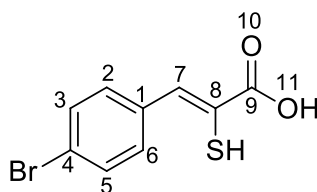
Interface residues	vdW (≤ 4.0 Å)	Distance/Å and Atoms involved (residue:74)	
Leu122			
Val125	5	3.42 (Cγ1 - **10**)	3.75 (Cγ1 - **5**)
		3.53 (Cγ1 - **4**)	3.87 (Cγ1 - Br)
		3.92 (Cγ1 - **3**)	
Val126			
Arg128	2	3.70 (Nη1 - **10**)	3.79 (Nη1 - **11**)
His 129	2	3.99 (Nε2 - **3**)	3.93 (Nε2 - **2**)
Leu132			
Phe137			
Trp166	3	3.57 (Cζ2 - **6**)	3.84 (Cη2 - **5**)
		3.83 (Cζ2 - **5**)	
Ile169			
Lys170	1	3.40 (Cγ - **7**)	
Gln173	5	3.40 (Cγ - **3**)	3.95 (Cδ – **3**)
		3.71 (Cγ - **2**)	3.70 (Oε1 - **2**)
		3.74 (Cδ - **2**)	
Phe224	1	3.84 (Cε2-Br)	

Table 4.5: *A table listing the residues found at the interface with* ***74*** *and the interactions that occur between the molecule and the amino acids. Interface residues were calculated with PDBePISA.[235] Van der Waals (vdW) distances calculated with the program Contact in the CCP4 package.[234]*
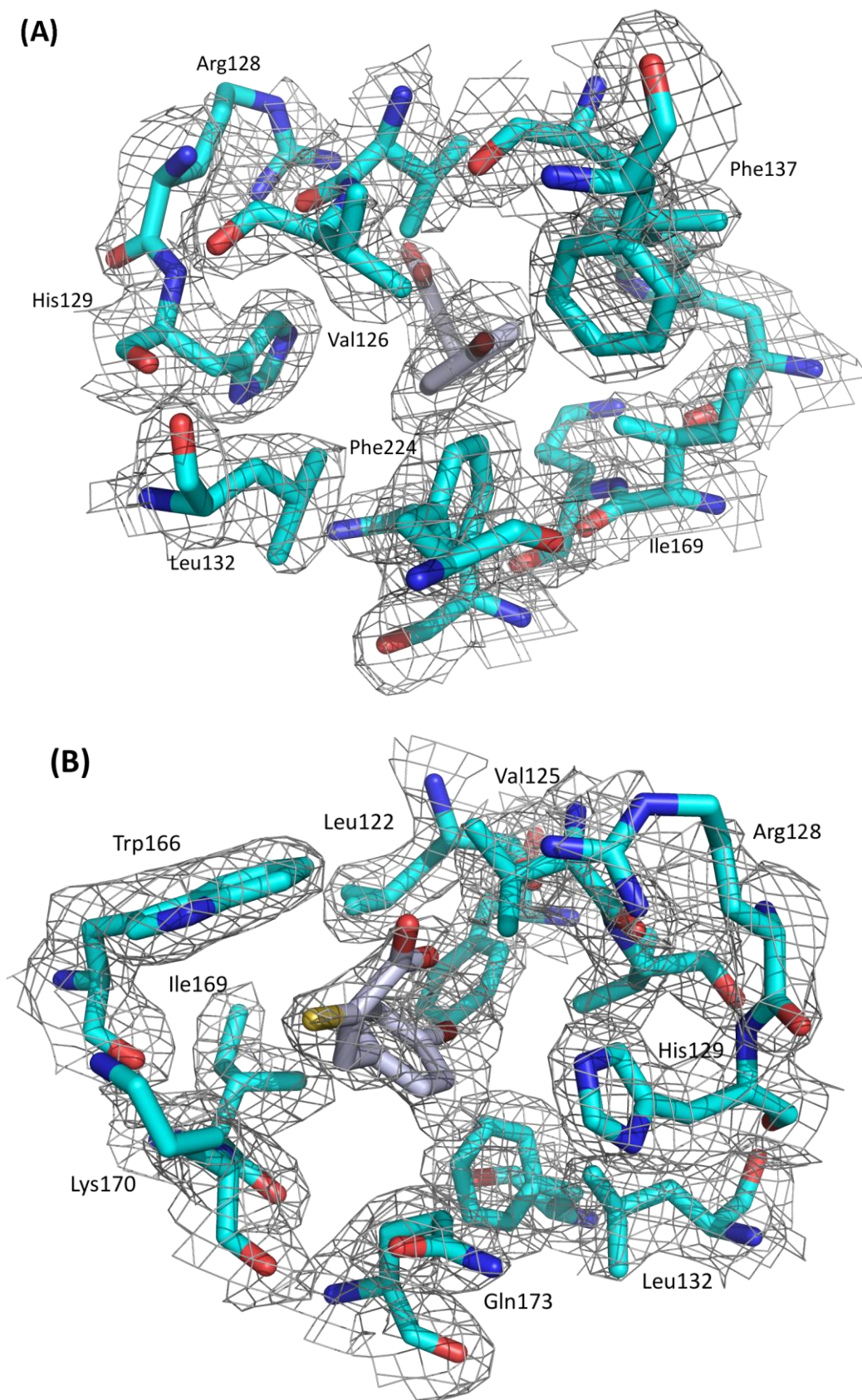
Figure 4.10: *(A) A stick representation of **74** bound to the PEF(S) highlighting the hydrophobic binding pocket and (B) the α-mercaptoacrylic acid interacting with the residues on the surface of PEF(S). Density map is contoured at 1.0σ.*

Comparison of PEF(S)-74 structure with *holo*-PEF(S)

When comparing the structure of the complex with that of the *holo* protein there is very little overall movement of the residues that interact **74**. Only a slight displacement of the hydrophobic residues within the cavity is required for the inhibitor to bind. The hydrophilic residues that bind to the α-mercaptoacrylic acid demonstrate the greatest deviation, with Arg128 and Gln173 moving most to accommodate the ligand (Figure 4.11). The sidechain of Arg128 moves to interact with the α-mercaptoacrylic acid.
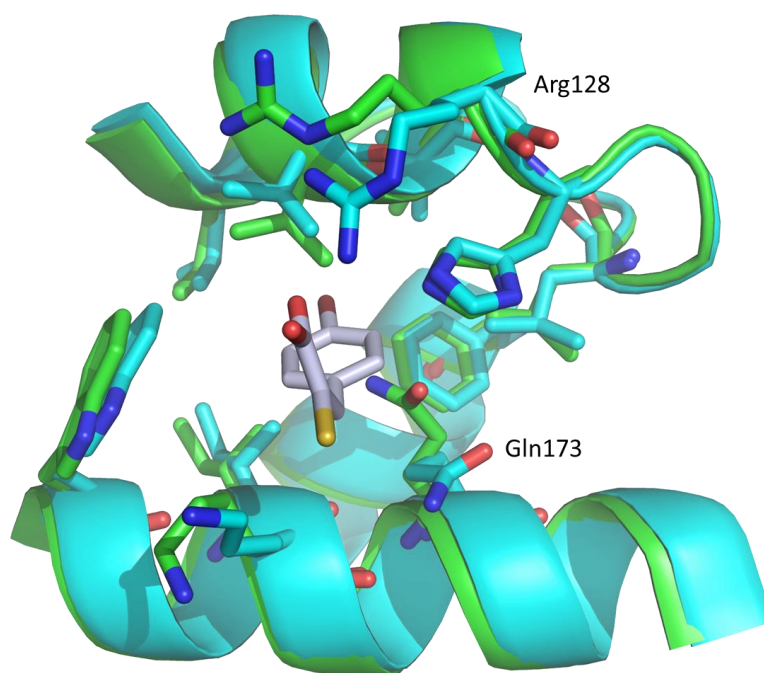
Figure 4.11: *Aligned secondary structure representations of* holo-*PEF(S) (green) and the complex of **74** with PEF(S) (cyan). Arg128 and Gln173 demonstrate the greatest deviation.*

To determine how significantly the residues move upon binding **74** into the hydrophobic pocket, an r.m.s.d. plot of the residues was carried out with Superpose in CCP4, with all the atoms in each residue taken into account.[234] The r.m.s.d. values were calculated between monomer (chain A) of the dimeric structure of both the holo structure and the inhibitor complex. The plot revealed that at the binding site residue Gln173 and Lys170 have r.m.s.d. values of 2.8 Å (Figure 4.12). Other residues show higher r.m.s.d values at ~4 Å, these are either located at flexible loops within the protein or are comprised of long alkyl side chains, such as Arg101.
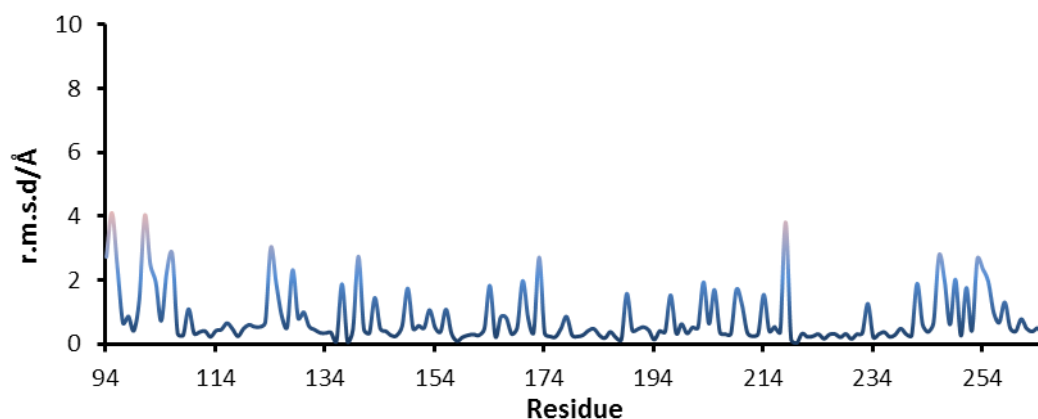
Figure 4.12: A graph representing the r.m.s.d. of the residues as calculated by Superpose in CCP4, the comparison was carried out between chain A of holo*-PEF(S) and the complex of PEF(S) and* **74***.[234]*

Comparison of PEF(S)-74 with PEF(S)-PD150606
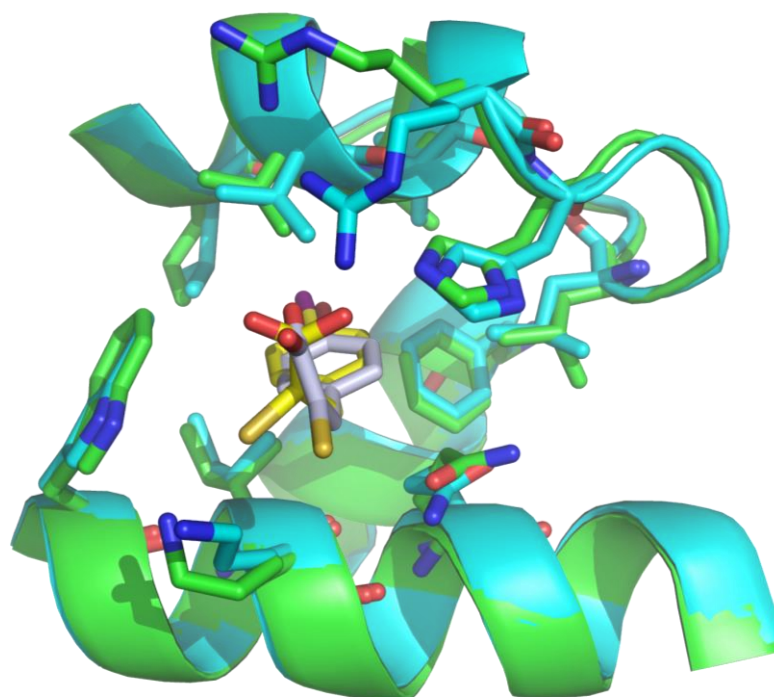
*Figure 4.13: Aligned secondary structure representations of the structure of **74** (white) bound to PEF(S) (cyan) and the published PD150606 (yellow) bound to PEF(S) (green) (PDB 1NX3).[186]*

The structure of **74** bound to the protein is very similar to the published structure of PD150606 bound to PEF(S) (Figure 4.13).[186] The inhibitor resides in the same hydrophobic pocket occupied by PD150606 and hence interacts with the same residues. The side chain of Arg128 also moves to interact with the carboxylate group of the α-mercaptoacrylic acid in the published structure (Figure 4.13).

4.5.2 Structure of indole derivative 84 bound to PEF(S)

Bromine-substituted indole based α-mercaptoacrylic acid derivatives were found to be highly potent inhibitors of calpain-I. (*Z*)-3-(6-Bromoindol-3-yl)-2-mercaptoacrylic acid (**84**) in particular gave an IC_{50} value of 6 nM. Therefore to determine the molecular basis for the improved activity of **84**, this compound was soaked into preformed PEF(S) crystals for 24 hours at 1 mM and data was collected. The diffraction pattern extended to a resolution of 1.96 Å with the same $P2_1$ space group (Table 4.6). Upon refinement with COOT and REFMAC5, the R factors and R free yielded values of 18.0% and 22.6%, respectively (Table 4.6).

Crystallographic Data	PEF(S) with 84
Unit cell dimensions/Å	a = 49.75, b = 70.36, c = 58.59
Unit cell angles/°	a = 90.00, b = 108.72, c = 90.00
Space Group	P 1 2_1 1
Resolution/Å	39.15 – 1.96 Å
Total Reflections	102564
Unique Reflections	27012
Completeness/% (last shell)	98.0 (98.9)
I/σ (last shell)	17.0 (2.3)
R-merge/% (last shell)	3.9 (50.4)
B(iso) from Wilson/Å^2	30.88
Refinement Statisitics	
Protein atoms excluding H	2790
Calcium atoms	8
Ligand atoms	32
Solvent atoms	206
R_{merge}/%	18.02
R_{free}/%	22.64
R.m.s. bond length/Å	0.019
R.m.s. bond angle/°	1.985

Table 4.6: *The data statistics from the structure of PEF(S) soaked with **84**, with the final refined data statistics.*

During refinement a volume of density appeared within the same hydrophobic pocket where both **74** and PD150606 bind. This unoccupied density was found in both monomers of the protein (Figure 4.14). A model of **84** was placed into the unoccupied density and further refinement was carried out. Once again,

only partial occupancy by the inhibitor was found with the density only accounting for 50% of the model inserted at 1σ. This occupancy is sufficient to summarise the interactions that occur between the protein and this inhibitor.
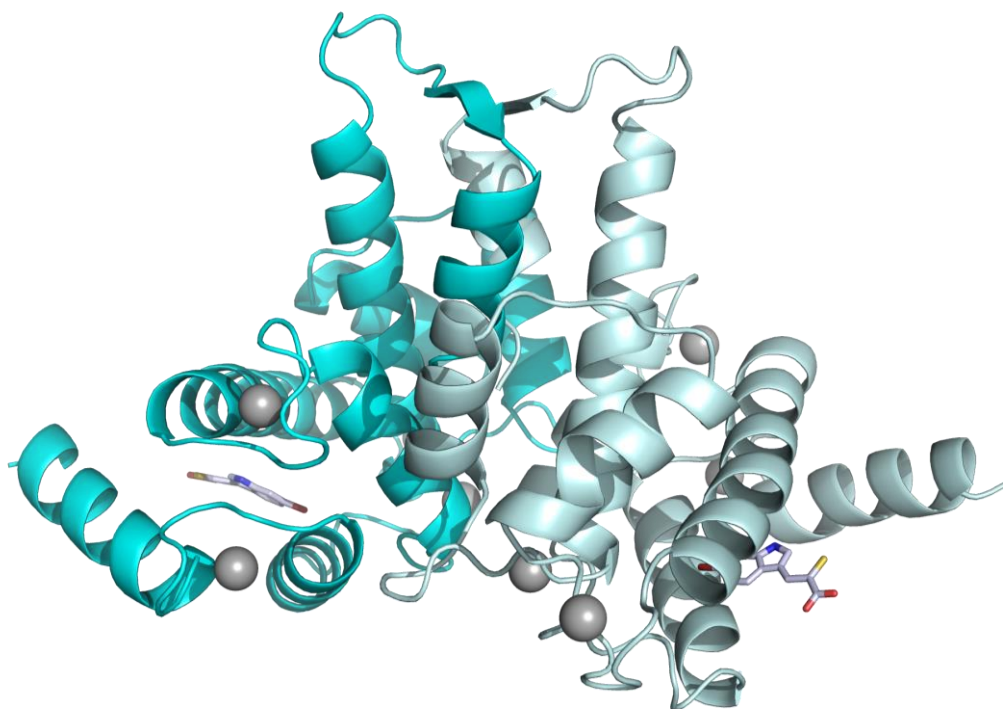
Figure 4.14: *A secondary structure representation of two molecules of **84** bound to the dimer of PEF(S).*

Compound **84** interacts with the same aromatic and alkyl residues within the hydrophobic pocket as **74**. The distances between the aliphatic residues and the indole ring of the inhibitor reveal that there are a greater number of van der Waals (vdW) interactions than observed in the **74**-PEF(S) complex (Section 4.5.1) (Table 4.7). Arg128, one of the residues found to interact with **74**, is not found at the interface with **84**. The distance between the NH_2 and the carboxylate increases from 3.70 Å to 6.85 Å. Lys177 is a residue involved in the interface although this residue does not appear to have any integral hydrogen bonds or vdW contacts with **84**, due to the distance of the ligand from the residue (6.0 Å) (Figure 4.15). Lys170 and Gln173 demonstrate many vdW interactions with the ligand, both with the hydrophilic parts and with the alkyl chains of these residues (Table 4.7). The hydrophilic residues that interact with the α-mercaptoacrylic acid moiety in **74** were not close enough for hydrogen bonds to form with the carboxylate or sulfhydryl groups. In the case of Lys170

this may be deceptive as the side chain of this residue is flexible and can move towards the α-mercaptoacrylate and eventually form a strong hydrogen bond with the ligand.
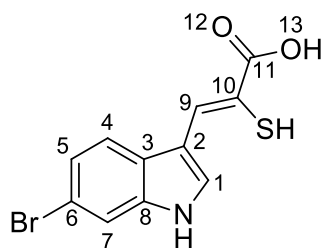
Interface residues	vdW (≤ 4.0 Å)	Distance/Å and Atoms involved (residue:84)	
Leu122	2	3.96 (Cα - Br)	3.46 (O - Br)
Val125	8	3.72 (Cγ1 - **3**)	3.42 (Cγ1 - **4**)
		3.18 (Cγ1 - **5**)	3.73 (Cβ - **6**)
		3.28 (Cγ1 - **6**)	3.95 (Cβ - Br)
		3.59 (Cγ1 - **7**)	3.79 (Cγ1 - **8**)
Val126	1	3.56 (Cγ1 - Br)	
His 129	2	3.79 (Nε2 - **4**)	3.99 (Cδ2 - **4**)
Leu132			
Phe137			
Trp166	3	3.83 (Cε2 - NH)	3.74 (Cη2 - NH)
		3.55 (Cζ2 - NH)	
Ile169	2	3.56 (Cγ2 - **7**)	3.95 (Cγ2 - **8**)
Lys170	7	3.87 (Nζ - **13**)	3.90 (Cγ - SH)
		3.70 (Cδ - SH)	3.98 (Cε - SH)
		3.56 (Nζ - SH)	3.92 (Cα - **1**)
		3.80 (Cβ - **1**)	
Gln173	6	3.70 (Nε2 - **12**)	3.94 (Cδ - **12**)
		3.74 (Oε1 - **12**)	3.54 (Nε2 - **9**)
		3.79 (Cδ - **9**)	3.73 (Cγ - **4**)
Lys177			
Phe224	1	3.85 (Cε2-Br)	

Table 4.7: *A table listing the residues found at the interface with **84** and the interactions that occur between the molecule and the amino acids. Interface residues were calculated with PDBePISA.[235] Van der Waals (vdW) distances calculated with the program Contact in the CCP4 package.[234]*
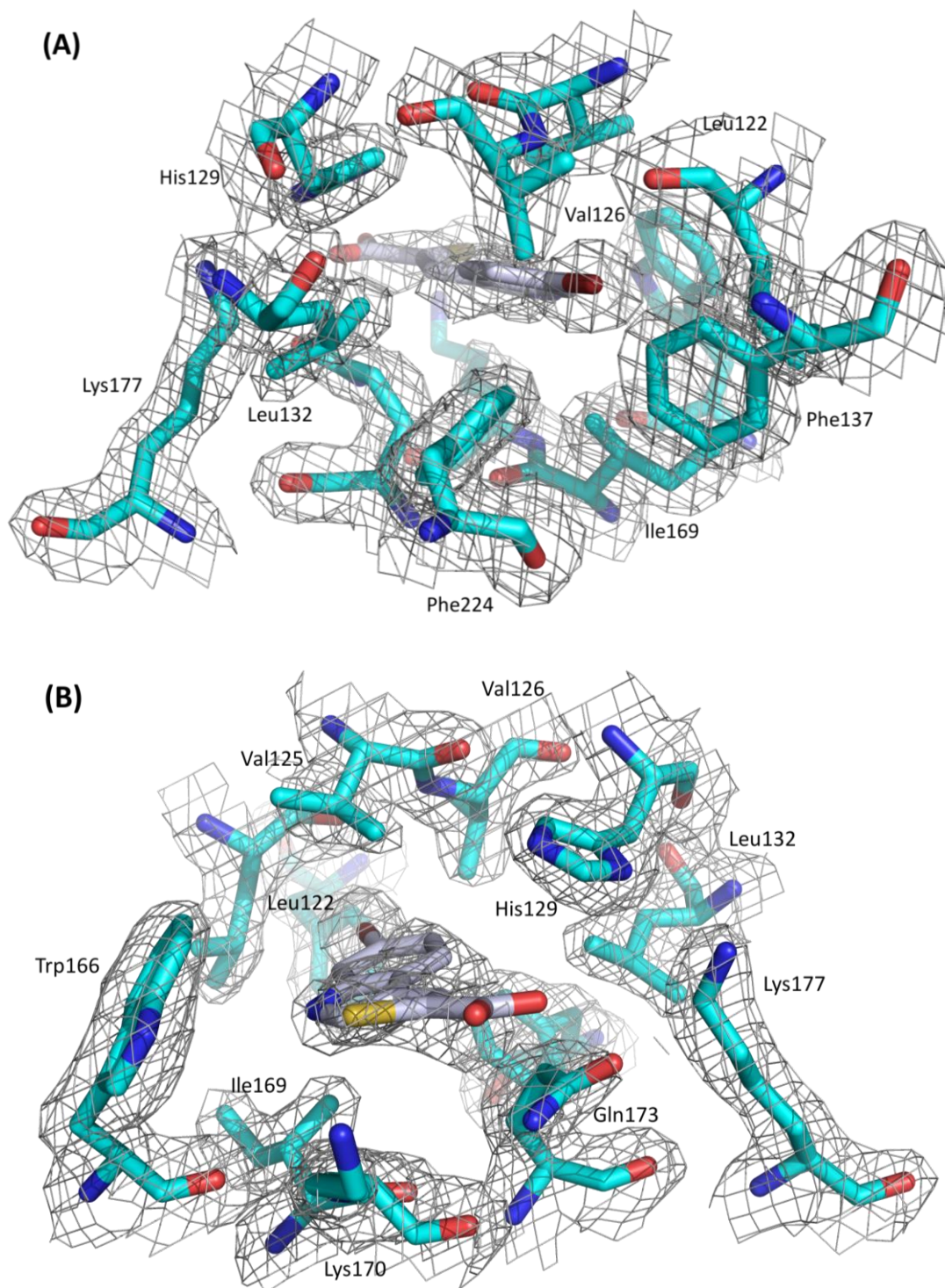
Figure 4.15: *(A) A stick representation of **84** bound to the PEF(S) highlighting the hydrophobic pocket and (B) the α-mercaptoacrylic acid interacts with the residues on the surface of PEF(S). Density map is contoured to 1.0σ.*

Comparison of PEF(S)-84 structure with *holo*-PEF(S)

Very little reorganisation is required for **84** to occupy the binding pocket. The most significant movement occurs with Gln173, where the side chain moves out of the hydrophobic pocket to accommodate the ligand (Figure 4.16). The amine of Lys170 angles towards the thiol of **84** enabling hydrophilic interactions to occur (Figure 4.16).
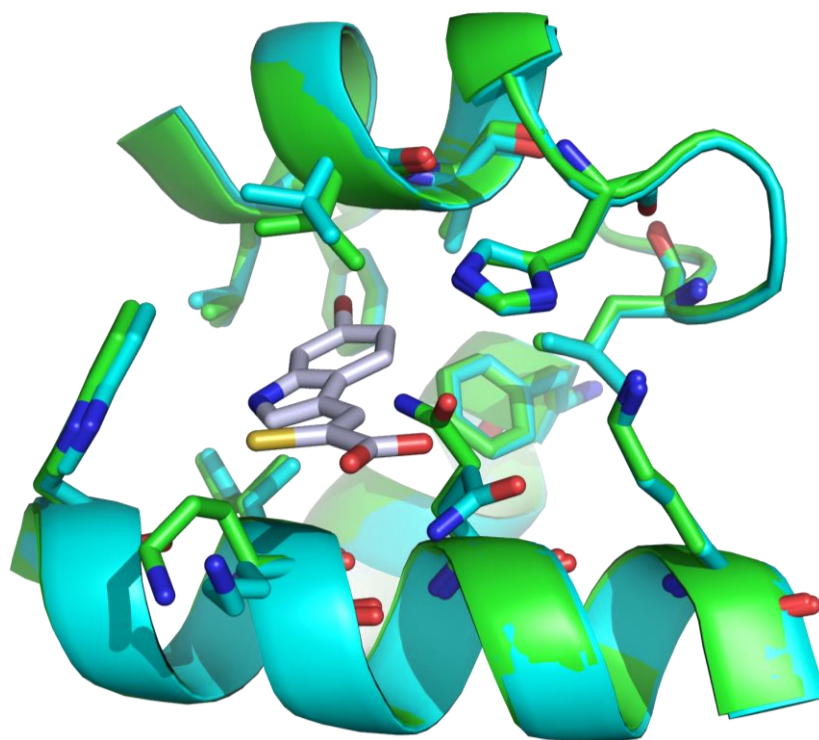
Figure 4.16: *Aligned secondary structure of* holo-*PEF(S) (green) and the complex of **84** (white) bound to PEF(S) (cyan) demonstrating the movement of the side chains upon ligand binding.*

The r.m.s.d values of all the residues of one monomer, within the homodimer, were calculated with Superpose where a monomer of **84** bound to PEF(S) (chain A) was compared to *holo*-PEF(S) (chain A) (Figure 4.17).[234] The values obtained corroborate visual analysis of the binding site; there is very little movement of the residues within the binding site of the ligand. The residues that gave r.m.s.d. values of ∼4 Å are mainly located at loop regions or contain highly mobile side chains (Figure 4.17).
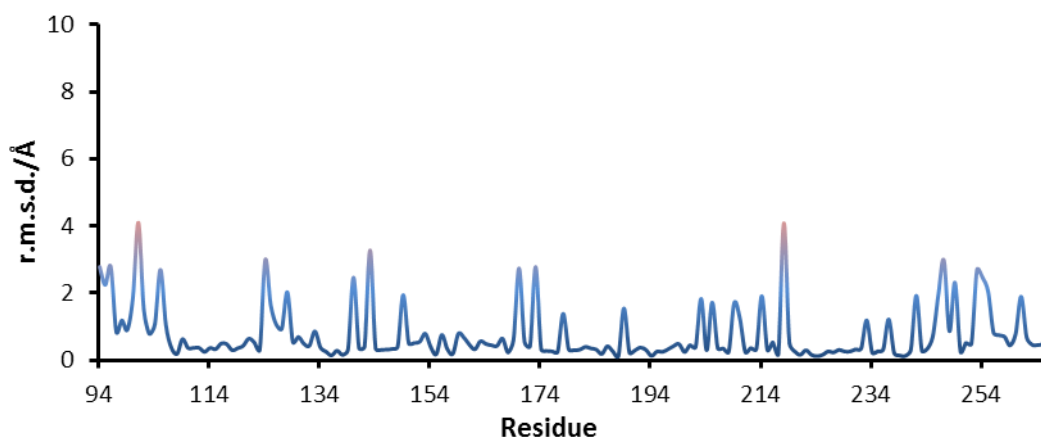
Figure 4.17: *A graph representing the r.m.s.d. of the residues as calculated by Superpose in CCP4, the comparison was carried out between chain A of* holo-*PEF(S) and the complex of PEF(S) and* ***84***.[234]

Comparison of PEF(S)-84 with PEF(S)-PD150606
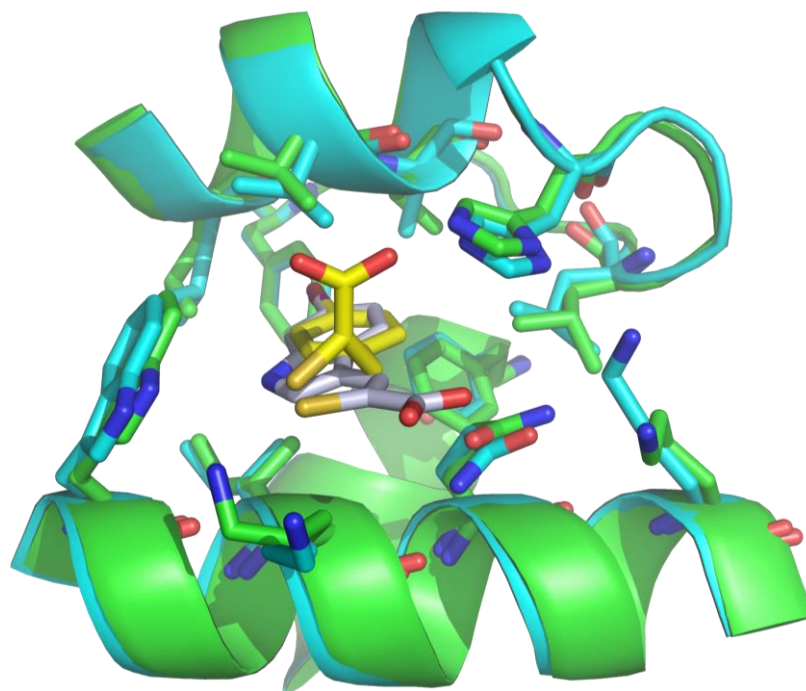
Figure 4.18: *Aligned secondary structure of* ***84*** *(white) bound to PEF(S) (cyan) and the published structure of PD150606 (yellow) bound to PEF(S) (green) (PDB 1NX3).*[186]

Structural analysis of the **84** complex in comparison to the PD150606 complex reveals that the indole based inhibitor appears to bind more deeply into the hydrophobic pocket than the phenyl based inhibitor (Figure 4.18).[186] The α-mercaptoacrylic acid interacts with different residues within the structure. PD150606 interacts with the NH of the indole of Trp166, whereas in the case of

84 the interaction is formed between the carboxylate group of the inhibitor and Lys170 and Gln173.

4.5.3   Comparison of the new ligand bound structures
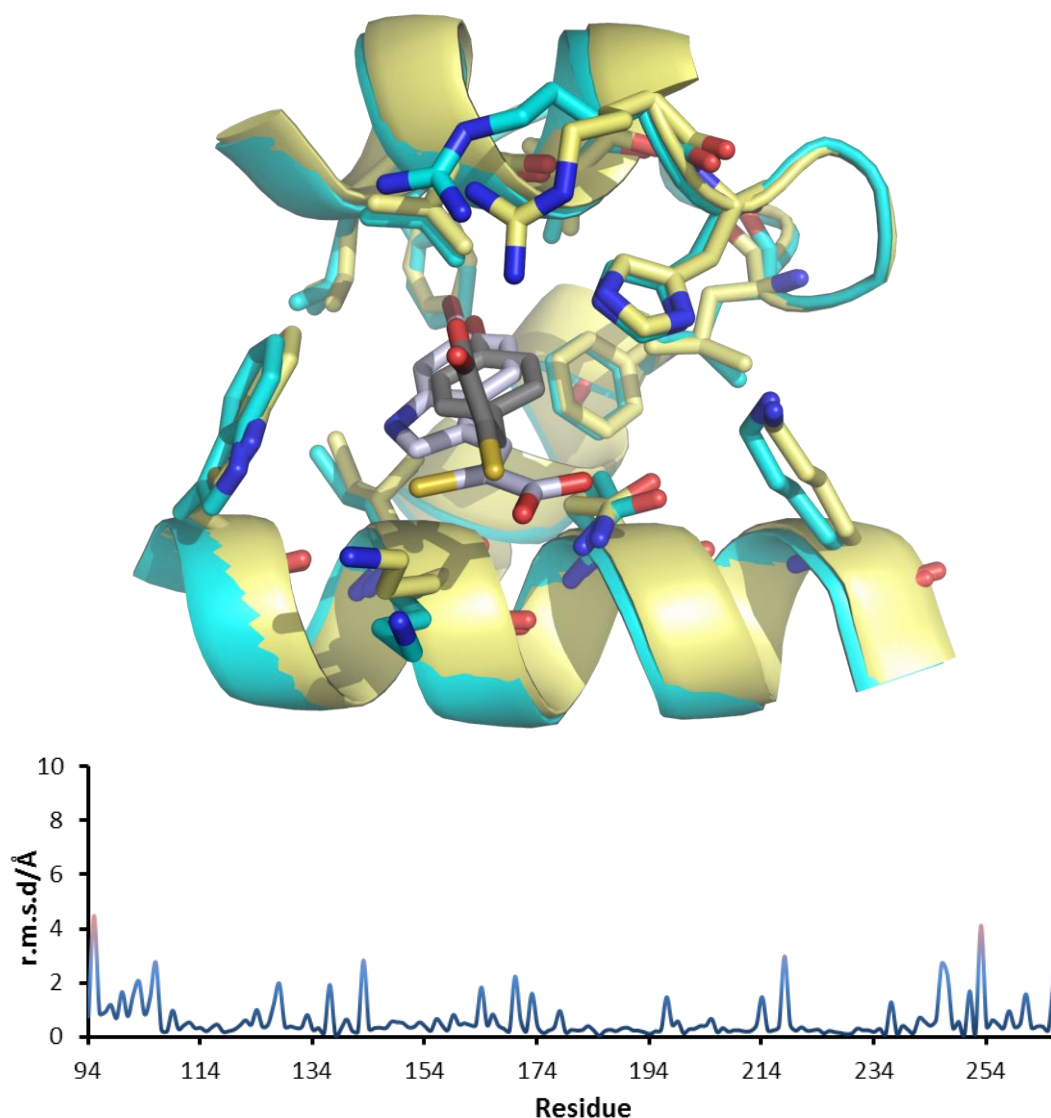
Figure 4.19: *Aligned secondary structure of **74** (grey) bound to PEF(S) (yellow) and the structure of **84** (white) bound to PEF(S) (cyan). A graph representing the r.m.s.d. values of the residues in chain A of the **74**/PEF(S) complex and the **84**/PEF(S) complex, values obtained with Superpose in the CCP4 package.[234]*

The aligned structures of the **74**-PEF(S) complex and **84**-PEF(S) complex show little movement of the side chains of the residues involved in the binding of each ligand. The residues within the hydrophobic pocket maintain very similar conformations upon binding either ligand. The main differences in side chain conformation are those interacting with the α-mercaptoacrylic acid, with

Arg128 interacting with the carboxylate upon the binding of **74** (Figure 4.19). When **84** binds to the pocket the α-mercaptoacrylic acid substituent interacts with Gln173 and Lys170 (Figure 4.19). The different hydrophilic interactions are due to the orientation of the α-mercaptoacrylic acid relative to the aromatic component of the inhibitor. Where the indole is completely planar in the case of **84**, the hydrophilic part of **74** is perpendicular to the phenyl ring and therefore interacts with different hydrophilic residues. The r.m.s.d. plot comparing the two complexes revealed only very small alterations in the residue conformation between both structures (Figure 4.19).[234]

The fact that **84** appears to bind further into the hydrophobic pocket than **74** might indicate a better fit for the pocket. The tighter binding to the hydrophobic pocket in combination with the residues that interact with the α-mercaptoacrylic acid may be the reason why the indole based derivative is a more potent inhibitor than the phenyl based α-mercaptoacrylic acid.

4.6 Conclusion

The two compounds that were soaked into PEF(S), **74** ((*Z*)-3-(4-bromophenyl)-2-mercaptoacrylic acid) and **84** ((*Z*)-3-(6-bromoindol-3-yl)-2-mercaptoacrylic acid), bind into the same hydrophobic pocket in the protein. Though the interactions of these ligands within the binding site are slightly different, the differences are more significant in the hydrophilic interactions with the α-mercaptoacrylic acid. When (*Z*)-3-(6-bromoindol-3-yl)-2-mercaptoacrylic acid binds to the pocket, the residues that interact with the α-mercaptoacrylic acid are Lys170 and Gln173. With (*Z*)-3-(4-bromophenyl)-2-mercaptoacrylic acid this interaction is with Arg128. The different hydrophilic interactions are due to the orientation of the α-mercaptoacrylic acid relative to the aromatic moiety which varies between the compounds. The α-mercaptoacrylic acid group is perpendicular to the phenyl group in the structure of (*Z*)-3-(4-bromophenyl)-2-mercaptoacrylic acid whereas the hydrophilic group is coplanar with the indole of (*Z*)-3-(6-bromoindol-3-yl)-2-mercaptoacrylic acid. The hydrophobic pocket that binds the two analogues is composed of a variety of non-polar residues; Leu122, Val125, Val126, His129, Leu132, Phe137, Trp166, Ile169 and Phe224.

(*Z*)-3-(6-bromoindol-3-yl)-2-mercaptoacrylic acid appears to bind more deeply within the hydrophobic pocket than (*Z*)-3-(4-bromophenyl)-2-mercaptoacrylic acid.

Therefore the difference in the potency of the indole based derivatives (as compared to the phenyl based compounds) may be explained through the increased hydrophilic and electrostatic interactions between the α-mercaptoacrylic acid substituents and hydrophilic residues. The fact that the aromatic component of the indole is bulkier may explain the increased number of contacts. The larger size may also explain why it is positioned more deeply into the hydrophobic pocket.

5 Summary and Outlook

Summary

Routes were identified to allow the synthesis of 24 novel mono-halogenated aromatic α-mercaptoacrylic acids in good yields. These compounds readily form dimers *via* disulfide bonds, but the addition of 10 mM DTT restores the mercaptoacrylate group within 5 minutes allowing the monomeric form to be tested for activity against isolated calpain and live cell motility.

Phenyl α-mercaptoacrylates all displayed low micromolar IC_{50} values towards calpain-I with the nature and position of halogen substituents around the aromatic ring proving important for potency. Indole-based α-mercaptoacrylates were much more potent inhibitors, with IC_{50} values in the nanomolar range. Compounds with a bromine substituent displayed the greatest efficacy (2-7 nM). Testing these compounds against calpain-II revealed them to be moderately selective towards calpain-I.

Single crystal soaks of the PEF(S) domain bound to (*Z*)-3-(6-bromoindol-3-yl)-2-mercaptoacrylic acid and (*Z*)-3-(4-bromophenyl)-2-mercaptoacrylic acid diffracted to resolutions of 2.0 Å and 2.1 Å, respectively. The inhibitors bound in a pocket known to host hydrophobic side chains of the endogenous calpain inhibitor calpastatin. The X-ray structure of the 6-bromoindole-PEF(S) complex revealed that the indole-based inhibitor formed more interactions with alkyl and aromatic residues within the pocket and hydrophilic residues on the surface of the protein than the phenyl-based compound. The indole based inhibitor also appeared to sit more deeply within the pocket than the phenyl ring of the 4-bromophenyl derivative. The combination of these effects provides a likely explanation for the difference in the potency of the indole and phenyl based inhibitors.

When live neutrophils were treated with these inhibitors their ability to spread in response to chemical stimuli, a process mediated by calpain-I, was reduced, with (*Z*)-3-(6-bromoindol-3-yl)-2-mercaptoacrylic acid subduing this process by ~70% over a 3 minute period. Irradiation of live cells containing an inhibitor, with 410 nm light, completely inhibited the cell spreading process,

implying that these compounds have an additional photo-reactive mode of calpain-I inhibition.

Outlook

These compounds are allosteric in their mode of inhibition and are capable of passing through cell membranes. Their activity has also shown that calpain-I is integral for the rapid response of neutrophils to damage cues within the body. Therefore these compounds can be used *in vivo* to define the role of neutrophils in rheumatoid arthritis as well as other diseases where these cells are involved in the pathology.

Calpain-I has also been linked to many other pathological conditions such as cancer metastasis, ischemic cell death and Alzheimer's disease. The α-mercaptoacrylic acid compounds can therefore be used both in cells and *in vivo* to elucidate the exact consequences of calpain-I over-activation within these disease states. The inhibition of calpain-I may prove to be an important component in halting the progression of these diseases and therefore may prove to be a highly valuable therapeutic target.

These inhibitors were shown to bind to the common calcium binding domain, PEF(S), but also demonstrated a greater affinity towards calpain-I over calpain-II. These results imply further interactions must be made with the discriminatory calcium binding domain, PEF(L). Since inhibitor binding to PEF(S) complicates whole-enzyme assays, domains of PEF(L) must be examined individually to determine further inhibitor binding sites. Crystallisation of PEF(L) from both calpain-I and calpain-II (and/or the whole large subunit), with mercaptoacrylic acids will elucidate the molecular basis of the observed differences in affinity. Crystal structures of inhibitors bound to PEF(L) will inform the design of a second generation of inhibitors, that are more stable and interact with increased selectivity with calpain-I or II. Even though the inhibition of calpain-II is embryonically lethal the development of molecular probes for the selective inhibition of this protease will lead to a greater understanding of why this occurs.[171] The role that calpain-II has demonstrated

in mitosis could be exploited to prevent the progression of cancer through its inhibition.[127]

In addition to their use as inhibitors, the α-mercaptoacrylic acid derivatives used in this work have the potential to be used for the development of a new screening assay. Fluorescent derivatives of these compounds may be used as probes that bind to the calcium binding domains of calpain-I and libraries of known compounds can then be used to displace the probe. This would enable rapid screening and discovery of novel classes of calpain inhibitors.

Both of the major isoforms of calpain are involved in many fundamental cellular processes but their individual roles are still poorly understood. Armed with specific inhibitors for each isoform, researchers in many areas of cell biology will be able to make major advances in the understanding and control of cell behaviour linked to calpains I and II within living systems.

6 Materials and Methods

6.1 Organic Synthesis

General Synthetic Methods

All chemicals were purchased from Sigma-Aldrich unless otherwise stated. Anhydrous THF, diethyl ether, toluene and MeCN were obtained from a MBraun SPS800 solvent purification system. All other chemicals were of analytical quality or better and used as received unless otherwise stated. Reactions were stirred at room temperature in air unless otherwise stated. All glassware was clean and dry before use.

^{1}H and ^{13}C NMR spectra were measured on a Bruker Avance 500 NMR spectrometer, a Bruker Avance DPX400 NMR spectrometer, a Bruker Fourier 300 or a Bruker Avance DPX250 NMR spectrometer and are reported as chemical shifts in parts per million downfield from tetramethylsilane, multiplicity (s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet), coupling constant (to the nearest 0.5 Hz) and assignment, respectively. Assignments are made to the limitations of COSY, DEPT 90/135, gradient HSQC and gradient HMBC spectra. ^{19}F NMR spectra were recorded on a Jeol Eclipse +300 NMR spectrometer and are reported in chemical shift downfield from $CFCl_3$ followed by multiplicity and coupling constant (to the nearest 0.5 Hz) if appropriate. IR spectra were recorded on a Perkin-Elmer 1600 series FTIR spectrometer and samples were prepared as KBr disks. Mass spectra were recorded by the UK National Mass spectrometry service in Swansea. High performance liquid chromatography was performed on a Dionex Ultimate 3000 HPLC system fitted with a Dionex Acclaim analytic C18 reverse phase HPLC column (3 µm, 120 Å, 4.6 x 150 mm). Melting points were recorded on a Gallenlamp and are quoted uncorrected.

6.1.1 Synthesis

4-Iodoindole[193] (15)
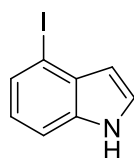
To a stirred suspension of thallium(III)trifluoroacetate (9.13 g, 16.19 mmol) in trifluoroacetic acid (40 ml), methyl indole-3-carboxylate (2.0 g, 11.42 mmol) in trifluoroacetic acid (2 ml) was added dropwise under argon. The suspension was left to stir for 2 hr at room temperature. The suspension was concentrated under reduced pressure then water (128 ml) and potassium iodide (5.7 g, 34.34 ml) were added. This mixture was left to stir for 2 hr. The suspension was extracted with dichloromethane:methanol (95:5, 68 ml) and filtered to remove the solid. The organic layer was extracted and washed with aq. Na_2SO_3 (5% (w/v) 2 × 50 ml) and brine (50 ml). The separated organic layer was dried over anhydrous $MgSO_4$, filtered and concentrated under reduced pressure. The resulting brown oil was passed through a pad of flash silica using hexane:ethyl acetate (1:1) resulting in a crude product which was hydrolysed without further purification.

To a solution of the crude material (2.0 g, 6.62 mmol) in methanol (45 ml), 40% aqueous sodium hydroxide solution was added (45 ml) and the solution was heated under reflux for 1.5 hr. The solution was cooled to room temperature and concentrated under reduced pressure to remove the methanol. The suspension was extracted with dichloromethane:methanol (95:5, 3 x 50 ml ml) and the organic layer was washed with brine (50 ml). The organic layer was dried over anhydrous $MgSO_4$, filtered and concentrated under reduced pressure. The resulting brown oil was purified by flash chromatography on silica gel (hexane:ethyl acetate 3:1) to give the title compound as pale yellow crystals (0.90 g, 33%). mp. 96 - 99 °C (lit. 98.5 - 99.5 °C)[236]. δ_H (400 MHz. CDCl$_3$): 8.36 (1 H, s, NH), 7.56 (1 H, d, J = 7.5, ArCH), 7.39 (1 H, d, J = 8.5, ArCH), 7.29 (1 H, d, J = 2.5, ArCH), 6.96 (1 H, t, J = 7.5, ArCH), 6.52 (1 H, td, J_1 = 2.5, J_2 = 1.0, ArCH). δ_C (100 MHz, CDCl$_3$): 134.73 (ArC), 132.53 (ArC), 129.45 (ArCH), 124.49

(ArCH), 123.36 (ArCH), 111.11 (ArCH), 106.39 (ArCH), 87.49 (ArCI). Mass spectrum: HRMS (EI⁺) found 242.9544, C_8H_6IN calculated 242.9545. υ cm⁻¹: 3101.9, 1733.7, 1623.3, 1605.5, 1558.2, 1500.4, 1483.0, 1474.3, 1424.7, 1411.2, 1330.2, 1268.5, 1175.9, 1138.3, 1094.9, 1067.4, 1042.8, 902.0, 881.8, 799.8, 777.2, 757.9, 745.8, 702.8, 619.0, 562.6, 515.4, 484.5, 437.3.

N-Acetyl-6-nitroindoline[191] (17)
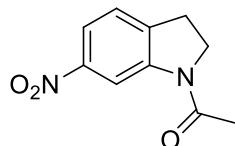
A stirred suspension of 6-nitroindoline (1.0 g, 6.09 mmol) in acetic anhydride (5 ml) was heated to 130 °C for 15 min. The reaction mixture was cooled to room temperature and water was added (20 ml). The resulting suspension was heated under reflux for 45 min. The reaction mixture was left to cool to room temperature; the precipitate that formed was collected by filtration under reduced pressure and washed with water (20 ml). The solid was then air dried under reduced pressure to give the title compound as a yellow solid (1.19 g, 5.77 mmol, 95%). mp. 167 - 169 °C (lit. 152 - 156 °C)[191]. δ_H (400 MHz, DMSO-d₆): 8.73 (1H, d, *J* = 2.0, ArCH), 7.87 (1 H, dd, J_1 = 8.5, J_2 = 2.0, ArCH), 7.45 (1 H, d, *J* = 8.5, ArCH), 4.18 (2 H, t, *J* = 8.5, CH₂), 3.24 (2 H, t, *J* = 8.5, CH₂), 2.19 (3H, s, CH₃). δ_C (125 MHz, DMSO-d₆): 170.00 (*C*OCH₃), 147.33 (ArC), 144.23 (ArC), 140.87 (ArCNO₂), 125.77 (ArCH), 119.11 (ArCH), 110.15 (ArCH), 49.26 (CH₂), 27.96 (CH₂), 24.31 (CH₃). Mass spectrum: HRMS (EI⁺) found 206.0694, $C_{10}H_{10}N_2O_3$ calculated 206.0691. υ cm⁻¹ : 3135.2, 1666.7, 1592.4, 1526.4, 1515.8, 1481.1, 1430.4, 1401.5, 1334.5, 1103.1, 1069.8, 1029.8, 892.9, 813.3, 745.4, 636.4, 565.5, 516.8.

N-Acetyl-6-aminoindoline[191] (18)
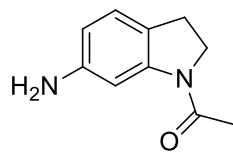
A suspension of *N*-acetyl-6-nitroindoline (1.00 g, 4.86 mmol), 10% Pd/C (0.1 g, 0.09 mmol Pd) in methanol (160 ml) was stirred vigorously under an atmosphere of hydrogen for 24 hrs. The Pd/C was removed by filtration under reduced pressure through a plug of Celite©. The filtrate was concentrated *in vacuo* to leave the title compound as a white solid (0.68 g, 3.89 mmol, 80%). mp. 172 - 177 °C (lit. 187 - 191 °C)[191]. δ_H (400 MHz, DMSO-d_6): 7.43 (1 H, d, *J* = 2.0, ArCH), 6.83 (1 H, d, *J* = 8.0, ArCH), 6.20 (1 H, dd, *J_1* = 8.0, *J_2* = 2.0, ArCH), 4.97 (2H, s, NH$_2$), 3.99 (2 H, t, *J* = 8.0, CH$_2$), 2.92 (2 H, t, *J* = 8.0, CH$_2$), 2.11 (3 H, s, CH$_3$). δ_C (62.5 MHz, DMSO-d_6): 168.05 (*C*OCH$_3$), 147.86 (ArC), 143.52 (ArC), 124.51 (ArCH), 118.28 (ArCH), 108.79 (ArCH), 102.59 (ArCNH$_2$), 48.73 (CH$_2$), 26.50 (CH$_2$), 24.12 (CH$_3$). Mass spectrum: HRMS (APCI$^+$) found 177.1031, $C_{10}H_{13}N_2O$ calculated 177.1028. υ cm^{-1}: 3413.4, 3320.8, 3113.5, 2916.8, 1644.5, 1605.9, 1497.0, 1456.5, 1417.4, 1362.5, 1330.2, 1308.5, 1243.9, 1204.8, 1167.2, 1030.3, 954.6, 876.5, 806.1, 631.6, 550.1.

N-Acetyl-6-iodoindoline[191] (19)
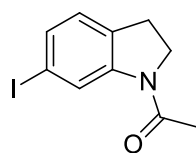
To a stirred solution of *N*-acetyl-6-aminoindoline (0.70 g, 4.04 mmol) in 80% acetic acid (31 ml) under argon at 0 °C, sodium nitrite (0.32 g, 4.64 mmol) in water (1.4 ml) was added and the mixture was stirred for 15 min. Potassium iodide (3.3 g, 19.88 mmol) in water (7.0 ml) was added to the reaction mixture which was then stirred for 24 hrs at 0 °C. The solution was concentrated *in vacuo* and the black residue was dissolved in ethyl acetate (45 ml) and washed with water (3 × 45 ml). The organic layer was separated and dried over anhydrous MgSO$_4$ then filtered and concentrated under reduced pressure. The

brown residue was purified using flash chromatography on silica gel (toluene:acetone 20:1) to give the title compound as yellow crystals. (0.70 g, 2.42 mmol, 60%). mp. 131 - 135 °C (lit. 102 - 118 °C)[191]. δ_H (400 MHz, CDCl$_3$): 8.55 (1 H, d, *J* = 2.0, ArCH), 7.30 (1 H, dd, *J$_1$* = 8.0, *J$_2$* = 2.0, ArCH), 6.87 (1 H, d, *J* = 8.0, ArCH), 4.07 (2 H, t, *J* = 8.5, CH$_2$), 3.12 (2 H, t, *J* = 8.5, CH$_2$), 2.19 (3 H, s, CH$_3$). δ_C (125 MHz, CDCl$_3$): 168.87 (*C*OCH$_3$), 144.14 (ArC), 132.55 (ArCH), 130.96 (ArC), 126.09 (ArCH), 125.56 (ArCH), 92.19 (ArCI), 48.98 (CH$_2$), 28.01 (CH$_2$), 24.21 (CH$_3$). Mass spectrum: HRMS (EI$^+$) found 286.9800, C$_{10}$H$_{13}$INO calculated 287.9807. υ cm^{-1}: 3113.5, 2960.2, 2360.0, 1659.5, 1594.8, 1482.0, 1437.7, 1413.1, 1340.3, 1309.4, 1264.6, 1125.7, 1101.6, 1031.3, 927.1, 882.8, 787.3, 768.0, 624.3, 598.3, 558.3, 508.6, 425.7.

6-Iodoindoline[191] (20)
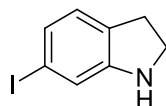
A 40% aqueous sodium hydroxide solution (12 ml) was added to a solution of *N*-acetyl-6-iodoindoline (0.2 g, 0.68 mmol) in methanol (12 ml). The reaction mixture was heated to 75 °C for 2.5 hrs. The cooled reaction mixture was concentrated under reduced pressure. The residue was extracted with ethyl acetate (3 × 30 ml) and the organic layer was dried over anhydrous MgSO$_4$, filtered and then concentrated under reduced pressure. The resulting yellow oil was purified using flash chromatography on silica gel (hexane:ethyl acetate 10:1) to give the title compound as white crystals, (0.11 g, 0.44 mmol, 64%). mp. 93 - 96 °C (lit. 83 - 90 °C)[191]. δ_H (250 MHz, CDCl$_3$): 6.91 (1 H, dd, *J$_1$* = 7.5, *J$_2$* = 1.5, ArCH), 6.85 (1 H, d, *J* = 1.5, ArCH), 6.74 (1 H, d, *J* = 7.5, ArCH), 3.45 (2 H, t, *J*=8.5, CH$_2$), 2.89 (2H, t, *J* = 8.5, CH$_2$). δ_C (100 MHz, CDCl$_3$): 153.41 (ArC), 129.19 (ArC), 127.35 (ArCH), 126.24 (ArCH), 117.97 (ArCH), 91.95 (ArCI), 47.49 (CH$_2$), 29.36 (CH$_2$). Mass spectrum: HRMS (EI$^+$) found 244.9707, C$_8$H$_8$IN calculated 244.9702. υ cm^{-1}: 3378.2, 2931.3, 2864.3, 1590.0, 1488.3, 1474.3, 1438.2, 1321.0, 1309.4, 1250.6, 1239.0, 1160.9, 1048.1, 1021.6, 929.0, 881.8, 845.2, 798.4, 724.6, 699.1, 573.7, 551.5, 513.9, 422.3, 409.3.

6-Iodoindole[191] (21)
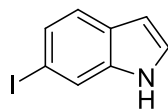
Air was bubbled through a stirred solution of 6-iodoindoline (0.23 g, 0.94 mmol) and *N,N*-bis(salicylidene)ethylenediaminocobalt(II) (0.034 g, 0.10 mmol) in methanol (30 ml) for 24 hr. The reaction mixture was concentrated under reduced pressure. The residue was then purified by flash chromatography on silica gel (hexane:ethyl acetate 8:1) to give the title compound as pale brown crystals, (0.18 g, 0.73 mmol, 77%). mp. 108 - 113 °C (lit. 111.0 – 111.5 °C)[237]. δ_H (400 MHz, CDCl$_3$): 7.96 (1 H, s, NH), 7.62 (1 H, d, *J* = 1.0, ArCH), 7.31 (2 H, s, ArCH), 7.02 (1 H, t, *J* = 3.0, ArCH), 6.43 (1 H, q, *J* = 1.0, ArCH). δ_C (125 MHz, CDCl$_3$): 137.14 (ArC), 128.67 (ArCH), 127.25 (ArC), 124.66 (ArCH), 122.40 (ArCH), 120.06 (ArCH), 102.90 (ArCH), 85.80 (ArCI). Mass spectrum: HRMS (EI$^+$) found 242.9538, C$_8$H$_6$IN calculated 242.9545. υ cm^{-1}: 3412.4 1624.3, 1598.7, 1494.6, 1451.7, 1440.1, 1394.3, 1333.1, 1311.4, 1232.8, 1201.4, 1089.6, 1042.8, 996.1, 890.0, 878.9, 859.1, 810.0, 758.4, 729.9, 609.4.

6.1.2 General method for the preparation of 7 substituted indoles[192]

7-Bromoindole (23)
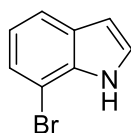
A stirred solution of 1-bromo-2-nitrobenzene (1.0 g, 4.95 mmol) in THF (50 ml) under argon was cooled to -40 °C (dry-ice - acetone bath). To this solution was quickly added vinylmagnesium bromide (1 M in THF, 14.76 ml, 14.76 mmol). The reaction was left to stir for 20 minutes at -40 °C. The reaction mixture was quenched with saturated aq. NH$_4$Cl (50 ml) then extracted with diethyl ether (2 × 50 ml). The pooled organic extracts were dried with MgSO$_4$, filtered and concentrated under reduced pressure. The black oil that resulted was purified using flash chromatography on silica gel (hexane:ethyl acetate 6:1) giving the title compound as an orange solid (0.43 g, 44%). mp. 39 - 41 °C (lit. 43 - 44

°C)[192]. δ_H (400 MHz. CDCl$_3$): 8.37 (1 H, s, NH), 7.62 (1 H, d, *J* = 8.0, ArCH), 7.29 (1 H, t, *J* = 3.0, ArCH), 7.04 (1 H, t, *J* = 8.0, ArCH), 6.67 (1 H, dt, *J$_1$* = 3.0, *J$_2$* = 1.0, ArCH). δ_C (100 MHz, CDCl$_3$): 134.59 (ArC), 129.00 (ArC), 124.74 (ArCH), 124.34 (ArCH), 121.02 (ArCH), 119.98 (ArCH), 104.67 (ArCBr), 103.86 (ArCH). Mass spectrum: HRMS (EI$^+$) found 194.9686, C$_8$H$_6$^{79}BrN calculated 194.9684. υ cm^{-1}: 3400.9, 1901.0, 1613.6, 1560.1, 1483.0, 1428.0, 1408.3, 1329.7, 1282.0, 1214.0, 1190.4, 1135.9, 1094.4, 1065.5, 1032.2, 1032.2, 920.4, 878.4, 817.2.

7-Fluoroindole (25)
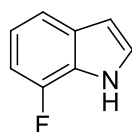
Yield; 14%. mp. 43 - 52 °C (lit. 61 - 63 °C)[192]. δ_H (400 MHz. CDCl$_3$): 8.17 (1 H, s, NH), 7.31 (1 H, td, *J$_1$* = 8.0, *J$_2$* = 1.0, ArCH), 7.07 (1 H, t, *J* = 3.0, ArCH), 6.92 (1 H, td, *J$_1$* = 8.0, *J$_2$* = 2.5, ArCH), 6.81 (1 H, m, ArCH), 6.49 (1 H, dd, *J$_1$* = 2.5, *J$_2$* = 1.0, ArCH). δ_C (100 MHz, CDCl$_3$): 149.71 (ArCF, d, *J* = 24.0), 131.52 (ArC, d, *J* = 5.0), 124.91 (ArCH, s), 124.27 (ArC, d, *J* = 13.0), 120.06 (ArCH, d, *J* = 6.0), 116.46 (ArCH, d, *J* = 13.0), 106.79 (ArCH, d, *J* = 16.0), 103.35 (ArCH, d, *J* = 2.0). Mass spectrum: HRMS (EI$^+$) 135.0481, found C$_8$H$_6$FN calculated 135.0484. υ cm^{-1}: 3390.7, 3111.1, 3071.1, 2956.3, 2923.6, 2689.8, 2627.1, 2516.2, 2361.9, 1895.7, 1823.9, 1726.9, 1640.6, 1610.3, 1579.4, 1521.6, 1489.7, 1443.0, 1415.0, 1342.7, 1286.8, 1235.2, 1214.9, 1160.9, 1106.0, 1066.0, 1046.2, 1023.5, 949.3, 891.9, 878.9, 867.0, 849.0, 841.3.

6.1.3   General method for the preparation of indole-3-carboxaldehydes[189]

4-Fluoroindole-3-carboxaldehyde (26)
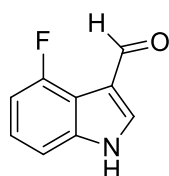
A stirred mixture of 4-fluoroindole (0.50 g, 3.71 mmol) and *N,N*-dimethylformamide (1.37 g, 18.78 mmol) was placed in an ice bath until the

solution reached 5 °C. Phosphorous oxychloride (0.69 g, 4.50 mmol) was then added dropwise. The mixture was left to stand for 15 minutes at room temperature. The stirred mixture was then heated to 40 °C for 1 hr and was then cooled. Ice (5 g) was added along with sodium hydroxide solution (2 M, 6 ml) and then the mixture was heated under reflux for 16 hr. After cooling to room temperature; the precipitate that formed was collected by filtration under reduced pressure and washed with water (2 × 10 ml). The solid was air dried under reduced pressure to give the required product as a pink solid (0.59 g, 3.61 mmol, 98%). mp. 175 - 181 °C. δ_H (400 MHz, DMSO-d$_6$): 12.51 (1 H, s, NH), 10.00 (1 H, dd, J_1 = 8.0, J_2 = 3.5, CHO), 8.31 (1 H, d, J = 3.5, ArCH), 7.35 (1 H, t, J = 8.0, ArCH), 7.24 (1 H, td, J_1 = 8.0, J_2 = 5.0, ArCH), 7.01 (1 H, td, J_1 = 8.0, J_2 = 3.0, ArCH). δ_C (125 MHz, DMSO-d$_6$): 184.07 (CHO), 156.48 (ArCF, d, J = 248.5), 140.21 (ArCH, d, J = 12.0), 136.42 (ArC, s), 124.39 (ArCH, d, J = 8.0), 117.43 (ArC, d, J = 6.0), 113.39 (ArCH, d, J = 21.9), 109.55 (ArCH, d, J = 4.0), 107.71 (ArCH, d, J = 19.0). Mass spectrum: HRMS (EI$^+$) found 162.0434, C_9H_5FNO calculated 162.0433. υ cm^{-1}: 3231.2, 2819.4, 2748.1, 1653.7, 1635.3, 1581.3, 1513.4, 1455.0, 1437.7, 1428.0, 1403.4, 1364.9, 1347.0, 1301.2, 1251.6, 1222.7, 1161.9, 1140.2, 1111.3, 1146.7, 1030.3, 849.0, 789.7, 783.0, 764.6, 734.3, 638.3, 618.1, 588.2, 531.3, 519.2.

4-Chloroindole-3-carboxaldehyde (27)
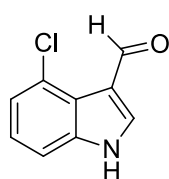
Yield; 80%. mp. 158 - 162 °C (lit. 166 - 167 °C)[238]. δ_H (400 MHz, DMSO-d$_6$): 12.59 (1 H, s, NH), 10.50 (1 H, s, CHO), 8.31 (1 H, s, ArCH), 7.53 (1 H, d, J = 8.0, ArCH), 7.30 (1 H, d, J = 8.0, ArCH), 7.23 (1 H, t, J = 8.0, ArCH). δ_C (125 MHz, DMSO-d$_6$): 185.20 (CHO), 138.75 (ArC), 134.47 (ArCH), 125.03 (ArCCl), 124.01 (ArCH), 123.37 (ArC), 123.05 (ArCH), 118.27 (ArC), 112.45 (ArCH). Mass Spectrum HRMS [EI$^+$] found 179.0138, $C_9H_7{}^{35}ClNO$ calculated 179.0138. υ cm^{-1}: 3163.7, 3127.5, 3006.5, 2939.5, 2870.5, 1894.7, 1818.5, 1747.2, 1639.4, 1570.3,

1521.1, 1485.4, 1459.4, 1430.0, 1388.5, 1341.3, 1331.6, 1293.0, 1258.8, 1189.4, 1147.4, 1082.4, 1072.2, 1048.6, 937.7, 872.6, 845.2, 773.8, 729.9, 665.3, 602.6.

4-Bromoindole-3-carboxaldehyde (28)
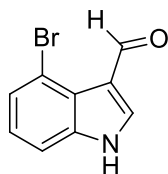
Yield; 94 %. mp. 175 - 177 °C (lit. 185 - 187 °C)[238]. δ_H (400 MHz, DMSO-d_6): 12.59 (1 H, s, NH), 10.69 (1 H, s, CHO), 8.30 (1 H, s, ArCH), 7.58 (1 H, d, J = 8.0, ArCH), 7.48 (1H, d, J=8.0, ArCH), 7.17 (1 H, t, J = 8.0, ArCH). δ_C (125 MHz, DMSO-d_6): 185.03 (CHO), 138.71 (ArC), 134.31 (ArCH), 126.45 (ArCH), 125.16 (ArC), 124.27 (ArCH), 118.29 (ArC), 112.92 (ArCH), 112.75 (ArCBr). Mass Spectrum HRMS [EI+] found 222.9635, $C_9H_7{}^{79}BrNO$ calculated 222.9633. υ cm^{-1}: 3166.1, 3065.3, 2865.3, 1902.0, 1761.7, 1635.8, 1518.2, 1483.0, 1457.0, 1428.5, 1385.6, 1330.6, 1293.5, 1260.7, 1191.3, 1147.9, 1130.1, 1075.1, 1045.2, 913.6, 888.1, 841.3, 773.3, 732.3, 658.6, 608.9.

4-Iodoindole-3-carboxaldehyde (29)
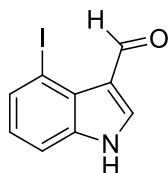
Yield; 92%. mp. 186 - 188 °C (lit. 182 – 183 °C)[239]. δ_H (400 MHz, DMSO-d_6): 12.50 (1 H, s, NH), 10.95 (1 H, s, CHO), 8.30 (1 H, s, ArCH), 7.73 (1 H, d, J = 8.0, ArCH), 7.59 (1 H, d, J = 8.0, ArCH), 7.01 (1 H, t, J = 8.0, ArCH). δ_C (125 MHz, DMSO-d_6): 184.25 (CHO), 138.29 (ArC), 134.40 (ArCH), 133.29 (ArCH), 128.75 (ArC), 124.59 (ArCH), 118.09 (ArC), 113.41 (ArCH), 83.89 (ArCI). Mass Spectrum HRMS [APCI+] found 271.9572, C_9H_7INO calculated 271.9572. υ cm^{-1}: 3155.9, 2929.8, 2863.8, 1767.4, 1628.1, 1515.3, 1480.6, 1456.0, 1424.2, 1383.7, 1337.4, 1292.1, 1259.3, 1191.3, 1147.9, 1125.7, 1068.9, 1040.9, 897.2, 836.0, 799.8, 773.8, 764.2, 734.3, 655.2, 609.9, 600.2, 561.7, 487.9.

5-Iodoindole-3-carboxaldehyde (30)
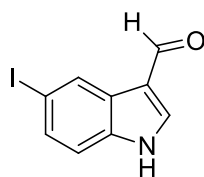
Yield: 95%. mp. 232 - 234 °C (lit. 220 °C)[240]. δ_H (400 MHz, DMSO-d_6): 12.29 (1 H, s, NH), 9.92 (1 H, s, CHO), 8.44 (1 H, s, ArCH), 8.30 (1 H, d, J = 2.0, ArCH), 7.54 (1 H, d, J = 8.5, ArCH), 7.38 (1 H, d, J = 8.5, ArCH). δ_C (62.5 MHz, DMSO-d_6): 185.10 (COH), 138.84 (ArC, s), 136.11 (ArC, s), 131.49 (ArCH, s), 129.10 (ArCH, s), 126.17 (ArC, s), 117.12 (ArCH, s), 114.88 (ArCH, s), 86.55 (ArCI, s). Mass spectrum: HRMS (ES$^-$) found 269.9405, C_9H_5INO calculated 269.9416. υ cm^{-1}: 2894.6, 1622.8, 1565.9, 1523.0, 1478.7, 1435.3, 1384.2, 1318.1, 1284.4, 1232.3, 1134.4, 1091.0, 1042.3, 877.5, 795.0, 769.9, 748.3, 661.5, 608.4, 562.2.

6-Fluoroindole-3-carboxaldehyde (31)
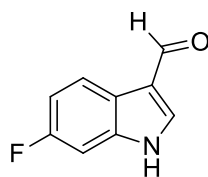
Yield; 77%. mp. 173 - 178 °C (lit. 173 - 174 °C)[241]. δ_H (400 MHz, DMSO-d_6): 12.19 (1 H, s, NH), 9.92 (1 H, s, CHO), 8.30 (1 H, d, J = 2.0, ArCH), 8.08 (1 H, dd, J_1 = 9.0, J_2 = 2.5, ArCH), 7.32 (1 H, dd, J_1 = 9.0, J_2=2.5, ArCH), 7.08 (1 H, td, J_1=9.0, J_2=2.5, ArCH). δ_C (125 MHz, DMSO-d_6): 185.42 (CHO), 160.02 (ArCF, d, J = 236.0), 139.54 (ArCH), 137.64 (ArC, d, J = 12.5), 122.39 (ArCH, d, J = 10.0), 121.28 (ArC), 118.53 (ArC), 110.90 (ArCH, d, J=25.0), 99.25 (ArCH, d, J=26.5). Mass spectrum: HRMS (EI$^+$) found 162.0431, C_9H_5FNO calculated 162.0433. υ cm^{-1}: 3868.5, 3852.1, 3819.8, 3800.0, 3748.0, 3743.2, 3674.2, 3445.7, 3130.9, 2812.7, 1626.7, 1594.8, 1532.2, 1505.2, 1448.8, 1389.5, 1335.0, 1283.4, 1230.4, 1149.4, 1117.6, 1082.4, 951.2, 833.1, 812.9, 718.4, 633.5, 620.0, 604.1, 509.6.

6-Chloroindole-3-carboxaldehyde (32)
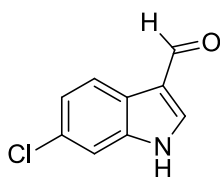
Yield; 96%. mp. 212 - 214 °C (lit. 210 - 211°C)[242]. δ_H (400 MHz, DMSO-d$_6$): 12.23 (1 H, s, NH), 9.94 (1 H, s, COH), 8.40 (1 H, d, *J* = 2.0, ArCH), 8.08 (1 H, d, *J* = 8.0, ArCH), 7.57 (1 H, s, ArCH), 7.24 (1 H, d, *J* = 8.0, ArCH). δ_C (62.5 MHz, DMSO-d$_6$): 185.06 (CHO), 139.20 (ArCH), 137.49 (ArCH), 127.90 (ArCCl), 122.86 (ArC), 122.41 (ArCH), 122.05 (ArCH), 117.96 (ArC), 112.97 (ArCH). Mass spectrum: HRMS (APCI⁻) found 178.0055, $C_9H_5{}^{35}ClNO$ calculated 178.0060. υ cm⁻¹: 3434.1, 3108.2, 3041.9, 2968.9, 2915.4, 2853.2, 1634.4, 1575.6, 1525.4, 1492.2, 1451.7, 1430.9, 1384.6, 1327.8, 1278.6, 1243.4, 1145.0, 1132.0, 1090.1, 1061.1, 906.4, 841.8, 811.4, 752.6, 683.6, 605.5, 504.8.

6-Iodoindole-3-carboxaldehyde (33)
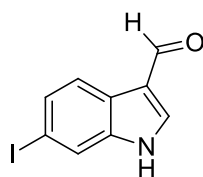
Yield; 95%. mp. 213 - 215 °C (lit. 216 °C)[237]. δ_H (400 MHz, DMSO-d$_6$): 12.19 (1 H, s, NH), 9.93 (1 H, s, CHO), 8.27 (1 H, d, *J*=3.0, ArCH), 7.90 (1 H, d, *J* = 8.5, ArCH), 7.88 (1 H, s, ArCH), 7.51 (1 H, dd, *J$_1$* = 8.5, *J$_2$* = 1.5, ArCH). δ_C (100 MHz, DMSO-d$_6$): 185.61 (CHO), 139.31 (ArCH), 138.84 (ArC), 131.02 (ArCH), 123.20 (ArCH), 121.45 (ArCH), 123.90 (ArC), 118.44 (ArC), 88.20 (ArCI). Mass Spectrum HRMS [ES⁻] found 269.9409, C_9H_7INO calculated 269.9416. υ cm⁻¹: 3434.6, 2846.9, 1683.6, 1582.3, 1563.0, 1510.0, 1446.4, 1355.2, 1327.8, 1302.7, 1232.8, 1219.3, 1141.7, 1129.6, 1073.7, 1054.4, 847.6, 794.1, 774.8, 725.1, 712.1, 679.3, 661.5.

7-Fluoroindole-3-carboxaldehyde (34)
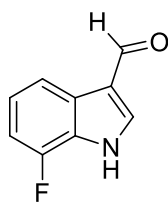
Yield; 54%. mp. 136 - 142 °C. δ_H (400 MHz, DMSO-d$_6$): 12.71 (1 H, s, NH), 9.98 (1 H, s, COH), 8.37 (1 H, d, J = 3.0, ArCH), 7.91 (1 H, d, J = 7.5, ArCH), 7.20 (1 H, dt, J_1 = 7.5, J_2 = 4.5 , ArCH), 7.17 (1 H, dd, J_1 = 11.5, J_2 = 7.5, ArCH). δ_C (125 MHz, DMSO-d$_6$): 185.67 (CHO), 149.53 (ArCF, d, J = 243.5), 139.41 (ArCH), 128.24 (ArC, d, J = 5.0), 125.26 (ArC, d, J = 12.5), 123.33 (ArCH, d, J = 6.0), 119.26 (ArC), 117.43 (ArCH, d, J = 4.0), 109.00 (ArCH, d, J = 16.0). Mass Spectrum HRMS [EI$^+$] found 163.0431, C$_9$H$_7$FNO calculated 163.0433. υ cm^{-1}: 3744.6, 3675.7, 3629.4, 3117.9, 3035.9, 2877.3, 2362.9, 1617.5, 1529.8, 1499.4, 1467.1, 1400.6, 1344.6, 1282.0, 1229.9, 1179.7, 1151.3, 1133.0, 1044.3, 966.2, 821.5.

7-Chloroindole-3-carboxaldehyde (35)
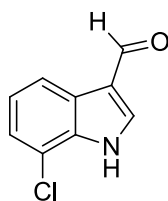
Yield; 89%. mp. 177 - 180 °C (lit. 181 - 182 °C)[243]. δ_H (400 MHz, DMSO-d$_6$): 12.56 (1 H, s, NH), 9.97 (1 H, s, COH), 8.39 (1 H, d, J=3.0, ArCH), 8.06 (1 H, d, J = 8.0, ArCH), 7.36 (1 H, d, J = 8.0, ArCH), 7.23 (1 H, t, J = 8.0, ArCH). δ_C (125 MHz, DMSO-d$_6$): 185.81 (CHO), 139.61 (ArCH), 134.39 (ArCCl), 126.48 (ArC), 123.79 (ArCH), 123.54 (ArCH), 120.25 (ArCH), 119.34 (ArC), 117.13 (ArC). Mass Spectrum HRMS [EI$^+$] found 179.0134, C$_9$H$_7$^{35}ClNO calculated 179.0138. υ cm^{-1}: 3434.6, 3135.2, 3111.1, 3064.8, 3034.9, 2938.5, 2888.8, 2825.2, 2698.4, 2577.9, 1938.1 1642.6, 1621.4, 1526.9, 1499.9, 1453.3, 1390.4, 1340.3, 1275.7, 1248.2, 1206.7, 1160.0, 1130.1, 1111.8, 1048.6, 976.8, 890.5, 848.0, 780.1, 740.5, 658.1, 610.36.

7-Bromoindole-3-carboxaldehyde (36)
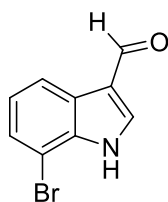
Yield; 82%. mp. 167 - 170 °C (lit. 171 – 172)[237]. δ_H (400 MHz, DMSO-d$_6$): 12.40 (1 H, s, NH), 9.97 (1 H, s, CHO), 8.38 (1 H, s, ArCH), 8.11 (1 H, d, J = 8.0, ArCH), 7.50 (1H, dd, J_1=8.0, J_2 = 1.0, ArCH), 7.23 (1H, t, J = 8.0, ArCH). δ_C (125 MHz, DMSO-d$_6$): 185.93 (CHO), 139.67 (ArCH), 135.92 (ArC), 126.62 (ArCH), 126.23 (ArC), 124.18 (ArCH), 120.72 (ArCH), 119.34 (ArC), 105.30 (ArCBr). Mass Spectrum HRMS [EI$^+$] found 221.9552, $C_9H_7{}^{79}BrNO$ calculated 221.9554. υ cm^{-1}: 3434.1, 3143.9, 3034.9, 2929.3, 2881.1, 2358.0, 1638.7, 1617.5, 1568.3, 1522.0, 1447.3, 1426.6, 1389.9, 1337.9, 1271.3, 1242.9, 1203.9, 1158.5, 1125.3, 1098.8, 901.6, 879.4, 831.7, 777.7, 740.1, 652.3, 606.0, 561.7, 496.1, 454.2.

7-Iodoindole-3-carboxaldehyde (37)
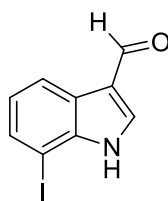
Yield; 92%. mp. 104 - 109 °C. δ_H (400 MHz, DMSO-d$_6$): 12.11 (1 H, s, NH), 9.96 (1 H, s, CHO), 8.33 (1 H, d, J = 2.5 ArCH), 8.12 (1 H, d, J = 8.0, ArCH), 7.67 (1 H, d, J = 8.0, ArCH), 7.03 (1H, t, J = 8.0, ArCH). δ_C (125 MHz, DMSO-d$_6$): 185.98 (CHO), 139.43 (ArCH), 139.25 (ArC), 132.94 (ArCH), 125.19 (ArC), 124.44 (ArCH), 121.24 (ArCH), 119.44 (ArC), 78.04 (ArCI). Mass Spectrum HRMS [EI$^+$] found 270.9490, C_9H_6INO calculated 270.9494. υ cm^{-1}: 3433.6, 3063.4, 1638.2, 1610.8, 1558.7, 1522.5, 1488.3, 1442.0, 1384.6, 1331.1, 1277.6, 1237.1, 1204.8, 1149.4, 1127.2, 1092.5, 872.6, 817.2, 773.8, 736.7, 650.4, 608.0, 554.0.

6.1.4 General method for the condensation of aromatic aldehydes with rhodanine.[188]

(*Z*)-5-((4-Fluoroindol-3-yl)methylene)-2-thioxothiazolidin-4-one (49)
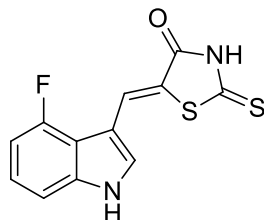
A suspension of 4-fluoroindole-3-carboxaldehyde (0.40 g, 2.46 mmol), 2-thioxothiazoolidin-4-one (0.33 g, 2.49 mmol) and sodium acetate (0.59 g, 7.21 mmol) in glacial acetic acid (5 ml) was heated under reflux for 2 hr. The reaction mixture was then cooled to room temperature and water (15 ml) was added to the solution. The precipitate that formed was collected by filtration under reduced pressure and washed with water (2 × 10 ml). The solid was air dried under reduced pressure and washed with diethyl ether (5 ml) to give the title compound as a dark red solid (0.54 g, 1.94 mmol, 78%). mp. ~340 °C (dec.). δ_H (400 MHz, DMSO-d_6): 13.61 (1 H, s, NH), 12.55 (1 H, s, NH), 7.94 (1 H, s, CH=CS), 7.81 (1 H, d, *J* = 3.5, ArCH), 7.50 (1 H, d, *J* = 8.0, ArCH), 7.22 (1 H, dt, *J* = 8.0, *J* = 5.0, ArCH), 7.00 (1H, ddd, J_1=12.0, J_2=8.0, J_3= 3.5, ArCH). δ_C (125 MHz, DMSO-d_6): 194.63 (CS), 169.08 (CO), 156.49 (ArCF, d, *J* = 244.5), 138.94 (ArC, d, *J* = 10.0), 129.89 (ArCH), 125.19 (ArCH, d, *J* = 5.0), 123.89 (ArCH, d, *J* = 8.0), 118.95 (ArC), 114.83 (ArC, d, *J* = 18.0), 109.24 (ArCH, d, *J* = 4.0), 109.16 (ArC, d, *J* = 3.0), 106.68 (ArCH, d, *J* = 19.0). Mass spectrum: HRMS (EI[+]) found 277.9988, $C_{12}H_7FN_2OS_2$ calculated 277.9984. υ cm[-1]: 3242.2, 1685.0, 1634.4, 1589.5, 1575.6, 1511.9, 1430.0, 1345.1, 1324.4, 1296.4, 1277.1, 1224.1, 1205.8, 1158.5, 1148.4, 1033.7, 841.8, 808.0, 777.2, 726.6, 690.9, 670.6, 614.2, 597.8, 576.6, 556.4, 495.1, 481.6.

Benzylidene Derivatives

(*Z*)-5-(2-Fluorobenzylidene)-2-thioxothazolidene-4-one (38)
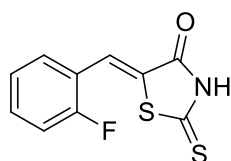
Yield; 92%. mp. 202 - 206 °C. δ_H (400 MHz, DMSO-d_6): 13.93 (1 H, s, NH), 7.58 (1 H, s, CH=CS), 7.55 (1 H, dt, J_1 = 7.5, J_2=1.0, ArCH), 7.49 (1 H, dt, J_1 = 7.5, J_2=1.0, ArCH), 7.39 (2H, m, ArCH). δ_C (125 MHz, DMSO-d_6): 195.88 (CS), 169.64 (CO), 161.12 (ArCF, d, J = 251.5), 133.50 (ArCH, d, J = 10.0), 129.82 (ArCH), 128.58 (ArC, d, J = 2.5), 126.01 (ArC, d, J = 4.0), 122.77 (ArCH, d, J = 6.5), 121.35 (ArCH, d, J = 11.5), 116.73 (ArCH, d, J = 21.5). δ_F (282 MHz, DMSO-d_6): -114.01. Mass spectrum: HRMS (ES$^-$) found 237.9796, $C_{10}H_5FNOS_2$ calculated 237.9797. υ cm^{-1}: 3054.7, 1695.1, 1608.8, 1593.4, 1482.0, 1456.0, 1430.0, 1309.9, 1211.1, 1158.0, 1068.9, 798.9, 751.1, 677.4, 548.7.

(*Z*)-5-(2-Chlorobenzylidene)-2-thioxothazolidene-4-one (39)
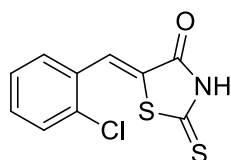
Yield; 22.1%. mp. 189 - 191 °C. δ_H (400 MHz, DMSO-d_6): 13.96 (1 H, s, NH), 7.75 (1 H, s, CH=CS), 7.64 (1 H, dd, J_1 = 8.5, J_2 = 1.0, ArCH), 7.52 (3H, m, ArCH). δ_C (125 MHz, DMSO-d_6): 195.55 (CS), 169.18 (CO), 134.73 (ArCCl), 132.05 (ArCH), 130.83 (ArC), 130.41 (ArCH), 129.26 (ArCH), 129.11 (ArC), 128.23 (ArCH), 126.02 (ArCH). Mass spectrum: HRMS (EI$^+$) 254.9586, $C_{10}H_5{}^{35}ClNOS_2$ calculated 254.9579, *m/z* (ES$^+$): 168.0 (50 % [M-CSNHCO]$^+$). υ cm^{-1}: 3440.9, 3066.7, 2868.6, 1734.7, 1699.0, 1594.6, 1584.7, 1456.5, 1435.7, 1336.0, 1306.5, 1280.5, 1239.5, 1197.6, 1161.9, 1069.8, 1041.4, 805.6, 750.2, 723.7, 701.0, 675.5, 632.5, 532.7, 467.2.

(*Z*)-5-(2-Bromobenzylidene)-2-thioxothazolidene-4-one (40)
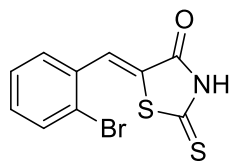
Yield; 93%. mp. 189 - 192 °C. δ_H (500 MHz, DMSO-d_6): 14.00 (1 H, s, NH), 7.88 (1 H, d, J = 8.0, ArCH), 7.63 (1 H, s, CH=CS), 7.61 (1 H, d, J = 7.0 ArCH), 7.49 (1H, t, J_1 = 8.0, J_2 = 7.0, ArCH). δ_C (100 MHz, DMSO-d_6): 196.02 (CS), 169.50 (CO), 134.10 (ArCH), 132.91 (ArC), 132.62 (ArCH), 129.79 (ArCH), 129.53 (ArC), 129.21 (ArCH), 129.18 (ArCH), 126.07 (ArCBr). Mass spectrum: HRMS (APCI$^-$) found 297.8987, $C_{10}H_5{}^{79}BrNOS_2$ calculated 297.8996. υ cm^{-1}: 3093.7, 2848.9, 1732.3, 1605.9, 1452.1, 1429.5, 1277.1, 1233.3, 1196.1, 1065.5, 1002.8, 890.5, 752.6, 724.6, 681.2, 525.5.

(*Z*)-5-(2-Iodobenzylidene)-2-thioxothazolidene-4-one (41)
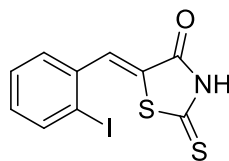
Yield; 71%. mp. 188 - 190 °C. δ_H (500 MHz, MeOD-d_4): 14.00 (1 H, s, NH), 7.75 (1 H, s, CH=CS), 7.65 (1 H, dd, J_1 = 8.0, J_2 = 1.0, ArCH), 7.45 (1 H, dd, J_1 = 8.0, J_2 = 2.0, ArCH), 7.40 (1 H, t, J = 8.0, ArCH), 7.24 (1 H, td, J_1 = 8.0, J_2 = 2.0, ArCH). δ_C (62.5 MHz, THF-d_8): 195.11 (CS), 168.36 (CO), 140.43 (ArCH), 136.90 (ArC), 134.04 (ArCH), 131.32 (ArCH), 129.60 (ArC), 128.83 (ArCH), 128.63 (ArCH), 101.98 (ArCI). Mass spectrum: HRMS (EI$^+$) found 346.8947, $C_{10}H_5INOS_2$ calculated 346.8936. υ cm^{-1}: 3142.9, 3051.8, 1694.6, 1589.5, 1455.5, 1428.0, 1311.4, 1281.0, 1230.4, 1198.5, 1068.9, 1014.9, 753.1, 669.2, 651.8, 537.6.

(*Z*)-5-(3-Fluorobenzylidene)-2-thioxothazolidene-4-one (42)
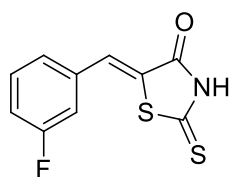
Yield; 68%. mp. 203 - 204 °C. δ_H (400 MHz, DMSO-d_6): 13.91 (1 H, s, NH), 7.63 (1 H, s, CH=CS), 7.58 (1 H, q, J = 8.0, ArCH), 7.45 (1 H, dd, J_1 = 10.0, J_2 = 2.0, ArCH), 7.40 (1 H, d, J = 8.0 ArCH), 7.33 (1 H, dt, J = 8.0, J = 2.0, ArCH). δ_C (125 MHz, DMSO-d_6): 195.85 (CS), 169.68 (CO), 162.78 (ArCF, d, J = 245.5), 135.88 (ArC, d, J = 8.0), 131.93 (ArCH, d, J = 8.0), 130.51 (ArCH, d, J = 3.0), 127.66 (ArC, s), 126.39 (ArCH, d, J = 3.0), 117.90 (ArCH, d, J = 21.0), 117.55 (ArCH, d, J = 22.0). δ_F (282 MHz, DMSO-d_6): -111.75; Mass spectrum: HRMS (ES⁻) found 237.9802, $C_{10}H_5FNOS_2$ calculated 237.9797. υ cm⁻¹: 3445.7, 3193.5, 2360.0, 1705.3, 1612.2, 1600.6, 1577.0, 1489.7, 1432.9, 1288.2, 1253.5, 1220.7, 952.2, 775.2, 672.1, 529.4.

(*Z*)-5-(3-Chlorobenzylidene)-2-thioxothazolidene-4-one (43)
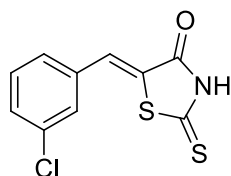
Yield; 86%. mp. 240 - 242 °C. δ_H (400 MHz, DMSO-d_6): 13.90 (1 H, s, NH), 7.66 (1 H, s, ArCH), 7.61 (1 H, s, CH=CS), 7.55 (1 H, t, J = 3.0, ArCH), 7.53 (1 H, d, J = 2.0, ArCH), 7.50 (1 H, dd, J_1 = 6.0, J_2 = 2.0, ArCH). δ_C (125 MHz, DMSO-d_6): 195.74 (CS), 169.61 (CO), 135.57 (ArC), 134.49 (ArCCl), 131.63 (ArCH), 130.66 (2ArCH), 130.24 (ArCH), 128.57 (ArCH), 127.78 (ArC). Mass spectrum: HRMS found (ES⁻) 253.9500, $C_{10}H_5{}^{35}ClNOS_2$ calculated 253.9501. υ cm⁻¹: 3111.1, 2848.4, 1717.3 1608.3, 1559.7, 1478.7, 1438.6, 1412.6, 1324.4, 1299.8, 1284.8, 1222.2, 1180.7, 1094.4, 1078.5, 1058.3, 1000.9, 911.2, 895.8, 860.6, 782.0, 739.1, 704.4, 681.2, 630.1, 547.2, 518.3.

(*Z*)-5-(3-Bromobenzylidene)-2-thioxothazolidene-4-one (44)
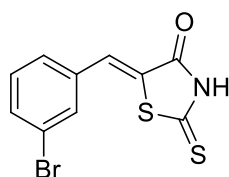
Yield; 86%. mp. 250 - 254 °C. δ_H (400 MHz, DMSO-d$_6$): 13.92 (1 H, s, NH), 7.80 (1 H, s, ArCH), 7.68 (1 H, d, *J* = 8.0, ArCH), 7.61 (1 H, s, CH=CS), 7.48 (1 H, d, *J*=8.0, ArCH), 7.48 (1 H, t, *J*=8.0, ArCH). δ_C (125 MHz, DMSO-d$_6$): 195.64 (CS), 169.53 (CO), 135.77 (ArC), 133.48 (ArCH), 131.77 (ArCH), 130.13 (ArCH), 128.89 (ArCH), 127.69 (ArC), 122.93 (ArCBr). Mass spectrum: HRMS (ES$^-$) found 297.8999, C$_{10}$H$_5{}^{79}$BrNOS$_2$ calculated 297.8996. υ cm^{-1}: 3422.6, 3108.7, 2845.0, 1716.8, 1606.4, 1552.9, 1476.2, 1437.7, 1408.7, 1283.4, 1220.2, 1178.3, 1072.2, 1059.2, 1002.8, 910.2, 881.3, 862.0, 779.6, 737.6, 693.3, 673.0, 627.7, 546.7, 517.8.

(*Z*)-5-(3-Iodobenzylidene)-2-thioxothazolidene-4-one (45)
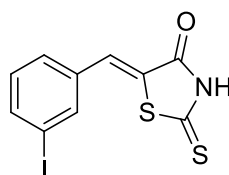
Yield; 75%. mp. 236 - 239 °C. δ_H (400 MHz, DMSO-d$_6$): 13.90 (1 H, s, NH), 7.96 (1 H, s, ArCH), 7.84 (1 H, d, *J* = 8.0, ArCH), 7.57 (1 H, s, CH=CS) 7.57 (1 H, d, *J* = 8.0, ArCH), 7.32 (1 H, t, *J* = 8.0, ArCH). δ_C (100 MHz, DMSO-d$_6$): 195.73 (CS), 169.56 (CO), 139.31 (2 ArCH), 135.60 (ArC), 131.67 (ArCH), 130.22 (ArCH), 129.27 (ArCH), 127.27 (ArC), 96.18 (ArCI). Mass spectrum: HRMS (APCI$^-$) found 345.8861, C$_{10}$H$_5$INOS$_2$ calculated 345.8857. υ cm^{-1}: 3396.0, 3147.7, 3055.2, 2847.4, 2349.8, 1690.3, 1600.6, 1563.5, 1550.5, 1472.9, 1437.7, 1404.4, 1310.9, 1280.5, 1230.8, 1193.7, 1179.3, 1099.2, 1066.9, 1013.9, 992.2, 911.2, 885.6, 876.0, 864.9, 805.1, 772.8, 714.0, 685.6, 669.2, 626.8, 546.7, 530.3, 429.1, 414.6.

(*Z*)-5-(4-Fluorobenzylidene)-2-thioxothazolidene-4-one (46)
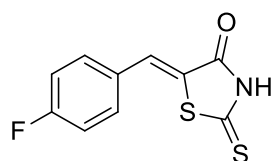
Yield; 88%. mp. 226 – 231 °C. δ_H (400 MHz, DMSO-d_6): 13.86 (1 H, s, NH), 7.66 (3 H, m, ArCH and CH=CS), 7.37 (2 H, t, *J* = 8.5, ArCH). δ_C (125 MHz, DMSO-d_6): 195.45 (CS), 169.26 (CO), 163.01 (ArCF, d, *J* = 251.5), 132.91 (ArCH, d, *J* = 9.0), 130.47 (ArCH), 129.63 (ArC, d, *J* = 4.0), 125.20 (ArCH), 116.58 (2ArCH, d, *J* = 22.0). δ_F (300 MHz, DMSO-d_6): -108.34; Mass spectrum: HRMS (ES⁻) found 237.9794, $C_{10}H_5FNOS_2$ calculated 237.9797. υ cm⁻¹: 3098.1, 1724.1, 1586.7, 1508.1, 1446.8, 1290.6, 1228.4, 1179.3, 1160.9, 1063.6, 921.3, 828.3, 794.1, 748.3, 689.0, 597.3, 549.6, 524.5.

(*Z*)-5-(4-Chlorobenzylidene)-2-thioxothazolidene-4-one (47)
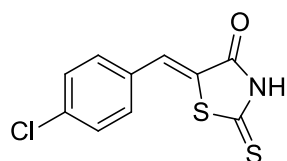
Yield; 76%. mp. 233 - 235 °C. δ_H (400 MHz, DMSO-d_6): 13.90 (1 H, s, NH), 7.61 (1 H, s, CH=CS), 7.58 (4 H, d, *J* = 8.5, ArCH). δ_C (125 MHz, DMSO-d_6): 195.91 (CS), 169.90 (CO), 135.80 (ArCCl), 132.48 (2 ArCH), 132.39 (ArC), 130.54 (ArCH), 129.95 (2 ArCH), 126.92 (ArC). Mass spectrum: HRMS (ES⁻) found 253.9509, $C_{10}H_5{}^{35}ClNOS_2$ calculated 253.9501. υ cm⁻¹: 3413.9, 3066.7, 2848.8, 1704.8, 1596.3, 1582.3, 1558.2, 1487.3, 1444.9, 1404.4, 1293.0, 1281.5, 1232.3, 1187.5, 1177.3, 1086.7, 1061.6, 1007.6, 917.5, 827.3, 796.9, 756.4, 723.2, 684.1, 572.8, 545.8, 518.8.

(*Z*)-5-(4-Bromobenzylidene)-2-thioxothazolidene-4-one (48)
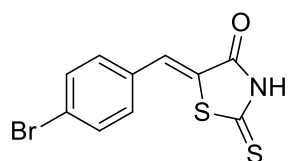
Yield; 85%. mp. 243 - 248 °C. δ_H (400 MHz, DMSO-d$_6$): 13.90 (1 H, s, NH), 7.74 (2 H, d, *J* = 8.5, ArCH), 7.62 (1 H, s, CH=CS), 7.53 (2 H, d, *J* = 8.5, ArCH). δ_C (125 MHz, DMSO-d$_6$): 195.84 (CS), 169.78 (CO), 132.90 (2 ArCH), 132.69 (ArC), 132.64 (2 ArCH), 130.72 (ArCH), 126.91 (ArCBr), 124.76 (ArC). Mass spectrum: HRMS (ES⁻) found 297.8991, $C_{10}H_5{}^{79}BrNOS_2$ calculated 297.8996. υ cm⁻¹: 3403.3, 3076.9, 2847.9, 1706.2, 1595.8, 1577.5, 1553.9, 1483.5, 1441.5, 1399.1, 1303.6, 1291.6, 1279.1, 1231.3, 1177.8, 1070.8, 1059.2, 1006.7, 918.0, 825.9, 800.8, 750.2, 717.9, 686.1, 665.3, 561.7, 540.0, 516.3.

Indole Derivatives

(*Z*)-5-((4-Chloroindol-3-yl)methylene)-2-thioxothiazolidin-4-one (50)
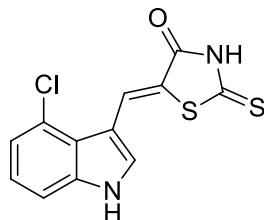
Yield; 86%. mp. ~ 350 °C (dec.). δ_H (400 MHz, DMSO-d$_6$): 13.62 (1 H, s, NH), 12.63 (1 H, s, NH), 8.52 (1 H, s, C*H*=CS), 7.90 (1 H, d, *J* = 2.5, ArCH), 7.50 (1 H, dd, *J$_1$* = 7.0, *J$_2$* = 2.5, ArCH), 7.24 (2 H, m, 2 x ArCH). δ_C (125 MHz, DMSO-d$_6$): 195.35 (CS), 169.70 (CO), 138.54 (ArC), 131.63 (ArCH), 125.99 (ArCH), 125.34 (ArCCl), 124.43 (ArCH), 123.03 (ArCH), 122.93 (ArC), 119.10 (ArC), 112.61 (ArCH), 111.24 (ArC). Mass Spectrum HRMS [EI⁺] found 293.9691, $C_{12}H_7{}^{35}ClN_2OS_2$ calculated 293.9688. υ cm⁻¹: 3821.7, 3816.0, 3217.7, 2837.7, 2360.0, 1849.7, 1833.0, 1692.7, 1616.1, 1578.9, 1513.9, 1487.8, 1443.9, 1421.8, 1349.0, 1340.3, 1322.9, 1275.2, 1258.3, 1258.3, 1212.0, 1194.2, 1155.6, 1079.5, 1047.2, 1023.5, 945.4, 936.8, 918.0, 828.3, 801.8, 787.8, 7752, 727.0, 694.7, 648.4, 608.9, 571.3, 540.5, 506.7, 477.3, 455.6, 420.4.

(*Z*)-5-((4-Bromoindol-3-yl)methylene)-2-thioxothiazolidin-4-one (51)
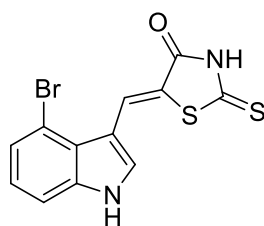
Yield; 36%. mp. ~350 °C (dec.). δ_H (400 MHz, DMSO-d$_6$): 12.39 (2 H, s, 2 NH), 8.48 (1 H, s, C*H*=CS), 7.75 (1 H, s, ArCH), 7.54 (1 H, d, *J* = 8.0, ArCH), 7.38 (1 H, d, *J* = 8.0, ArCH), 7.13 (1 H, t, *J* = 8.0, ArCH). δ_C (100 MHz, DMSO-d$_6$): 183.92 (CS), 172.03 (CO), 137.89 (ArC), 129.94 (ArCH), 125.36 (2ArCH), 123.78 (ArCH), 123.72 (ArC), 120.79 (ArC), 112.97 (ArCBr), 112.33 (ArCH), 112.09 (ArC). Mass Spectrum HRMS [EI$^+$] found 337.9186, C$_{12}$H$_7$^{79}BrN$_2$OS$_2$ calculated 337.9183. υ cm^{-1}: 3209.5, 3108.2, 2838.2, 1892.8, 1829.6, 1691.8, 1577.0, 1562.5, 1512.9, 1484.9, 1443.9, 1417.9, 1354.3, 1336.0, 1324.4, 1300.8, 1274.7, 1257.8, 1213.0, 1192.8, 1155.6, 1082.4, 1045.7, 1021.6, 944.0, 914.1, 797.4, 786.3, 772.4, 726.6, 694.2, 642.2, 608.4, 564.6, 534.7, 498.5, 472.5, 453.7, 419.9.

(*Z*)-5-((4-Iodoindol-3-yl)methylene)-2-thioxothiazolidin-4-one (52)
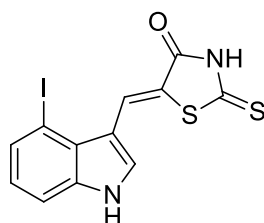
Yield; 54%. mp. ~ 320 °C (dec.). δ_H (400 MHz, DMSO-d$_6$): 13.66 (1 H, s, NH), 12.53 (1 H, s, NH), 8.94 (1 H, s, C*H*=CS), 7.92 (1 H, d, *J* = 2.0 ArCH), 7.70 (1 H, d, *J* = 7.5, ArCH), 7.58 (1 H, d, *J* = 7.5, ArCH), 7.01 (1 H, t, *J* = 7.5, ArCH). δ_C (100 MHz, DMSO-d$_6$): 194.23 (CS), 169.73 (CO), 137.79 (ArC), 133.64 (ArCH), 132.15 (ArCH), 126.72 (ArC), 125.30 (ArCH), 125.15 (ArCH), 117.98 (ArC), 113.63 (ArCH), 112.14 (ArC), 85.36 (ArCI). Mass Spectrum HRMS [ES$^-$] found 384.8966, C$_{12}$H$_7$IN$_2$OS$_2$ calculated 384.8961. υ cm^{-1}: 3434.1, 3236.9, 3041.2, 2851.7, 1676.8, 1569.8, 1509.5, 1477.2, 1449.2, 1410.2, 1350.9, 1320.0, 1270.4, 1215.9, 1194.2, 1153.2, 1128.2, 1045.2, 1019.7, 877.9, 783.4, 771.9, 736.2, 694.3, 640.7, 608.4.

(*Z*)-5-((5-Chloroindol-3-yl)methylene)-2-thioxothiazolidin-4-one (53)
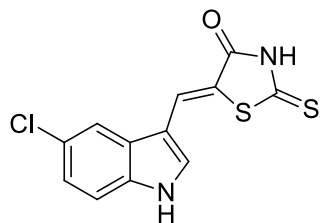
Yield; 79%. mp. ~ 330 °C (dec.). δ_H (500 MHz, DMSO-d_6): 13.57 (1 H, s, NH), 12.40 (1 H, s, NH), 8.06 (1 H, s, CH=CS), 7.95 (1 H, s, ArCH), 7.86 (1 H, s, ArCH), 7.51 (1 H, d, J = 8.5, ArCH), 7.26 (1 H, d, J = 8.5, ArCH). δ_C (125 MHz, DMSO-d_6): 195.13 (CS), 169.47 (CO), 135.34 (ArC), 131.79 (ArCH), 128.52 (ArC), 126.56 (ArC), 124.66 (ArCH), 123.85 (ArCH), 119.19 (ArC) 118.76 (ArCH), 114.48 (ArCH), 111.16 (ArCCl). Mass spectrum: HRMS (EI⁺) found 293.9688, $C_{12}H_6{}^{35}ClN_2OS_2$ calculated 293.9688. υ cm⁻¹: 3259.6, 3110.6, 2837.3, 2360.0, 1824.3, 1688.4, 1622.8, 1584.2, 1564.0, 1510.0, 1457.0, 1441.1, 1423.2, 1355.7, 1332.1, 1305.6, 1271.8, 1225.5, 1202.9, 1157.6, 1128.2, 1093.9, 1045.7, 1015.3, 914.6, 895.7, 864.4, 819.6, 797.9, 766.1, 747.8, 729.4, 693.8, 683.2, 609.9, 571.3, 558.8, 531.3, 468.2, 418.5.

(*Z*)-5-((5-Bromoindol-3-yl)methylene)-2-thioxothiazolidin-4-one (54)
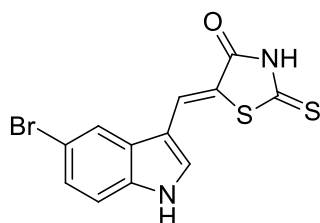
Yield; 86.2%. mp. ~ 300 °C (dec.). δ_H (400 MHz, DMSO-d_6): 13.60 (1 H, s, NH), 12.44 (1 H, s, NH), 8.23 (1 H, d, J = 2.0, ArCH), 7.96 (1 H, s, CH=CS), 7.86 (1 H, d, J = 3.0, ArCH), 7.47 (1 H, d, J = 8.5, ArCH), 7.38 (1 H, dd, J_1 =8.5, J_2 = 2.0, ArCH). δ_C (100 MHz, DMSO-d_6): 195.17 (CS), 169.57 (CO), 135.57 (ArC), 131.36 (ArCH), 129.11 (ArC), 126.31 (ArCH), 124.90 (ArCH), 121.77 (ArCH), 119.28 (ArCH), 114.92 (ArCBr), 114.55 (ArCH), 111.04 (ArC). Mass spectrum: HRMS (EI⁺) found 337.9180, $C_{12}H_6{}^{79}BrN_2OS_2$ calculated 337.9183. υ cm⁻¹: 3254.8, 3109.2, 3036.9, 2849.3, 2725.9, 1868.2, 1810.8, 1678.7, 1616.1, 1580.9, 1565.9, 1511.4, 149.2, 1349.0, 1330.2, 1308.0, 1268.5, 1213.5, 1146.5, 1092.0, 1080.0, 1045.2,

1019.7, 936.7, 908.3, 886.6, 859.1, 803.2, 772.4, 748.2, 723.2, 702.0, 687.5, 668.2, 613.7, 595.4, 570.8, 560.2, 542.9, 528.9, 471.0, 461.4, 461.4, 418.0.

(*Z*)-5-((5-Iodoindol-3-yl)methylene)-2-thioxothiazolidin-4-one (55)
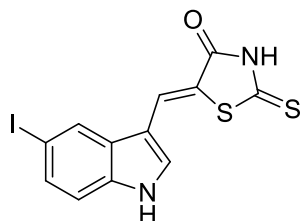
Yield; 74%. mp. ~ 310 °C (dec.). δ_H (500 MHz, DMSO-d_6): 13.56 (1 H, s, NH), 12.38 (1 H, s, NH), 8.37 (1 H, s, CH=CS), 7.94 (1 H, s, ArCH), 7.79 (1 H, s, ArCH), 7.52 (1 H, d, *J* = 8.0, ArCH), 7.34 (1 H, d, *J* = 8.0, ArCH). δ_C (100 MHz, DMSO-d_6): 195.14 (CS), 169.56 (CO), 135.93 (ArC), 134.31 (ArC), 131.77 (ArCH), 130.90 (ArCH), 129.75 (ArC), 127.76 (ArCH), 124.93 (ArCH), 115.27 (ArCH), 110.69 (ArC), 86.13 (ArCI). Mass spectrum: HRMS (EI⁺) found 385.9038, $C_{12}H_7IN_2OS_2$ calculated 385.9045. υ cm⁻¹: 3262.5, 3037.8, 2850.8, 1678.3, 1576.5, 1562.5, 1510.0, 1447.8, 1349.0, 1327.3, 1305.1, 1269.4, 1212.0, 1147.9, 1076.6, 908.3, 879.4, 861.5, 802.7, 769.9, 688.0, 609.9, 590.1, 528.9.

(*Z*)-5-((6-Fluoroindol-3-yl)methylene)-2-thioxothiazolidin-4-one (56)
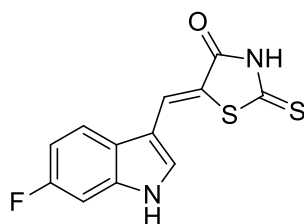
Yield; 70%. mp. ~ 330 °C (dec.). δ_H (400 MHz, DMSO-d_6): 13.57 (1 H, s, NH), 12.30 (1 H, s, NH), 7.94 (1 H, dd J_1 = 8.5, J_2 = 5.5, ArCH), 7.89 (1 H, s, CH=CS), 7.77 (1 H, d, *J* = 3.0, ArCH), 7.27 (1 H, dd. J_1 = 9.5, J_2 = 2.0, ArCH), 7.41 (1 H, dt, J_1=8.5, J_2 = 2.0, ArCH). δ_C (100 MHz, DMSO-d_6): 194.98 (CS), 169.43 (CO), 160.01 (ArCF, d, *J* = 236.0), 136.72 (ArC d, *J* = 12.0), 130.78 (ArC), 124.77 (ArCH), 123.82 (ArC), 120.29 (ArCH, d, *J* = 10.0), 119.01 (ArCH), 111.32 (ArCH, s), 110.10 (ArCH d, *J* = 25.0), 98.96 (ArCH, d, *J* = 26.0). Mass spectrum: HRMS (APCI⁻) found 276.9900, $C_{12}H_6FN_2OS_2$ calculated 276.9906. υ cm⁻¹: 3446.2, 3214.3, 3107.7,

2851.2, 1811.8, 1678.7, 1626.7, 1599.2, 1573.2, 1516.3, 1497.0, 1448.8, 1438.6, 1419.8, 1357.6, 1340.8, 1301.2, 1190.4, 1146.5, 1107.9, 1072.2, 1052.9, 1019.7, 952.7, 909.3, 854.3, 816.7, 806.6, 733.8, 703.9, 691.4, 614.2, 605.1, 594.5, 571.8, 542.9.

(*Z*)-5-((6-Chloroindol-3-yl)methylene)-2-thioxothiazolidin-4-one (57)
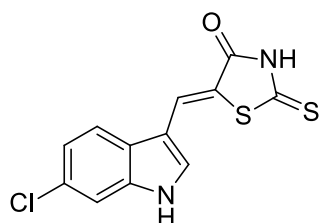
Yield; 85%. mp. ~ 350 °C (dec.). δ_H (400 MHz, DMSO-d_6): 13.59 (1 H, s, NH), 12.35 (1 H, s, NH), 7.94 (1 H, d, *J* = 8.5, ArCH), 7.88 (1 H, s, CH=CS), 7.82 (1 H, d, *J* = 2.5, ArCH), 7.53 (1 H, s, ArCH), 7.19 (1 H, d, *J* = 8.5, ArCH). δ_C (100 MHz, DMSO-d_6): 194.60 (CS), 169.03 (CO), 136.76 (ArCH), 130.66 (ArC), 127.82 (ArCCl), 125.50 (ArCH), 124.17 (ArCH), 121.50 (ArCH), 120.08 (ArC), 118.91 (ArC), 112.09 (ArCH), 110.93 (ArCH). Mass spectrum: HRMS (ES$^-$) found 292.9601, $C_{12}H_6{}^{35}ClN_2OS_2$ calculated 292.9610. υ cm^{-1}: 3231.6, 3103.9, 2849.3, 2350.3, 1676.8, 1584.7, 1570.3, 1513.4, 1488.3, 1437.2, 1416.5, 1353.3, 1333.1, 1295.5, 1228.0, 1192.8, 1146.5, 1072.2, 1058.3, 1021.1, 931.9, 905.4, 851.9, 798.9, 785.9, 749.7, 735.2, 688.0, 608.9, 594.5, 575.7, 565.0, 538.0, 467.2, 418.5.

(*Z*)-5-((6-Bromoindol-3-yl)methylene)-2-thioxothiazolidin-4-one (58)
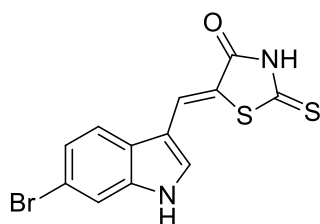
Yield; 72%. mp. ~ 320 °C (dec.). δ_H (400 MHz, DMSO-d_6): 13.59 (1 H, s, NH), 12.34 (1 H, s, NH), 7.91 (2 H, m, ArCH and CH=CS), 7.81 (1 H, d, *J* = 2.5, ArCH), 7.67 (1 H, s, ArCH), 7.31 (1 H, d, *J* = 8.5, ArCH). δ_C (125 MHz, DMSO-d_6): 195.07 (CS), 169.49 (CO), 137.70 (ArC), 131.70 (ArCH), 126.25 (ArC), 124.55 (2ArCH), 120.09 (ArCH), 119.47 (ArC), 116.31 (ArC), 115.53 (ArCH), 111.46 (ArCBr).

Mass spectrum: HRMS (ES⁻) found 336.9098, $C_{12}H_6{}^{79}BrN_2OS_2$ calculated 336.9105. υ cm⁻¹: 3233.1, 3102.4, 2843.0, 1676.8, 1582.8, 1568.8, 1511.9, 1484.0, 1436.2, 1355.2, 1330.6, 1295.0, 1227.5, 1194.7, 1147.4, 1072.2, 1049.1, 1021.1, 931.5, 893.8, 852.4, 797.4, 779.6, 748.2, 687.0, 669.2, 607.5, 587.7, 570.8, 534.2, 466.7, 418.0.

(*Z*)-5-((6-Iodoindol-3-yl)methylene)-2-thioxothiazolidin-4-one (59)
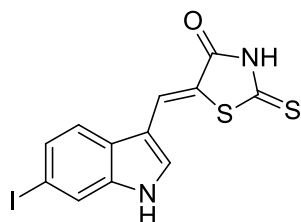
Yield; 69%. mp. ~ 320 °C (dec.). δ_H (400 MHz, DMSO-d₆): 13.59 (1 H, s, NH), 12.32 (1 H, s, NH), 7.88 (1 H, s, CH=CS), 7.85 (1 H, s, ArCH), 7.78 (1 H, d, *J* = 6.5, ArCH), 7.47 (1 H, d, *J* = 8.5, ArCH). δ_C (125 MHz, DMSO-d₆): 194.62 (CS), 169.05 (CO), 137.27 (ArC), 130.27 (ArCH), 129.58 (ArCH), 126.11 (ArC), 124.09 (ArCH), 120.96 (ArCH), 120.70 (ArCH), 118.89 (ArC), 110.97 (ArCH), 87.43 (ArCI). Mass spectrum: HRMS (ES⁻) found 384.8966, $C_{12}H_6IN_2OS_2$ calculated 384.8966, *m/z* (ES⁻): 297.9156 (25 % [M-CSNHCO]⁻). υ cm⁻¹: 3231.2, 2998.3, 2837.7, 1683.6, 1581.3, 1568.3, 1509.5, 1479.6, 1445.4, 1354.3, 1327.3, 1302.2, 1232.3, 1218.3, 1141.2, 1128.6, 1073.2, 1053.9, 1024.0, 887.1, 847.1, 805.1, 793.6, 775.8, 724.6, 712.1, 678.8, 661.0.

(*Z*)-5-((7-Fluoroindol-3-yl)methylene)-2-thioxothiazolidin-4-one (60)
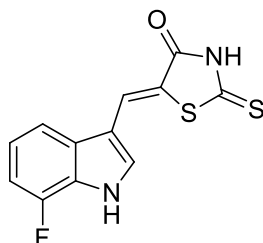
Yield; 54%. mp. ~ 350 °C (dec.). δ_H (400 MHz, DMSO-d₆): 13.59 (1 H, s, NH), 12.82 (1 H, s, NH), 7.90 (1 H, s, CH=CS), 7.80 (1 H, d, *J* = 2.5, ArCH), 7.77 (1 H, d, *J* = 7.5, ArCH), 7.35 (1 H, dt, *J₁* = 7.5, *J₂* = 5.0, ArCH), 7.21 (1 H, dd, *J₁* = 11.5, *J₂* = 8.0, ArCH). δ_C (125 MHz, DMSO-d₆): 195.25 (CS), 169.76 (CO), 147.65 (ArCF, d, *J* =

273.5), 130.93 (ArC), 130.71 (ArCH), 124.88 (ArC, d, *J* = 13.0), 124.51 (ArCH), 122.33 (ArCH, d, *J*= 6.0), 120.10 (ArC), 115.21 (ArCH), 112.17 (ArC), 108.67 (ArC, d, *J* = 16.0). Mass Spectrum HRMS [ES$^-$] found 276.9901, $C_{12}H_7FN_2OS_2$ calculated 276.9906, *m/z* (ES$^-$): 190.0099 (100 % [M-CONHCS]$^-$). υ cm^{-1}: 3275.5, 3042.2, 2851.2, 1685.5, 1641.1, 1576.5, 1514.8, 1445.9, 1361.0, 1342.2, 1320.5, 1295.9, 1256.9, 1229.9, 1216.9, 1171.1, 1141.7, 1113.2, 1063.6, 1051.5, 1019.2, 975.8, 910.2, 865.4.

(*Z*)-5-((7-Chloroindol-3-yl)methylene)-2-thioxothiazolidin-4-one (61)
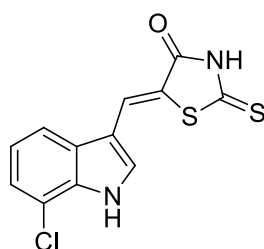
Yield; 63%. mp. ~ 350 °C (dec.). δ$_H$ (400 MHz, DMSO-d$_6$): 13.64 (1 H, s, NH), 12.70 (1 H, s, NH), 7.94 (1 H, d, *J* = 8.0, ArCH), 7.90 (1 H, s, CH=CS), 7.75 (1 H, d, *J* = 3.0, ArCH), 7.35 (1 H, d, *J* = 8.0, ArCH), 7.21 (1 H, t, *J* = 8.0, ArCH). δ$_C$ (125 MHz, DMSO-d$_6$): 195.12 (CS), 169.58 (CO), 133.74 (ArC), 130.68 (ArCH), 129.10 (ArCCl), 124.44 (ArCH), 123.33 (ArCH), 122.78 (ArCH), 120.19 (ArC), 118.23 (ArCH), 117.22 (ArC), 112.39 (ArC). Mass Spectrum HRMS [EI$^+$] found 293.9688, $C_{12}H_7{}^{35}ClN_2OS_2$ calculated 293.9681. υ cm^{-1}: 3357.9, 3066.7, 2837.7, 1896.7, 1806.97, 1696.6, 1645.0, 1620.9, 1586.7, 1564.5, 1513.9, 1490.2, 1436.7, 1358.6, 1339.3, 1313.8, 1295.5, 1247.2, 1218.8, 1194.2, 1149.4, 1134.4, 1087.2, 1051.0, 1019.7, 899.6, 856.7, 823.9, 785.4, 779.1, 772.8, 734.8, 698.1, 632.1, 609.9.

(*Z*)-5-((7-Bromoindol-3-yl)methylene)-2-thioxothiazolidin-4-one (62)
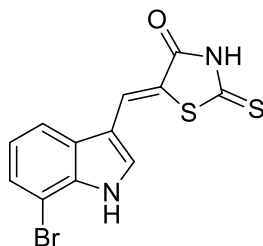
Yield; 91%. mp. ~330 °C (dec.). δ_H (400 MHz, DMSO-d_6): 13.64 (1 H, s, NH), 12.54 (1 H, s, NH), 7.99 (1 H, d, J = 7.5, ArCH), 7.90 (1 H, s, CH=CS), 7.72 (1 H, d, J = 3.0, ArCH), 7.50 (1 H, d, J = 7.5, ArCH), 7.16 (1 H, t, J = 7.5, ArCH). δ_C (125 MHz, DMSO-d_6): 195.36 (CS), 168.98 (CO), 134.89 (ArC), 129.96 (ArC), 128.45 (ArCH), 125.79 (ArCH), 123.89 (ArCH), 122.76 (ArCH), 119.73 (ArC), 118.22 (ArCH), 111.78 (ArC), 104.96 (ArBr). Mass Spectrum HRMS [EI$^+$] found 337.9184, $C_{12}H_7{}^{79}BrN_2OS_2$ calculated 337.9183. υ cm^{-1}: 3349.8, 3070.1, 2837.6, 1695.1, 1584.2, 1560.6, 1512.9, 1487.8, 1434.5, 1356.7, 1336.0, 1312.8, 1295.0, 1245.3, 1217.8, 1192.4, 1144.6, 1129.6, 1018.2, 882.3, 846.6.

(*Z*)-5-((7-Iodoindol-3-yl)methylene)-2-thioxothiazolidin-4-one (63)
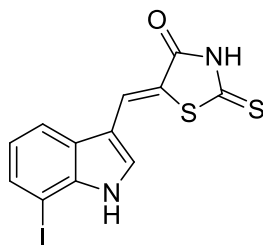
Yield; 79%. mp. ~ 320 °C (dec.). δ_H (250 MHz, DMSO-d_6): 13.63 (1 H, s, NH), 12.24 (1 H, s, NH), 7.97 (1 H, d, J = 7.5, ArCH), 7.86 (1 H, s, CH=CS), 7.67 (1 H, d, J = 7.5, ArCH), 7.66 (1 H, s, ArCH), 7.01 (1 H, t, J = 7.5, ArCH). δ_C (125 MHz, DMSO-d_6): 194.86 (CS), 169.48 (CO), 139.17 (ArC), 132.76 (ArCH), 130.22 (ArCH), 127.67 (ArC), 124.56 (ArCH), 123.43 (ArCH), 119.95 (ArC), 119.10 (ArCH), 112.49 (ArC), 78.08 (ArCI). Mass Spectrum HRMS [ES$^-$] found 384.8966, $C_{12}H_7IN_2OS_2$ calculated 384.8967. υ cm^{-1}: 3270.2, 3057.1, 2872.9, 1733.7, 1672.0, 1590.5, 1576.0, 1556.8, 1509.5, 1484.4, 1464.7, 1440.6, 1425.6, 1351.4, 1332.1, 1309.0, 1241.0, 1195.7, 1148.9, 1129.1, 1073.2, 1017.3, 876.5, 802.7, 773.3, 738.6, 707.7, 685.6, 629.2, 584.8, 556.8, 530.8.

6.1.5   General method for the preparation of α-mercaptoacrylic acids[188]

(*Z*)-3-(4-Fluoroindol-3-yl)-2-mercaptoacrylic acid (75)
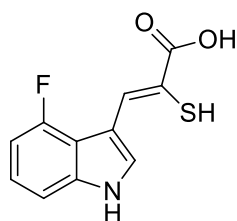
To (*Z*)-5-((4-fluoroindol-3-yl)methylene)-2-thioxothiazolidin-4-one (0.40 g, 1.44 mmol) was added 15% aqueous sodium hydroxide (4 ml). The stirred reaction mixture was heated under reflux under an N_2 atmosphere for 1.5 hr. The reaction mixture was then cooled in an ice bath and hydrochloric acid (1 M, 10 ml) added. The resulting precipitate was collected by filtration under reduced pressure and washed with water (2 × 10 ml). The solid was air dried under reduced pressure to give the title compound as a light brown coloured solid (0.30 g, 1.26 mmol, 88%). mp. 193 - 196 °C. δ_H (400 MHz, DMSO-d_6): 12.12 (1 H, s, NH), 8.26 (1 H, s, C*H*=CSH), 7.97 (1 H, d, *J* = 2.5, ArCH), 7.33 (1 H, d, *J* = 8.0, ArCH), 7.17 (1 H, dd, J_1 = 5.0, J_2 = 8.0, ArCH), 6.92 (1 H, dt, J_1 =8.0, J_2 = 5.0, ArCH), 5.20 (1 H, s, SH). δ_C (100 MHz, MeOD-d_4): 167.00 (COOH), 157.11 (ArCF, d, *J* = 243.5) 138.86 (ArC, d, *J* = 11.0), 128.64 (ArCH, d, *J* = 4.5), 127.70 (ArCH), 123.60 (ArCH, d, *J* = 8.0), 117.73 (ArC), 115.53 (ArC), 113.47 (ArC), 109.27 (ArCH), 106.15 (ArCH, d, *J* = 20.0).   Mass spectrum: HRMS (NSI⁻) found 236.0190, $C_{11}H_7FNO_2S$ calculated 236.0187. υ cm⁻¹: 3428.8, 2932.7, 2792.4, 2569.7, 1657.0, 1595.8, 1505.2, 1448.8, 1413.1, 1325.3, 1291.6, 1270.9, 1252.5, 1225.5, 1155.2, 1137.8, 1089.6, 1030.3, 901.6, 841.3, 804.2, 775.7, 727.0, 694.2, 648.9, 594.0, 547.7, 519.2, 492.2, 418.5.

Phenyl Derivatives

(*Z*)-3-(2-Fluorophenyl)-2-mercaptoacrylic acid (64)
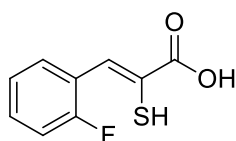
Yield; 91%. mp. 135 - 140 °C. δ_H (400 MHz, MeOD-d$_4$): 7.91 (2 H, d, *J* = 10.0, ArCH and C*H*=CSH), 7.41 (1 H, q, *J$_1$* = 7.5, *J$_2$* = 6.0, ArCH), 7.29 (1 H, t, *J$_1$* = 8.0, *J$_2$* = 7.5, ArCH), 7.18 (1 H, t, *J$_1$* = 10.0, *J$_2$* = 8.0, ArCH). δ_C (125 MHz, MeOD-d$_4$): 166.27 (COOH), 160.42 (ArCF, d, *J* = 250.5), 130.46 (ArCH, d, *J* = 8.0), 129.20 (ArCH, d, *J* = 2.0), 126.49 (ArC), 125.42 (ArCH, d, *J* = 7.0), 123.83 (ArCH, d, *J* = 4.0), 123.22 (ArC, d, *J* = 12.0), 115.09 (ArCH, d, *J* = 22.0). Mass spectrum: HRMS (NSI$^-$) found 197.0074, C$_9$H$_6$O$_2$FS calculated 197.0078. υ cm^{-1}: 2567.3, 1677.8, 1610.3, 1573.6, 1480.1, 1455.5, 1415.0, 1302.7, 1270.9, 1155.2, 1104.1, 1049.1, 1035.6, 892.9, 846.1, 801.3, 762.7, 721.6, 669.7, 602.2, 540.5, 527.0.

(*Z*)-3-(2-Chlorophenyl)-2-mercaptoacrylic acid (65)
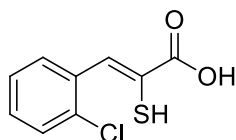
Yield; 63%. mp. 145 - 148 °C. δ_H (400 MHz, MeOD-d$_4$): 7.96 (1 H, s, C*H*=CSH), 7.83 (1 H, dd, *J$_1$* = 7.5, *J$_2$* = 1.5, ArCH), 7.51 (1 H, dd, *J$_1$* = 7.5, *J$_2$* = 1.5, ArCH), 7.41 (1 H, td, *J$_1$* = 7.5, *J$_2$* = 1.5, ArCH), 7.36 (1 H, td, *J$_1$* = 7.5, *J$_2$* = 1.5, ArCH). δ_C (125 MHz, MeOD-d$_4$): 166.20 (COOH), 133.80 (ArC), 133.62 (ArCCl), 130.50 (ArCH), 129.74 (ArCH), 129.48 (ArCH), 129.42 (ArCH), 127.09 (ArC), 126.55 (ArCH). Mass spectrum: HRMS (EI$^+$) found 213.9861, C$_9$H$_7$O$_2$^{35}ClS calculated 213.9855. υ cm^{-1}: 3058.55, 2973.21, 2639.11, 2561.97, 1680.66, 1590.02, 1563.50, 1465.63, 1437.67, 1411.64, 1295.45, 1260.25, 1212.52, 1049.09, 1034.62, 887.58, 853.35, 761.26, 743.90, 737.16, 716.91, 686.53, 663.88, 602.16, 534.19, 493.21, 468.62, 418.96.

(*Z*)-3-(2-Bromophenyl)-2-mercaptoacrylic acid (66)
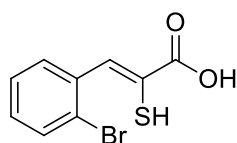
Yield; 84%. mp. 142 - 145 °C. δ_H (400 MHz, MeOD-d$_4$): 7.89 (1 H, s, C*H*=CSH), 7.77 (1 H, dd, J_1 = 8.0, J_2 = 1.5, ArCH), 7.69 (1 H, dd, J_1 = 8.5, J_2 = 1.5, ArCH), 7.45 (1 H, dt, J_1 = 8.0, J_2 = 1.5, ArCH), 7.27 (1 H, dt, J_1 = 8.0, J_2 = 1.5, ArCH). δ_C (100 MHz, MeOD-d$_4$) 166.13 (COOH), 135.45 (ArC), 133.00 (ArCH), 132.74 (ArCH), 129.92 (ArCH), 129.59 (ArCH), 127.16 (ArCH), 127.00 (ArC), 123.94 (ArCBr). Mass spectrum: HRMS (EI$^+$) found 257.9344, C$_9$H$_6$^{79}BrO$_2$S calculated 257.9350. υ cm^{-1}: 3445.7, 2828.6, 2649.7, 2570.7, 2548.0, 2360.4, 1683.1, 1600.2, 1560.1, 1463.7, 1434.8, 1416.0, 1318.6, 1289.7, 1264.6, 1212.0, 1047.2, 1028.4, 908.3, 883.2, 850.0, 758.4, 740.5, 733.3, 718.8, 663.4, 600.2, 527.9.

Benzothiophene-2-carboxylic acid (67)
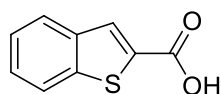
Yield; 72%. mp. 241 - 244 °C. δ_H (400 MHz, MeOD-d$_4$): 8.07 (1 H, s, C*H*=CSH), 7.94 (2 H, t, J_1 = 7.0, J_2 = 7.0, ArCH), 7.49 (1 H, td, J_1 = 7.0, J_2 = 1.5, ArCH), 7.44 (1 H, dt, J_1 = 7.0, J_2 = 1.5, ArCH). δ_C (125 MHz, MeOD-d$_4$): 164.37 (COOH), 142.22 (ArC), 138.96 (ArC), 130.18 (ArCH), 126.66 (ArCH), 125.21 (ArCH), 124.32 (ArCH), 122.32 (ArCH), 100.00 (ArC). Mass spectrum: HRMS (NSI$^-$) found 177.0019, C$_9$H$_5$O$_2$S calculated 177.0016; υ cm^{-1}: 2817.49, 2578.36, 1666.20, 1593.40, 1557.72, 1521.56, 1460.81, 1439.60, 1415.49, 1338.36, 1318.11, 1306.54, 1276.16, 1253.02, 1184,08, 1162.87, 1155.15, 1136.35, 1079.46, 1051.98, 1011.96, 949.77, 924.70, 886.13, 858.17, 761.74, 750.66, 725.10, 711.60, 584.81, 554.43, 504.29.

(*Z*)-3-(3-Fluorophenyl)-2-mercaptoacrylic acid (68)
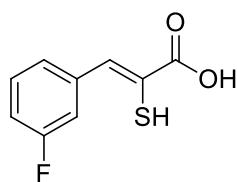
Yield; 68%. mp. 137 - 143 °C. δ_H (400 MHz, MeOD-d_4): 7.77 (1 H, s, C*H*=CSH), 7.45 (3 H, m, ArCH), 7.10 (1 H, m, ArCH). δ_C (100 MHz, MeOD-d_4): 166.40 (COOH), 162.72 (ArCF, d, *J* = 244.5), 137.60 (ArC, d, *J* = 8.1), 132.55 (ArCH), 129.98 (ArCH, d, *J* = 8.0), 125.57 (ArCH, d, *J* = 3.0), 125.52 (ArC, s), 115.58 (ArCH, d, *J* = 22.5), 115.07 (ArCH, d, *J* = 22.0). Mass spectrum: HRMS (ES⁻) found 197.0077, $C_9H_6O_2FS$ calculated 197.0073. υ cm⁻¹: 2567.3, 1677.8, 1610.3, 1573.6, 1480.1, 1455.5, 1415.0, 1302.7, 1270.9, 1155.2, 1104.1, 1049.1, 1035.6, 892.9, 846.1, 801.3, 762.7, 724.6, 669.7, 602.2, 540.5, 527.0.

(*Z*)-3-(3-Chlorophenyl)-2-mercaptoacrylic acid (69)
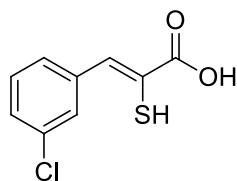
Yield; 80%. mp. 141 - 142 °C. δ_H (400 MHz, MeOD-d_4): 7.75 (1 H, s, C*H*=CSH*)*, 7.68 (1 H, s, ArCH), 7.59 (1 H, d, *J* = 7.5, ArCH), 7.44 (1 H, t, *J* = 8.0, ArCH), 7.37 (1 H, d, *J* = 8.0, ArCH). δ_C (125 MHz, MeOD-d_4): 166.36 (COOH), 137.32 (ArC), 134.17 (ArC), 132.31 (ArCH), 129.71 (ArCH), 128.95 (ArCH), 128.22 (ArCH), 127.74 (ArCH), 125.77 (ArCCl). Mass spectrum: HRMS (ES⁻) found 212.9769, $C_9H_6O_2{}^{35}ClS$ calculated 212.9777. υ cm⁻¹: 2934.2, 2818.9, 2702.3, 2636.2, 2558.1, 2510.9, 1675.8, 1587.1, 1558.7, 1480.6, 1423.7, 1347.0, 1294.5, 1268.0, 1204.8, 1098.3, 1084.3, 1045.7, 932.4, 903.5, 890.0, 779.6, 748.7, 704.4, 675.9, 658.6, 607.5, 563.6, 495.6, 456.6, 427.6.

(*Z*)-3-(3-Bromophenyl)-2-mercaptoacrylic acid (70)
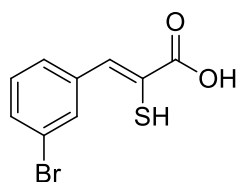
Yield; 75%. mp. 128 - 135 °C. δ_H (500 MHz, MeOD-d_4): 7.84 (1 H, s, ArCH), 7.73 (1 H, s, C*H*=CSH), 7.63 (1 H, d, *J* = 8.0, ArCH), 7.51 (1 H, d, *J* = 8.0, ArCH), 7.37 (1 H, t, *J* = 8.0, ArCH). δ_C (125 MHz, MeOD-d_4): 166.48 (COOH), 137.65 (ArC), 132.05 (ArCH), 131.90 (ArCH), 131.13 (ArCH), 129.92 (ArCH), 128.13 (ArCH), 126.11 (ArC), 122.18 (ArCBr). Mass spectrum: HRMS (ES⁻) found 256.9270, $C_9H_6O_2{}^{79}BrS$ calculated 256.9272.υ cm⁻¹: 2961.6, 2823.8, 2638.1, 2557.6, 1675.4, 1590.0, 1552.9, 1476.2, 1417.9, 1345.6, 1291.6, 1261.7, 1205.8, 1093.92, 1074.2, 1042.3, 992.7, 905.4, 889.0, 872.6, 779.6, 737.6, 692.8, 675.4, 650.4, 604.1, 555.9, 486.9, 451.3.

(*Z*)-3-(3-Iodophenyl)-2-mercaptoacrylic acid (71)
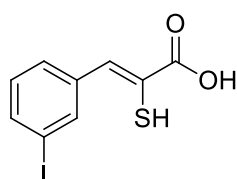
Yield; 68%. mp. 147 - 149 °C. δ_H (500 MHz, MeOD-d_4): 8.02 (1 H, s, ArCH), 7.71 (1 H, d, *J* = 7.0, ArCH), 7.70 (1 H, s, C*H*=CSH), 7.66 (1 H, d, *J* = 8.0, ArCH), 7.22 (1H, t, *J* = 8.0, ArCH). δ_C (125 MHz, MeOD-d_4): 166.36 (COOH), 138.03 (ArCH), 137.53 (ArC), 137.23 (ArCH), 132.20 (ArCH), 129.93 (ArCH), 128.61 (ArCH), 125.59 (ArC), 93.44 (ArCI). Mass spectrum: HRMS (EI⁺) found 305.9212, $C_9H_6O_2IS$ calculated 305.9212 υ cm⁻¹: 2990.6, 2577.9, 1693.7, 1588.1, 1552.4, 1473.4, 1414.5, 1345.6, 1285.8, 1251.1, 1207.7, 1067.4, 1042.3, 989.8, 921.8, 674.0, 644.6, 604.6, 549.1, 488.9, 447.9, 419.0.

(*Z*)-3-(4-Fluorophenyl)-2-mercaptoacrylic acid (72)
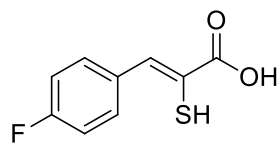
Yield; 88%. mp. 138 - 143 °C. δ_H (400 MHz, MeOD-d$_4$): 7.79 (1 H, s, C*H*=CSH), 7.72 (2 H, dt, J_1 = 8.5, J_{HF} = 3.5, ArCH), 7.20 (2 H, t, *J* = 8.5, ArCH). δ_C (100 MHz, MeOD-d$_4$): 167.12 (COOH), 163.57 (ArCF, d, *J* = 250.5), 145.20 (ArCH), 133.12 (ArCH, d, *J* = 8.0), 131.67 (ArCH, d, *J* = 8.0), 127.88 (ArC), 123.48 (ArC), 115.12 (ArCH, d, *J* = 22.0), 114.89 (d, *J* = 22.0, ArCH). Mass spectrum: HRMS (EI$^+$) found 198.0152, C$_9$H$_6$O$_2$FS calculated 198.0151. υ cm^{-1}: 2822.8, 2577.4, 1667.6, 1596.8, 1580.9, 1505.7, 1421.8, 1307.5, 1298.3, 1264.1, 1233.7, 1165.8, 1105.5, 1039.9, 1014.4, 892.9, 829.2, 798.4, 714.5, 678.3, 621.4, 542.4, 526.0.

(*Z*)-3-(4-Chlorophenyl)-2-mercaptoacrylic acid (73)
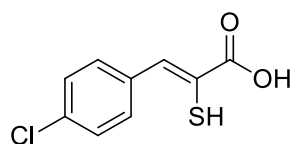
Yield; 58%. mp. 163 – 173 °C. δ_H (400 MHz, MeOD-d$_4$): 7.77 (1 H, s, C*H*=CSH), 7.66 (2 H, d, *J* = 8.0, ArCH), 7.46 (2 H, dd, J_1 = 6.5, J_2 = 2.0, ArCH). δ_C (125 MHz, DMSO-d$_6$): 166.52 (COOH), 134.02 (ArC), 132.63 (ArCH), 130.92 (2ArCH), 128.37 (2ArCH), 124.70 (ArCCl). Mass spectrum: HRMS (ES$^-$) found 212.9786, C$_9$H$_6$O$_2$^{35}ClS calculated 212.9777. υ cm^{-1}: 2945.7, 2824.2, 2644.9, 2571.1, 2506.5, 2346.0, 1665.2, 1586.2, 1559.2, 1489.3, 1422.2, 1400.1, 1349.0, 1309.9, 1287.7, 123.6, 1207.7, 1115.6, 1092.0, 1037.5, 1013.4, 905.4, 887.1, 845.6, 811.4, 738.1, 702.4, 682.2, 664.8, 623.4, 521.7, 505.3, 446.0.

(*Z*)-3-(4-Bromophenyl)-2-mercaptoacrylic acid (74)
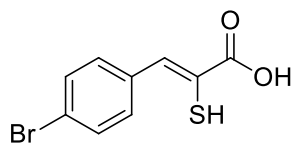
Yield; 82%. mp. 174 – 185 °C. δ_H (400 MHz, MeOD-d$_4$): 7.75 (1 H, s, C*H*=CSH), 7.60 (4 H, d, *J* = 2.7, ArCH). δ_C (125 MHz, MeOD-d$_4$): 166.58 (COOH), 134.41 (ArC), 132.62 (ArCH), 131.40 (2ArCH), 131.13 (2 ArCH), 124.97 (ArC), 122.29 (ArCBr). Mass spectrum: HRMS (ES$^-$) found 256.9281, $C_9H_6O_2{}^{79}BrS$ calculated 256.9272. υ cm^{-1}: 3445.2, 2951.0, 2823.8, 2638.6, 2587.0, 2558.6, 1674.4, 1580.4, 1557.2, 1485.9, 1418.4, 1395.7, 1342.7, 1308.5, 1288.7, 1270.4, 1206.3, 1181.2, 1115.6, 1077.1, 1037.0, 1009.6, 909.8, 886.6, 850.9, 814.3, 803.2, 768.5, 750.7, 734.8, 699.6, 677.4, 518.8, 500.0, 446.4.

Indole Derivatives

(*Z*)-3-(4-Chloroindol-3-yl)-2-mercaptoacrylic acid (76)
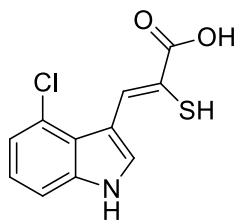
Yield; 81%. mp. 188 - 192 °C. δ_H (400 MHz, DMSO-d$_6$): 12.17 (1 H, s, NH), 8.81 (1 H, s, C*H*=CSH), 8.06 (1 H, s, ArCH), 7.49 (1 H, m, ArCH), 7.25 (2 H, m, ArCH), 5.20 (1 H, s, SH). δ_C (125 MHz DMSO-d$_6$): 167.63 (COOH), 139.90 (ArCH), 137.82 (ArCCl), 132.05 (ArCH), 125.03 (ArC), 123.63 (ArCH), 123.54 (ArC), 122.71 (ArCH), 120.71 (ArC), 112.27 (ArCH), 110.85 (ArC). Mass spectrum: HRMS (NSI$^-$) found 251.9892, $C_{11}H_7{}^{35}ClNO_2S$ calculated 251.9892. υ cm^{-1}: 3331.9, 3064.8, 2563.9, 1668.6, 1538.3, 1513.4, 1485.4, 1404.9, 1339.8, 1280.5, 1240.5, 1190.8, 1146.0, 1122.4, 1085.2, 1051.5, 1037.5, 938.7, 890.0, 793.1, 776.7, 734.7, 676.9, 570.3, 505.3, 482.1.

(*Z*)-3-(4-Bromoindol-3-yl)-2-mercaptoacrylic acid (77)
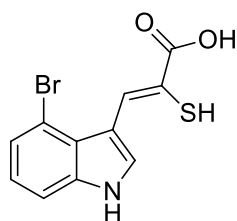
Yield; 71%. mp. 188 - 191 °C. δ_H (400 MHz, DMSO-d$_6$): 12.22 (1 H, s, NH), 8.94 (1 H, s, C*H*=CSH), 8.05 (1 H, d, *J* = 2.5, ArCH), 7.54 (1 H, d, *J* = 7.5, ArCH), 7.36 (1 H, d, *J* = 7.5, ArCH), 7.12 (1 H, t, *J* = 7.5, ArCH), 5.13 (1 H, s, SH). δ_C (125 MHz DMSO-d$_6$): 167.69 (COOH), 139.55 (ArCH), 137.80 (ArC), 132.20 (ArCH), 126.10 (ArCH), 124.81 (ArC), 123.97 (ArCH), 120.49 (ArC), 113.11 (ArCH), 112.76 (ArC), 111.13 (ArCBr). Mass spectrum: HRMS (NSI⁻) found 297.9353, $C_{11}H_7{}^{79}BrNO_2S$ calculated 297.9365. υ cm^{-1}: 3403.7, 2559.1, 1661.4, 1579.9, 1558.7, 1509.0, 1481.5, 1403.9, 1335.5, 1281.0, 1238.6, 1188.9, 1142.1, 1080.4, 1048.6, 913.1, 820.6, 774.3, 734.3, 673.0, 607.0, 563.1, 497.1.

5*H*-Thiopyrano[4,3,2-*cd*]indole-2-carboxylic acid (78)
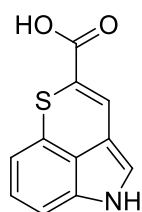
Yield; 68%. mp. ~ 320 °C (dec.). δ_H (400 MHz, DMSO-d$_6$): 11.14 (1 H, s, NH), 7.44 (1 H, s, C*H*=CS), 7.17 (1 H, d, *J* = 2.5, C=C*H*NH), 6.83 (1 H, t, *J* = 8.0, ArCH), 6.78 (1 H, d, *J* = 8.0, ArCH), 6.43 (1 H, d, *J* = 8.0, ArCH). δ_C (125 MHz DMSO-d$_6$): 165.21 (COOH), 135.54 (ArC), 128.33 (ArCH), 128.09 (ArC), 125.87 (ArCH), 125.35 (ArC), 122.22 (ArCH), 121.97 (ArC), 112.99 (ArC), 112.30 (ArCH), 108.58 (ArCH). Mass spectrum: HRMS (EI⁺) found 217.0197, $C_{11}H_7NO_2S$ calculated 217.0198. υ cm^{-1}: 3385.9, 2919.2, 2525.8, 2362.9, 1874.0, 1653.2, 1617.0, 1595.3, 1559.7, 1517.2, 1487.3. 1420.8, 1383.7, 1329.7, 1296.9, 1254.5, 1215.9, 1194.2, 1148.9, 1126.2, 1094.4, 1039.9, 1022.1, 948.8, 893.4, 850.0, 825.4, 785.4, 771.4, 733.8, 717.9, 696.7, 629.6, 614.2, 587.2, 554.4, 494.2, 466.2, 419.0.

(*Z*)-3-(5-Chloroindol-3-yl)-2-mercaptoacrylic acid (79)
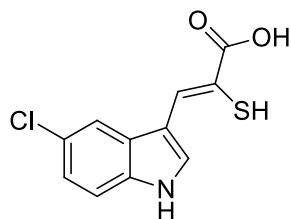
Yield; 97%. mp. 196 – 199 °C. δ_H (400 MHz, DMSO-d$_6$): 12.08 (1 H, s, NH), 8.02 (1 H, s, C*H*=CSH), 8.00 (1 H, d, *J* = 2.0, ArCH), 7.82 (1 H, d, *J* = 2.0, ArCH), 7.51 (1 H, d, *J* = 8.5, ArCH), 7.22 (1 H, dd, *J$_1$* =8.5, *J$_2$*=2.0, ArCH) 5.17 (1 H, s, SH). δ_C (100 MHz DMSO-d$_6$): 167.45 (COOH), 138.26 (ArCH), 134.53 (ArC), 132.07 (ArCH), 129.26 (ArC), 126.09 (ArCCl), 123.03 (ArCH), 121.02 (ArC), 117.86 (ArCH), 114.30 (ArCH), 110.52 (ArC). Mass spectrum: HRMS (NSI$^-$) found 251.9890, C$_{11}$H$_7$^{35}ClNO$_2$S calculated 251.9892; υ cm^{-1}: 3344.0, 3027.7, 2925.5, 2802.1, 2558.1, 2361.4, 2341.2, 1660.4, 1618.0, 1585.2, 167.8, 1511.0, 1457.9, 1405.8, 1322.9, 1276.6,1258.3, 1225.5, 1139.7,1117.6, 1054.9, 1034.6, 892.9, 852.34, 794.5, 737.5, 740.5, 723.2, 683.6, 656.6, 589.1, 498.5, 475.4, 417.5, 372.2, 354.8, 340.4, 322.1, 313.4.

(*Z*)-3-(5-Bromoindol-3-yl)-2-mercaptoacrylic acid (80)
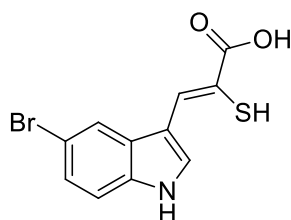
Yield; 89%. mp. 209 - 213 °C. δ_H (400 MHz, DMSO-d$_6$): 12.05 (1 H, s, NH), 8.05 (1 H, s, C*H*=CSH), 8.03 (1 H, d, *J* = 2.0, ArCH), 8.00 (1 H, d, *J* = 2.0, ArCH), 7.50 (1 H, d, *J* = 9.0, ArCH), 7.37 (1 H, dd, *J$_1$* = 9.0, *J$_2$* = 2.0, ArCH). δ_C (100 MHz DMSO-d$_6$): 167.05 (COOH), 134.87 (ArC), 129.43 (ArCH), 129.29 (ArC), 125.67 (ArCH), 125.59 (ArC), 121.05 (ArC), 118.01 (ArCH), 114.56 (ArCH), 113.43 (ArC), 111.29 (ArCBr). Mass spectrum: HRMS (NSI$^-$) found 297.9359, C$_{11}$H$_7$^{79}BrNO$_2$S calculated 297.9365. υ cm^{-1}: 3377.7, 2925.0, 1656.1, 1615.1, 1579.4, 1563.5, 1505.7, 1453.6, 1403.4, 1321.0, 1274.2, 1255.4, 1221.7, 1139.2, 1111.3, 1050.5,

881.3, 793.6, 768.5, 739.6, 718.8, 645.6, 718.8, 645.5, 593.5, 554.0, 496.6, 468.1, 414.1.

(*Z*)-3-(5-Iodoindol-3-yl)-2-mercaptoacrylic acid (81)
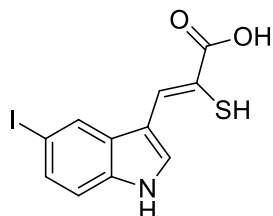
Yield; 29%. mp. 202 – 219 °C. δ_H (250 MHz, DMSO-d_6): 13.00 (1 H, s, COOH), 12.00 (1 H, s, NH), 8.13 (1 H, s, C*H*=CSH), 8.01 (1 H, s, ArCH), 7.95 (1 H, s, ArCH), 7.49 (1 H, d, *J* = 8.0, ArCH), 7.35 (1 H, d, *J* = 8.0, ArCH). δ_C (100 MHz DMSO-d_6): 167.07 (COOH), 135.23 (ArC), 131.94 (ArCH), 130.98 (ArCH), 130.02 (ArC), 128.87 (ArCH), 127.04 (ArCH), 125.50 (ArC), 114.99 (ArC), 110.95 (ArCH), 84.74 (ArCI). Mass spectrum: HRMS (APCI$^+$) found 345.9388, $C_{11}H_9INO_2S$ calculated 345.9393. υ cm^{-1}: 3418.2, 2921.2, 2251.5, 1671.0, 1612.2, 1577.0, 1559.2, 1504.7, 1446.8, 1328.7, 1226.0, 1137.3, 1043.8, 878.9, 795.5, 764.6, 699.1, 606.1.

(*Z*)-3-(6-Fluoroindol-3-yl)-2-mercaptoacrylic acid (82)
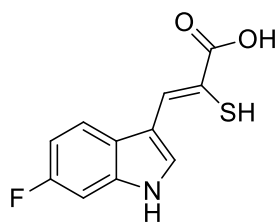
Yield; 56%. mp. 194 - 196 °C. δ_H (400 MHz, DMSO-d_6): 11.87 (1 H, s, NH), 8.01 (1 H, s, C*H*=CSH), 7.94 (1 H, d, *J* = 2.5, ArCH), 7.76 (1 H, dd, J_1 = 5.5, J_2 = 3.5, ArCH), 7.26 (1 H, dt, J_1 =9.5, J_2 = 2.5, ArCH), 7.00 (1H, td, J_1 = 9.5, J_2 = 2.5, ArCH). δ_C (100 MHz DMSO-d_6): 167.02 (COOH), 159.87 (ArCF, d, *J* = 236.5), 136.05 (ArC, d, *J* = 13.5), 128.76 (ArCH), 125.85 (ArCH), 124.22 (ArC), 119.75 (ArCH, d, *J* = 10.0), 117.80 (ArC), 111.65 (ArC), 109.19 (ArCH, d, *J* = 24.5), 98.58 (ArCH, d, *J* = 25.5). Mass spectrum: HRMS (NSI$^-$) found 236.0190, $C_{11}H_7FNO_2S$ calculated 236.0187, *m/z* (NSI$^-$). υ cm^{-1}: 3632.8, 3223.9, 1583.3, 1561.6, 1506.1, 1452.1, 1408.8,

1333.5, 1298.8, 1256.4, 1228.4, 1144.1, 1069.8, 907.8, 800.8, 783.4, 725.6, 656.2, 577.6, 491.8.

(*Z*)-3-(6-Chloroindol-3-yl)-2-mercaptoacrylic acid (83)
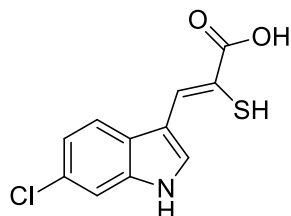
Yield; 52%. mp. 201 - 202 °C. δ_H (400 MHz, DMSO-d$_6$): 12.93 (1 H, s, COOH), 11.93 (1 H, s, NH), 8.02 (1 H, s, C*H*=CSH), 7.99 (1 H, d, *J* = 2.5, ArCH), 7.78 (1 H, d, *J* = 8.5, ArCH), 7.54 (1 H, t, *J* = 12.5, ArCH), 7.14 (1 H, d, *J* = 8.5, ArCH). δ_C (100 MHz DMSO-d$_6$): 167.42 (COOH), 138.10 (ArCH), 136.42 (ArC), 131.61 (ArCH), 127.60 (ArCCl), 126.79 (ArC), 121.55 (ArCH), 121.28 (ArC), 119.90 (ArCH), 112.36 (ArCH), 110.82 (ArC). Mass spectrum: HRMS (NSI$^-$) found 251.9892, $C_{11}H_7{}^{35}ClNO_2S$ calculated 251.9890. υ cm^{-1}: 3632.8, 3223.9, 1583.3, 1561.6, 1506.1, 1452.1, 1408.8, 1333.5, 1298.8, 1256.4, 1228.4, 1144.1, 1069.8, 907.8, 800.8, 783.4, 725.6, 656.2, 577.6, 491.8.

(*Z*)-3-(6-Bromoindol-3-yl)-2-mercaptoacrylic acid (84)
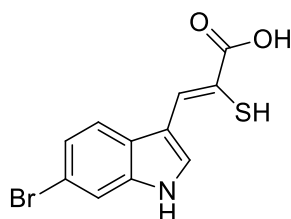
Yield; 78%. mp. 193 – 203 °C. δ_H (400 MHz, DMSO-d$_6$): 13.11 (1 H, s, COOH), 11.94 (1 H, s, NH), 8.01 (1 H, s, C*H*=CSH), 7.97 (1 H, d, *J* = 2.0, ArCH), 7.73 (1 H, d, *J* = 8.5, ArCH), 7.68 (1 H, d, *J* =1.5, ArCH), 7.27 (1 H, dd, *J$_1$* = 8.5, *J$_2$* = 2.0, ArCH), 5.19 (1 H, s, SH). δ_C (100 MHz DMSO-d$_6$): 166.99 (COOH), 136.98 (ArC), 129.00 (ArCH), 126.53 (ArC), 125.58 (ArCH), 123.56 (ArCH), 120.43 (ArCH), 118.11 (ArC), 115.65 (ArCBr), 115.10 (ArCH), 111.71 (ArC). Mass spectrum: HRMS (NSI$^-$) found 295.9378, $C_{11}H_7{}^{79}BrNO_2S$ calculated 295.9386. υ cm^{-1}: 3432.7, 2923.1, 2562.5, 1672.0, 1590.5, 1516.3, 1454.1, 1423.7, 1388.0, 1332.6, 1312.8,

1297.4, 1268.5, 1141.2, 1111.3, 1071.3, 895.8, 851.9, 811.9, 775.2, 726.6, 678.3, 517.3, 419.9.

(*Z*)-3-(6-Iodoindol-3-yl)-2-mercaptoacrylic acid (85)
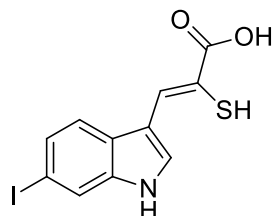
Yield; 67%. mp. 185 - 187 °C. δ_H (400 MHz, DMSO-d_6): 13.02 (1 H, s, COOH), 11.89 (1 H, s, NH), 8.00 (1 H, s, C*H*=CSH), 7.91 (1 H, s, ArCH), 7.85 (1 H, s, ArCH), 7.60 (1 H, d, *J* = 7.5, ArCH), 7.41 (1 H, d, *J* = 7.5, ArCH). δ_C (125 MHz DMSO-d_6): 167.42 (COOH), 138.01 (ArCH), 137.45 (ArC), 131.17 (ArCH), 129.59 (ArCH), 127.42 (ArC), 123.20 (ArC), 121.17 (ArCH), 120.55 (ArCH), 110.87 (ArC), 87.11 (ArCI). Mass spectrum: HRMS (NSI$^-$) found 343.9241, $C_{11}H_7INO_2S$ calculated 343.9241. υ cm^{-1}: 3423.5, 2815.6, 2561.5, 1670.1, 1603.5, 1588.6, 1563.5, 1512.9, 1449.7, 1419.4, 1385.1, 1329.7, 1312.8, 1295.5, 1267.0, 1230.8, 1144.6, 1109.8, 1069.8, 1037.0, 901.1, 886.1, 851.9, 808.5, 772.4, 756.0, 742.5, 723.2, 674.0, 635.4, 622.9.

(*Z*)-3-(7-Fluoroindol-3-yl)-2-mercaptoacrylic acid (86)
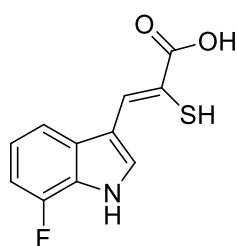
Yield; 78%. mp. 181 - 183 °C. δ_H (400 MHz, DMSO-d_6): 12.38 (1 H, s, NH), 8.03 (1 H, s, C*H*=CSH), 7.95 (1 H, d, *J* = 2.5, ArCH), 7.59 (1 H, d, *J* = 8.0, ArCH), 7.09 (3 H, m, ArCH). δ_C (100 MHz DMSO-d_6): 167.26 (COOH, s), 149.63 (ArCF, d, *J* = 244.5), 138.20 (ArCH), 131.72 (ArC, d, *J* = 5.5), 131.13 (ArCH), 123.75 (ArC, d, *J* = 4.0), 121.75 (ArCH, d, *J* = 6.5), 121.43 (ArC), 114.45 (ArCH), 111.45 (ArC), 107.94 (ArCH, d, *J* = 16.5). Mass spectrum: HRMS (EI$^+$) found 237.0252, $C_{11}H_8FNO_2S$ calculated 237.0260. υ cm^{-1}: 3318.4, 3044.1, 2563.4, 1665.7, 1577.0, 1518.2,

1498.4, 1447.3, 1409.2, 1338.4, 1307.5, 1262.2, 1228.9, 1139.2, 1106.0, 1041.9, 975.8, 896.3, 866.4.

(*Z*)-3-(7-Chloroindol-3-yl)-2-mercaptoacrylic acid (87)
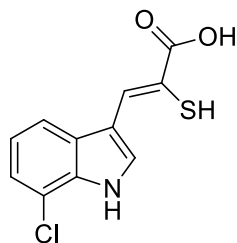
Yield; 82%. mp. 179 - 180 °C. δ_H (400 MHz, DMSO-d_6): 12.24 (1 H, s, NH), 8.01 (1 H, s, C*H*=CSH), 7.93 (1 H, s, ArCH), 7.76 (1 H, d, *J* = 8.0, ArCH), 7.31 (1 H, d, *J* = 8.0, ArCH) 7.18 (1 H, t, *J* = 8.0, ArCH). δ_C (125 MHz DMSO-d_6): 167.64 (COOH), 139.92 (ArCH), 137.82 (ArC), 132.06 (ArCH), 125.03 (ArC), 123.63 (ArCH), 123.54 (ArC), 122.72 (ArCH), 120.70 (ArCCl), 112.27 (ArCH), 110.85 (ArC). Mass spectrum: HRMS (NSI⁻) found 251.9887, $C_{11}H_7{}^{35}ClNO_2S$ calculated 251.9881. υ cm⁻¹: 3344.93, 2810.7, 2567.3, 1669.1, 1623.3, 1596.3, 1569.8, 1518.2, 1491.2, 1435.7, 1422.7, 1405.9, 1338.4, 1306.1, 1265.1, 1243.9, 1200.0, 1155.2, 1119.0, 1090.1, 1042.3, 894.3, 857.7, 780.5, 751.6, 733.8, 718.8, 682.2, 582.4, 540.9, 505.3, 488.4.

(*Z*)-3-(7-Bromoindol-3-yl)-2-mercaptoacrylic acid (88)
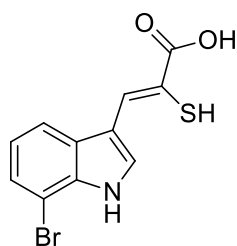
Yield; 86%. mp. 165 - 168 °C. δ_H (400 MHz, DMSO-d_6): 12.09 (1 H, s, NH), 8.01 (1 H, s, C*H*=CSH), 7.92 (1 H, d, *J* = 3.0, ArCH), 7.42 (1 H, dd, *J₁* = 8.0, *J₂* = 3.0, ArCH), 7.10 (1 H, dt, *J₁* = 8.0, *J₂* = 3.0, ArCH). δ_C (125 MHz DMSO-d_6): 167.26 (COOH), 138.09 (ArCH), 134.48 (ArC), 131.34 (ArCH), 129.74 (ArC), 125.69 (ArCH), 122.66 (ArCH), 121.97 (ArC), 117.97 (ArCH), 111.86 (ArC), 105.29 (ArCBr). Mass spectrum: HRMS (EI⁺) found 296.9466, $C_{11}H_8{}^{79}BrNO_2S$ calculated

296.9459. υ cm^{-1}: 3351.7, 2946.7, 2812.2, 2625.6, 2563.4, 1670.1, 1617.5, 1594.4, 1563.0, 1515.8, 1487.8, 1434.8, 1420.3, 1403.4, 1335.0, 1304.6, 1262.2, 1081.4, 1040.9, 881.3, 847.1, 807.6.

(*Z*)-3-(7-Iodoindol-3-yl)-2-mercaptoacrylic acid (89)
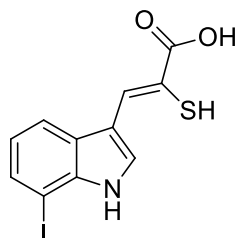
Yield; 81%. mp. 188 - 189 °C. δ_H (400 MHz, DMSO-d$_6$): 11.78 (1 H, s, NH), 8.24 (1 H, d, *J* = 2.5, ArCH), 8.01 (1 H, s, C*H*=CSH), 7.59 (1 H, d, *J* = 8.0, ArCH), 7.45 (1 H, d, *J* = 8.0, ArCH), 6.80 (1 H, t, *J* = 8.0, ArCH). δ_C (100 MHz DMSO-d$_6$): 166.94 (COOH), 137.99 (ArCH), 132.02 (ArC), 128.49 (ArCH), 127.93 (ArC), 125.65 (ArCH), 122.61 (ArCH), 118.72 (ArC), 118.59 (ArCH), 112.86 (ArC), 77.96 (ArCI). Mass spectrum: HRMS (NSI$^-$) found 345.9391, C$_{11}$H$_9$INO$_2$S calculated 345.9393. υ cm^{-1}: 3384.0, 2920.2, 2571.1, 1679.2, 1588.6, 1555.3, 1514.8, 1481.1, 1419.8, 1396.2, 1331.6, 1303.2, 1260.7, 1239.0, 1197.1, 1143.1, 1112.2, 1072.7, 1039.4, 875.5, 780.5, 752.6, 719.3, 675.9, 605.5, 570.3, 522.6.

6.2 Testing for disulfide formation

6.2.1 UV-Vis Analysis

1 μl of a stock solution of (*Z*)-3-(4-bromophenyl)-2-mercaptoacrylic acid (**74**), 100 mM in DMSO, was added to 1 ml of a 1:1 solution of acetonitrile and 100 mM K$_i$HPO$_4$ solution (pH 7.4), to create an overall concentration 100 μM. A UV-Vis spectrum was measured over the range 450 nm to 250 nm at 20 °C and every 20 minutes subsequently to that for a total of 300 minutes. A Jasco V-660 Spectrophotometer (Jasco-UK, Essex, UK) was used with the temperature regulated with a Peltier element. To the solution 10 μl of a 1.0 M solution of dithiothreitol (DTT) was added and a spectrum was obtained and every 1 minute subsequently for 5 minutes until the compound was fully reduced, using the same wavelength and temperature parameters as above.

The same proceedure was applied for the UV-Vis spectroscopy analysis of the indole derivative (Z)-3-(7-bromoindol-3- yl)-2-mercaptoacrylic acid (**88**)

6.2.2   HPLC Analysis

HPLC was performed on a Dionex UltiMate 3000 (ThermoFisher, UK) pump with an UltiMate 3000 variable wavelength detector. 1.0 mg of **88** was dissolved in 0.5 ml of a 1:1 mixture of acetonitrile and 100 mM K_iHPO_4 (at pH 7.4), 20 μl of this solution was injected onto an analytical HPLC column [Acclaim® 120, C18, 3 μm, 120 Å, 150 × 4.6 mm (ThermoFisher, Surrey, UK)]. The sample was eluted using a gradient from 100% water (0.1% TFA) to 100% acetonitrile (0.1% TFA) over the course of 60 min - 1 ml a minute. The eluent was analysed at 4 different wavelengths (210, 260, 290 and 350 nm). The sample was left to oxidise over the course of 5 hours whereupon 20 μl of the sample was injected onto the HPLC column and the same analysis was repeated. To the stock solution 10 μl of 1.0 M of DTT was added, making the total solution to 10 mM DTT, the solution was then injected onto the analytical HPLC column and the same analysis was repeated once more.

The same procedure was applied for the HPLC analysis of **74**, where 1.5 mg was dissolved in 1 ml of a 1:1 solution of acetonitrile and 100 mM K_iPO_4 (at pH 7.4). After 5 hrs the compound had not fully oxidised so the solution was left for a further 19 hrs to fully oxidise.

6.2.3   NMR analysis

5 mg of **88** was dissolved in 0.6 ml of DMSO-d_6 and a ^1H NMR spectrum was collected immediately. The solution was then left to stand for 5 hrs at room temperature. After this time period another spectrum was collected. A third spectrum was collected after 24 hrs at room temperature.

6.3 Testing of α-mercaptoacrylic acid derivatives (Dr C. Parr, Dr E. Robinson and Prof. M. B. Hallett, School of Medicine)

The final α-mercaptoacrylic acid derivatives were tested by Dr Christian Parr, Dr Emma Robinson and Prof. Maurice B. Hallett in the School of Medicine at Cardiff University (School of Medicine, Heath Hospital, Cardiff University, Cardiff, UK).

6.3.1 Enzymatic assays

Assay buffer 1

10 mM HEPES, 10 mM DTT (dithiothreitol), 0.5 mM EDTA (ethylenediaminetetraacetic acid) and 0.1% (w/v) bovine serum albumin (BSA), at pH 7.2.

FAM based calpain-I specific fluorogenic assay

A calpain-I specific substrate; α-fodrin was utilised for this assay, with fluorescein amidite (FAM) attached to the N-terminal lysine which is internally quenched by 4-([4-(dimethylamino)phenyl]azo)benzoic acid (DABCYL) attached to the C-terminal lysine; (H-K(FAM)-EVYGMMK(DABCYL)-OH).[206] Cleavage by calpain-I occurs between the tyrosine and the glycine removing the quenching effect of the DABCYL substituent and thereby leading to an increase in fluorescence at 520 nm.

Assays were performed with 25 nM human calpain-I (CalBiochem) and 1 mM fluorogenic calpain-I substrate in assay buffer 1 (*vide supra*). Each assay (100 μl) was performed at 37 °C within a fluorescent plate reader (BMG Optistar), with the excitation wavelength set to 490 nm and emission wavelength detected at 520 nm. Potential inhibitors were added as a solution in DMSO (*vide infra*) to the assay mixture prior to the reaction and assays were initiated by addition of 5 mM $CaCl_2$. The reaction was incubated for 1 hr at 37 °C and the fluorescence emission was detected continuously over that time period.

A stock solution of each inhibitor in DMSO (50 mM) was diluted into assay buffer to give a range of solutions from 100 nM to 10 mM. In each assay the effect of DMSO alone on calpain-I was measured; at lower concentrations no effect was observed but at 0.01-0.1% some inhibitory effect was observed. The effect of DMSO upon calpain-I was subtracted from the inhibition data collected.

AMC based fluorogenic assay

The substrate succinyl-LLVT-7-amido-4-coumarin was used as a non-specific calpain fluorogenic substrate for testing with calpain-I and calpain-II. AMC fluoresces even when attached to the substrate but upon cleavage the fluorescence of AMC increases and is detected at 460 nm.

Assays (100 μl) were performed with 25 nM human calpain-I or 25 nM porcine calpain-II (CalBiochem) and 1 mM fluorogenic AMC substrate in assay buffer 1 (*vide supra*). Each assay (100 μl) was performed a 37 °C within a fluorescent plate reader (BMG Optistar), with the excitation wavelength set to 390 nm and emission wavelength detected at 460 nm. Potential inhibitors were added as a solution in DMSO (*vide supra*) to the assay mixture prior to the reaction and assays were initiated by addition of 5 mM $CaCl_2$. The reaction was incubated for 1 hour at 37 °C and the fluorescence emission was detected continuously over that time period through the fluorescence detector in the plate reader

6.3.2   Cell spreading assay

Neutrophils were isolated from donated blood,[244] and re-suspended in Krebs medium at pH 7.4; 120 mM NaCl, 4.9 mM KH_2PO_4, 1.2 mM $MgSO_4$, 1.2 mM $CaCl_2$, 1.3 mM KCl, 25 mM HEPES and 0.1% BSA. 100 μl of cell-free Krebs medium was placed on an untreated coverslip which was warmed to 37 °C by the heated stage of a confocal laser scanning microscope (Leica). 100 μl of the cell suspension ($\sim 10^6$/ml) was added and the cells were allowed to settle upon the glass surface, where upon contact with the surface the cells begin to spontaneously spread. $\sim$50 cells were analysed for 3 minutes, following the contact with the glass surface. Images were collected with Leica software, the boundaries of the cell were identified through contrast enhancement and the 2D area of the cell was determined with Image J software.

To allow the maximum time for the inhibitors to enter the cell and take effect, the inhibitors were added to the cell solution prior to addition to the coverslip, to give a total concentration of 10 µM,

6.3.3 Photo-reaction

Cell spreading assay

Neutrophils were isolated from donated blood,[244] and re-suspended in Krebs medium (*vide supra*). 100 µl of cell-free Krebs medium was placed on an untreated coverslip which was warmed to 37 °C by the heated stage of a convention phase contrast light microscope (Leica). 100 µl of the cell suspension (~10^6/ml) was added and the cells were allowed to settle upon the glass surface. The cells were imaged using a transmitted light (488 nm) and the 410 nm diode laser (UV laser) restricted to a few cells. The UV laser was activated for 1 min and each cell was exposed for 5-15 sec (depending on the percentage of the scan zaone occupied by the cell). Formyl-methionyl-leucyl-phenylalanine (fMLP) was added and the irradiated cells were analysed for 5 minutes. Images were collected with Leica software and analysed using Image J software (NIH, USA).

Inhibitors or scam addition (equivalent DMSO concentrations) were addad to the cell solutions prior to addition to the coverslip (total concentration of 10 µM).

AMC based fluorogenic assay

The substrate succinyl-LLVT-7-amido-4-coumarin was used as a non-specific calpain fluorogenic substrate for testing with calpain-I.

Assays (100 µl) were performed with 20 nM porcine calpain-I (CalBiochem) and 1 mM fluorogenic AMC substrate in assay buffer 1 (*vide supra*). Prior to the addition to the 96 well plate the two control solutions were kept in the dark, the two light samples were irradiated with a 100 W light source with a 360 nm band pass filter. Each assay (100 µl) was performed a 37 °C within a fluorescent plate reader (BMG Optistar), with the excitation wavelength set to 390 nm and emission wavelength detected at 460 nm. Potential inhibitors were added as a

solution in DMSO (*vide supra*) to the assay mixture prior to the reaction and assays were initiated by addition of 5 mM $CaCl_2$. The reaction was incubated for 1 hour at 37 °C and the fluorescence emission was detected continuously over that time period.

6.4 Protein Synthesis

6.4.1 Materials

All chemicals were purchased from Sigma Aldrich and Fisher Scientific, unless otherwise stated. IPTG and ampicillin were purchased from Melford. All restriction enzymes were purchased from New England Biolabs. The genes encoding for porcine PEF(S) was purchased from Epoch Gene Synthesis (USA). All chemicals were used in accordance with their material safety sheets, COSHH and risk assessments.

6.4.2 Media

Luria-Bertani medium

Tryptone (10.0 g), yeast extract (5.0 g) and sodium chloride (10.0 g) were dissolved in 1 l of deionised water. The solution was sterilised in an autoclave at 121 °C, 15 lb in^{-2} for 20 minutes.

LB agar

Tryptone (10.0 g), yeast extract (5.0 g) and sodium chloride (10.0 g) and agar (15.0 g) were dissolved in 1 l of deionised water for preparation of culture plates. The solution was sterilised in an autoclave at 121 °C, 15 lb.(sq. in)$^{-1}$ for 20 minutes.

6.4.3 Sterile Solutions

Antibiotics

Ampicillin was dissolved in deionised water to a concentration of 50 mg/ml and used at a working concentration of 0.1 mg/ml. The antibiotic was sterilised utilising a 0.2 µm syringe filter, aliquoted and stored at -20 °C.

Isopropyl-β-D-1-thiogalactopyranoside (IPTG)

IPTG (360 mg) was dissolved in 1 ml of deionised water for use at a working concentration of 0.12 mg/ml. The solution was sterilised using a 0.2 μm syringe filter, aliquoted and stored at -20 °C.

Competent cell solutions

Rubidium chloride preparation of ultra-competent cells

Rubidium chloride solution 1 (Tbf1) was prepared by dissolving potassium acetate (0.29 g), rubidium chloride (RbCl) (1.21 g), calcium chloride (CaCl$_2$) (0.11 g), manganese(II)chloride (MnCl$_2$) (0.69 g) and glycerol (15 ml) in 100 ml of deionised water. The total concentrations within the solution were 30 mM potassium acetate, 100 mM rubidium chloride, 10 mM CaCl$_2$, 50 mM manganese chloride and 15% v/v glycerol. The pH of the solution was adjusted to 5.8 with dilute acetic acid. The solution was sterilised utilising a 0.2 μm syringe filter and stored at 4 °C.

Rubidium chloride solution 2 (Tbf2) was prepared by dissolving 3-(N-morpholino)propanesulphonic acid (MOPS) (0.21 mg), CaCl$_2$ (0.83 g), RbCl (0.12 g) and glycerol (15% v/v) in 100 ml of deionised water. The pH of the solution was adjusted to 6.5 with dilute sodium hydroxide and sterilised utilising a 0.2 μm syringe filter. The solution was stored at 4 °C.

Calcium chloride preparation of competent cells

Calcium chloride solution 1 (Ca1) was prepared by dissolving CaCl$_2$ (1.11 g) in 100 ml of deionised water, which gave a concentration of 100 mM. The solution was sterilised *via* autoclave at 121 °C (15 lb in^{-2}) for 20 minutes and stored at 4 °C.

Calcium chloride solution 2 (Ca2) was prepared by dissolving CaCl$_2$ (1.11 g) and glycerol (15 ml) in 100 ml deionised water, which gave concentrations of 100 mM CaCl$_2$ and 15% w/v glycerol. The solution was sterilised *via* autoclave at 121 °C (15 lb in^{-2}) for 20 minutes and was stored at 4 °C.

6.4.4 Non-sterile solutions

Ethidium Bromide

A 25 mM stock ethidium bromide solution was prepared by dissolving 8.1 g of the solid in 100 ml of deionised water. The solution was stored in the dark at 4 °C. Immediately prior to gel staining, a solution with a working concentration of 6 µM was prepared by adding 48 µl of the stock solution to 200 ml of deionised water.

10× DNA loading dye

0.5% (w/v) Bromophenol blue (2 µl) and glycerol (300 µl) were dissolved in 0.7 ml of deionised water, yielding total concentrations of 0.001% (w/v) bromophenol blue and 30% (v/v) glycerol. The solution was stored at room temperature. The dye was diluted 1:9 with each DNA sample immediately prior to use.

TAE buffer stock (50×) for agarose gels

A 0.5 M EDTA solution was prepared by dissolving EDTA (14.61 g) in 100 ml of deionised water, the pH of the solution was then adjusted to 8.0 with 5 M NaOH. The solution was then immediately used in the preparation of TAE buffer stock (50 ×).

Tris base (242 g), glacial acetic acid (57 ml) and 0.5 M EDTA (100 ml, pH 8.0) were dissolved in 1 l with deionised water, which gave total concentrations of 2 M tris base, 5.7% acetic acid and 50 mM EDTA. The solution was stored at room temperature. The solution was diluted 1:49 with deionised water immediately prior to use.

SDS stacking buffer

Tris base (6.0 g) was dissolved in 100 ml of deionised water to give a total concentration of 0.5 M. The pH of the solution was adjusted to 6.8 with 6 M HCl. The solution was stored at room temperature.

SDS resolving buffer

Tris base (27.73 g) was dissolved in 150 ml of deionised water, to give a total concentration of 1.5 M. The pH of the solution was adjusted to 8.8 and the resulting solution was stored at room temperature.

10% (w/v) Sodium dodecyl sulfate (SDS)

Sodium dodecyl sulfate (10.0 g) was dissolved in 100 ml of water. The resulting solution was stored at room temperature.

10% (w/v) Ammonium persulfate

Ammonium persulfate (100 mg) was dissolved in 1 ml of deionised water. The solution was freshly made prior to use.

SDS electrode running buffer (10×)

Tris base (30.3 g), glycine (144.0 g), SDS (10.0 g) were dissolved in 1 l of deionised water, yielding total concentrations of 0.25 M Tris base, 1.92 M glycine and 1% (w/v) SDS. The pH (~8.3) of the solution was not adjusted. The solution was stored at room temperature. Immediately prior to use the solution was diluted 1:9 with deionised water.

Protein loading dye

Tris-HCl, (1.25 ml, pH 6.8) (SDS stacking buffer), glycerol (3 ml), bromophenol blue (1 ml) and SDS (2 ml) were mixed and the total volume taken to 10 ml with deionised water, to make total concentrations of 0.5 M Tris-HCl, 30% glycerol, 0.6% (w/v) bromophenol blue, 10% SDS. β-Mercaptoethanol (β-ME) was added to a final concentration of 10% prior to use.

Protein purification buffer A

Tris base (4.8 g), EDTA (1.3 g) and β-ME (0.7 ml) were dissolved in 2 l of deionised water, to yield total concentrations of 20 mM Tris base, 2 mM EDTA and 5 mM β-ME. The pH of the solution was adjusted to 8.0 using 5 M HCl. The solution was filtered and degassed *via* vacuum pump and stored at room temperature.

Protein purification buffer B

Tris base (4.8 g), EDTA (1.3 g), β-ME (0.7 ml) and NaCl (120 g) were dissolved in 2 l of deionised water, which gave total concentrations of 20 mM Tris base, 2 mM EDTA, 5 mM β-ME and 1 M NaCl. The pH of the solution was adjusted to 8.0 using 5 M HCl. The solution was filtered and degassed *via* vacuum pump) and stored at room temperature.

Protein purification buffer C

Tris base (4.8 g), EDTA (1.3 g), β-ME (0.7 ml) and NaCl (18 g) were dissolved in 2 l of deionised water, which gave total concentrations of 20 mM Tris base, 2 mM EDTA, 5 mM β-ME and 0.15 M NaCl. The pH of the solution was adjusted to 8.0 using 5 M HCl. The solution was filtered and degassed *via* vacuum pump and stored at room temperature.

Dialysis buffer

KH_2PO_4 (174.2 g) and K_2HPO_4 (136.1 g) were separately dissolved in 1 l of deionised water for use as 1 M stock solutions.

The stock solutions were used to prepare the dialysis buffer for the protein.

12.3 ml of K_2HPO_4 stock solution and 7.7 ml of KH_2PO_4 stock solution were diluted in 4 l of deionised water, along with 2.6 g of EDTA and 1.4 ml of β-ME to give final concentrations of 5 mM, 2 mM and 5 mM respectively. The pH of the solution was adjusted to 7.0. The buffer was prepared fresh each time and the dialysis was performed at 4 °C.

DNA miniprep buffers

Buffer P1 (suspension buffer)

Tris-HCl (157.6 mg) and EDTA (58.4 mg) and were dissolved in 100 ml of deionised water, to give total concentrations of 50 mM and 10 mM. The pH of the solution was then adjusted to 8.0. RNAase (50 µg/ml final concentration) was dissolved and the solution was stored at 4 °C.

Buffer P2 (lysis buffer)

NaOH (4 g) and SDS (5 g) were dissolved separately in 250 ml of deionised water, to give concentrations of 0.4 M and 2% (w/v) these solutions were then mixed 1:1. The solution was stored at room temperature.

Buffer N3 (neutralisation and binding buffer)

Guanidine hydrochloride (7.6 g) and potassium acetate (981.4 mg) were dissolved in 20 ml of deionised water, to give concentrations of 4 M and 0.5 M respectively. The pH of the solution was adjusted to 4.2 and was stored at room temperature.

Buffer PB (wash buffer)

Guanidine hydrochloride (95.5 g) and Tris-HCl (6.3 mg) were dissolved in 125 ml of deionised water. The pH of the solution was adjusted to 6.6 and 76 ml of ethanol was added, to give total concentrations of 5 M and 2 mM. The solution was stored at room temperature.

Buffer PE (wash buffer)

NaCl (0.117 g) and Tris-HCl (3.1 mg) were dissolved in 20 ml of deionised water. The pH of the solution was adjusted to 7.5 and 80 ml of ethanol was added, where the total concentrations are 20 mM NaCl and 3 mM Tris-HCl,. The solution was stored at room temperature.

Buffer EB (elution buffer)

Tris-HCl (32 mg,) was dissolved in 20 ml of deionised water and the pH of the solution was adjusted to 8.5, where the concentration of Tris-HCl is 10 mM. The solution was stored at room temperature.

Agarose gel DNA isolation buffers

Buffer QG (gel solubilisation buffer)

Guanidine thiocyanate (13.0 g) and Tris-HCl (6.3 mg) were dissolved in 20 ml of deioised water, yielding total concentrations of 5.5 M guanidine thiocyanate and 20 mM Tris-HCl. The pH of the solution was adjusted to 6.6 and stored at room temperature.

Buffer PE (wash buffer)

NaCl (0.117 g) and Tris-HCl (3.1 mg) were dissolved in 20 ml of deionised water, where the total concentrations are 20 mM NaCl and 2 mM Tris-HCl. The pH of the solution was adjusted to 7.5 and 80 ml of ethanol was added. The solution was stored at room temperature.

Buffer EB (elution buffer)

Tris-HCl (32 mg) was dissolved in 20 ml of deionised water, the total concentration was 10 mM Tris-HCl. The pH of the solution was adjusted to 8.5. The solution was stored at room temperature.

6.4.5 *E. coli* strains and their preparation

Cloning strains

XL1-Blue ® Super competent cells (Stratagene, CA, USA) were utilised for the preparation of new plasmid constructs since they are highly proficient in replicating DNA, creating high concentrations of the desired plasmid, with an efficiency of $\geq 1 \times 10^8$ colony forming units per μg.[245]

Expression strains

BL21-CodonPlus(DE3)-RP competent cells and BL21-CodonPlus(DE3)-RIL competent cells (Stratagene, CA, USA) were used for the expression of the porcine *pPEF(S)* gene. These two strains of *E. coli* contain extra genes that encode for arginine and proline tRNA in the case of the BL21-CodonPlus(DE3)-RP cells and with BL21-CodonPlus(DE3)-RIL cells there are genes that encode for the tRNA of arginine, isoleucine and leucine.[246] The excess tRNA allows for the efficient expression of proteins that contain a large quantity of these amino acids.[246]

Preparation of super-competent cells

50 μl of cells in calcium chloride solution 2 (Section 6.4.3) was added to 100 ml of non-selective LB medium, the culture was incubated overnight at 37 °C. 1 ml of this culture was utilised to inoculate 100 ml of fresh non-selective LB media. The culture was allowed to grow until an OD of 0.6 at 600 nm was reached after which time the cells were chilled on ice for at least 20 min. The cells were

harvested by centrifugation (Eppendorf centrifuge 5810R) at 3220 RCF for 10 min at 4 °C and the supernatant solution was discarded. The pellet was then re-suspended in 20 ml of Tbf1 solution and incubated on ice for 20 minutes; the centrifugation step was then repeated and the supernatant solution was discarded once more. The pellet was then resuspended in 5 ml of Tbf2 solution, the suspension was then left on ice for 20 min and aliquoted (50 μl) into sterile Eppendorf tubes and flash frozen in liquid nitrogen for storage at -80 °C.

Preparation of competent cells for protein expression

50 μl of cells in calcium chloride solution 2 (Section 6.4.3) was added to 100 ml of non-selective LB medium, the culture was incubated overnight at 37 °C. 1 ml of this culture was utilised to inoculate 100 ml of fresh non-selective LB medium. The culture was allowed to grow until an OD of 0.6 at 600 nm was reached after which time the cells were chilled on ice for at least 20 min. The cells were harvested by cenrtrifugation (Eppendorf centrifuge 5810R) at 3220 RCF for 10 min at 4 °C and the supernatant solution was discarded. The pellet was then re-suspended in 20 ml of Ca1 solution and incubated on ice for 20 min, the centrifugation step was then repeated and the supernatant solution discarded once more. The pellet was resuspended in 5 ml of Ca2 solution and the solution was chilled on ice for 20 min. The suspension was then aliquoted (50 μl) into sterile Eppendorf tubes and flash frozen in liquid nitrogen for storage at -80 °C.

Transformation protocols and controls

The appropriate competent cells were allowed to thaw on ice. DNA solution was thawed on ice and 1 μl of this solution was transferred into the cell suspension, mixing with the pipette tip under sterile conditions. The cell/DNA mixture was incubated for 20 minutes on ice, then heat shocked at 40 °C for 40 seconds and returned to the ice for a minimum of 2 minutes. Non-selective LB medium was added to the solution and the cells were incubated for 1 hour 37 °C. The cells were harvested utilising a centrifugation in (Eppendorf centrifuge 5415R) at 16,100 RCF for 1 minute, the supernatant solution was discarded. The pellet was re-suspended in 100 μl of LB medium and plated onto agar plates that

contained the appropriate antibiotic. The plated cells were incubated at 37 °C overnight.

A negative control experiment was also carried out, whereby water was substituted for the DNA solution. No cells grew on these experiments.

6.4.6   DNA Manipulation

Growth and Purification of Plasmid DNA

A single colony was extracted from the ampicillin selective plate (*vide supra*) and placed into 10 ml LB medium containing ampicillin (1 µg/ml), the culture was incubated overnight at 37 °C with shaking at 150 rpm in an Innova® 43 shaker (New Brunswick Scientific, Hertfordshire, UK). The cells were then harvested *via* centrifugation (Eppendorf centrifuge 5810R) for 10 minutes at 3220 RCF, the supernatant solution was discarded. The plasmids were isolated *via* the QIAprep spin miniprep kit with the pellet initially re-suspended in buffer P1 (250 µl). Cells underwent lysis *via* the alkaline lysis buffer. Buffer P2 (250 µl) was then added, followed by buffer N3 (350 µl) the solution was mixed then cells were harvested by centrifugation (Eppendorf centrifuge 5415R) for 10 min at 16100 RCF. The supernatant solution was collected and added to a spin column where the plasmid then binds to the filter. The centrifugation step was then repeated for 1 minute at 16,100 RCF and the flow through solution was discarded. To the spin column PB buffer (0.5 ml) was added and the previous centrifugation step repeated, the flow through solution was once again discarded. Buffer PE (0.75 ml) was then added to the spin column for the washing step, the centrifugation step was repeated, the flow through solution was discarded and to ensure all the ethanol was removed the centrifugation step was repeated once more. The miniprep filter was then placed in a clean Eppendorf tube, EB buffer (50 µl) was added and this was left to stand for a minimum for 1 minute, centrifugation was then repeated, during this final step the plasmid DNA was washed from the column and collected in an Eppendorf tube. The purified DNA solution was then stored at -20 °C.

Restriction Digestion of DNA

For each restriction digest the manufacturer's guidelines were followed with regards to operating buffers, temperatures and denaturation steps. *Nco*I and *Bam*HI were used for the excision of the *PEF(S)* gene from the pBSK vector. *Nco*I and *Bam*HI were incubated at the same temperature (37 °C) for 3 hours and 4 hours respectively. The restriction endonucleases were removed by agarose gel electrophoresis.

DNA Visualisation, Isolation and Purification

The agarose gels were prepared with a w/v ratio of 1% with 1 × TAE buffer (*vide supra*). Gels were run at a constant voltage of 100 V for 60 minutes, they were then stained with 6 μM ethidium bromide for 20 minutes whilst stirring. The DNA was visualised utilising a Syngene GeneFlash UV light box (Syngene, Cambridge, UK), this ensured limited exposure to UV light. The bands that were visualised were compared to the 1 kb DNA ladder (0.5 kb, 1.0 kb, 1.5 kb, 2.0 kb, 3.0 kb, 4.0 kb, 5.0 kb, 6.0 kb, 8.0 kb and 10.0 kb), the desired bands were then identified and where appropriate they were cut from the gel with a clean scalpel blade and placed into an Eppendorf tube. In order to extract the DNA from the agarose gel the QIA gel extraction kit (QIAGEN, Crawley, UK) (Section 6.4.4) was used. Initially the gel was suspended in QG buffer the volume of which was 3 times that of the weight of the gel within the Eppendorf tube. The suspension was placed in a heat block for 10 min at 50 °C. After this a volume equal to the calculated weight of the gel of isopropanol was added to the Eppendorf tube, the solution was then placed into a spin column and centrifuged (Eppendorf centrifuge 5415R), for 1 minute at 16,100 RCF, the flow through solution was discarded. The DNA that was bound to the membrane was washed with 0.5 ml of QG buffer and the centrifuge step repeated followed addition of 0.75 ml of PE buffer. The centrifuge step was repeated twice and the DNA bound to the membrane was then inserted into a 1 ml Eppendorf tube and 50 μl of EB buffer was added and the solution left to stand for 1 minute. The sample was collected through centrifugation using the conditions mentioned above and stored at -20 °C.

Cloning of PEF(S) into expression vector pET21d

The *PEF(S)* gene in the pBSK vector was subjected to sequential *Nco1* and *BamH1* restriction endonuclease treatment (*vide supra*). A pET21d vector containing a gene encoding for wtAS (wild type Aristolochene Synthase, from lab stocks) was also subjected to a sequential *Nco*I and *Bam*HI restriction endonuclease digest. Following each digestion the restriction endonucleases, plasmid and insert were separated using agarose gel electrophoresis, the gel was then visualised and the bands indicating the expression vector and the *PEF(S)* gene were isolated and purified (*vide supra*).

Digested pET21d and the *PEF(S)* gene were ligated under two different conditions indicated in Table 6.1.

Ligation	Component (µl)				
	Plasmid	**Gene**	**T4 ligase Buffer**	**dH₂O**	**T4 Ligase**
A	1	3	1.5	9	1
B	1	6	1.5	6	1

Table 6.1: *Components used for the ligation of the PEF(S) gene and the pET21d expression vector.*

The ligation temperature was varied between 4 °C and 22 °C, with 20 seconds for each 1 °C increase and decrease, with 4 cycles in total. Super-competant *E. coli* XL1 Blue cells were then transformed with 5 µl of each ligation mixture.

DNA Sequencing

Upon isolation of the plasmid DNA from the appropriate cloning strain using the QIA spin miniprep kit (QIAGEN, Crawley, UK) (Section 6.4.4) to give pure DNA at a concentration of ~100 ng/µl. DNA sequencing of the plasmid constructs was carried out by School of Biosciences, Cardiff University, Cardiff, UK.

6.4.7 Protein production and purification

Purification of the protein was carried out utilising an ÄKTA fast protein liquid chromatography (FPLC) (GE Healthcare, Buckinghamshire, UK). The pump that was used for the buffers was an UP-920 and the buffers were mixed to achieve the desired gradient with a M-925 gradient controller. The samples were

injected on to the column with an INV-907 injection valve and the UV analysis that was carried out with a UPC-900 detector.

Small scale test-expression of the *PEF(S)* gene

A small scale test expression was carried out prior to large scale production of protein, this was to ensure that the cells grew to an optical density (OD) of 0.6 at 600 nm and to ensure that upon induction of the gene, with IPTG, a sufficient quantity of protein was produced. To ampicillin (1 μg/ml) selective LB medium (500 ml) 5 ml of overnight culture was added of BL21-CodonPlus(DE3)-RP cells containing thepET21d-*PEF(S)* construct. The culture was allowed to reach the desired OD at 37 °C. IPTG was used to induce gene expression, which continued for 16 hours in total. 1 ml samples were collected prior to induction and every hour subsequently, these samples were analysed utilising SDS-PAGE.

This process was also carried out with the BL21-CodonPlus(DE3)-RIL cells that contained the pET21d-*PEF(S)* construct.

Large scale expression of the *PEF(S)* gene

BL21-CodonPlus-(DE3)-RP competent cells were transformed with the appropriate DNA solution and incubated overnight on ampicillin selective agar plates at 37 °C. A single colony was retrieved in order to inoculate 100 ml of ampicillin selective (1 μg/ml) LB medium, which was incubated overnight at 37 °C whilst shaking at 150 rpm in an Innova® 43 shaker (New Brunswick Scientific, Hertfordshire, UK). 6 × 5 ml of the overnight pre-culture was utilised to inoculate 6 x 500 ml of ampicillin selective (1 μg/ml) LB medium. The culture was incubated at 37 °C whilst shaking as previously described until an OD of 0.6 a.u. at 600 nm was reached and IPTG (0.5 mM) was then added and the incubation was continued for a further 3 hours.

After this time period the cells were then harvested *via* centrifugation in a Sorvall RC6 Plus centrifuge (Thermo Fisher Scientific, Inc, MA, USA) using an SLA-3000 rotor at 6080 RCF for 15 minutes. The pellets were then stored at -20 °C.

Anion exchange chromatography

Frozen cell pellets were allowed to thaw on ice. The cells were then re-suspended in protein purification buffer A (Section 6.4.4) and lysed by sonication on ice (5 seconds sonication period followed by 10 seconds off, this sequence was run for 3 minutes in total), The solution was clarified through centrifugation at 5 °C for 40 minutes at 30310 RCF in a Sorvall RC6 Plus centrifuge. The supernatant solution was then removed from the debris and filtered with a 0.2 μm syringe tip filter. The supernatant solution was then applied to a diethylaminoethyl (DEAE) anion exchange column which was attached to a Fast Protein Liquid Chromatography instrument (FPLC) (ÄKTA FPLC). The DEAE column had a column volume (CV) of ~60 ml. The flow-through was collected and the column was washed with protein purification buffer A (Section 6.4.4) until the UV absorbance (280 nm) reached approximately the baseline absorbance. A two-stage NaCl gradient was applied to the column, with 0.0-0.5 M NaCl (50 % protein purification buffer B, section 6.4.4) over 4 CV, followed by 0.5-1.0 M NaCl (50-100% protein purification buffer B, Section 6.4.4) over 1 CV. The column was washed with protein purification buffer B (Section 6.4.4) for 1 CV and then stored under high (1 M) salt conditions at 4 °C. The flow rate was set to 10 ml/min throughout the purification process. The absorbance was monitored at 280 nm and fractions (6 ml) were collected from 0-50% buffer B. SDS-PAGE was carried out on each of the fractions that demonstrated a high absorbance from the UV monitoring.

The fractions containing the desired protein as indicated by the SDS-PAGE analysis were pooled in preparation for purification with size exclusion chromatography (*vide infra*)

Size exclusion chromatography

The eluent from the DEAE column which contained PEF(S) was concentrated *via* Amicon to around 15 ml. The concentrated sample was loaded onto a SuperDex 75 column (320 ml column volume) which had been equilibrated with protein purification buffer C (Section 6.4.4). The column was eluted with 1.5 CV (480 ml) of protein purification buffer C (at 2.5 ml/min) and the

absorbance was monitored at 280 nm. 10 ml fractions were collected and SDS-PAGE analysis was used to determine the point of elution of the desired protein.

The fractions containing the desired protein were then pooled and dialysed against 3 changes of 4 l of dialysis buffer 2 (Section 6.4.4) at 4 °C.

Analytical size exclusion chromatography

For the calibration of the SuperDex 75 column (26 ml column volume) three proteins were used; bovine serum albumin (MW 66,000), trypsin inhibitor from soya beans (MW 20,090) and RNAase A (MW 13,700). A 0.5 ml sample containing each protein at a concentration of 2.5 mg/ml was loaded onto the column which had been equilibrated with protein purification buffer C. The column was eluted with 1.5 CV (39 ml) of protein purification buffer C (at 0.5 ml/min) and the absorbance was monitored at 280 nm.

A 0.5 ml sample was taken from the eluted fractions of PEF(S) from the preparative SuperDex 75 (320 ml) and was loaded on to the analytical SuperDex 75 (26 ml), which had been equilibrated with protein purification buffer C. The column was eluted with 1.5 CV (39 ml) of protein purification buffer C (at 0.5 ml/min) and the absorbance was monitored at 280 nm.

Protein storage

The protein was stored in a buffer solution with a final concentration of 5 mM K_iPO_4, 2 mM EDTA and 5 mM βME at pH 7.0. The solution was stored at 4 °C.

Protein concentration determination

Potassium phosphate (100 mM, pH 7.0) was filtered and degassed using a vacuum pump. The Beer-Lambert law was used to calculate the concentration (mg/ml) of the protein solution (Equation 6.1).

$$A = \varepsilon \times c \times l \text{, and therefore } c = \frac{A \times df}{\varepsilon \times l}$$

Equation 6.1: *The equation for measuring the concentration of protein within solution through UV-Vis analysis. Where A = absorbance at a given wavelength, ε = extinction coefficient, c = the concentration, l = the path length of the cuvette (cm) and df is the dilution factor.*

Average extinction coefficients for proteins at the wavelengths of 210, 215 and 220nm are 20, 15 and 11.[217] These were utilised to determine the concentration

of the protein. The measurements were obtained using a Jasco V-660 UV-Vis spectrophotometer (Jasco UK, Essex, UK), with the temperature controlled with a Peltier cooler.

The concentrations were converted into micromolar concentrations, this was done using the concentration in mg/ml that was obtained from the equation was divided by the molecular weight of the protein (19,993 for pPEF(S)), multiplied by 10^6.

SDS-PAGE protocol

The gel, composed of two parts, the resolving and stacking solutions were prepared as described in Table 6.2.

Component	Resolving solution (ml)	Stacking solution (ml)
H_2O	3.4	5.7
Gel buffer	2.5	2.5
30% Acrylamide solution (with bis-acrylamide)	4.0	1.7
10% SDS solution	0.1	0.1

Table 6.2: *Components in the resolving and stacking solutions for SDS-PAGE gel preparation.*

Prior to pouring the resolving solution freshly prepared ammonium persulfate solution (APS, 100 μl) (10% *w/v* in dH$_2$O) and 15 μl of *N,N,N',N'*-tetramethylenediamine (TEMED) were added to the resolving solution, and for the stacking solution 100 μl of 10% APS and 20 μl of TEMED. Solutions were stirred in order to initiate polymerisation. When the resolving solution was poured it was covered with a layer of isopropanol and allowed to polymerise. Once the resolving solution had polymerised, the isopropanol was then removed with blotting paper, the stacking gel was poured on top and a comb was inserted to create the wells for sample loading.

6.4.8 MALDI-TOF mass spectrometry

The standard matrix solution used as a carrier for the protein of interest was a 10 mg/ml solution of sinapinic acid in a 1:3 acetonitrile to water (0.1% trifluoroacetic acid). 1.75 μl of the carrier solution was mixed with 1.75 μl of

protein solution. Various concentrations of the protein solution were set-up from the concentration that was determined from the UV-Vis measurements (*vide supra*), these ranged from 5-100 µM.

A MALDI-TOF Micro MX spectrophotometer (Waters, Manchester, UK) was used in linear mode, positive polarity and the mass range was 5,000-30,000.

6.4.9   Circular Dichroism spectroscopy

All circular dichroism spectroscopy was out on the Chirascan™ apparatus (Applied Photophysics Limited, UK).

CD spectroscopy Buffers

A 10 mM K_iPO_4 buffer at pH 7.0 was used for the denaturation experiments, the buffer was sterilised using a 0.2 µm syringe and it was then degassed under reduced pressure. A blank measurement was taken of the buffer alone prior to the experiment.

Calculation of the Mean Residue Ellipticity (MRE)

In order to calculate the MRE from the signal obtained from the CD experiments Equation 6.2 was used.

$$\theta_{MRE} = \frac{\theta}{10 \times n \times c \times l}$$

Equation 6.2: The equation for calculating the mean residue epilicity. Where: Θ = CD signal (millidegrees), n = number of back bone peptide bonds (i.e. number of amino acid residues minus 1), c = molecular concentration and l = the path length of the cuvette used (cm).

The cuvette that was used had a path length of 0.1 cm and a total volume of 350 µl. The CD spectra of PEF(S) were collected from 190-350 nm. CD spectra were collected at 20 °C and the total concentration of PEF(S) was ~ 20 µm.

6.4.10 Analysis of protein-inhibitor affinity by fluorsecence spectroscopy

The fluorescence spectra were obtained with the use of a Perkin-Elmer LS-22 Luminescence Spectrometer (Perkin-Elmer, Buckinghamshire, UK), which was attached to a Julabo F25 water bath (Julabo, Seelbach, Germany).

3 ml samples containing; 20 µM TNS (6-(*p*-toluidino)-2-naphthanlenesulfonic acid) in 20 mM Tris.HCl, 0.1 mM EDTA and 20 µM PEF(S) were incubated for 5 minutes in a fluorescence cuvette at 20 °C (pH 7.4). The incubations took place both with and without 1.1 mM CaCl$_2$. The fluorescence spectra were obtained with an excitation wavelength of 340 nm and the data collected over a range of 350 nm to 600 nm, with the slit widths of 15 nm and 5 nm for excitation and the emission slits respectively. 5 µl of a 50 mM solution of (*Z*)-3-(-4-bromophenyl)-2-mercaptoacrylic acid was added to the solution, which gave an overall concentration of 83 µM. The sample was analysed in the fluorimeter with the same conditions mentioned previously in order to detect a change.

6.5 Protein crystallography

All consumables for the purpose of protein crystallography were purchased from Molecular Dimensions Ltd (Suffolk, UK).

6.5.1 Screening solutions

General screen solutions

All general screens were purchased from Molecular Dimensions Ltd (Suffolk, UK), these screens were used to obtain the optimum crystallisation conditions for PEF(S). The 3 screens that were used were; JCSG-*plus*™ HT-96, PACT *premier*™ HT-96 and *ProPlex* HT-96.[223, 224, 226]

Crystallisation precipitant 1

HEPES (1.19 g), 1.4 M sodium citrate (20.59 g) with 20 mM CaCl$_2$ (0.11 g) was dissolved in 50 ml of deionised water resulting in concentrations of 0.1 M HEPES, 1.4 M sodium citrate and 20 mM CaCl$_2$. The pH of the solution was adjusted to 7.0 with 1 M NaOH. The solution was filtered with a syringe tip filter (0.2 µm) and stored at 20 °C.

Crystallisation precipitant 2

Sodium cacodylate (0.54 g), PEG6000 (6.25 g) and CaCl$_2$ (0.11 g) were dissolved in 50 ml of deionised water, resulting in final concentrations of 50 mM sodium cacodylate, 12.5% (w/v)PEG6000 and 20 mM CaCl$_2$. The pH of the solution was

adjusted to 7.4 with 5 M NaOH. The solution was filtered with a 0.22 μm syringe tip filter and stored at 20 °C.

6.5.2   Protein preparation

Protein solution 1

The protein was purified and dialysed into dialysis buffer 1 (Section 6.4.4), then concentrated in an Amicon with a MW cut-off of 10,000, until the concentration became at least 5 mg/ml.

Protein solution 2

Sodium cacodylate (1.07 g), EDTA (0.65 g), β-ME (350 μl) and $CaCl_2$ (0.11 g)were dissolved in 1 l of deionised water, resulting in the final concentrations of 5 mM sodium cacodylate, 2 mM EDTA, 5 mM β-ME and 1 mM $CaCl_2$. The pH of the solution was adjusted to 7.0 with 5 M NaOH. The solution was then stored at 20 °C.

PEF(S) was purified (Section 6.4.7) and concentrated using a protein spin column with a molecular weight cut-off of 10,000. The protein was then washed (4 times) with the protein solution 2 (*vide supra*), to ensure that the buffer was fully exchanged. The protein was concentrated to ~10 mg/ml.

6.5.3   Crystal plate set-up

General screen setup

An Art-Robbins Phoenix dispensing robot (Alpha Biotech, Glasgow, UK) was used in the process of setting up the screens. The robot was initially cleaned with 0.5 M NaOH which is followed by a washing step with deionised water. A 40 μl aliquot of protein sample was placed into a 200 μl Eppendorf tube, the solution was centrifuged at 3381 RCF for 30 seconds. The protein sample was then placed in the robot, along with the general screen solutions (Section 6.5.1) and a 96 well Intelliwell II crystallisation plate. 60 μl volumes of the broad screen solutions were placed into each deep well of the crystallisation plate and 0.2 μl of the broad screen solution was placed into the sitting drop depression. The protein solution was dispensed in 0.2 μl aliquots into each of the sitting drop wells on the crystallisation plate. The plate was then removed from the

robot and sealed with Easy Seal DWB sheets (Molecular Dimensions Ltd, Suffolk, UK).

Before the crystal plate was stored in an incubator at 20 °C each drop photographed with a microscope at a magnification of 3.0 (Leica Microsystems, Buckinghamshire UK) equipped with a camera (Leica Microsystems, Buckinghamshire UK). The plates were scanned 24 hrs after set-up and every 48 hrs subsequently to that until crystals formed.

Crystallisation plate set-up 1

Using a 48 well sitting drop plate (Molecular Dimensions Ltd, Suffolk, UK), 300 µl of crystallisation precipitant 1 (Section 6.5.1) was manually added to the deep well and 1 µl of crystallisation precipitant 1 (Section 6.5.1) in the sitting drop, in each well of the 48 well crystallisation plate. To each drop, 1 µl of protein solution 1 (Section 6.5.2) was added. The plate was then sealed with Easy Seal DWB sheets (Molecular Dimensions Ltd, Suffolk, UK). The plates were viewed under a microscope (Zeiss Vert.A1. AX10 (Cambridge, UK)) at 5 × magnification straight after set-up. The plates were stored at 20 °C and viewed 24 hrs after set-up and every 48 hrs subsequently, until crystals form.

Crystallisation plate set-up 2

The EasyXtal 24 well plates (Qiagen, Crawley, UK) were set-up using the hanging drop vapour diffusion method. 1 ml of crystallisation precipitant 2 (Section 6.5.1) was placed in the well. The drops contained 3 µl of the crystallisation precipitant 2 (Section 6.5.1), 1 µl of deionised H_2O and 3 µl of protein solution 2 (Section 6.5.2) onto the screw cap lid. The lid of the wells were sealed and the plate stored at 20 °C until crystals formed. The plates were viewed under a microscope (Zeiss Vert.A1. AX10 (Cambridge, UK)) at 5 × magnification straight after set-up (images were collected with a Ziess Axio Cam MRm camera (Cambridge, UK)). The plates were stored at 20 °C and viewed 24 hours after set-up and every 48 hours subsequently, until crystals form.The crystals normally took 3 days to appear within the hanging drop.

6.5.4   Soak set-up

Soaks of the crystals formed in crystallisation precipitant 1

Concentration of the ligand/mM	Volume /µl	Concentration of the protein/mM	Ratio /Ligand:PEF(S)
0.5	0.5	0.25	2:1
0.75	0.75	0.25	3:1
1.0	1	0.25	4:1

Table 6.3: The concentrations of the inhibitor used in the soaking expreriments of the inhibitors into the crystallised PEF(S), using a stock solution of 2 mM of the ligands.

50 mM stock solutions (1 ml) of the mercaptoacrylic acid derivatives (Section 6.1.5) in DMSO were used to make the solutions for soaking the protein. The stock solution was diluted with crystallisation precipitant 1 to 2 mM. The screw caps were removed and the diluted stock solutions were added to the sitting drop containing the crystals in the concentrations listed in Table 6.3. The soaks were carried out both with and without 10 mM DTT.

Soaks of the crystals formed in crystallisation precipitant 2

With crystallisation precipitant 2 the stock solutions were diluted to 8 mM. The screw caps were then removed from the wells containing crystals and 1 µl of the solution was added to the 7 µl drop containing the crystals, taking the total concentration of the inhibitor to 1 mM. Therefore the ratio of PEF(S) to the ligand was 1:2, as the concentration of the protein in the drop solution was ~200 µM (~4 mg/ml). The wells were re-sealed and the plates were stored at 20 °C for 16 hours until the crystals were collected.

6.5.5   Crystal harvesting and data collection

Crystal collection

The cover was removed from each well where crystals were found, 100 to 300 µm loops were used to harvest the crystals. Each crystal collected by the loop was then flash frozen in liquid nitrogen and placed into a position in the transport puck. The puck contains spaces for 16 samples and was kept in liquid nitrogen until the samples reached the synchrotron (Diamond Light Source, Oxford, UK).

Data collection and refinement

Data was collected at Diamond Light Source (Oxfordshire, UK) at 100 K; all data were collected at 0.976 Å with a Pilatus pixelated detector. The raw diffraction images were processed through the xia2 data-reduction system. The data were scaled, reduced and analysed using the Scala program from the CCP4i package (Collaborative Computational Project number 4).[227]

Molecular replacement was done with Phaser (CCP4i) with a model derived from the structure of PEF(S)(PDB: 1NX2 and 1ALV).[81, 186, 228] The model obtained was adjusted with the COOT program (Crystallographic Object-Oreinted Toolkit) for molecular model building, and completion, the model was refined further with the Refmac5 refinement program.[229, 230]

7 References

1.   M. Feldmann, F. M. Brennan and R. N. Maini, *Cell*, 1996, **85**, 307-310.

2.   J. P. Pelletier, J. Martel-Pelletier and S. B. Abramson, *Arthritis Rheum.*, 2001, **44**, 1237-1247.

3.   A. Davidson and B. Diamond, *N. Engl. J. Med.*, 2001, **345**, 340-350.

4.   D. L. Scott, F. Wolfe and T. W. J. Huizinga, *Lancet*, 2010, **376**, 1094-1108.

5.   F. Wolfe, D. M. Mitchell, J. T. Sibley, J. F. Fries, D. A. Bloch, C. A. Williams, P. W. Spitz, M. Haga, S. M. Kleinheksel and M. A. Cathey, *Arthritis Rheum.*, 1994, **37**, 481-494.

6.   D. M. Lee and M. E. Weinblatt, *Lancet*, 2001, **358**, 903-911.

7.   K. Devlin, in *The Daily Telegraph*, United Kingdom, 31/03/2010.

8.   W. P. Arend, *Arthritis Rheum.*, 1997, **40**, 595-597.

9.   N. Hogg, D. G. Palmer and P. A. Revell, *Immunology*, 1985, **56**, 673-681.

10.   G. S. Firestein, G. S. Panayi and F. A. Wollheim, *Rheumatoid Arthritis*, 2nd edn., Oxford University Press, United States, 2006.

11.   C. Gerard and B. J. Rollins, *Nat. Immunol.*, 2001, **2**, 108-115.

12.   E. D. Harris, *N. Engl. J. Med.*, 1990, **322**, 1277-1289.

13.   S. Shiozawa, K. Shiozawa and T. Fujita, *Arthritis Rheum.*, 1983, **26**, 472-478.

14.   E. M. Gravallese, J. M. Darling, A. L. Ladd, J. N. Katz and L. H. Glimcher, *Arthritis Rheum.*, 1991, **34**, 1076-1084.

15.   S. S. McCachren, B. F. Haynes and J. E. Niedel, *J. Clin. Immunol.*, 1990, **10**, 19-27.

16.   S. Lefevre, A. Knedla, C. Tennie, A. Kampmann, C. Wunrau, R. Dinser, A. Korb, E. M. Schnaker, I. H. Tarner, P. D. Robbins, C. H. Evans, H. Sturz, J. Steinmeyer, S. Gay, J. Scholmerich, T. Pap, U. Muller-Ladner and E. Neumann, *Nat. Med.*, 2009, **15**, 1414-1420.

17.   S. W. Edwards and M. B. Hallett, *Immunol. Today*, 1997, **18**, 320-324.

18.   A. Cross, T. Barnes, R. C. Bucknall, S. W. Edwards and R. J. Moots, *J. Leukoc. Biol.*, 2006, **80**, 521-528.

19.   H. L. Nurcombe, R. C. Bucknall and S. W. Edwards, *Ann. Rheum. Dis.*, 1991, **50**, 147-153.

20.   W. W. Chatham, R. Swaim, H. Frohsin, Jr., L. W. Heck, E. J. Miller and W. D. Blackburn, Jr., *Arthritis Rheum.*, 1993, **36**, 51-58.

21.	A. W. Segal, in *Annu. Rev. Immunol.*, 2005, vol. 23, pp. 197-223.

22.	S. W. Edwards, *Biochemistry and physiology of the neutrophil*, 1st edn., Cambridge University Press, UK, 1994.

23.	M. B. Hallett and S. Dewitt, *Trends Cell Biol.*, 2007, **17**, 209-214.

24.	W. L. Lee, R. E. Harrison and S. Grinstein, *Microbes Infect.*, 2003, **5**, 1299-1306.

25.	R. Appelberg, *Trends Microbiol.*, 2007, **15**, 87-92.

26.	B. J. Rollins, *Blood*, 1997, **90**, 909-928.

27.	T. A. Springer, *Cell*, 1994, **76**, 301-314.

28.	S. L. Erlandsen, S. R. Hasslen and R. D. Nelson, *J. Histochem. Cytochem.*, 1993, **41**, 327-333.

29.	A. Ivetic, J. Deka, A. Ridley and A. Ager, *J. Biol. Chem.*, 2002, **277**, 2321-2329.

30.	S. Dewitt and M. B. Hallett, *J. Cell Biol.*, 2002, **159**, 181-189.

31.	E. V. Davies and M. B. Hallett, *Biochem. Biophys. Res. Commun.*, 1998, **248**, 679-683.

32.	A. J. Wiemer, M. A. Lokuta, J. C. Surfus, S. A. Wernimont and A. Huttenlocher, *Mol. Immunol.*, 2010, **47**, 894-902.

33.	M. Hayashi, H. Suzuki, S. Kawashima, T. C. Saido and M. Inomata, *Arch. Biochem. Biophys.*, 1999, **371**, 133-141.

34.	X. B. Yao, A. Thibodeau and J. G. Forte, *Am. J. Physiol.*, 1993, **265**, C36-C46.

35.	P. N. Devreotes and S. H. Zigmond, *Annu. Rev. Cell Biol.*, 1988, **4**, 649-686.

36.	D. A. Lauffenburger and A. F. Horwitz, *Cell*, 1996, **84**, 359-369.

37.	M. A. Lawson and F. R. Maxfield, *Nature*, 1995, **377**, 75-79.

38.	P. J. Berti and A. C. Storer, *J. Mol. Biol.*, 1995, **246**, 273-283.

39.	T. Murachi, K. Tanaka, M. Hatanaka and T. Murakami, *Adv. Enzyme Regul.*, 1980, **19**, 407-424.

40.	H. Sorimachi, S. Hata and Y. Ono, *J. Biochem.*, 2011, **150**, 23-37.

41.	Y. Ono and H. Sorimachi, *BBA-Proteins Proteom.*, 2012, **1824**, 224-236.

42.	P. Johnson, *Int. J. Biochem.*, 1990, **22**, 811-822.

43.   H. Sorimachi, S. Imajohohmi, Y. Emori, H. Kawasaki, S. Ohno, Y. Minami and K. Suzuki, *J. Biol. Chem.*, 1989, **264**, 20106-20111.

44.   H. Sorimachi, S. Hata and Y. Ono, *P. Jpn. Acad. B-Phys.*, 2011, **87**, 287-327.

45.   A. Fersht, *Structure and Mechanism in Protein Science*, 1st edn., W. H. Freeman and Company, 1999.

46.   L. Polgar, *Mechanism of Protease Action*, 1st edn., CRC Press, United States, 1989.

47.   K. Aoki, S. Imajoh, S. Ohno, Y. Emori, M. Koike, G. Kosaki and K. Suzuki, *Febs Lett.*, 1986, **205**, 313-317.

48.   S. Imajoh, K. Aoki, S. Ohno, Y. Emori, H. Kawasaki, H. Sugihara and K. Suzuki, *Biochemistry-US*, 1988, **27**, 8122-8128.

49.   S. Ohno, Y. Emori and K. Suzuki, *Nucleic Acids Res*, 1986, **14**, 5559.

50.   D. E. Goll, V. F. Thompson, H. Li, W. Wei and J. Cong, *Physiol. Rev.*, 2003, **83**, 731-801.

51.   R. L. Campbell and P. L. Davies, *Biochem. J.*, 2012, **447**, 335-351.

52.   E. A. Nalefski and J. J. Falke, *Protein Sci.*, 1996, **5**, 2375-2390.

53.   S. Strobl, C. Fernandez-Catalan, M. Braun, R. Huber, H. Masumoto, K. Nakagawa, A. Irie, H. Sorimachi, G. Bourenkow, H. Bartunik, K. Suzuki and W. Bode, *Proc. Natl. Acad. Sci. USA*, 2000, **97**, 588-592.

54.   W. Delano, www.pymol.org, 2006.

55.   E. Melloni, M. Averna, F. Salamino, B. Sparatore, R. Minafra and S. Pontremoli, *J. Biol. Chem.*, 2000, **275**, 82-86.

56.   A. Wendt, V. F. Thompson and D. E. Goll, *Biol. Chem.*, 2004, **385**, 465-472.

57.   G. P. Pal, T. De Veyra, J. S. Elce and Z. C. Jia, *Structure*, 2003, **11**, 1521-1526.

58.   D. Reverter, S. Strobl, C. Fernandez-Catalan, H. Sorimachi, K. Suzuki and W. Bode, *Biol. Chem.*, 2001, **382**, 753-766.

59.   P. Dutt, C. N. Spriggs, P. L. Davies, Z. Jia and J. S. Elce, *Biochem. J.*, 2002, **367**, 263-269.

60.   C. M. Hosfield, J. S. Elce, P. L. Davies and Z. C. Jia, *Embo J.*, 1999, **18**, 6880-6889.

61.   S. Imajoh, H. Kawasaki and K. Suzuki, *J. Biochem.*, 1986, **100**, 633-642.

62. K. Nakagawa, H. Masumoto, H. Sorimachi and K. Suzuki, *J. Biochem.*, 2001, **130**, 605-611.

63. J. S. Arthur, S. Gauthier and J. S. Elce, *Febs Lett.*, 1995, **368**, 397-400.

64. A. Fernandez-Montalvan, I. Assfalg-Machleidt, D. Pfeiler, H. Fritz, M. Jochum and W. Machleidt, *Biochem. J.*, 2004, **382**, 607-617.

65. T. Moldoveanu, C. M. Hosfield, D. Lim, J. S. Elce, Z. C. Jia and P. L. Davies, *Cell*, 2002, **108**, 649-660.

66. J. Rizo and T. C. Sudhof, *J. Biol. Chem.*, 1998, **273**, 15879-15882.

67. P. Tompa, Y. Emori, H. Sorimachi, K. Suzuki and P. Friedrich, *Biochem. Biophys. Res. Commun.*, 2001, **280**, 1333-1339.

68. S. Gil-Parrado, O. Popp, T. A. Knoch, S. Zahler, F. Bestvater, M. Felgentrager, A. Holloschi, A. Fernandez-Montalvan, E. A. Auerswald, H. Fritz, P. Fuentes-Prior, W. Machleidt and E. Spiess, *J. Biol. Chem.*, 2003, **278**, 16336-16346.

69. R. A. Hanna, R. L. Campbell and P. L. Davies, *Nature*, 2008, **456**, 409-412.

70. C. M. Hosfield, T. Moldoveanu, P. L. Davies, J. S. Elce and Z. C. Jia, *J. Biol. Chem.*, 2001, **276**, 7404-7407.

71. Kretsing.Rh and C. E. Nockolds, *J. Biol. Chem.*, 1973, **248**, 3313-3326.

72. A. Lewit-Bentley and S. Rety, *Curr. Opin. Struct. Biol.*, 2000, **10**, 637-643.

73. J. L. Gifford, M. P. Walsh and H. J. Vogel, *Biochem. J.*, 2007, **405**, 199-221.

74. H. Blanchard, P. Grochulski, Y. Li, J. S. C. Arthur, P. L. Davies, J. S. Elce and M. Cygler, *Nat. Struct. Biol.*, 1997, **4**, 532-538.

75. O. Herzberg and M. N. G. James, *Biochemistry-US*, 1985, **24**, 5298-5302.

76. R. H. Kretsinger, *Nat. Struct. Biol.*, 1997, **4**, 514-516.

77. P. Dutt, J. S. Arthur, P. Grochulski, M. Cygler and J. S. Elce, *Biochem. J.*, 2000, **348**, 37-43.

78. G. P. Pal, J. S. Elce and Z. Jia, *J. Biol. Chem.*, 2001, **276**, 47233-47238.

79. S. Imajoh, H. Kawasaki and K. Suzuki, *J. Biochem.*, 1987, **101**, 447-452.

80. H. Blanchard, Y. G. Li, M. Cygler, C. M. Kay, J. S. C. Arthur, P. L. Davies and J. S. Elce, *Protein Sci.*, 1996, **5**, 535-537.

81. G. D. Lin, D. Chattopadhyay, M. Maki, K. K. W. Wang, M. Carson, L. Jin, P. Yuen, E. Takano, M. Hatanaka, L. J. DeLucas and S. V. L. Narayana, *Nat. Struct. Biol.*, 1997, **4**, 539-547.

82. W. Zhang and R. L. Mellgren, *Biochem. Biophys. Res. Commun.*, 1996, **227**, 891-896.

83. K. Brandenburg, F. Harris, S. Dennison, U. Seydel and D. A. Phoenix, *Eur. J. Biochem.*, 2002, **269**, 5414-5422.

84. S. R. Dennison, S. Dante, T. Hauss, K. Brandenburg, F. Harris and D. A. Phoenix, *Biophys. J.*, 2005, **88**, 3008-3017.

85. Y. Emori, H. Kawasaki, S. Imajoh, K. Imahori and K. Suzuki, *Proc. Natl. Acad. Sci. USA*, 1987, **84**, 3590-3594.

86. M. Maki, E. Takano, H. Mori, A. Sato, T. Murachi and M. Hatanaka, *Febs Lett.*, 1987, **223**, 174-180.

87. R. A. Hanna, B. E. Garcia-Diaz and P. L. Davies, *Febs Lett.*, 2007, **581**, 2894-2898.

88. R. Betts, S. Weinsheimer, G. E. Blouse and J. Anagli, *J. Biol. Chem.*, 2003, **278**, 7800-7809.

89. P. Tompa, Z. Mucsi, G. Orosz and P. Friedrich, *J. Biol. Chem.*, 2002, **277**, 9022-9026.

90. A. Alexa, Z. Bozoky, A. Farkas, P. Tompa and P. Friedrich, *J. Biol. Chem.*, 2004, **279**, 20118-20126.

91. C. M. Hosfield, J. S. Elce and Z. Jia, *J. Mol. Biol.*, 2004, **343**, 1049-1053.

92. T. Moldoveanu, Z. C. Jia and P. L. Davies, *J. Biol. Chem.*, 2004, **279**, 6106-6114.

93. S. Hata, H. Sorimachi, K. Nakagawa, T. Maeda, K. Abe and K. Suzuki, *Febs Lett.*, 2001, **501**, 111-114.

94. T. Moldoveanu, C. M. Hosfield, D. Lim, Z. C. Jia and P. L. Davies, *Nat. Struct. Biol.*, 2003, **10**, 371-378.

95. A. Khorchid and M. Ikura, *Nat. Struct. Biol.*, 2002, **9**, 239-241.

96. T. C. Saido, M. Shibata, T. Takenawa, H. Murofushi and K. Suzuki, *J. Biol. Chem.*, 1992, **267**, 24585-24590.

97. J. S. C. Arthur and C. Crawford, *BBA-Protein Struct. M.*, 1996, **1293**, 201-206.

98. E. Melloni, M. Michetti, F. Salamino and S. Pontremoli, *J. Biol. Chem.*, 1998, **273**, 12827-12831.

99. S. Pontremoli, E. Melloni, M. Michetti, F. Salamino, B. Sparatore and B. L. Horecker, *Proc. Natl. Acad. Sci. USA*, 1988, **85**, 1740-1743.

100. A. Glading, R. J. Bodnar, I. J. Reynolds, H. Shiraha, L. Satish, D. A. Potter, H. C. Blair and A. Wells, *Mol. Cell. Biol.*, 2004, **24**, 2499-2512.

101. R. L. Mellgren, *Faseb J.*, 1987, **1**, 110-115.

102. M. Molinari and E. Carafoli, *J. Membr. Biol.*, 1997, **156**, 1-8.

103. E. Melloni, M. Michetti, F. Salamino, R. Minafra and S. Pontremoli, *Biochem. Biophys. Res. Commun.*, 1996, **229**, 193-197.

104. J. S. Elce, C. Hegadorn, J. Simon and C. Arthur, *J. Biol. Chem.*, 1997, **272**, 11268-11275.

105. A. Baki, P. Tompa, A. Alexa, O. Molnar and P. Friedrich, *Biochem. J.*, 1996, **318**, 897-901.

106. J. Y. Cong, D. E. Goll, A. M. Peterson and H. P. Kapprell, *J. Biol. Chem.*, 1989, **264**, 10096-10103.

107. A. Farkas, G. Nardai, P. Csermely, P. Tompa and P. Friedrich, *Biochem. J.*, 2004, **383**, 165-170.

108. G. V. W. Johnson and R. P. Guttmann, *Bioessays*, 1997, **19**, 1011-1018.

109. E. Carafoli and M. Molinari, *Biochem. Biophys. Res. Commun.*, 1998, **247**, 193-203.

110. K. K. W. Wang, A. Villalobo and B. D. Roufogalis, *Biochem. J.*, 1989, **262**, 693-706.

111. G. Y. Huh, S. B. Glantz, S. J. Je, J. S. Morrow and J. H. Kim, *Neurosci. Lett.*, 2001, **316**, 41-44.

112. S. J. Franco, M. A. Rodgers, B. J. Perrin, J. W. Han, D. A. Bennin, D. R. Critchley and A. Huttenlocher, *Nat. Cell Biol.*, 2004, **6**, 977-983.

113. A. Shcherbina, A. Bretscher, D. M. Kenney and E. Remold-O'Donnell, *Febs Lett.*, 1999, **443**, 31-36.

114. K. S. Au, *Biochim. Biophys. Acta.*, 1987, **905**, 273-278.

115. E. A. Tallant, L. M. Brumley and R. W. Wallace, *Biochemistry-US*, 1988, **27**, 2205-2211.

116. T. Nakagawa and J. Y. Yuan, *J. Cell Biol.*, 2000, **150**, 887-894.

117. P. Tompa, P. Buzder-Lantos, A. Tantos, A. Farkas, A. Szilagyi, Z. Banoczi, F. Hudecz and P. Friedrich, *J. Biol. Chem.*, 2004, **279**, 20775-20785.

118. D. Cuerrier, T. Moldoveanu and P. L. Davies, *J. Biol. Chem.*, 2005, **280**, 40632-40641.

119. H. Sorimachi, H. Mamitsuka and Y. Ono, *Biol. Chem.*, 2012, **393**, 853-871.

120. T. Moldoveanu, R. L. Campbell, D. Cuerrier and P. L. Davies, *J. Mol. Biol.*, 2004, **343**, 1313-1326.

121. K. Sato and S. Kawashima, *Biol. Chem.*, 2001, **382**, 743-751.

122. L. Santella, K. Kyozuka, L. De Riso and E. Carafoli, *Cell Calcium*, 1998, **23**, 123-130.

123. V. Randriamboavonjy and I. Fleming, *Vasc. Pharmacol.*, 2012, **56**, 210-215.

124. A. Wieschhaus, A. Khan, A. Zaidi, H. Rogalin, T. Hanada, F. Liu, L. De Franceschi, C. Brugnara, A. Rivera and A. H. Chishti, *Biochem. J.*, 2012, **448**, 141-152.

125. M. A. Smith and R. G. Schnellmann, *Cardiovasc. Res.*, 2012, **96**, 32-37.

126. S. J. Franco and A. Huttenlocher, *J. Cell Sci.*, 2005, **118**, 3829-3838.

127. J. E. Schollmeyer, *Science*, 1988, **240**, 911-913.

128. W. L. Zhang, Q. Lu, Z. J. Xie and R. L. Mellgren, *Oncogene*, 1997, **14**, 255-263.

129. R. Flaumenhaft, *Arterioscler. Thromb. Vasc. Biol.*, 2003, **23**, 1152-1160.

130. S. O. Sage, *Exp. Physiol.*, 1997, **82**, 807-823.

131. J. E. B. Fox, C. C. Reynolds, J. S. Morrow and D. R. Phillips, *Blood*, 1987, **69**, 537-545.

132. K. Croce, R. Flaumenhaft, M. Rivers, B. Furie, B. C. Furie, I. M. Herman and D. A. Potter, *J. Biol. Chem.*, 1999, **274**, 36321-36327.

133. M. Azam, S. S. Andrabi, K. E. Sahr, L. Kamath, A. Kuliopulos and A. H. Chishti, *Mol. Cell. Biol.*, 2001, **21**, 2213-2220.

134. A. H. Schmaier, H. N. Bradford, D. Lundberg, A. Farber and R. W. Colman, *Blood*, 1990, **75**, 1273-1281.

135. D. E. Croall and G. N. DeMartino, *Physiol. Rev.*, 1991, **71**, 813-847.

136. J. S. Elce, L. Sigmund and M. J. Fox, *Biochem. J.*, 1989, **261**, 1039-1042.

137. S. M. Harwood, M. M. Yaqoob and D. A. Allen, *Ann. Clin. Biochem.*, 2005, **42**, 415-431.

138. S. Orrenius, B. Zhivotovsky and P. Nicotera, *Nat. Rev. Mol. Cell. Biol.*, 2003, **4**, 552-565.

139. M. Chen, H. P. He, S. X. Zhan, S. Krajewski, J. C. Reed and R. A. Gottlieb, *J. Biol. Chem.*, 2001, **276**, 30724-30728.

140. G. D. Cao, J. Xing, X. Xiao, A. K. F. Liou, Y. Q. Gao, X. M. Yin, R. S. B. Clark, S. H. Graham and J. Chen, *J. Neurosci.*, 2007, **27**, 9278-9293.

141. P. Kar, K. Samanta, S. Shaikh, A. Chowdhury, T. Chakraborti and S. Chakraborti, *Arch. Biochem. Biophys.*, 2010, **495**, 1-7.

142. B. M. Polster, G. Basanez, A. Etxebarria, J. M. Hardwick and D. G. Nicholls, *J. Biol. Chem.*, 2005, **280**, 6447-6454.

143. A. Joshi, V. Bondada and J. W. Geddes, *Exp. Neurol.*, 2009, **218**, 221-227.

144. B. J. Perrin and A. Huttenlocher, *Int. J. Biochem. Cell Biol.*, 2002, **34**, 722-725.

145. A. Glading, D. A. Lauffenburger and A. Wells, *Trends Cell Biol.*, 2002, **12**, 46-54.

146. S. Miyamoto, H. Teramoto, O. A. Coso, J. S. Gutkind, P. D. Burbelo, S. K. Akiyama and K. M. Yamada, *J. Cell Biol.*, 1995, **131**, 791-805.

147. M. Pfaff, X. P. Du and M. H. Ginsberg, *Febs Lett.*, 1999, **460**, 17-22.

148. X. Du, T. C. Saido, S. Tsubuki, F. E. Indig, M. J. Williams and M. H. Ginsberg, *J. Biol. Chem.*, 1995, **270**, 26146-26151.

149. M. T. Rock, A. R. Dix, W. H. Brooks and T. L. Roszman, *Exp. Cell Res.*, 2000, **261**, 260-270.

150. D. A. Potter, J. S. Tirnauer, R. Janssen, D. E. Croall, C. N. Hughes, K. A. Fiacco, J. W. Mier, M. Maki and I. M. Herman, *J. Cell Biol.*, 1998, **141**, 647-662.

151. N. Dourdin, A. K. Bhatt, P. Dutt, P. A. Greer, J. S. C. Arthur, J. S. Elce and A. Huttenlocher, *J. Biol. Chem.*, 2001, **276**, 48382-48388.

152. K. I. Saito, J. S. Elce, J. E. Hamos and R. A. Nixon, *Proc. Natl. Acad. Sci. USA*, 1993, **90**, 2628-2632.

153. F. Raynaud and A. Marcilhac, *Febs J.*, 2006, **273**, 3437-3443.

154. A. B. Tontchev and T. Yamashima, *Neuropathology*, 1999, **19**, 356-365.

155. S. J. Storr, N. O. Carragher, M. C. Frame, T. Parr and S. G. Martin, *Nat. Rev. Cancer*, 2011, **11**, 364-374.

156. H. Y. Lee, J. D. Morton, L. J. Robertson, J. D. McDermott, R. Bickerstaffe, A. D. Abell, M. A. Jones, J. M. Mehrtens and J. M. Coxon, *Clin. Exp. Ophthalmol.*, 2008, **36**, 852-860.

157. D. C. Shields, K. E. Schaecher, T. C. Saido and N. L. Banik, *Proc. Natl. Acad. Sci. USA*, 1999, **96**, 11486-11491.

158. R. Chandramohanadas, P. H. Davis, D. P. Beiting, M. B. Harbut, C. Darling, G. Velmourougane, M. Y. Lee, P. A. Greer, D. S. Roos and D. C. Greenbaum, *Science*, 2009, **324**, 794-797.

159. Y. H. Huang and K. K. W. Wang, *Trends Mol. Med.*, 2001, **7**, 355-362.

160. L. Buee, T. Bussiere, V. Buee-Scherrer, A. Delacourte and P. R. Hof, *Brain Res. Rev.*, 2000, **33**, 95-130.

161. M. Chen and H. L. Fernandez, *Biochem. Biophys. Res. Commun.*, 2005, **330**, 714-721.

162. S. Shimohama, T. Suenaga, W. Araki, Y. Yamaoaka, K. Shimizu and J. Kimura, *Brain Res.*, 1991, **558**, 105-108.

163. T. Tsuji, S. Shimohama, J. Kimura and K. Shimizu, *Neurosci. Lett.*, 1998, **248**, 109-112.

164. T. Yamashima, *Cell Calcium*, 2004, **36**, 285-293.

165. K. K. Wang and P. W. Yuen, *Trends Pharmacol. Sci.*, 1994, **15**, 412-419.

166. S. A. Lipton and P. A. Rosenberg, *N. Engl. J. Med.*, 1994, **330**, 613-622.

167. Q. Chen, M. Paillard, L. Gomez, T. Ross, Y. Hu, A. J. Xu and E. J. Lesnefsky, *Biochem. Biophys. Res. Commun.*, 2011, **415**, 533-538.

168. C. Braun, M. Engel, M. Seifert, B. Theisinger, G. Seitz, K. D. Zang and C. Welter, *Int. J. Cancer*, 1999, **84**, 6-9.

169. A. Mamoune, J. H. Luo, D. A. Lauffenburger and A. Wells, *Cancer Res.*, 2003, **63**, 4632-4640.

170. A. Lakshmikuttyamma, P. Selvakumar, R. Kanthan, S. C. Kanthan and R. K. Sharma, *Cancer Epidemiol. Biomarkers Prev.*, 2004, **13**, 1604-1609.

171. P. Dutt, D. E. Croall, J. S. C. Arthur, T. De Veyra, K. Williams, J. S. Elce and P. A. Greer, *BMC Dev. Biol.*, 2006, **6**.

172. M. Pietsch, K. C. H. Chua and A. D. Abell, *Curr. Top. Med. Chem.*, 2010, **10**, 270-293.

173. I. O. Donkor, *Expert Opin. Ther. Patents*, 2011, **21**, 601-636.

174. S. Gil-Parrado, I. Assfalg-Machleidt, F. Fiorino, D. Deluca, D. Pfeiler, N. Schaschke, L. Moroder and W. Machleidt, *Biol. Chem.*, 2003, **384**, 395-402.

175. F. Fiorino, S. Gil-Parrado, I. Assfalg-Machleidt, W. Machleidt and L. Moroder, *J. Pept. Sci.*, 2007, **13**, 70-73.

176. C. Parkes, A. A. Kembhavi and A. J. Barrett, *Biochem. J.*, 1985, **230**, 509-516.

177. A. J. Barrett, A. A. Kembhavi, M. A. Brown, H. Kirschke, C. G. Knight, M. Tamai and K. Hanada, *Biochem. J.*, 1982, **201**, 189-198.

178. K. Saito and R. A. Nixon, *Neurochem Res*, 1993, **18**, 231-233.

179. D. Cuerrier, T. Moldoveanu, R. L. Campbell, J. Kelly, B. Yoruk, S. H. L. Verhelst, D. Greenbaum, M. Bogyo and P. L. Davies, *J. Biol. Chem.*, 2007, **282**, 9600-9611.

180. K. K. Wang, *Trends Pharmacol. Sci.*, 1990, **11**, 139-142.

181. I. O. Donkor, *Curr. Med. Chem.*, 2000, **7**, 1171-1188.

182. S. Mehdi, *Trends Biochem Sci*, 1991, **16**, 150-153.

183. A. Montero, E. Mann, A. Chana and B. Herradon, *Chem. Biodivers.*, 2004, **1**, 442-457.

184. T. L. Graybill, R. E. Dolle, I. K. Osifo, S. J. Schmidt, J. S. Gregory, A. L. Harris and M. S. Miller, *Bioorg. Med. Chem. Lett.*, 1995, **5**, 387-392.

185. D. H. Kang, K. Y. Jun, J. P. Lee, C. S. Pak, Y. Na and Y. Kwon, *J. Med. Chem.*, 2009, **52**, 3093-3097.

186. B. Todd, D. Moore, C. C. S. Deivanayagam, G. D. Lin, D. Chattopadhyay, M. Maki, K. K. W. Wang and S. V. L. Narayana, *J. Mol. Biol.*, 2003, **328**, 131-146.

187. K. K. W. Wang, R. Nath, A. Posner, K. J. Raser, M. BurokerKilgore, I. Hajimohammadreza, A. W. Probert, F. W. Marcoux, Q. H. Ye, E. Takano, M. Hatanaka, M. Maki, H. Caner, J. L. Collins, A. Fergus, K. S. Lee, E. A. Lunney, S. J. Hays and P. W. Yuen, *Proc. Natl. Acad. Sci. USA*, 1996, **93**, 6687-6692.

188. K. K. W. Wang and P. W. Yuen, *US 5760048*, 1998, 546-547.

189. I. M. Downie, M. J. Earle, H. Heaney and K. F. Shuhaibar, *Tetrahedron*, 1993, **49**, 4015-4034.

190. J. B. Hendrickson and J. Wang, *Org. Lett.*, 2004, **6**, 3-5.

191. Y. Yamada, A. Akiba, S. Arima, C. Okada, K. Yoshida, F. Itou, T. Kai, T. Satoua, K. Takeda and Y. Harigaya, *Chem. Pharm. Bull.*, 2005, **53**, 1277-1290.

192. G. Bartoli, G. Palmieri, M. Bosco and R. Dalpozzo, *Tetrahedron Lett.*, 1989, **30**, 2129-2132.

193. M. Somei, T. Hasegawa and C. Kaneko, *Heterocycles*, 1983, **20**, 1983-1985.

194. M. Somei, Y. Saida, T. Funamoto and T. Ohta, *Chem. Pharm. Bull.*, 1987, **35**, 3146-3154.

195. E. C. Taylor and A. McKillop, *Accounts Chem. Res.*, 1970, **3**, 338-346.

196. K. Irie, M. Iguchi, T. Oda, Y. Suzuki, S. Okuno, H. Ohigashi, K. Koshimizu, H. Hayashi, M. Arai, H. Nishino and A. Iwashima, *Tetrahedron*, 1995, **51**, 6255-6266.

197. A. Inada, Y. Nakamura and Y. Morita, *Chem. Lett*, 1980, 1287-1290.

198. H. H. Hodgson, *Chem Rev*, 1947, **40**, 251-277.

199. J. K. Kochi, *J. Am. Chem. Soc.*, 1957, **79**, 2942-2948.

200. R. Dalpozzo and G. Bartoli, *Curr. Org. Chem.*, 2005, **9**, 163-178.

201. C. M. Marson, *Tetrahedron*, 1992, **48**, 3659-3726.

202. A. Meister and M. E. Anderson, *Annu. Rev. Biochem.*, 1983, **52**, 711-760.

203. J. M. Robertslewis, M. J. Savage, V. R. Marcy, L. R. Pinsker and R. Siman, *J. Neurosci.*, 1994, **14**, 3934-3944.

204. E. Bednarski, P. Vanderklish, C. Gall, T. C. Saido, B. A. Bahr and G. Lynch, *Brain Res.*, 1995, **694**, 147-157.

205. V. T. Marchesi, *Annu. Rev. Cell Biol.*, 1985, **1**, 531-561.

206. S. Mittoo, L. E. Sundstrom and M. Bradley, *Anal. Biochem.*, 2003, **319**, 234-238.

207. T. Sasaki, T. Kikuchi, N. Yumoto, N. Yoshimura and T. Murachi, *J. Biol. Chem.*, 1984, **259**, 12489-12494.

208. J. W. Walker, J. Feeney and D. R. Trentham, *Biochemistry-US*, 1989, **28**, 3272-3280.

209. M. Iino and M. Endo, *Nature*, 1992, **360**, 76-78.

210. J. A. McCray and D. R. Trentham, *Annu. Rev. Biophys. Biophys. Chem.*, 1989, **18**, 239-270.

211. I. Laffafian and M. B. Hallett, *Biophys. J.*, 1998, **75**, 2558-2563.

212. J. A. Nick, N. J. Avdi, S. K. Young, C. Knall, P. Gerwins, G. L. Johnson and G. S. Worthen, *J. Clin. Invest.*, 1997, **99**, 975-986.

213. I. Laffafian and M. B. Hallett, *J. Cell Sci.*, 1995, **108**, 3199-3205.

214. Agilent-Technologies, BL21-CodonPlus Competent Cells, 2010.

215. GE-Healthcare, HiLoad 16/600 and 26/600 Superdex 75 prep grade 2011.

216. L. K. Kostanski, D. M. Keller and A. E. Hamielec, *J. Biochem. Biophys. Methods*, 2004, **58**, 159-186.

217. R. K. Scopes, *Protein Purification: principles and practice*, 3rd edn., Springer-Verlag, New York, U.S.A, 1994.

218. J. T. Pelton and L. R. McLean, *Anal. Biochem.*, 2000, **277**, 167-176.

219. N. Greenfie and G. D. Fasman, *Biochemistry-US*, 1969, **8**, 4108-4116.

220. L. Whitmore and B. A. Wallace, DichroWeb, London, UK, 2013.

221. W. C. Johnson, *Proteins*, 1999, **35**, 307-312.

222. L. Whitmore and B. A. Wallace, *Nucleic Acids Res.*, 2004, **32**, W668-W673.

223. Molecular-Dimensions, JCSG-plus HT-96, Suffolk, UK, 2013, p. 4.

224. Molecular-Dimensions, Proplex HT-96, Suffolk, UK, 2013, p. 4.

225. J. Newman, D. Egan, T. S. Walter, R. Meged, I. Berry, M. Ben Jelloul, J. L. Sussman, D. I. Stuart and A. Perrakis, *Acta Crystallogr. D*, 2005, **61**, 1426-1431.

226. Molecular-Dimensions, PACT premier HT-96, Suffolk, UK, 2013, p. 5.

227. E. Potterton, P. Briggs, M. Turkenburg and E. Dodson, *Acta Crystallogr. D*, 2003, **59**, 1131-1137.

228. A. J. McCoy, *Acta Crystallogr. D*, 2007, **63**, 32-41.

229. P. Emsley and K. Cowtan, *Acta Crystallogr. D*, 2004, **60**, 2126-2132.

230. G. N. Murshudov, A. A. Vagin and E. J. Dodson, *Acta Crystallogr. D*, 1997, **53**, 240-255.

231. A. Wlodawer, W. Minor, Z. Dauter and M. Jaskolski, *Febs J.*, 2008, **275**, 1-21.

232. A. T. Brunger, *Nature*, 1992, **355**, 472-475.

233. G. D. Lin, D. Chattopadhyay, M. Maki, E. Takano, M. Hatanaka, L. DeLucas and S. V. L. Narayana, *Acta Crystallogr. D*, 1997, **53**, 474-476.

234. S. Bailey, *Acta Crystallogr. D*, 1994, **50**, 760-763.

235. PDBePISA, EMBL-EBI, Wellcome Trust, Cambridge, UK, 1.47 edn., 2013.

236. M. Somei and M. Tsuchiya, *Chem. Pharm. Bull.*, 1981, **29**, 3145-3157.

237. T. Ohta, Y. Yamato, H. Tahira and M. Somei, *Heterocycles*, 1987, **26**, 2817-2822.

238. M. Somei, K. Kizu, M. Kunimoto and F. Yamada, *Chem. Pharm. Bull.*, 1985, **33**, 3696-3708.

239. R. A. Hollins, L. A. Colnago, V. M. Salim and M. C. Seidl, *J. Heterocycl. Chem.*, 1979, **16**, 993-996.

240. L. C. Heda, R. Sharma, C. Pareek and P. B. Chaudhari, *E-Journal of Chemistry*, 2009, **6**, 770-774.

241. R. Guillon, C. Loge, F. Pagniez, V. Ferchaud-Roucher, M. Duflos, C. Picot and P. Le Pape, *J. Enzyme Inhib. Med. Chem.*, 2011, **26**, 261-269.

242. T. Hara, S. R. Durell, M. C. Myers and D. H. Appella, *J. Am. Chem. Soc.*, 2006, **128**, 1995-2004.

243. T. F. Hofmann A., *BE 628441*, 1963, 14.

244. E. J. Pettit and M. B. Hallett, *J. Cell Sci.*, 1996, **109**, 1689-1694.

245. Stratagene, XL1-Blue Competent Cells, 2004.

246. C. P. Carstens, J. Bonnardel, R. Allen and A. Waesche, *Strategies*, 2001, **14**, 50-52.