



Ward, David John (2019) *The development of thiosulfonates as cysteine protease inhibitors*. PhD thesis.

<https://theses.gla.ac.uk/78987/>

Copyright and moral rights for this work are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This work cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Enlighten: Theses

<https://theses.gla.ac.uk/>  
[research-enlighten@glasgow.ac.uk](mailto:research-enlighten@glasgow.ac.uk)

# The Development of Thiosulfonates as Cysteine Protease Inhibitors

David John Ward

Submitted in fulfilment of the requirements for the  
Degree of Doctor of Philosophy

School of Chemistry  
College of Science and Engineering  
University of Glasgow



University  
of Glasgow

August 2019

# Abstract

Cysteine proteases are ubiquitous throughout nature as proteolytic machinery that are responsible for key physiological processes. Unregulated, uncontrolled or undesired proteolysis is often a key process in many disease states. As such, specific inhibitors of cysteine proteases offer a unique therapeutic target for chemotherapeutic intervention. This is particularly true in many diseases caused by parasitic infections, such as Malaria, Schistosomiasis and Chagas disease, as the parasitic lifecycle is often highly dependent upon cysteine proteases.

The majority of attempts at conferring specificity of covalent inhibitors have been derived from classical structure activity relationship (SAR) studies. Such studies place emphasis on the primary non-covalent interaction with little optimisation of the electrophilic trap, which forms the covalent bond, being attempted. In this work a new class of electrophilic traps, the thiosulfonates, will be developed as cysteine protease inhibitors. This aims to take advantage of the differing chemistry observed for cysteine proteases, with a sulfur centred nucleophile, rather than the oxygen centred nucleophile which is present in all other protease classes. When combined with classical SAR this two pronged attack should greatly increase inhibitor specificity and reduce off target effects, adding to the chemical toolkit available to medicinal chemists.

Throughout this work the synthesis, reactivity profiles and biological evaluation of thiosulfonates as cysteine protease inhibitors will be explored. Computational modelling studies will be incorporated, highlighting the synergy between classical SAR and electrophilic trap optimisation. This will culminate in the application of thiosulfonates towards schistosomiasis, the second biggest neglected tropical disease after malaria.

# Contents

<b>Abstract</b>	<b>i</b>
<b>List of Abbreviations</b>	<b>v</b>
<b>Acknowledgements</b>	<b>xi</b>
<b>Declaration</b>	<b>xii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Cysteine Proteases: Structure, Mechanism and Role in Disease . . . . .	1
1.1.1 Structure . . . . .	2
1.1.2 Mechanism . . . . .	10
1.1.3 Cysteine Proteases Role in Disease . . . . .	11
1.2 Current Cysteine Protease Inhibitors . . . . .	21
1.3 The Thiosulfonate Moiety . . . . .	24
<b>2 <i>DrawToDock</i>: Development of a Virtual Screening Platform to Guide Rational Design.</b>	<b>26</b>
2.1 Developing <i>DrawToDock</i> : A Virtual Screening Platform Directly From Chemdraw Files . . . . .	27
2.2 Summary and Future Work . . . . .	35
<b>3 First Generation Thiosulfonate Inhibitors</b>	<b>37</b>
3.1 Docking Study: Are Thiosulfonates Viable Warheads? . . . . .	37
3.2 Exploring Routes Towards an Efficient Warhead synthesis . . . . .	42
3.2.1 A Retrosynthetic Approach . . . . .	42
3.2.2 Warhead Synthesis: From Thiosulfonate Salts . . . . .	43
3.2.3 Warhead Synthesis: Oxidative Cleavage of Disulfides . . . . .	47
3.3 Backbone Synthesis and Coupling to Warheads . . . . .	50
3.4 Synthesis of K11777: A Reference Compound . . . . .	51
3.5 Stability Testing Towards Common Nucleophiles . . . . .	52

3.5.1	pH Dependent Aqueous Stability . . . . .	53
3.5.2	Stability Towards Thiols and Amines - The Benzyl Thiol and Benzyl Amine Test System . . . . .	55
3.6	Biological Testing - Papain Enzyme Assay . . . . .	57
3.6.1	Developing a Thiol Independent Assay . . . . .	58
3.6.2	Testing Initial Inhibitor Series . . . . .	59
3.6.3	Future work on papain assay . . . . .	62
3.7	Summary . . . . .	63
<b>4</b>	<b>Second Generation Inhibitors: A Convergent Strategy to Improved Bind- ing</b>	<b>65</b>
4.1	Docking Study: Targeting Residues Introduced with Warhead Fragments .	65
4.2	Arginine Derived Thiosulfonates: Exploiting a Flexible Synthesis . . . . .	67
4.2.1	Synthesis of Arginine Derived Warheads . . . . .	67
4.2.2	Final Inhibitors: Coupling Warheads to the Backbone . . . . .	68
4.2.3	Biological Testing: Papain Enzyme Assay . . . . .	69
4.3	Summary . . . . .	71
<b>5</b>	<b>Third Generation Inhibitors: Modulating Leaving Group Ability to Con- trol Reactivity</b>	<b>73</b>
5.1	Improved Stability by Further Substitution: Tris-Methoxybenzene Analogues	74
5.2	Improved Stability by Loss of Aromaticity . . . . .	77
5.2.1	The Cyclohexane Warhead . . . . .	77
5.2.2	Aliphatic Linker to Aromatic Ring: The Best of Both Worlds? . . .	80
5.2.3	Stability Testing Towards Common Nucleophiles . . . . .	84
5.2.4	Biological Testing: Papain Enzyme Assay . . . . .	87
5.3	Summary . . . . .	90
<b>6</b>	<b>Fourth Generation inhibitors: A Steric Approach for Increased Warhead Stability</b>	<b>91</b>
6.1	Reasoning and Target Warheads . . . . .	91
6.2	Exploring a Route Towards $\beta$ -Methyl analogues . . . . .	92
6.2.1	Retrosynthetic Analysis - Two Key Disconnections . . . . .	92
6.2.2	Drawing From the Chiral Pool: A Sidechain Functionalisation Ap- proach . . . . .	94
6.2.3	Reduction & Diastereoselective Addition: Introduction of A Chiral Secondary Alcohol . . . . .	96
6.3	Exploring a Route Towards $\beta$ -Gem-DiMethyl analogues . . . . .	99
6.3.1	Substitution of a Hindered Tertiary Centre . . . . .	99

6.3.2	Stability Testing Towards Common nucleophiles . . . . .	104
6.4	Biological evaluation - Papain assay . . . . .	107
6.5	Summary and Future Work . . . . .	109
<b>7</b>	<b>Applying Thiosulfonates to target a Neglected Tropical Disease: Schistosomiasis</b>	<b>110</b>
7.1	Modelling Study: Applying Thiosulfonates Inhibitors to SmCB1 . . . . .	112
7.2	Viability Testing Against Schistosomula: The Larval Form of Schistosomiasis	115
7.3	Summary . . . . .	120
<b>8</b>	<b>Strategy to Improve Permeability: Towards a Benzyl Amine-Derived Arginine Mimic</b>	<b>121</b>
8.1	Attempted Synthesis of Benzyl Amine-Derived Warheads . . . . .	122
8.2	Summary . . . . .	124
<b>9</b>	<b>Conclusions</b>	<b>125</b>
<b>10</b>	<b>Experimental</b>	<b>127</b>
10.0.1	Materials and Methods . . . . .	127
10.0.2	General Procedures . . . . .	128
10.0.3	Chemistry . . . . .	131
10.0.4	Stability testing . . . . .	155
10.0.5	Biology . . . . .	156
	<b>Bibliography</b>	<b>159</b>
	<b>Appendix</b>	<b>182</b>

# List of Abbreviations

ACT	Artemisinin based Combination Therapy
Bn	Benzyl
Boc	<i>N</i> -tert-butoxycarbonyl
BOP	Benzotriazol-1-yloxy)tris(dimethylamino)phosphonium hexafluorophosphate
Bz	Benzoyl
CDC	Center for Disease Control
CP	Cysteine Protease
DABSO	1,4-Diazabicyclo[2.2.2]octane bis(sulfur dioxide)
DCM	Dichloromethane
DiPEA	<i>N,N</i> -Diisopropylethylamine
DMAP	4-Dimethylaminopyridine
DMF	Dimethylformamide
DMP	Dess-Martin periodinane
DMS	Dimethyl sulfide
DMSO	Dimethyl sulfoxide
Fmoc	Fluorenylmethyloxycarbonyl
GUI	Graphical User Interface
HCTU	<i>O</i> -(1H-6-Chlorobenzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate
HPLC	High Pressure Liquid Chromatography
IBX	2-Iodoxybenzoic acid
NMR	Nuclear Magnetic Resonance
Pbf	2,2,4,6,7-Pentamethyldihydrobenzofuran-5-sulfonyl chloride
TBDMS	<i>tert</i> -Butyldimethylsilyl
TFA	Trifluoroacetate (protecting group) or Trifluoroacetic acid
TIPS	Triisopropylsilane

# List of Figures

1.1	(a) Structure of pro Cathepsin L with the pro-peptide highlighted in yellow running up the active site cleft between cathepsin L's left and right domains and (b) Structure of Papain. . . . .	2
1.2	Papain subsite nomenclature and corresponding substrate residues. . . . .	4
1.3	Cathepsin B (PDB code: 1HUC) active site binding cleft highlighting subsites S <sub>3</sub> -S <sub>2</sub> ' . . . . .	5
1.4	S <sub>1</sub> site of cathepsin B highlighting key structural features. . . . .	6
1.5	S <sub>2</sub> site of cathepsin B highlighting key structural features. . . . .	7
1.6	The S <sub>1</sub> ' binding site of papain. . . . .	8
1.7	Cathepsin B S <sub>1</sub> ' binding site . . . . .	9
1.8	Catalytic cycle of cysteine proteases . . . . .	10
1.9	Role of cathepsins B and L in tumour progression and metastasis. . . . .	12
1.10	The Role of Osteoclasts and Cathepsin K in Bone Resorption . . . . .	14
1.11	Lifecycle of <i>Schistosomes</i> . . . . .	16
1.12	Structure of praziquantel, the only actively administered schistosomiasis medication. . . . .	17
1.13	Lifecycle of <i>Plasmodium</i> . . . . .	18
1.14	Structure of Quinine, Chloroquine and Mefloquine used for treatment of Malaria. . . . .	19
1.15	Lifecycle of <i>T.cruzi</i> . Protozoa . . . . .	20
1.16	Structure of Beznidazole, Nifurtimox and K11777. . . . .	20
1.17	Current cysteine protease inhibitors . . . . .	22
1.18	Example thiosulfonates from the literature . . . . .	24
1.19	Proposed thiosulfonate inhibitor constructs building on the success of K11777, with a warhead inspired by methyl methane thiosulfonate (MMTS). . . . .	25
2.1	<i>DrawtoDock</i> A Schematic Representation. . . . .	26
2.2	Generating Ligands for Docking with Chemdraw . . . . .	28
2.3	Generating 3D coordinated with Avogadro . . . . .	29
2.4	Generating 3D coordinated with Open Babel . . . . .	30



2.5	Converting Ligand Format from .pdb to .pdbqt with AutoDockTools GUI .	31
2.6	Defining the search space for docking with AutodockTools. . . . .	32
2.7	Running AutoDock Vina, an example. . . . .	33
2.8	The <i>DrawToDock</i> script. . . . .	35
3.1	Comparing the Site of Nucleophilic Attack of Peptides, Michael Acceptors and Thiosulfonates. . . . .	37
3.2	Vinyl sulfone inhibitor K11777 and the thiosulfonate mimic. . . . .	38
3.3	Docking binding mode of (a) K11777 <b>6</b> , (b) Thiosulfonate mimic <b>8</b> , (c) overlay of both <b>6</b> and <b>8</b> , (d) RMS calculation between binding modes of <b>6</b> and <b>8</b> , line representation with paired atoms highlighted in yellow. . . . .	39
3.4	<i>In situ</i> Activation of Warheads Through H-bonding. . . . .	41
3.5	pH Dependent Hydrolysis of Aromatic Thiosulfonates . . . . .	53
3.6	Chiral Sulfoxides Formed Upon Aqueous Hydrolysis of Thiosulfonate warheads via Sulfenic Acid Intermediate. . . . .	54
3.7	Stability of Aromatic Thiosulfonates Towards Thiols and Amines . . . . .	55
3.8	Asymmetric Disulfide Formed by Attack of Benzyl Mercaptan on Thiosulfonate Inhibitors . . . . .	56
3.9	Papain Assay Overview: Substrate Hydrolysis and Quantification . . . . .	58
3.10	Determining Papain Concentration for Thiol Independent Assay . . . . .	59
3.11	Papain Inhibition Curved for Homophenylalanine Derived Aromatic Thiosulfonates. . . . .	60
3.12	IC <sub>50</sub> Determination of Homophenylalanine Derived Aromatic Warheads. . . . .	61
4.1	Docked pose of arginine-derived thiosulfonate inhibitor <b>14</b> with papain . . . . .	66
4.2	Papain Inhibition Curved for Arginine Derived Aromatic Thiosulfonates. . . . .	69
4.3	IC <sub>50</sub> Determination of Arginine-Derived Aromatic Warheads. . . . .	70
5.1	Target Compounds for Improved Stability by Reduced Leaving Group Ability	73
5.2	Active Site Residues of Papain, Highlighting the Conserved Tryptophan Residue of Papain like Cysteine Proteases . . . . .	81
5.3	Varying the Number of Methylene Units Between the Thiosulfonate Warhead and Aromatic Ring in the S <sub>1</sub> ' Region . . . . .	82
5.4	Cyclohexane warhead stability towards benzyl thiol (left) and benzyl amine (right) . . . . .	85
5.5	pH dependent aqueous stability of aliphatic thiosulfonate warheads compared to the previous aromatic thiosulfonate warheads. . . . .	86
5.6	IC <sub>50</sub> determination of aliphatic warheads compared with the most potent aromatic warhead. . . . .	89

6.1	Key Building Blocks to Introduce Steric Arguments for Decreased Reactivity	91
6.2	Structure of primary thiosulfonate inhibitors and their secondary thiosulfonate counterparts. . . . .	104
6.3	Stability of Secondary Thiosulfonates Towards Amines and Thiols . . . . .	105
6.4	pH Dependent Aqueous Stability of Secondary Thiosulfonate Warheads Compared to the Primary aromatic and Aliphatic Thiosulfonate Warheads.	106
6.5	Papain IC <sub>50</sub> Values of the Secondary Thiosulfonate Warheads . . . . .	108
7.1	Structure of All Thiosulfonate Inhibitors Tested . . . . .	111
7.2	Crystal Structure of <i>S. mansoni</i> cathepsin B like Cysteine Protease SmCB1 with K11777 Bound in Active Site . . . . .	113
7.3	Docked pose of thiosulfonate inhibitor <b>7</b> with SmCB1. . . . .	114
7.4	Docked pose of thiosulfonate inhibitors with SmCB1. . . . .	115
7.5	In Vitro Thiosulfonate Susceptibility Sreen . . . . .	116
7.6	Schistosomula viability dose-response curves for compounds <b>7-11</b> . . . . .	117
7.7	Schistosomula viability dose-response curves for compounds <b>11</b> and <b>12</b> compared to <b>127</b> and <b>128</b> . . . . .	119
8.1	Warhead Building Block with a Benzyl Amine Arginine Mimic Side Chain.	121

# List of Schemes

3.1	Retrosynthetic Analysis of Amino Acid Derived Thiosulfonate Warheads . . . . .	42
3.2	Thiosulfonate Salt Synthesis from Sulfinic Salts . . . . .	43
3.3	Homophenylalanine Derived Aromatic Thiosulfonates from Thiosulfonate Salts . . . . .	44
3.4	Employing Finkelstein Conditions to Improve Thiosulfonate Formation from Thiosulfonate Salts . . . . .	45
3.5	Impurity Formation by Subsequent Substitution of Thiosulfonates . . . . .	45
3.6	Thiosulfonate Impurity Formation due to Impure Thiosulfonate Salts . . . . .	46
3.7	Thiosulfonate retrosynthesis S-S Disconnect with Synthon Options and Reagents . . . . .	47
3.8	Mechanism: Thiosulfonate Formation by Oxidative Cleavage of Disulfides . . . . .	48
3.9	Homophenylalanine Derived Aromatic Thiosulfonates from by Oxidative Cleavage of Disulfides . . . . .	49
3.10	Backbone Synthesis and Coupling to Homophenylalanine Derived Aromatic Warheads. . . . .	51
3.11	Synthesis of K11777 Reference Compound. . . . .	52
4.1	Arginine Derived Aromatic Thiosulfonates from by Oxidative Cleavage of Disulfides . . . . .	67
4.2	Backbone Synthesis and Coupling to Arginine Derived Aromatic Warheads. . . . .	68
5.1	2,4,6-Trimethoxybenzene Sodium Sulfinic: by Reduction of Sulfonyl Chloride . . . . .	74
5.2	2,4,6-Trimethoxybenzene Sodium Sulfinic: by Directed Ortho Lithiation and SO <sub>2</sub> . . . . .	75
5.3	2,4,6-Trimethoxybenzene Sodium Sulfinic: by Directed Ortho Lithiation with DABSO . . . . .	76
5.4	2,4,6-Trimethoxybenzene Thiosulfonate Warhead Synthesis by <i>in situ</i> Oxidative Cleavage of a Disulfide . . . . .	76
5.5	Cyclohexane Thiosulfonate Warhead Synthesis by Oxidative Cleavage of a Disulfide . . . . .	77

5.6	Proposed Synthesis of Cyclohexane Thiosulfonate from a Sulfonyl Chloride	78
5.7	Synthesis of the Cyclohexane Thiosulfonate Warhead . . . . .	79
5.8	Synthesis of Thiol by Reduction and Subsequent Thiosulfonate Formation.	79
5.9	Cyclohexane warhead coupling to backbone yielding final inhibitors <b>11</b> and <b>17</b> . . . . .	80
5.10	Original Intended Route Towards Benzyl Thiosulfonate Warhead Building Block <b>76</b> . . . . .	82
5.11	Proposed Radical Decomposition Pathways of Benzylsulfonyl Bromide . . .	83
5.12	Benzyl thiosulfonate Warhead SYnthesis from Benzylsulfonyl Chloride <i>via</i> a Sulfene Intermediate . . . . .	83
5.13	Final Inhibitor Coupling: Coupling the Benzyl Derived Thiosulfonate Warhead Building Block to Backbone. . . . .	84
6.1	Retrosynthetic Analysis: Towards the Key Secondary Alcohol Intermediate	92
6.2	Retrosynthetic Analysis: Towards Secondary Thiosulfonate Warheads by Diastereoselective Alkylation . . . . .	93
6.3	Retrosynthetic Analysis: Towards Secondary Thiosulfonate Warheads by Sidechain Functionalisation . . . . .	93
6.4	Synthesis Towards Secondary Thiosulfonates from D-threonine. . . . .	95
6.5	Methods of oxidation to aldehyde <b>112</b> . (a) Swern oxidation: Oxalyl chloride, DMSO, NEt <sub>3</sub> , dry DCM, -78 °C-RT. (b) Dess Martin oxidation: Dess Martin Periodinane, wet DCM (washed with water before use). . . . .	96
6.6	Wittig reaction and competing elimination . . . . .	96
6.7	$\beta$ -Methyl Cyclohexane Thiosulfonates by Diastereoselective Methylation . .	97
6.8	Oxazolidinone Formation Through Alcohol By-product . . . . .	97
6.9	Coupling the Secondary Thiosulfonate Warheads to their Backbone . . . .	98
6.10	Synthesis towards tertiary thiosulfonates starting from D-penicillamine ( <b>129</b> ). .	99
6.11	Synthesis towards tertiary thiosulfonates by a tertiary alcohol intermediate	100
6.12	Synthesis of tertiary thiosulfonates by di-methylation . . . . .	101
6.13	Attempted Synthesis of Tertiary Thiol by Direct Substitution of a Trifluoroacetate with Resin Bound Thiol . . . . .	102
6.14	Converting tertiary Alcohol to Tertiary Bromide Under Acidic Conditions .	102
6.15	Attempted Ring Opening of Oxazolidinone Under Acidic Conditions . . . .	103
8.1	Synthesis Towards an Arginine mimic with Increased Permiability . . . . .	122
8.2	Intramolecular Cyclisation and Aziridine Formation. . . . .	123
8.3	Revised Synthesis Towards an Arginine mimic with Increased Permiability	124

# Acknowledgements

Firstly, I would like to thank Prof. Rob Liskamp for the wonderful opportunity to work on this project. His guidance and support has proven invaluable. I will forever remember his amazingly positive attitude and enthusiasm, which proved to be contagious. It has been a pleasure to work with Rob over the past four years.

I would also like to thank all current and past members of the Liskamp group I have had the pleasure of working with. Thank you for being there to talk through problems and for making the lab a great place to be.

Finally I would like to thank my family for their support and encouragement over the past four years. I particularly appreciated their efforts to understand what I was doing!

# Declaration

I declare that, except where explicit reference is made to the contribution of others, the substance of this thesis is the result of my own work and has not been submitted for any other degree at the University of Glasgow or any other institution.

Signed: \_\_\_\_\_

David John Ward

Signed: \_\_\_\_\_

Prof. Rob M.J. Liskamp

Parts of this thesis have been published:

D.J. Ward, H. Van de Langemheen, E. Koehne, A. Kreidenweiss, and R.M.J. Liskamp. Highly tunable thiosulfonates as a novel class of cysteine protease inhibitors with anti-parasitic activity against schistosoma mansoni. *Bioorganic & Medicinal Chemistry*, 27(13):2857 - 2870, 2019.

# Chapter 1

## Introduction

### 1.1 Cysteine Proteases: Structure, Mechanism and Role in Disease

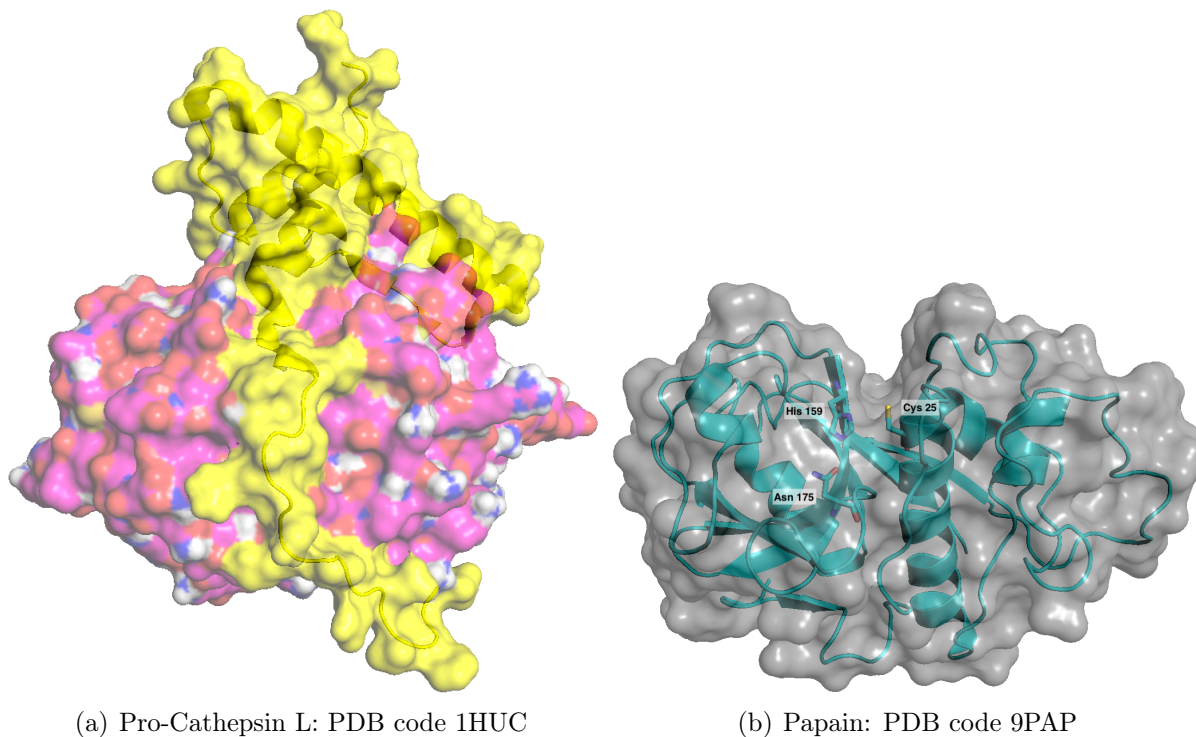
Proteases are ubiquitous throughout nature as one of the largest, most important enzyme classes. This proteolytic machinery is involved in a vast range of physiological processes including blood coagulation, digestion, fertilization, wound healing, cell differentiation and growth, the immune response, cell signalling, apoptosis and protein turnover. Undesired, unregulated and uncontrolled proteolysis can result in many disease states including stroke, emphysema, cancer, viral infections, Alzheimer's disease, arthritis and inflammation.<sup>1</sup> Furthermore, many pathogenic species depend upon proteases for a successful lifecycle, including those which cause neglected tropical diseases (NTDs) such as Malaria, Schistosomiasis, Leishmaniasis and Trypanosomiasis. As such, proteases have become interesting targets for therapeutic intervention in a range of disease states, especially where current therapies are limited and resistance is emerging. There are five main classes of protease which are classified according to their catalytic site: serine, threonine, cysteine, aspartic and metallo-proteases. This work focuses on the cysteine proteases, aiming to exploit subtle differences in their chemistry to generate novel, class specific, inhibitor constructs - the thiosulfonates.

Cysteine proteases (CPs) are ubiquitous in nature, being found in all living organisms. These CPs are categorised based on sequence homology into clans, containing one or more CP families that have arisen from a single evolutionary origin. Furthermore, the families are categorised based on three key features: (i) common sequence motifs surrounding the catalytic residue, (ii) the order of catalytic residues in their sequence and (iii) by their similar tertiary structure. The most abundant among all CPs is clan CA, encompassing

the calpain and papain families. The next most abundant clan is CD, which contains the caspase, separase, legumain, gingipain and clostripain families.<sup>2,3</sup> At present, 21 different families of CPs have been identified, with almost half of these being found in viruses.<sup>4,5</sup> Many of these proteases are also found in plants,<sup>6</sup> fungi,<sup>7</sup> bacteria and protozoa.<sup>8-11</sup> The mammalian CPs fall into two main groups; The cytosolic calpains and the lysosomal cathepsins, both members of clan CA.<sup>12,13</sup>

Some of the best characterised, most widely studied cysteine proteases are the papain like cysteine proteases (belonging to clan CA, family C1). The papain like cysteine proteases are structurally related to papain (figure 1.1), which is isolated from the latex of papaya fruit. The papain family includes CPs of mammals (e.g. the cathepsins) as well as CPs of parasites such as Leishmania, Trypanosomatids, Schistosoma, Fasciola, Giardia and other parasites.<sup>14,15</sup>

### 1.1.1 Structure



**Figure 1.1:** (a) Structure of pro Cathepsin L with the pro-peptide highlighted in yellow running up the active site cleft between cathepsin L's left and right domains and (b) Structure of Papain.

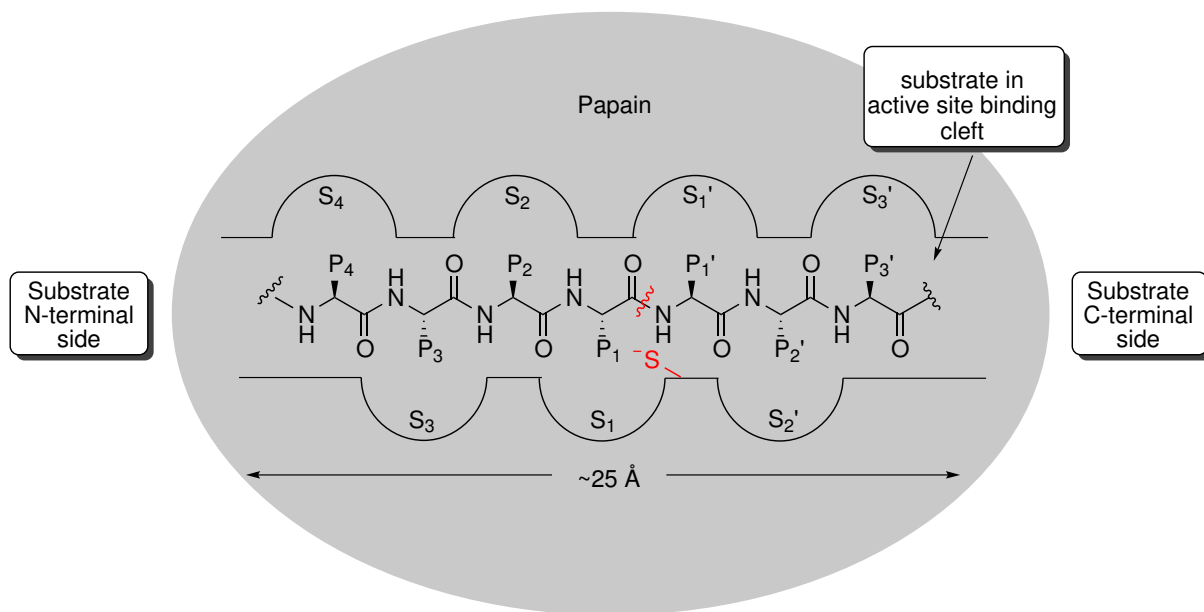
Looking towards papain, as a representative protease within the papain family, it is characterised by a single polypeptide chain of 212 amino acids, split into two domain of roughly equal size (figure 1.1, right). The space between between the right domain (R) and a left domain (L) forms the active site binding cleft, which is common to all members of the



papain family. This V-grooved cleft houses the active site catalytic dyad of Cys<sup>25</sup> (found on the L domain) and His<sup>159</sup> (found opposite on the R domain), which are evolutionary preserved in all cysteine proteases. The R domain forms a  $\beta$ -barrel like structure with a shorter  $\alpha$ -helical motif appended to it, whilst the most distinct feature of the L domain is the central  $\alpha$  helix comprising approximately 30 residues.<sup>16</sup>

To prevent undesired protein degradation, cysteine proteases are synthesised as zymogens (inactive precursors). The zymogens of CPs contain a prodomain that masks the active site, blocking access to substrates, as shown for pro-cathepsin L in figure 1.1.<sup>17</sup> It can be noted that, unlike papain, cathepsin L comprises a single polypeptide chain with two distinct domains forming the active site binding cleft between them. The pro-peptide of cathepsin L can be seen binding along this active site cleft, masking the catalytic function until activation takes place. Whilst these pro-peptides function as an endogenous inhibitor,<sup>18,19</sup> the prodomain may also have roles in intracellular sorting and / or protein folding.<sup>18,20,21</sup> Transformation of the zymogen to the enzymes active form usually takes place within a subcellular compartment, or the extracellular environment in which the protease fulfils its biological role, with multiple modes of activation being reported.<sup>22,23</sup> Zymogen conversion to the active form may occur auto catalytically (usually requiring a significant drop in pH),<sup>24</sup> by another protease or molecule of the same enzyme (trans-activation),<sup>25</sup> or by an uncommon mechanism such as that displayed for Schistosoma cathepsin B1 (SmCB1). SmCB1 exhibits a mode of activation that is part auto catalytic, beginning with autocatalysis of part of the prodomain, followed by a structural change induced by a sulfated polysaccharide, before final trans-activation by another protease (SmAE).<sup>26</sup>

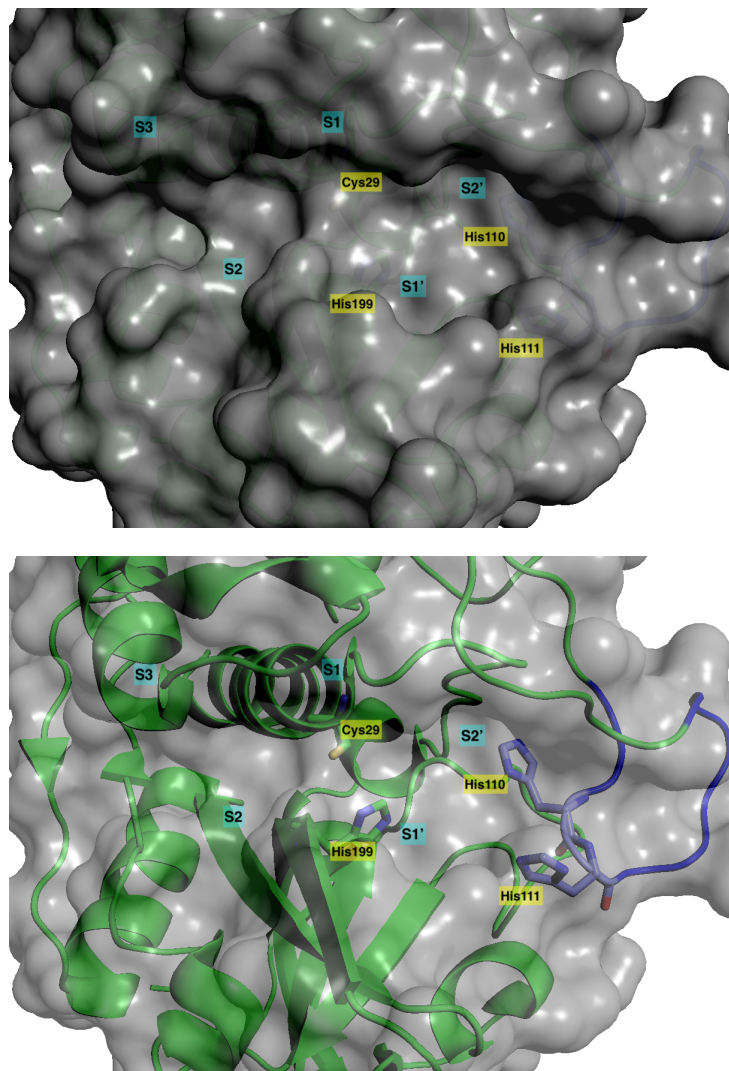
Based on the early proposal of Schechter & Berger,<sup>27</sup> the active site of papain (and related proteases) comprises seven subsites ( $S_4$ - $S_3'$ ) which can accommodate the corresponding amino acids of the substrate (commonly denoted  $P_4$ - $P_3'$ ). As shown in figure 1.2 the subsites on the C-terminal side of the scissile bond are denoted the primed ( $'$ ) sites, with the subsites on the N-terminal side denoted as non-primed. The later work of Turk *et al.* revised the definition of these substrate binding sites, highlighting that the three substrate binding sites of  $S_2$ - $S_1'$  were of the greatest importance, with binding contributing from both the substrate backbone as well as amino acid side chains.<sup>16</sup> Beyond these regions, interaction at the  $S_3$  and  $S_2'$  subsites with substrate is dependent on the substrate amino acid side chains, with no contribution from the substrate backbone. Furthermore, it was defined that beyond the  $S_3$  and  $S_2'$  subsites the location and definition of binding sites becomes questionable, which was in keeping with Schechter & Berger's original assessment.<sup>27</sup> Hence, it can be reasoned that (for inhibitor design) the key sites to consider for successful target engagement are the  $S_2$ - $S_1'$  regions.



**Figure 1.2:** Papain subsite nomenclature and corresponding substrate residues. Substrate N-terminal region is to the left as drawn, with the C-terminal region to the right. The active site thiolate is shown in red adjacent to the scissile bond.

Proteases can be further classified depending on where in the peptidic substrate the scissile bond lies. Endopeptidases represent the basis of the papain superfamily, cleaving a scissile bond within the main chain of the peptide. Conversely, exopeptidases cleave amino acids off the end of the peptide. Further to this, amino peptidases cleave from the N-terminus while carboxy peptidases cleave from the C-terminus. This is an important consideration as the mammalian papain like cysteine protease (cathepsin B) exhibits a dipeptidyl carboxypeptidase activity, cleaving selectively two amino acids back from the C-terminus. This activity is additional to the endopeptidase activity which is common amongst the papain like proteases. It is the presence of a unique structural feature, termed the occluding loop, which facilitates this function in cathepsin B. As shown in figure 1.3, the occluding loop is an amino acid chain which extends beyond the C-terminal region of the substrate (shown with a blue carbon backbone). Importantly, the occluding loop houses two key hydrogen bonding partners (His<sup>110</sup> and His<sup>111</sup>) which effectively anchor the C-terminal carboxylic acid through a  $-\text{COO}^- \cdots \text{H-Im}^+$  hydrogen bond pairing with both His residues. This activity is in addition to the endopeptidase activity exhibited for all other papain like CPs and operates in a highly pH dependant manner.<sup>28</sup>

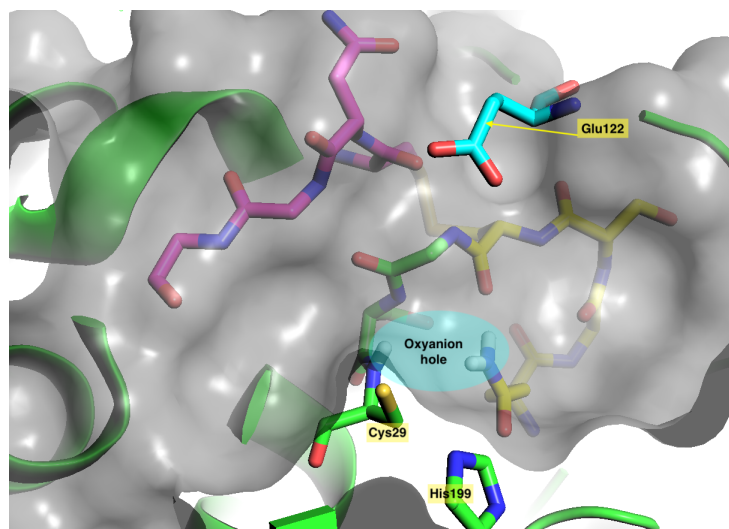
Rather than offer a full description of the structure of all papain-like cysteine proteases, papain will be discussed as a representative example. Furthermore, Cathepsin B and Cathepsin L will be discussed for comparison. Due to the high homology between papain like cysteine proteases, it is commonplace in literature to describe cysteine proteases as either cathepsin B or cathepsin L like, with the remaining structural features comparable



**Figure 1.3:** Cathepsin B (PDB code: 1HUC) active site binding cleft highlighting subsites  $S_3$ - $S_2'$  (cyan box), notable residues (yellow box) include the catalytic dyad of Cys29 and His199 and also His110 and His111 of the occluding loop. Occluding loop is displayed with a blue carbon backbone (towards the right of binding cleft as viewed)

to the parent papain enzyme. Thus, exploring the key structural elements of these three key proteases will give a good understanding of the specificity and activity of multiple papain like cysteine proteases. The three key regions will be discussed ( $S_2$ - $S_1'$ ), largely based on the work of Turk *et al.* which provides a good overview of all binding sites.<sup>16</sup> In the following description of the subsites, papain numbering has been used throughout, thus residue numbers in figures may vary slightly if different proteases are presented to highlight structural differences.

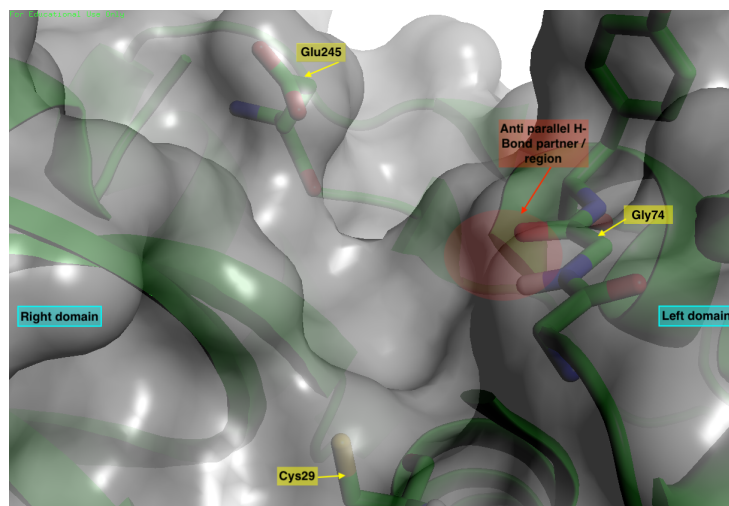
Beginning with the  $S_1$  site, this lies completely within the L domain of the enzyme. For reference, the  $S_1$  pocket of cathepsin B is shown in figure 1.4, which highlights both the general features of this region as well as Cathepsin B specific additions of interest. As this subsite is located directly next to the active site Cys<sup>29</sup> (Cys<sup>25</sup> in papain numbering) the



**Figure 1.4:**  $S_1$  site of cathepsin B highlighting key structural features. Based on papain numbering, the pocket is formed between residues 21-23 (pink carbon backbone) along the left side as pictured, connected by a Cys<sup>22</sup>-Cys<sup>63</sup> disulfide bridge at the base with the right side formed by the broad turn of residues 63-66 (yellow carbon backbone). The oxyanion hole is formed by the amide protons of Cys<sup>29</sup> (25 in papain numbering) and Glu<sup>19</sup> (papain numbering). Active site catalytic dyad of Cys<sup>29</sup> and His<sup>199</sup> can also be seen (residues 25 and 159 in papain numbering)

backbone carbonyl of the substrate is able to interact with a region known as the oxyanion hole. The oxyanion hole is a feature common to many protease classes, comprising a local H-bond network which is thought to stabilise the oxyanion tetrahedral intermediates during peptide hydrolysis (see the later discussion on the catalytic cycle, figure 1.8). As such, the backbone carbonyl of the substrate amide bond being hydrolysed offers an important interaction within this region. This H-bonding takes place through the amide protons of Cys<sup>25</sup> and the Gln<sup>19</sup> sidechain, forming the oxyanion hole. Moving along into the  $S_1$  subsite, the binding region for the  $P_1$  side chain is created by the Ser<sup>21</sup>, Cys<sup>22</sup> and Gly<sup>23</sup> loop along one side (left side as shown by pink carbon backbone in figure 1.4). The Cys<sup>22</sup> residue lies at the bottom of the  $S_1$  pocket, forming the base of the pocket with a disulfide bridge across to Cys<sup>63</sup>. Moving back down the other side of the  $S_1$  pocket is a broad turn, shown with a yellow carbon backbone in figure 1.4, which is formed from Cys<sup>63</sup>, Asn<sup>64</sup>, Gly<sup>65</sup> and Gly<sup>66</sup>. Combined, this forms a relatively shallow binding pocket for the majority of papain like proteases, with the  $P_1$  side chains running along the surface and pointing towards solvent facing regions. In the case of papain, the backbone carbonyls of the Ser<sup>21</sup> and Cys<sup>23</sup> are directed towards the  $P_1$  sidechain offering potential H-bond acceptors, which may account for the preference of an arginine side chain in this position.<sup>29</sup> In the case of cathepsin B, an additional structural feature known as the occluding loop is present, which further encloses this binding pocket. As shown in figure 1.4, Glu<sup>122</sup> (with cyan carbons) extends out from the occluding loop and folds over the top of the  $S_1$  pocket. This residue rests towards the bottom of the  $S_1$  pocket and further enhances

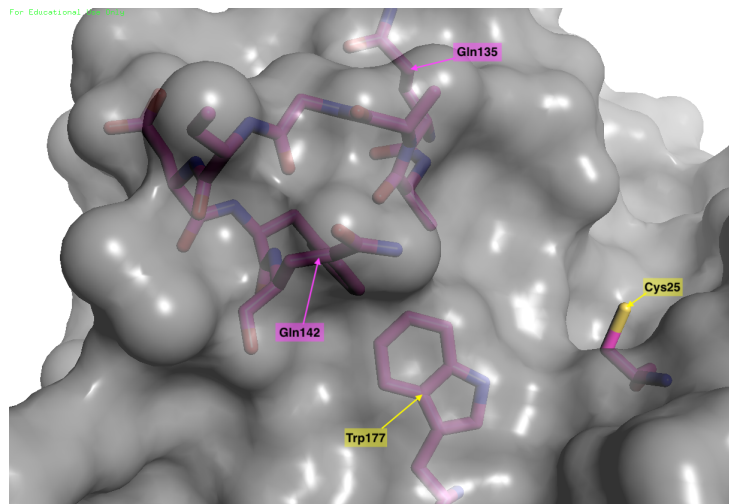
the electrostatically negative character of this binding site. As a result, cathepsin B has an increased preference towards basic residues binding in this region compared to papain, although both can process basic P<sub>1</sub> residues. Furthermore, the Glu<sup>122</sup> residue restricts the size of the S<sub>1</sub> binding pocket of cathepsin B, reducing the steric bulk of P<sub>1</sub> residues that can be accommodated when compared with other papain like cysteine proteases, such as papain or cathepsin L.<sup>30</sup>



**Figure 1.5:** S<sub>2</sub> site of cathepsin B highlighting key structural features. View is along the binding cleft with the active site Cys<sup>29</sup> in the foreground. Key H-bond partner of Gly<sup>74</sup> anchors peptidic backbone amides allowing the side chain to stretch into the S<sub>2</sub> binding pocket in the right domain. Glu<sup>245</sup> can be found at the bottom of the S<sub>2</sub> pocket, offering a H bonding partner for basic residues.

Moving on to the S<sub>2</sub> site this offers the most well defined subsite of papain like CPs. This binding site forms a deep pocket with contributions from both the L and R domains of the protease. Figure 1.5 displays the S<sub>2</sub> binding site for cathepsin B, which is representative of papain-like CPs with some notable additions which will be discussed. The S<sub>2</sub> site begins from Gly<sup>74</sup> (Gly<sup>66</sup> in papain numbering) in the L domain that presents its carbonyl and amide N-H towards the active site binding cleft, forming an excellent anti-parallel H bond partner for peptidic substrates. With the anti-parallel H bonding of Gly<sup>74</sup> anchoring the substrate backbone, the side chain can then extend down into the S<sub>2</sub> pocket, which takes the form of a cavity penetrating into the R domain (moving from right to left away from Gly<sup>74</sup> and into binding pocket in figure 1.5). In general, this pocket is hydrophobic in character being surrounded by aliphatic and methionine residues resulting in a preference for hydrophobic residues such as phenylalanine in the P<sub>2</sub> position.<sup>16</sup> In the case of cathepsin B, the base of the S<sub>2</sub> pocket houses a glutamic acid (Glu<sup>245</sup>, figure 1.5), replacing the serine residue of papain (Ser<sup>205</sup>, papain numbering) and offers an H-bonding partner for basic residues. As such, Cathepsin B and cathepsin B like proteases have been reported to efficiently process arginine as well as phenylalanine in this position.<sup>31,32</sup> Interestingly Turk

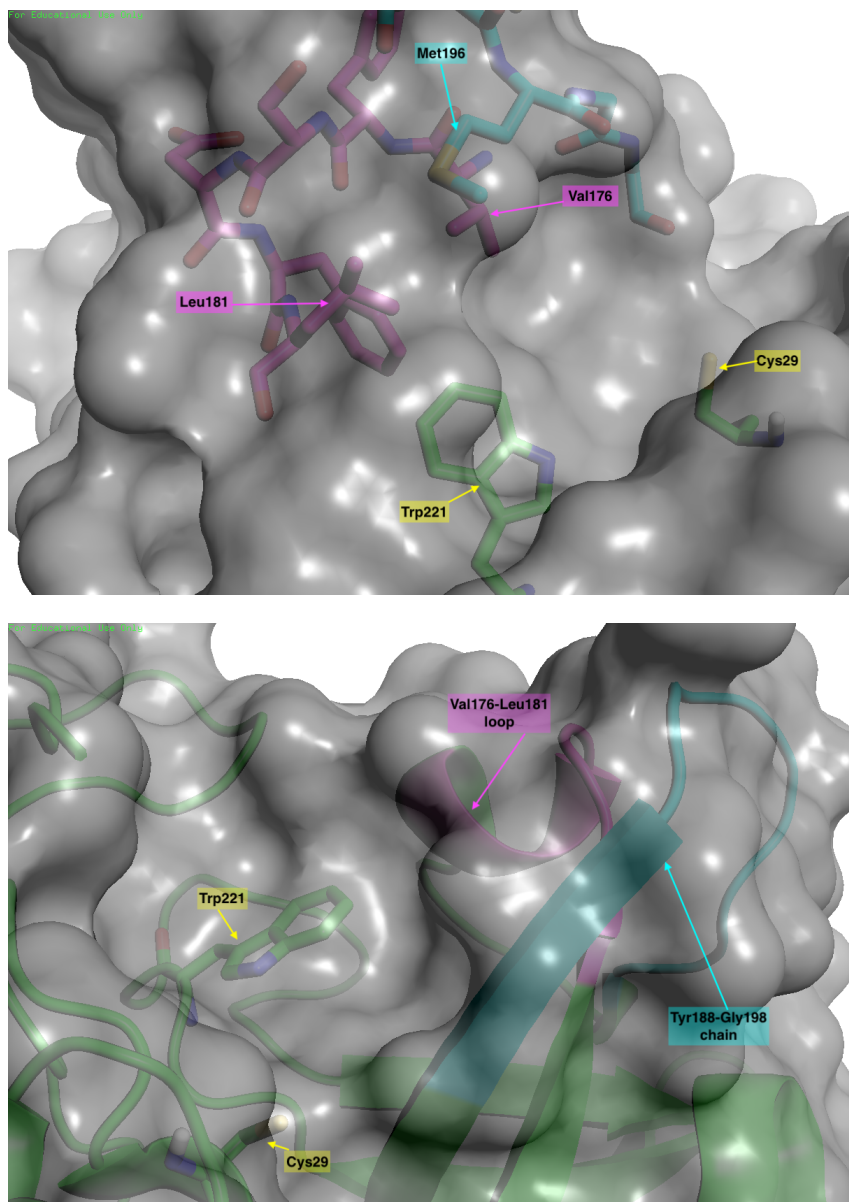
*et al.* concluded that for the parasitic protease cruzipain, which is cathepsin B like in this region, the glutamic acid adjusts its position in order to process different substrates in the P<sub>2</sub> position, swinging in towards the binding pocket to process basic residues, or outwards away from the binding pocket to process hydrophobic residues based on the structures published by Mcgrath *et al.* and Gilmor *et al.*<sup>32,33</sup>



**Figure 1.6:** The S<sub>1</sub>' binding site of papain.

Finally, the S<sub>1</sub>' site will be considered. The S<sub>1</sub>' site is formed entirely within the R domain, as shown in figure 1.6, lying directly within the base of the active site binding cleft is Trp<sup>177</sup> (in papain numbering), which is highly conserved throughout all papain like CPs. The NH of tryptophans NE1 atom offers a potential H bonding partner towards backbone carbonyl of a substrate binding along the active site cleft before the P<sub>1</sub>' sidechain extends out towards the S<sub>1</sub>' region. The main surface for interaction with P<sub>1</sub>' side chains is created by the loop between Gln<sup>135</sup> and Gln<sup>142</sup>. The character of this binding site accepts both small hydrophobic (up to three methylene units) as well as hydrophilic P<sub>1</sub>' side chains.<sup>34</sup> This dual functionality may be attributed to the hydrophobic contact patch created by Ala<sup>136</sup> and Ala<sup>137</sup>, while Gln<sub>135</sub> and Gln<sub>142</sub> are thought to contribute to the hydrophilic character. In the case of Cathepsin L there is a key alteration when compared with papain to include two acidic residues (Glu<sup>141</sup> and Asp<sup>137</sup>), which may explain its ability to bind basic residues such as lysine in this position.<sup>35</sup> Cathepsin L is also able to catalyse substrates with an arginine in the P<sub>1</sub>' position, which has been attributed to the (Asp<sup>137</sup>) residue.

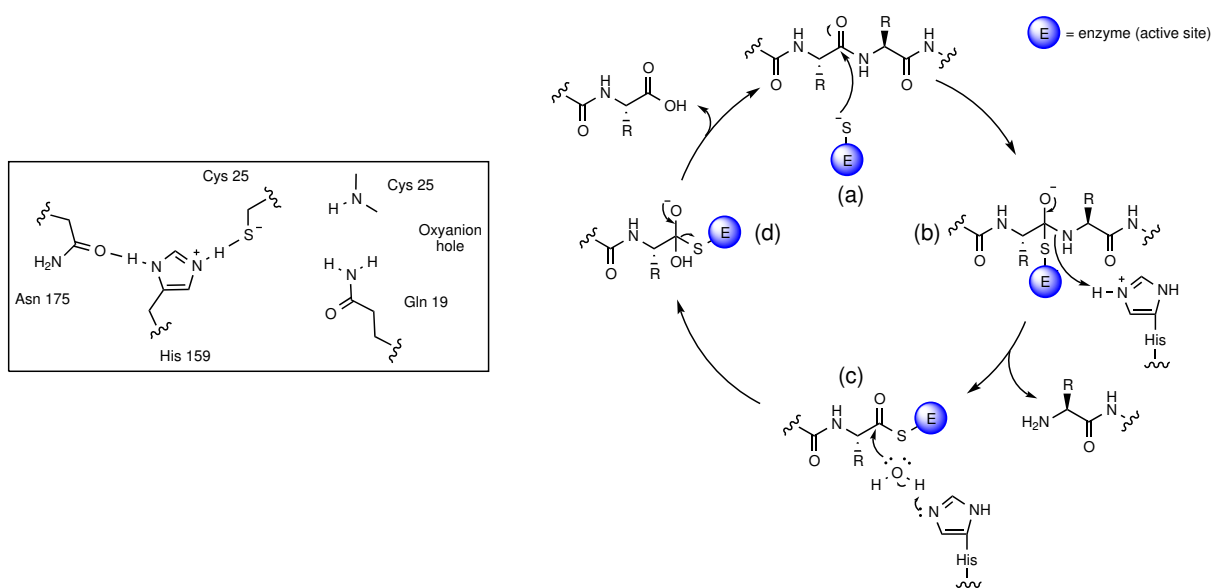
In the case of Cathepsin B this pocket is significantly larger and more hydrophobic in character. This is likely due to the key structural change induced by deletion of Cys<sup>153</sup> (papain numbering), which prevents disulfide formation to Cys<sup>240</sup> (Cys<sup>200</sup> in papain numbering).<sup>16</sup> The loss of this disulfide effectively allows the chain between Tyr<sup>188</sup> and Gly<sup>198</sup> to fold forward over the top of the S<sub>1</sub>' site (figure 1.7, increasing the size and contact patch



**Figure 1.7:** Cathepsin B S<sub>1</sub>' binding site. The conserved loop, common to papain, is shown with a pink carbon backbone. The additional peptide chain which overlaps the top of the binding site is shown with a cyan carbon backbone. The top image shows a stick representation of key residues, the bottom image shows a cartoon which highlights the nature of the additional peptide chain in overlapping the conserved binding region

of this binding region and allowing larger residues to be processed. This also enhances the hydrophobic character of Cathepsin B's  $S_1'$  region by overlapping it with Met<sup>196</sup>. Apart from the contribution of the additional peptide chain, the original binding surface also shows an increased hydrophobic character, replacing Gln<sup>135</sup> and Gln<sup>136</sup> (in papain) with Val<sup>176</sup> and Tyr<sup>177</sup> in cathepsin B. The combination of cathepsin B's key changes results in a larger, more hydrophobic region which is in keeping with the observation by Menard *et al.* that larger, hydrophobic  $P_1'$  residues (leucine, phenylalanine, tyrosine and tryptophan) are accepted in this region.<sup>35</sup>

## 1.1.2 Mechanism



**Figure 1.8:** Catalytic cycle of cysteine proteases

The enzymatic activity of cysteine proteases is highly dependent on the active site catalytic dyad between Cys<sup>25</sup> and His<sup>159</sup>, which exist as the ion pair  $-S^- \cdots +H-Im$ .<sup>36,37</sup> Figure 1.8 illustrates the catalytic cycle, which begins with the active site thiolate of Cys<sup>25</sup> performing a nucleophilic attack on the scissile bond (a) forming the tetrahedral intermediate (b), which is stabilised by a local hydrogen bond network termed the oxyanion hole (figure 1.8, left). Upon collapse of this intermediate protonated His<sup>159</sup> acts as an acid, protonating the amine moiety and forming the thioester-enzyme intermediate (c). The thioester then undergoes hydrolysis (d) generating the corresponding carboxylic acid and restoring the catalytic cysteine residue.

Mechanistically, the catalytic cycle of cysteine proteases shares many similarities with other protease classes, including the common sequence of nucleophilic attack, tetrahedral intermediate formation and collapse to the respective amine and carboxylic acid moieties.



In many instances, this also involves similar catalytic dyad, or triads utilising histidine as a catalytic residue. However, cysteine proteases are unique in that the catalytic histidine first acts as a general acid in the catalytic cycle (proton donor). This is in contrast to other protease classes, such as serine proteases, which first utilise their active site histidine as a general base. As such, cysteine proteases offer a unique hydrogen bond partner at the active site (protonated His), which can be targeted to enhance the selectivity towards cysteine proteases over other protease classes. Furthermore, all other protease classes operate through an oxygen centred nucleophile (i.e. the OH of serine and threonine, or by polarising a water molecule in metallo proteases). As such, cysteine proteases offer a softer nucleophile (sulfur centred) than the harder (oxygen centred) nucleophiles of other protease classes. By considering these unique catalytic features, it can be hypothesised that a softer electrophilic centre, which can take advantage of the the unique hydrogen bond network of a cysteine protease, will be intrinsically more selective towards cysteine proteases over other protease classes.

### 1.1.3 Cysteine Proteases Role in Disease

**Table 1.1:** Cysteine proteases, their host organism and associated disease state.

Cysteine Protease	Organism	Assosiated disease state
<u>Mammalian Proteases</u>		
Cathepsin B	Human	Cancer
Cathepsin L	Human	Cancer
Cathepsin K	Human	Osteoporosis / Rheumatoid arthritis
<u>Parasitic Proteases</u>		
<u><i>Plasmodium</i></u>		
Falcipain-1	<i>P. falciparum</i>	Malaria
Falcipain-2	<i>P. falciparum</i>	Malaria
Falcipain-3	<i>P. falciparum</i>	Malaria
<u><i>Trypanosoma</i></u>		
Cruzipain	<i>T. cruzi</i>	Chagas disease
Rhodizane	<i>T. rhodesiense</i>	Acute African trypanosomiasis
Brucipain	<i>T. brucei</i>	Chronic African trypanosomiasis
<u><i>Schistosoma</i></u>		
SmCB1	<i>S. mansoni</i>	Schistosomiasis
SmCB2	<i>S. mansoni</i>	Schistosomiasis

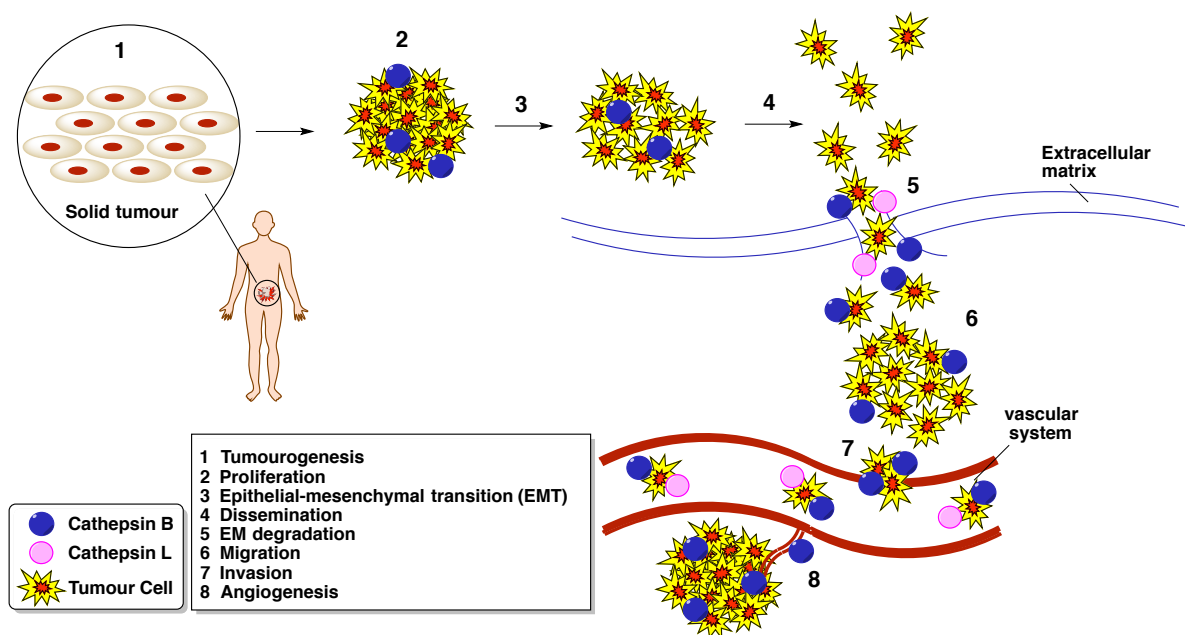
As papain-like CPs are ubiquitous throughout nature, they are responsible for a vast number of biochemical and physiological processes in living organisms. Members of the papain family (clan CA, family C1) have been identified from plant, viral, parasitic and mammalian origin. The primary function of CPs in their respective organisms is metabolic degradation of peptides and proteins. As such, dysregulation of cysteine proteases has

been implicated in many disease states. Mammalian CPs have been linked to the progression and development of diseases involving abnormal protein turnover.<sup>29,38-40</sup> Table 1.1 displays a few notable members of the papain superfamily including the some mammalian lysosomal cathepsin (Cathepsins B,L and K) and their associated disease states. Table 1.1 also displays some notable parasitic CPs related to plasmodium (responsible for Malaria), Trypanosoma (responsible for Chagas disease and African sleeping sickness) and schistosoma (responsible for schistosomiasis).

## Mammalian Cysteine Proteases in Disease

The main focus of this work will be on parasitic cysteine proteases, however there are some noteworthy mammalian proteases (the cathepsins) which will be briefly discussed. The cathepsins are implicated in a range of human disease states such as osteo-arthritis,<sup>41,42</sup> rheumatoid arthritis,<sup>43</sup> immune system and inflammation related diseases<sup>44</sup> and cancer.<sup>45</sup>

## The Role of Cathepsins in Cancer



**Figure 1.9:** Role of cathepsins B and L in tumour progression and metastasis. After initial tumour formation (Tumorigenesis, step 1), the following seven key steps of metastasis involve either cathepsin B, cathepsin L or both. These steps are; Proliferation (step 2), Epithelial-mesenchymal transition (EMT, step 3), dissemination (step 4) Extracellular matrix (EM) degradation (step 5), migration (step 6), invasion (step 7) and angiogenesis (step 8). Adapted from Tan et al.<sup>45</sup>

Cathepsins B and L are of particular interest as they show promise in the area of cancer treatment. Figure 1.9 displays some of the key areas in which cathepsins B and L have been implicated in cancer.<sup>45</sup> In general tumour metastasis and progression requires proteolysis of the surrounding environment for spatial expansion,<sup>46</sup> for angiogenesis (formation of blood vessels aiding tumour growth)<sup>47</sup> and for migration of transformed cells (metastasis).<sup>48</sup> The tumourigenesis stage is not thought to be reliant upon cathepsin B or L, although other cathepsins are thought to be involved.<sup>45</sup> Following on from this, the cancer cells proliferate and undergo a process known as epithelial-mesenchymal transition (EMT). EMT induces cellular changes in the cytoskeleton as well as protein expression and is common to processes such as wound healing, embryonic development and tumour metastasis.<sup>49</sup>

After EMT, dissemination of the cancer cells occurs. Cathepsin B has displayed a potential role in dissemination of squamous carcinoma cells and pancreatic cancer cells.<sup>50,51</sup> Vasiljeva *et al.* have shown that a reduction in invasion and migration of cancer cells correlated with reduced cathepsin B levels, which may offer circumstantial evidence of cathepsin B's role in dissemination of cancer cells.<sup>52</sup> According to the work of Tao *et al.* cathepsin L, along with multiple other lysosomal proteins, may be important for dissemination of cancer cells *via* the lymphatic system.<sup>53</sup>

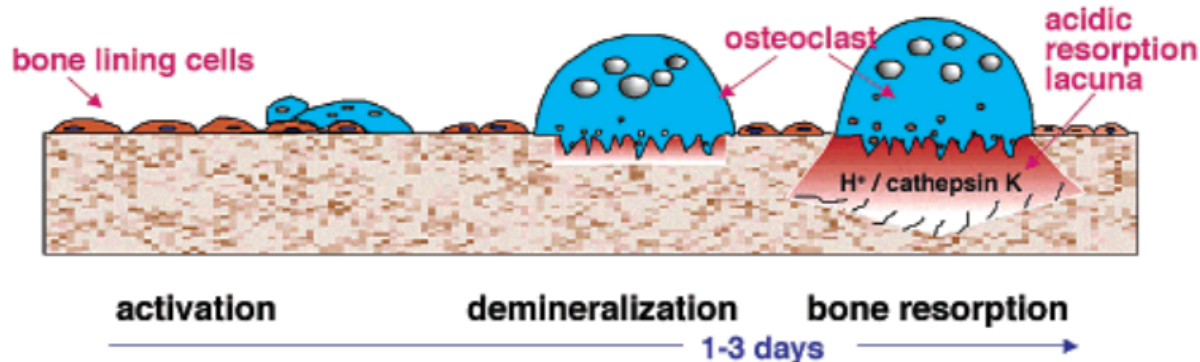
Dissemination increases the motility of the cancer cells, however, degradation of the extracellular matrix (ECM) is required to allow migration throughout the body. Cathepsin B can degrade components of the basement membrane and ECM, facilitating tumour progression,<sup>46</sup> whilst inhibition studies have demonstrated reduced cathepsin B activity correlated with reduced ECM degradation and, in turn, the invasiveness of inflammatory breast cancer.<sup>54</sup> Similarly, Withana *et al.* found that administration of a cathepsin B inhibitor (CA-074) reduced metastasis in tumour bearing animals.<sup>55</sup> It has also been proposed that the increased anaerobic glycolysis in cancer cells results in acidification of the local environment, which facilitates extracellular activity of cathepsin L.<sup>45</sup>

The final stages of invasion and angiogenesis complete the process of metastasis of cancer cells. Cathepsin B is necessary for fibroblast mediated invasion in esophageal cancer, with strong links between overexpression and angiogenesis in restricted colon adenocarcinoma,<sup>47,56</sup> while cathepsin L has been reported to increase the invasiveness of ovarian cancer cells *in vitro*.<sup>48</sup> Furthermore, a cathepsin L knockout study (murine model) significantly reduced tumour invasion and proliferation.<sup>57</sup>

As a whole, the data present in the literature supports the cathepsins as a potential therapeutic target in cancer.<sup>55,58</sup> Indeed, a recent study has linked Cathepsin B inhibition to reduced bone metastasis in breast cancer,<sup>55</sup> which is significant as metastasis to the bone is recognised as a major cause of morbidity. With this in mind, novel cysteine protease

inhibitors such as the thiosulfonates explored in this work may find future applications in this area.

## The Role of Cathepsins in Bone Resorption and Osteoporosis



**Figure 1.10:** The role of Osteoclasts and Cathepsin K in bone resorption.<sup>44</sup>

Bone development and homeostasis is a complex process, which is highly reliant on the balance of bone formation and resorption processes. Osteoblasts and osteoclasts are the key cells responsible for the fine balance between bone formation and resorption.<sup>59</sup> Osteoclasts are the specialised cells that attach to the bone surface during remodelling cycles and facilitate bone resorption as shown in figure 1.10. Osteoclasts form a cavity or depression in the bone (known as the lacuna) by demineralising the bone through creating a local acidic environment at their site of attachment. This unveils the organic matrix (approximately 90% type I collagen) to be degraded by proteases released by the osteoclasts, before osteoblasts refill the lacuna with new bone, completing the remodelling cycle.<sup>60</sup>

Early inhibition studies linked cysteine protease activity with bone matrix degradation.<sup>61,62</sup> For the most part, both cathepsins B and L have been found in bone tissue and are proposed to have roles in collagen degradation.<sup>63,64</sup> However, the later discovery that cathepsin K is selectively expressed by osteoclasts and secreted into the resorption lacunae,<sup>65,66</sup> suggests that cathepsin K is central in degrading the organic component of the bone matrix.<sup>67,68</sup> Additionally, Li *et al.* found that under physiologically relevant conditions cathepsin K represented the only lysosomal collagen degrading activity.<sup>69</sup> The observations of Li *et al.* also correlate with the finding that inhibiting cathepsin K, both *in vivo* and *in vitro*, reduced bone resorption.<sup>70</sup>

Understanding the role of cathepsin K led to questioning its role in disease states that are characterised by loss of bone density. Osteoporosis is a progressive disease characterised by loss of bone density and microstructural deterioration, resulting in an increased risk of bone fractures due to skeletal fragility.<sup>71</sup> The work of Kiviranta *et al.* displayed that

Cathepsin K accelerated the turnover of metaphyseal trabecular bone in a transgenic murine model, which is an indicator of osteoporosis.<sup>72</sup> Furthermore, Morko *et al.* used a transgenic mouse model to highlight the effect of overexpressing cathepsin K, which led to increased osteoclastic bone resorption. This resulted in an increased porosity of the diaphyseal cortical bone and osteopenia (a common osteoporosis pre-cursor) of metaphyseal trabecular bone.<sup>73</sup> Under conditions that increased bone resorption is known to occur, such as with bone resorption agents or during immobilisation (lack of movement, often related to paralysis or bed rest in humans), the expression of cathepsin K is seen to be up-regulated.<sup>74,75</sup> Leading on from this, osteoporosis patients have been found to have significantly higher cathepsin K levels than their corresponding controls.<sup>76</sup> In these patients the serum level of cathepsin K correlated with markers of bone turnover, bone mineral density and the incidence of nontraumatic fractures. As a whole, the data present in the literature validates cathepsin K as a suitable target for disease states involving increased bone resorption and degradation, such as osteoporosis.<sup>77,78</sup> Indeed, more recent clinical trials have unveiled the successful application of a cathepsin K inhibitor for treatment of osteoporosis,<sup>79</sup> with three other cathepsin K inhibitors reported for osteoporosis treatment (odanacatib, relacatib and balicatib).<sup>80–82</sup>

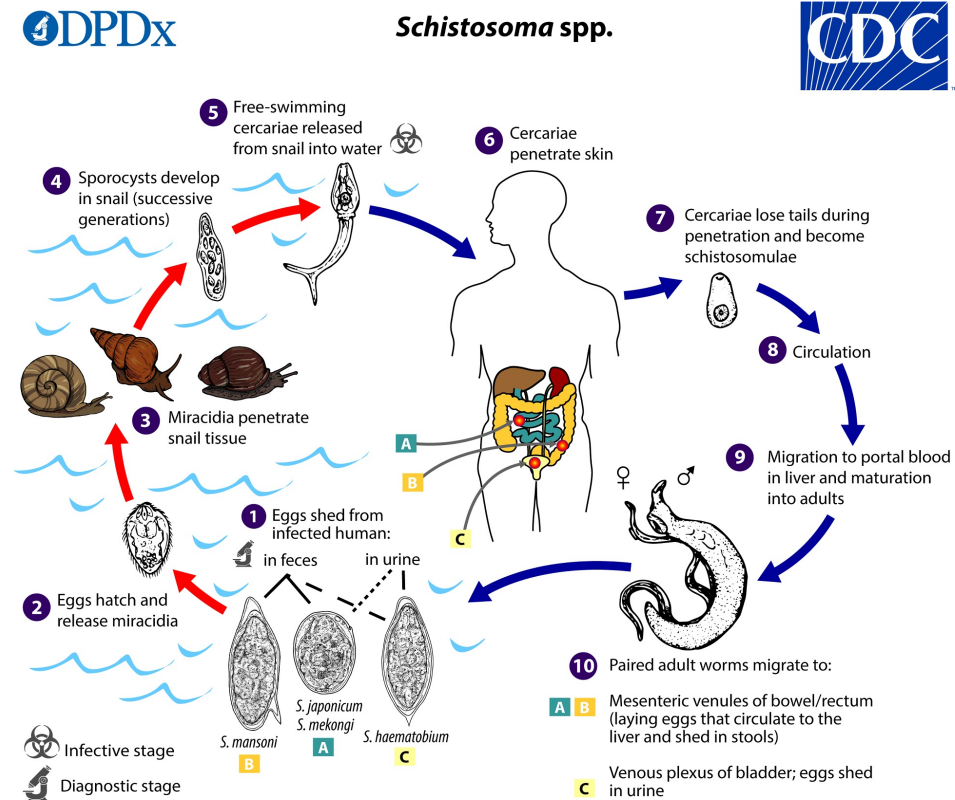
Although mammalian cysteine proteases (namely the cathepsins) will not be specifically targeted in this work, the involvement of cysteine proteases in such a broad range of human disease states, such as cancer and osteoporosis, has resulted in their interest as therapeutic targets.<sup>83</sup> Thus, this work on the development of thiosulfonates as cysteine protease inhibitors may find application in these areas in future.

## **Cysteine Proteases and Neglected Tropical Diseases**

Globally, parasitic disease has placed a burden on both domestic animals and humans alike. Combined with emerging drug resistance, the development of new chemotherapeutics is of the utmost importance. Cysteine proteases, which play a key role in the life cycle of helminth and protozoan parasites, offer a key target for the development of new chemotherapies. A common trend amongst these disease states is a high dependence on CPs within the papain-like family (Clan CA, family C1). In this section, three key neglected tropical diseases (NTD's) will be discussed. This will begin with Chagas disease, for which a cysteine protease inhibitor was developed playing a pivotal role in progressing cysteine protease inhibitors as therapeutic targets of NTD's. Next, the role of papain like CPs will be discussed in Malaria and finally Schistosomiasis. This will highlight the parasitic lifecycles and reliance on papain like cysteine proteases, validating CPs as suitable chemotherapeutics in each disease state. It is worth noting that multiple other (related) parasitic diseases such as Leishmaniasis, African sleeping sickness and Giardiasis offer

equally valid targets for similar reasons discussed herein. There are a number of reviews which cover the role of cysteine proteases in parasitic organisms, providing substantial background information and outlining the roles specific proteases play.<sup>11, 44, 84, 85</sup>

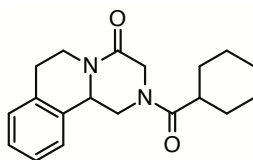
## The Role of Cathepsin Like Proteases in Schistosomiasis



**Figure 1.11:** Lifecycle of *Schistosoma* In both the snail vector (left, red) and human (right, blue) stages. (Centre for disease control and prevention [CDC], 2019).

Schistosomiasis (bilharzia) is a chronic, debilitating neglected tropical disease (NTD) that is second only to malaria in terms of public health importance. There are an estimated 200 million people infected, with a total of 400 million at risk. The disease is caused by blood flukes belonging to the genus *Schistosoma* with the majority of human infections caused by *Schistosoma japonicum*, *S. haematobium* and *S. mansoni*. In each case, the blood flukes have a complex life cycle (shown in figure 1.11) that involves an intermediate hosts (snails) and a final vertebrate host (often humans). During the life cycle, eggs are shed from their vertebrate host through urine or faeces to enter into fresh water. These eggs hatch (in an aqueous environment) and infect the intermediate snail host where they mature to the infective cercariae stage. The cercariae are free swimming, allowing them to come into contact with and infect their vertebrate host by penetration through the skin. Upon entering the final host the cercariae lose their tail, becoming schistosomulae

before maturation to adult schistosomes. The adult schistosomes lay eggs, completing the parasites life cycle. It is during the final stages that the most significant damage occurs to the host tissues, with schistosomes penetrating the walls of the small intestine or veins (*S. japonicum* and *S. mansoni*) or urinary bladder (*S. haematobium*). This results in chronic and debilitating symptoms such as growth stunting, cognitive impairment and lassitude.<sup>86</sup>



**Praziquantel**

**Figure 1.12:** Structure of praziquantel, the only actively administered schistosomiasis medication.

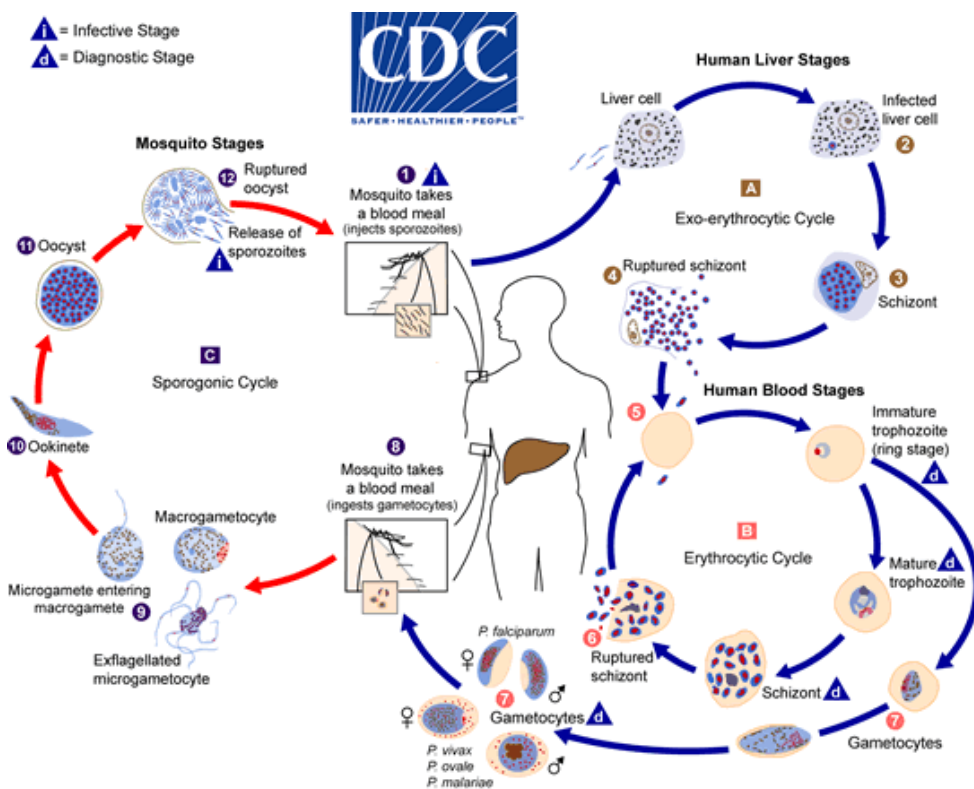
Large-scale elimination activities are ongoing and praziquantel, the only schistosomiasis medication, is widely deployed by mass drug administration programs (figure 1.12). Praziquantel is effective against all species of schistosomes which infect humans, however juvenile parasites are less susceptible to treatment than adults.<sup>87-89</sup> Despite the current success of praziquantel, relying on a single drug for treatment of 200 million people is concerning.<sup>90</sup> As such, new anti-schistosomal drug candidates are urgently needed, especially since reduced susceptibility to praziquantel has been reported for many decades.<sup>91,92</sup>

Many steps of the schistosomes life cycle are highly dependent upon papain-like cysteine proteases. For example, cathepsin B-like and cathepsin L-like activities have been described at different stages of trematode development.<sup>93</sup> Furthermore, the cercariae utilise proteolytic activities to penetrate the human skin and evade the immune response. In the case of *S. japonicum* cathepsin B activity has been found in the cercarial secretions,<sup>94</sup> with a cathepsin B2-like cysteine protease being identified as the main penetration tool.<sup>95</sup> Schistosomes have also been reported to utilise papain-like cysteine proteases such as legumain, cathepsin B, D, C and L to degrade the hemoglobin of hosts, which is required for nutrition growth, development and reproduction of the parasite.<sup>93</sup> In *S. mansoni* two isoforms of cathepsin L have been isolated (SmCL1 and SmCL2),<sup>96,97</sup> as well as a cathepsin B-like protease (SmCB1 and SmCB2).<sup>98,99</sup> SmCB1 has been reported as having high hemoglobinase activity in infected mammals, representing the major proteolytic activity of *S. mansoni*.<sup>100</sup>

Due to the parasites dependence on cysteine proteases throughout its life cycle, cysteine protease inhibitors offer valuable therapeutic targets. Indeed work by Abdulla *et al.* has highlighted that the cysteine protease inhibitor K11777 (structure shown in figure 1.16),

which was originally developed for treatment of Chagas disease, offers a novel chemotherapeutic for schistosomiasis.<sup>101</sup>

## The Role of Falcipains in Malaria

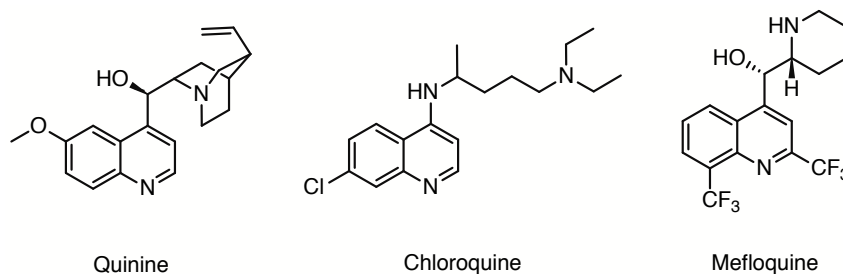


**Figure 1.13:** Lifecycle of *Plasmodium* In both mosquito (left, red) and human (right, blue) stages. Infective species of protozoa is trypanosome. (Centre for disease control and prevention [CDC], 2019).

Human malaria represents the most prevalent neglected tropical disease, with an estimated 219 million cases claiming more than 435,000 lives in 2017.<sup>102</sup> Malaria is caused by one of four *Plasmodium* protozoa (*Plasmodium malariae*, *P. vivax*, *P. ovale* and *P. falciparum*).<sup>103–106</sup> *P. falciparum* is responsible for the majority of malaria deaths, particularly in children under the age of 5.<sup>107</sup> The lifecycle of malaria is shown in figure 1.13, comprising both vector stages (in the female anopheles mosquito) and host stages (human). Infection begins when a mosquito vector takes a blood meal, injecting sporozoites into the host. These sporozoites relocate to and infect liver cells (hepatocytes), where they undergo maturation as part of the exo-erythrocytic cycle. Following on from this the hepatocytes rupture, which is likely to involve cysteine proteases,<sup>108</sup> releasing merozoites. The merozoites go on to infect red blood cells (erythrocytes), signifying the beginning of



the human blood stage of the infection. It is this erythrocytic life cycle (part of the human blood stage) which is responsible for the characteristic symptoms of malaria such as jaundice, fever, anemia and paroxysm.<sup>44</sup> A secondary stage of the life cycle occurs in tandem with the erythrocytic stage, whereby gametocytes replicate in the blood plasma. These gametocytes sustain the infection and allow transmission back to the vector stage (mosquito) completing the parasitic life cycle.

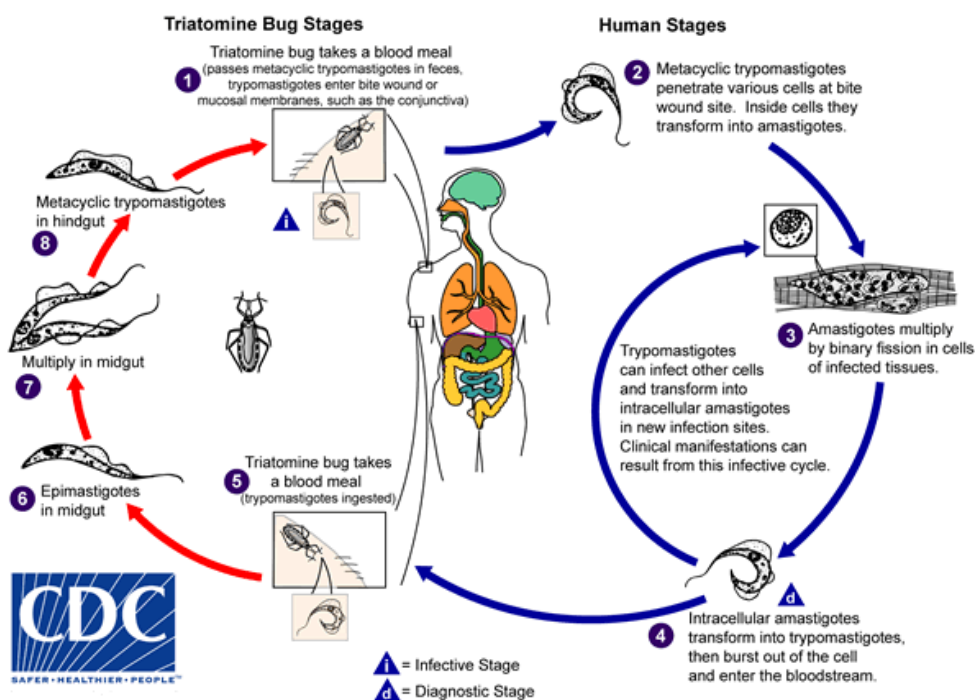


**Figure 1.14:** Structure of Quinine, Chloroquine and Mefloquine used for treatment of Malaria.

Current antimalarial drugs include quinine, chloroquine and mefloquine (figure 1.14), which *Plasmodium* are beginning to display resistance towards, especially in *P. falciparum*.<sup>109</sup> Methods to overcome resistance include combination therapies, with current artemisinin based combination therapies (ACTs) leading the way.<sup>110</sup> In addition to ACTs, haemoglobin degrading cysteine proteases are emerging as new promising drug targets in the treatment of *P. falciparum*.<sup>111,112</sup> *Plasmodium falciparum* expresses three key papain-like CPs, falcipain 1, 2, and 3, which are central in providing amino acids and nutrition through hemoglobin hydrolysis, facilitating parasitic development.<sup>113,114</sup> The falcipains which are most closely related to hemoglobin hydrolysis are falcipain 2 and 3, being expressed in the acidic food vacuole of the parasite where hydrolysis occurs.<sup>115,116</sup>

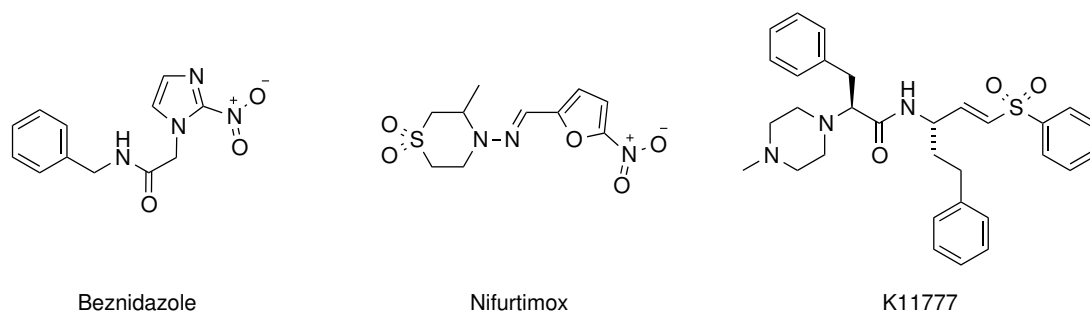
Due to the parasitic life cycles dependence on CPs, cysteine protease inhibitors have been recognised as potential chemotherapeutics.<sup>117</sup> After identifying the falcipains as key proteases for hemoglobin degradation,<sup>118</sup> Rosenthal *et al.* went on to show the antimalarial effects of vinyl sulfone based cysteine protease inhibitors.<sup>119</sup> As such, this work on the development of thiosulfonates as novel cysteine protease inhibitors may lead on to new anti-malarial chemotherapeutics.

## The Role of Cruzipain in Chagas disease



**Figure 1.15:** Lifecycle of *T. Cruzi*. In both triatomine bug (left, red) and human (right, blue). Infective species of protozoa is trypomastigote. (Centre for disease control and prevention [CDC], 2016).

Chagas disease is a neglected tropical disease that affects approximately 8 million people worldwide.<sup>120</sup> Chagas was previously confined to Latin America, where the disease is now considered endemic, however an increase in the number of worldwide cases has resulted in the disease being recognised as a global challenge.<sup>121</sup> Chagas is caused by the *Trypanosoma cruzi* protozoa, which is mainly transmitted through vector means (figure 1.15), namely the triatomine bug. After a brief acute phase, Chagas causes neurological disorders, results in megaorgans (enlargement of mainly the colon and oesophagus) and damages the heart muscle (leading cause of infectious myocarditis in latin america).<sup>122</sup>



**Figure 1.16:** Structure of Beznidazole, Nifurtimox and K11777.

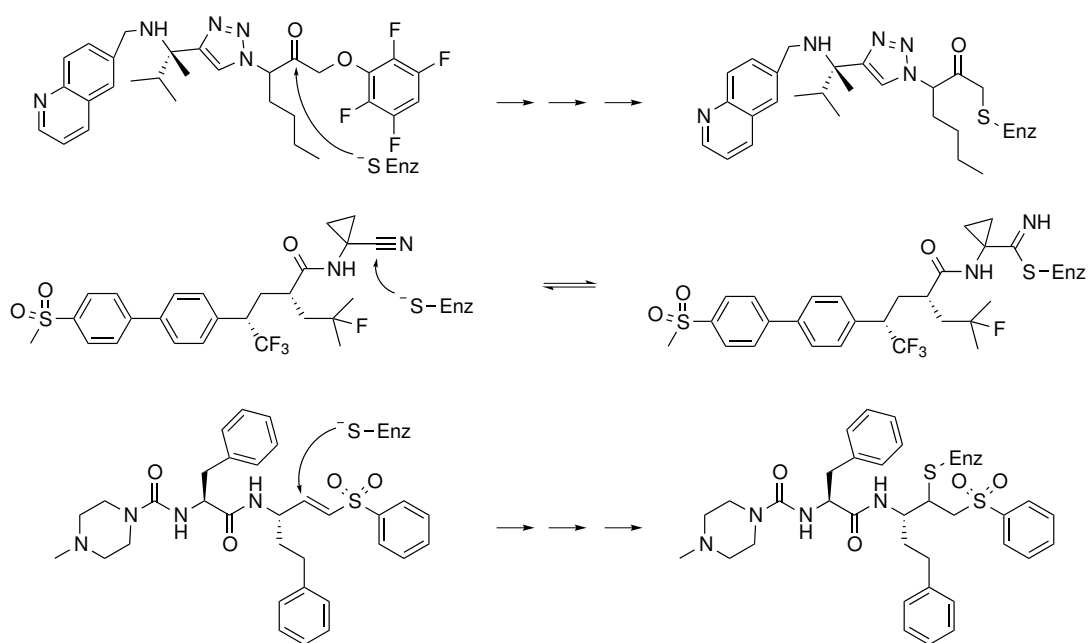
Current therapeutics, such as benznidazole and nifurtimox (figure 1.16), have limited efficacy and severe side effects hence alternatives are desired.<sup>123</sup> The cysteine protease cruzipain (also known as cruzain) is a cathepsin L-like protease and the major proteolytic enzyme present at all stages of the parasitic life cycle.<sup>124</sup> The highest levels of cruzipain are found in the epimastigote stage and the omnipresent nature of cruzipain makes this an enticing target, with potential to halt the parasite at any stage of its development.<sup>125</sup> Thomas *et al.* have shown that the overexpression of cruzipain is linked with metacyclogenesis (the process of parasitic maturation into the highly infective metacyclic trypomastigote form).<sup>126</sup> Cruzipain has also been implicated in the kinin cascade, which favours parasitic invasion in host cells that express kinin receptors.<sup>127</sup>

The realisation that *T. cruzi* was highly dependent on the cysteine protease cruzipain led on to inhibition studies. Inhibition of cruzipain has been shown to arrest cellular development by impairing *in vitro* host cell invasion, blocking amastigote replication and preventing amastigote and trypomastigote differentiation.<sup>128,129</sup> Following on from this, a vinyl sulfone inhibitor of cruzipain (K11777, figure 1.16) has been developed and toxicological, ADME and pharmacodynamics data has been reported in non-human primates.<sup>130,131</sup> One of the key attributes of K11777 is its selectivity for cruzipain over the mammalian cysteine proteases. Displaying this selectivity was a significant step forward in validating CPs as therapeutic targets of parasitic infections, due to the high homology between parasitic and mammalian CPs. With the current success of cysteine protease inhibitors in the treatment of Chagas disease, this work on the development of thiosulfonates as a novel class of cysteine protease inhibitors may find future applications in this area.

## 1.2 Current Cysteine Protease Inhibitors

As the catalytic mechanism is dependent upon an active site nucleophile, a common inhibitor design is to incorporate an electrophile on a small scaffold or ‘backbone’, often designed to confer specificity, in order to trap the active site nucleophile. This is often termed an electrophilic trap, or warhead. Many warheads are present in the literature, falling into three broad categories: (1) the 1,4-Michael acceptors (e.g. fumarates,<sup>134</sup> vinyl  $\alpha$ -keto esters, amides and acids<sup>135</sup> and vinyl sulfones).<sup>136</sup> (2) Irreversible inhibitors (e.g. tetrafluorophenoxymethyl ketones,<sup>137</sup> aziridine-2,3-dicarboxylates,<sup>138</sup> epoxyketones, epoxyamides,  $\alpha$ -alkoxyketones and diacyl-bis-hydrazides)<sup>139</sup> and finally (3) reversible inhibitors (e.g. nitriles,<sup>140</sup> aldehydes<sup>141</sup> and thiosemicarbazones).<sup>142</sup>

An example of an inhibitor from each of the three broad categories is given in figure 1.17. These include the Cathepsin K inhibitor Odanacatib, designed for treatment of osteoporosis, as well as the vinyl sulfone inhibitor K11777, designed for the treatment of Chagas

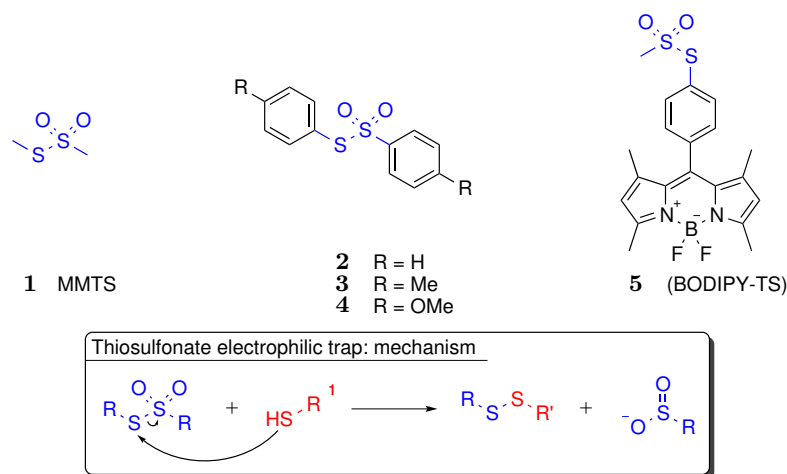


**Figure 1.17:** Current cysteine protease inhibitors, from top to bottom: The cruzipain inhibitor ( $\alpha$ -alkoxyketone warhead, irreversible inhibitor, top).<sup>132</sup> The cathepsin K inhibitor Odanacatib (Nitrile warhead, reversible inhibitor, middle).<sup>133</sup> The cruzipain inhibitor K11777 (Vinyl sulfone warhead, 1,4-Michael acceptor, bottom).<sup>130</sup>

disease. Interestingly, an analysis conducted in 2018 found that Michael acceptors dominate the landscape of cysteine protease inhibitors, accounting for approximately 70% of the published compounds.<sup>143,144</sup> Furthermore, the dependence on a specific warhead sub class is placing more emphasis on the initial non-covalent binding interactions, with warheads being viewed as ‘off the shelf’ reactive moieties for installation and generation of covalent inhibitors. This is resulting in a design paradigm whereby inhibitors are being designed purely on the basis of reversible inhibitors, with a known binding mode, placing reactive warheads in the correct spacial positioning. This is a process which placed all optimisation on the non-covalent interactions, preventing optimisation of the covalent interactions. As the overall inhibitor design has contributions from both the non-covalent scaffold and the warhead, current methods are preventing the parallel optimization of the non-covalent and covalent interactions. As a result, the literature is becoming increasingly biased towards proven chemistries for covalent bond formation, potentially ignoring valuable characteristics such as exploring new chemical space, chemically selective reactions and tunable reactivity profiles. This is exemplified when looking towards natural products, which offer a significantly wider range of covalent inhibition mechanisms such as oxidations, substitutions, nucleophilic additions, addition-eliminations, non-Michael additions as well as Michael additions.<sup>145</sup> This work aims to expand the chemical toolbox, focusing on the thiosulfonates as a novel electrophilic trap with inherent chemical selectivity towards its target, the cysteine proteases. In a similar manner to natural products, the thiosulfonates

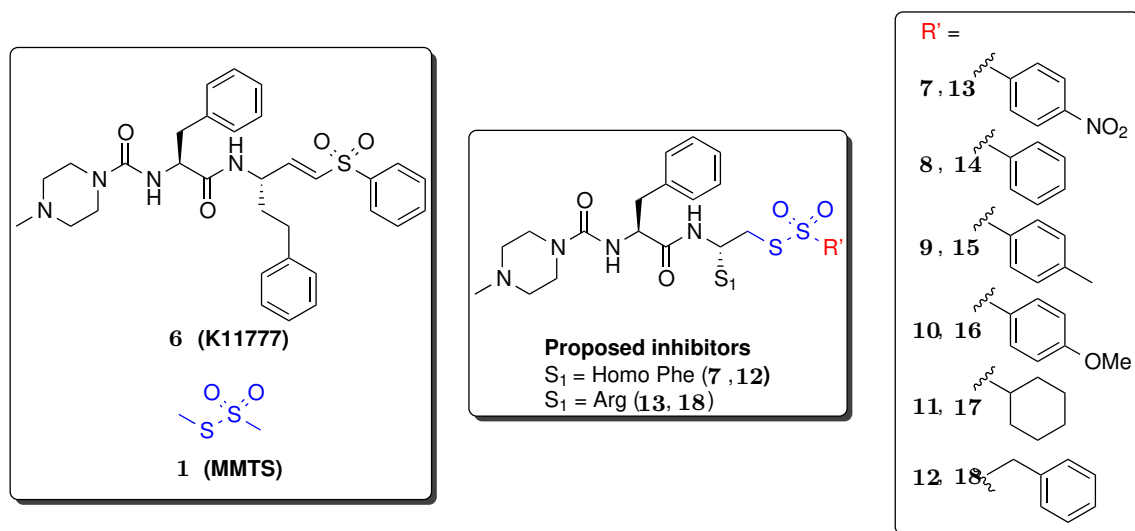
move into the underexplored chemical space of substitution-based electrophilic traps. Furthermore, the thiosulfonates display a highly tunable nature in their reactivity profiles. This is particularly interesting when considering the recent study by Abranyi *et al.*, which highlights that warheads operating through a substitution mechanism cover the same range of reactivity in a biological environment as the ubiquitous Michael acceptors. Furthermore, this study recognised the potential to access differing levels of selectivity towards biologically relevant nucleophiles by exploring under-developed chemistries and control of the reactivity profiles.<sup>143</sup>

## 1.3 The Thiosulfonate Moiety



**Figure 1.18:** The thiosulfonate, methyl methane thiosulfonate (MMTS, **1**), simple symmetrical thiosulfonates explored by Santos *et al.* (**2,3** and **4**)<sup>146</sup> and the 'switch on' thiol sensing fluorophore by Ge *et al.* (BODIPY-TS, **5**).<sup>147</sup>

There are few examples of the use of thiosulfonates as cysteine protease inhibitors in the literature other than the simple methyl methanethiosulfonate (MMTS). MMTS can be found in many plants and was isolated from cauliflower extracts by Nakamura *et al.* in 1996.<sup>148</sup> After its isolation Reddy *et al.* went on to show its chemo-preventive effect on colon cancer, demonstrating the link between thiosulfonates in diet and therapeutic potential.<sup>149</sup> There has also been further interest in the link between diet and cancer risk, such as exploring the various sulfur compounds (including thiosulfonates) of garlic and their cytotoxic effect in oesophageal cancer cells.<sup>150</sup> MMTS itself is regularly used as a reversible thiolating reagent in studies of the natural thiol-disulfide oxidation state of proteins,<sup>151</sup> for studying protein S-nitrosylation,<sup>151,152</sup> and for evaluation of the role of both catalytic and structural cysteine residues on enzyme activities.<sup>153–155</sup> A significant next step was the use of functionalised methane thiosulfonate derivatives by Matsumoto *et al.*<sup>156</sup> in post translational modifications of subtilisin mutants, which incorporated a cysteine by directed site mutagenesis. Since these early discoveries some exploration into thiosulfonates as protease inhibitors has been carried out with Santos *et al.* correlating insecticidal activity with reduced protease activity upon treatment with simple thiosulfonates, although no efforts were made to specifically target cysteine proteases (figure 1.18, compounds **2**, **3** and **4**).<sup>146</sup> The thiosulfonate moiety has also been explored as a bioisostere of the mesylate group for use in structure activity relationships by Gabriele *et al.* in order to improve the stability of compounds and avoid the multiple potential alkylations, facilitated by the mesylate group.<sup>157</sup> More recently, thiosulfonates have found an interesting application as 'turn on' fluorescent probes for thiol sensing (figure 1.18, BODIPY-thiosulfonate, **5**).<sup>147</sup>



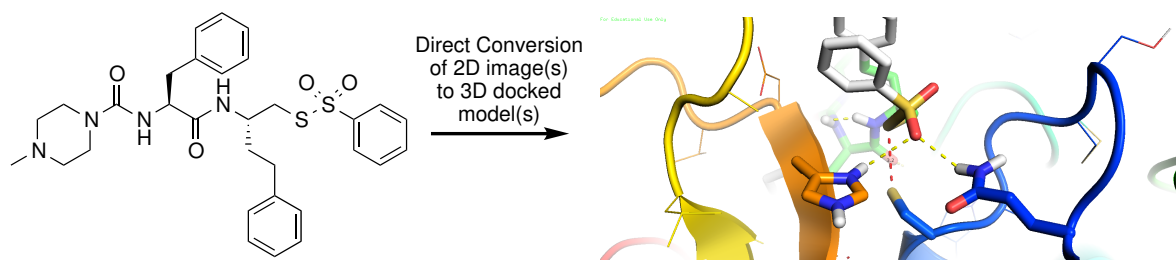
**Figure 1.19:** Proposed thiosulfonate inhibitor constructs building on the success of K11777, with a warhead inspired by methyl methane thiosulfonate (MMTS).

As a moiety that appears to be abundant in nature, it would be interesting to harness this potential electrophilic trap in the design of specific cysteine protease inhibitors. The aforementioned applications exploited the nature of the thiosulfonate as an excellent thiophile for highly chemoselective targeting of protein thiols. This is in part due to the soft electrophilic centre of the thiosulfonates (a bivalent sulfur), offering chemoselectivity towards sulfur centred nucleophiles (figure 1.18). In this work, inspiration has been drawn from MMTS, which serves as an SH specific electrophile, combined with the successful vinyl sulfone inhibitor K11777 (**6**) serving as a lead compound (figure 1.19). It was intended to use the backbone of K11777 as a foundation for thiosulfonate inhibitors, with K11777 its self offering a valid reference compound. From this starting point, rational modifications to both the backbone and warhead design will be made. It has been hypothesised that modification of the primed side residue R' (figure 1.19) group can be used to control reactivity resulting in highly tunable and specific cysteine protease inhibitors. The underlying hypothesis was that by modifying the substitution of aromatic analogues, the electrophilicity of the bivalent sulfur could be altered (compounds **7-10** and **13-16**). This trend will then be further explored by moving away from aromatic R' substituents and on to aliphatic ones, such as cyclohexane (compounds **11** and **17**) or the benzylic derivatives (compounds **12** and **18**).

## Chapter 2

# *DrawToDock*: Development of a Virtual Screening Platform to Guide Rational Design.

**Aim of *DrawToDock*: Making Computational Techniques More Accessible for Synthetic Chemists**



**Figure 2.1:** *DrawToDock*: A schematic representation. *DrawtoDock* allows for direct conversion of multiple two dimensional ligands, generated by common chemistry drawing software, to 3D docked models with target proteins.

Procedures which rapidly explore the complementarity between target binding sites and potential ligands present indispensable tools for medicinal chemists.<sup>158</sup> Medicinal chemists are living in an information age whereby 3D coordinates of potential drug targets, or closely related analogues, are readily available from crystal structures. As such, molecular modelling techniques allow for identification, optimisation and prioritisation of small molecule drug candidates. However, the knowledge required for the complex underlying protocols, and relevant experience in computational chemistry, may create a barrier between medicinal chemists and these highly valuable *in silico* methods. As such, a new tool has been developed named *DrawToDock* (Figure 2.1), which aims to overcome the barrier between these powerful predictive modelling techniques and the persons who require them most.



This is done by facilitating docking and virtual screening through recognisable chemistry drawing software, such as ChemDraw.

The fundamentals, advantages and limitations of docking are beyond the scope of this work with numerous reviews available to the interested reader.<sup>158–162</sup> Although many docking software packages are available, *DrawToDock* has been developed specifically to allow an intuitive link between chemistry drawing software, docking and virtual screening. This offers users a more direct route towards screening novel inhibitor constructs, which was desired in the development of thiosulfonates as cysteine protease inhibitors.

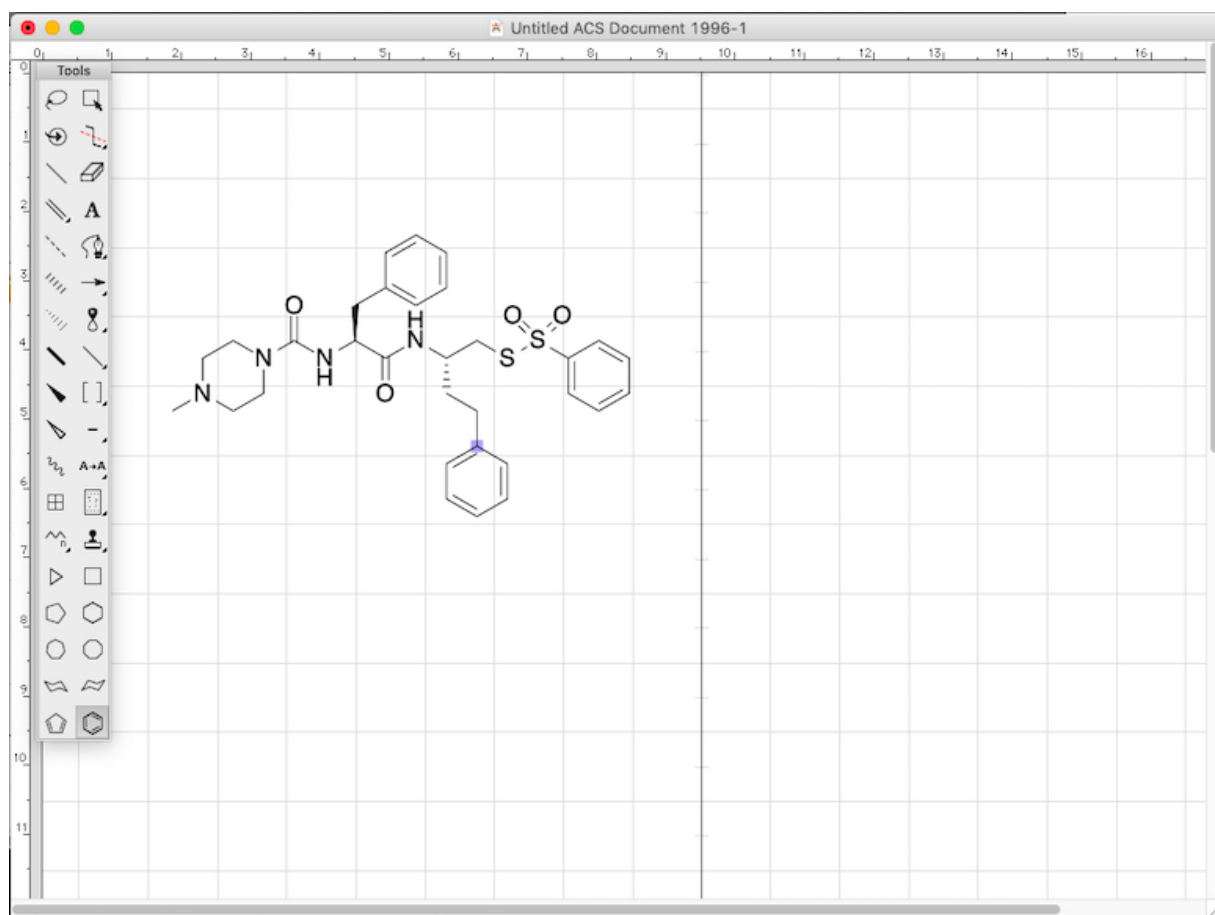
## 2.1 Developing *DrawToDock*: A Virtual Screening Platform Directly From Chemdraw Files

As *DrawToDock* is targeted towards chemists with little computational background a short note on computer specific nomenclature and the underlying principles is warranted. Bash shell scripting has been used to create the *DrawToDock* script, comprising a series of commands stored as a plain text document. The individual commands in a Bash script can also be executed individually through Command line on Windows operating systems (Win OS) or terminal for Macintosh Operating systems (Mac OS). These are in essence the process of operating a PC with strings of text rather than through a graphical user interface (GUI). A Graphical User Interface (GUI) is the collection of graphics, such as windows and toolbars, commonly used to control a software package. Through using Bash shell scripting, 2D images (such as a ChemDraw file) have been converted to 3D ligand constructs followed by docking of those ligands to a protein of interest. In order to achieve this, the correct software for each required process (i.e. transformation from 2D to 3D structures, docking, generating new directories for output files etc.) has been collated and contained within one convenient script. This avoids the need for users to jump between multiple software packages and GUI's to achieve the same outcome. Herein is presented the process of collating the required software and developing a script to automate the required processes, facilitating a virtual screen from 2D ChemDraw images, named *DrawToDock*.

After a survey of the freely available docking software, AutoDock Vina was identified as the optimal platform on which to base *DrawToDock*.<sup>163</sup> This was primarily due to the speed and accuracy with which Autodock Vina performs predictions, as well as the developer friendly way in which the software has been presented.<sup>163</sup> Moreover, Autodock Vina is executable through Bash shell scripting, allowing a direct route towards virtual screening with a Bash script. The required software for file conversions was identified

and combined into one Bash script, ending with the docking process which automatically performs iteratively for multiple ligands, offering what is commonly referred to as a virtual screen. There are three key steps to the docking process: (1) generating the ligands (small molecules) of interest, (2) obtaining the macromolecule of interest and (3) Bringing both together in the docking process. The following will outline each individual step to aid understanding, then the *DrawToDock* script, which automates the process, will be presented.

## Generating Ligands of Interest

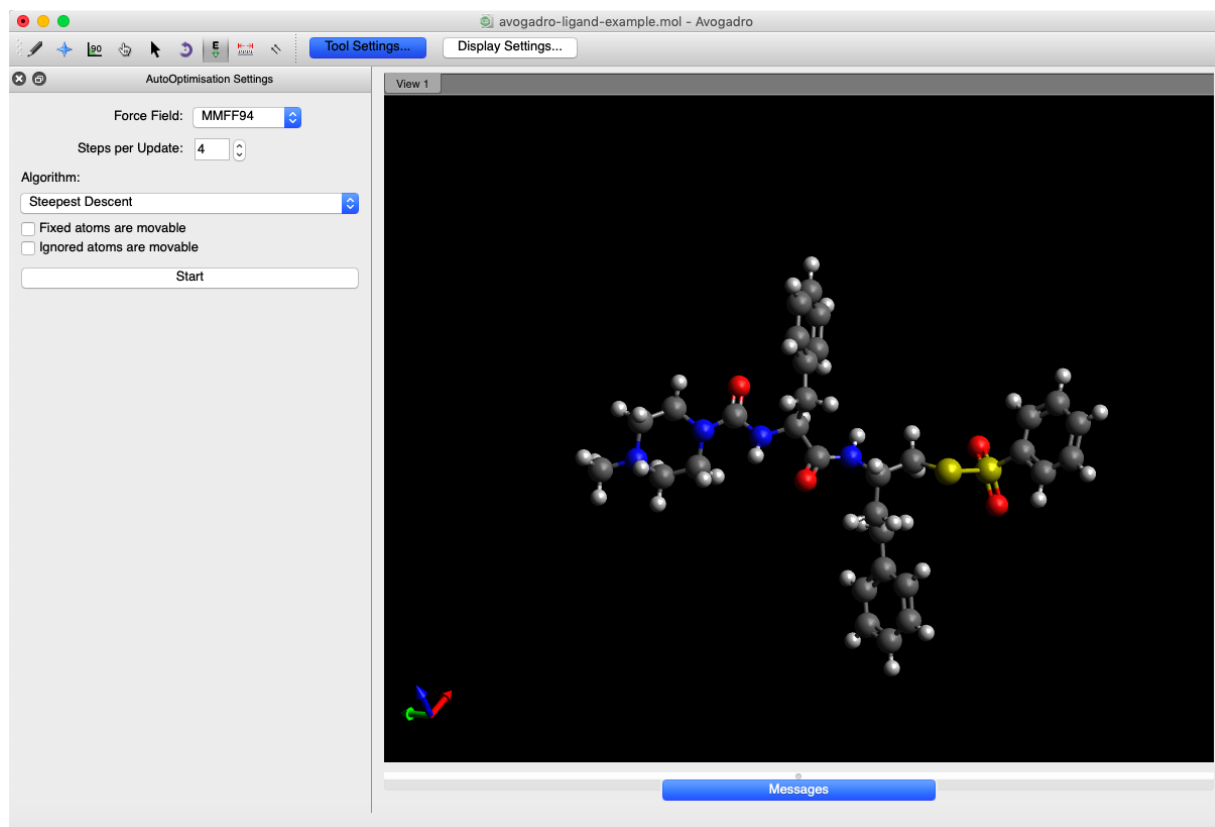


**Figure 2.2:** Generating ligands for docking with Chemdraw

The first step in any docking study involves obtaining ligands of interest. Often this can be achieved by downloading 3D structures from a database such as ZINC, the Cambridge structural database (CSD) or the Protein data bank (PDB).<sup>164-166</sup> Database methods are a convenient source of 3D ligand constructs however, they are limited to known small molecules and thus do not cover much of the chemical space available by newly developed chemistry and novel constructs. An alternative method is to generate 3D ligands by computational techniques, which requires greater user input and time. However, this

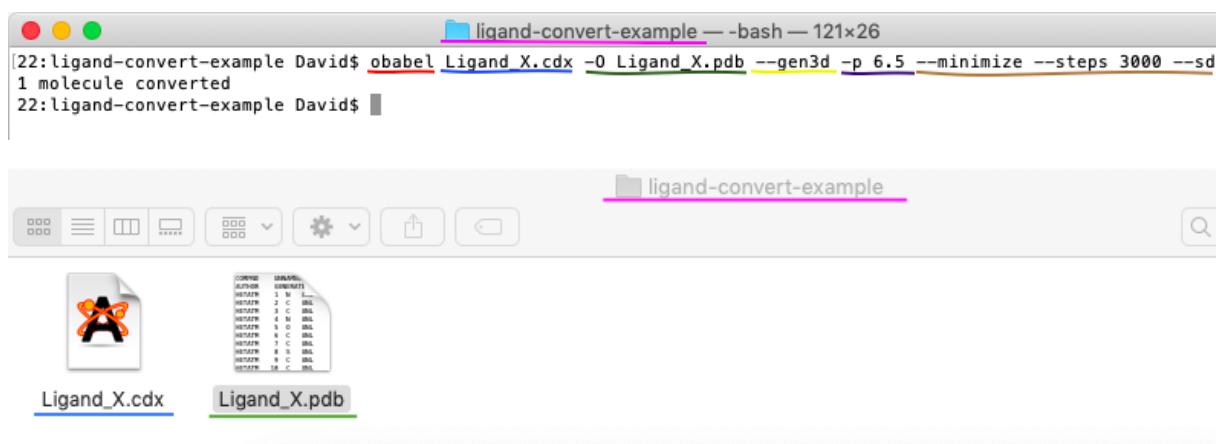
method allows for the introduction of novel chemical motifs and rational modifications to lead compounds. It was hypothesised that developing a method for generation of ligands by common chemistry drawing software such as ChemDraw (Figure 2.2) would bridge the gap between the convenience of database derived methods and the flexibility of computationally generated ligands.

With the easily available 2D images to hand generated through ChemDraw (figure 2.2) the next step was to generate 3D coordinates. The available methods for this fall into three main categories: Firstly, commercially available options such as advanced versions of ChemDraw, Corina and Discovery Studio.<sup>167,168</sup> Although these options undoubtedly function well they present both a financial barrier as well as the need to navigate additional software packages, increasing the time taken for ligand generation. Secondly, there are freely available web server based techniques such as PRODRG.<sup>169</sup> Such web based techniques are not suitable for use in virtual screening as they often place a limit on the daily number of submissions and incur a time delay when awaiting results to be sent back to the user. Third and finally are freely available, open source software packages such as Ghemical and Avogadro.<sup>170,171</sup> Not only do these remove the financial barrier imposed by commercial softwares but they often facilitate and encourage developers to modify and build on these techniques providing the ideal starting point for a virtual screening tool such as *DrawToDock*.



**Figure 2.3:** Using Avogadro to convert 2D .mol files to 3D .pdb files.

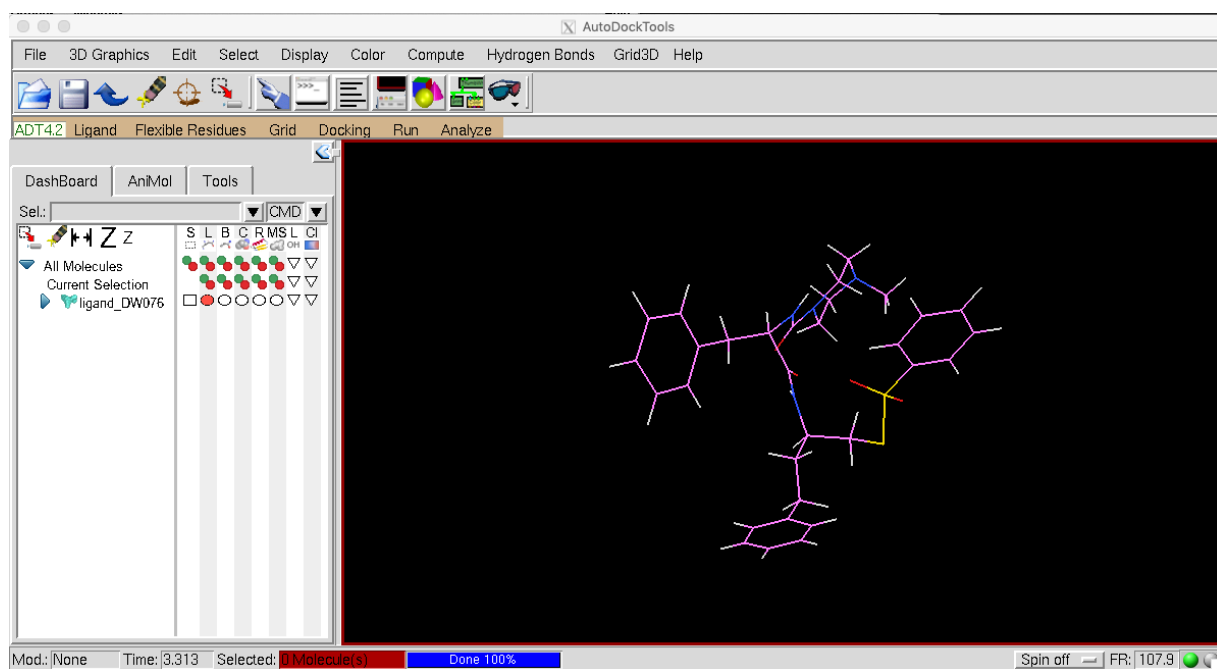
Initially, 3D coordinates were generated with the use of Avogadro, a freely available software which allows for the creation of 3D coordinates followed by energy minimisation from .mol files (figure 2.3).<sup>171</sup> Avogadro functioned well and allowed early docking studies to take place, however two main drawbacks were identified for improvement. Firstly input files were required in a .mol format, which required conversion from the native file format (.cdx in this instance), adding an additional step to the process. Secondly, the generation of 3D coordinates had to be conducted manually, through a graphical user interface (GUI), and then saved as a new output (.pdb) file. Although this process is possible, it becomes time consuming and laborious for multiple ligands.



**Figure 2.4:** Using Open Babel to convert 2D .cdx files to 3D .pdb files through terminal (MacOS). The terminal window (top) is overlaid with the file directory (bottom). Colour coding illustrated the function of each line of text in the terminal window and shows the corresponding file in the file directory where appropriate as follows: File directory (pink), Command to use Open Babel (red), input file name (blue), Output file name (green), Generate 3D coordinates (yellow), protonate ligand at pH 6.5 (purple), energy minimise over a defined number of steps with a steepest descending algorithm (brown)

Having identified the time consuming steps an approach to streamline the process was outlined. Looking into the inner workings of Avogadro it was apparent that Open Babel, a freely available chemistry file conversion software, was the underlying process responsible for generating 3D structures.<sup>172,173</sup> As such, Open Babel was explored and found to be executable directly through Bash scripting which is illustrated in figure 2.4. This removed the time consuming GUI dependent steps of Avogadro. Moreover, Open Babel offers a wide range of possible transformations which are well documented on the Open Babel Wiki.<sup>174</sup> This allowed for greater flexibility dependent on the desired docking parameters, in this instance pH dependent protonation states were employed. This allows the user to define a pH representative of the environment in which the target protein is located, thus accessing relevant ligand protonation states. The 3D coordinates were generated and then energy minimised with Molecular Force Fields (MFF94), allowing fast and accurate generation of 3D constructs.<sup>175</sup> Energy minimisation was performed to best

represent reasonable bond lengths and angles, and was conducted over 3000 steps with a steepest descending algorithm. Furthermore Open Babel facilitated direct conversion from ChemDraw (.cdx) files, removing additional conversion steps, which greatly decreased the time taken for file conversions. It is worth noting that Open Babel can also facilitate the above transformation from other common formats such as .mol or SMILES meaning that there is inherent flexibility for users of all backgrounds using differing chemistry drawing software (not just ChemDraw).



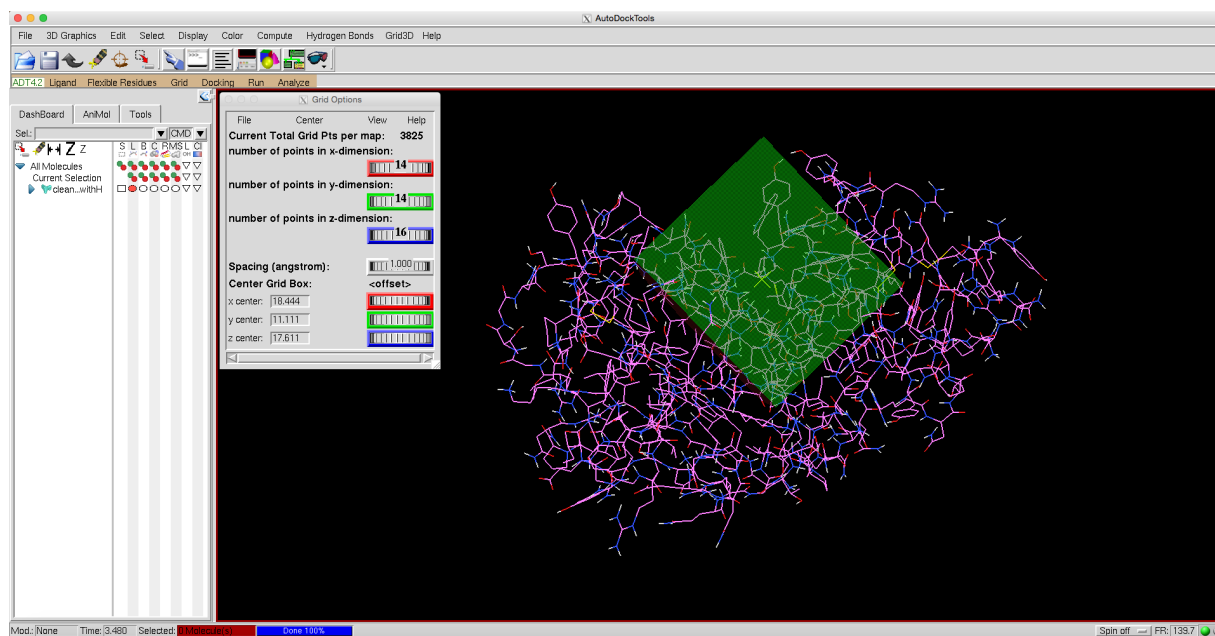
**Figure 2.5:** Converting ligand format from .pdb to .pdbqt with AutoDockTools GUI

Continuing on from generation of the 3D coordinates, software specific manipulations were required. Autodock Vina requires an input format known as .pdbqt, which is specific to AutoDock. As such this transformation was accomplished through AutoDockTools, a GUI provided by the makers of AutoDock shown in figure 2.5. Again, working through another GUI greatly increased the time taken to convert between file formats. As such, a python script provided by the developers of AutoDock, which can be found in the installation package of AutoDockTools, is available to convert from .pdb to .pdbqt formats. Executing this python script can be conducted through Bash scripting, again greatly decreasing the required user input time. This is achieved in a similar way to executing the previously discussed Open Babel command and is displayed in figure 2.4 by navigating through terminal to the correct directory and typing the following: `python ./prepare_ligand4.py -l Ligand_X.pdb`. Where 'python' calls on python to be used to execute the script (./) named `prepare_ligand4.py` with the input file `Ligand_X.pdb` which will generate the required .pdbqt output file.

An important feature with the above techniques is the ability to bypass GUI's. Although

GUI's are a welcome and intuitive method for conducting many computational tasks they come with associated time penalties such as installation, learning and execution. This poses a significant barrier towards their use for synthetic chemists and is one of the major drawbacks in exploring 'free' software packages. Thus users are faced with the option of either a 'computational' barrier, involving greater time taken searching for the correct selection of software and learning how to use it, or a financial barrier in paying for a software package which can address this issue. Either of these barriers are often enough to prevent many chemists from using computational techniques, which could potentially enhance their work. This is the fundamental issue *DrawToDock* ultimately aims to overcome by providing a free, fast and convenient docking solution based on open source software.

## Obtaining the Target Macromolecule

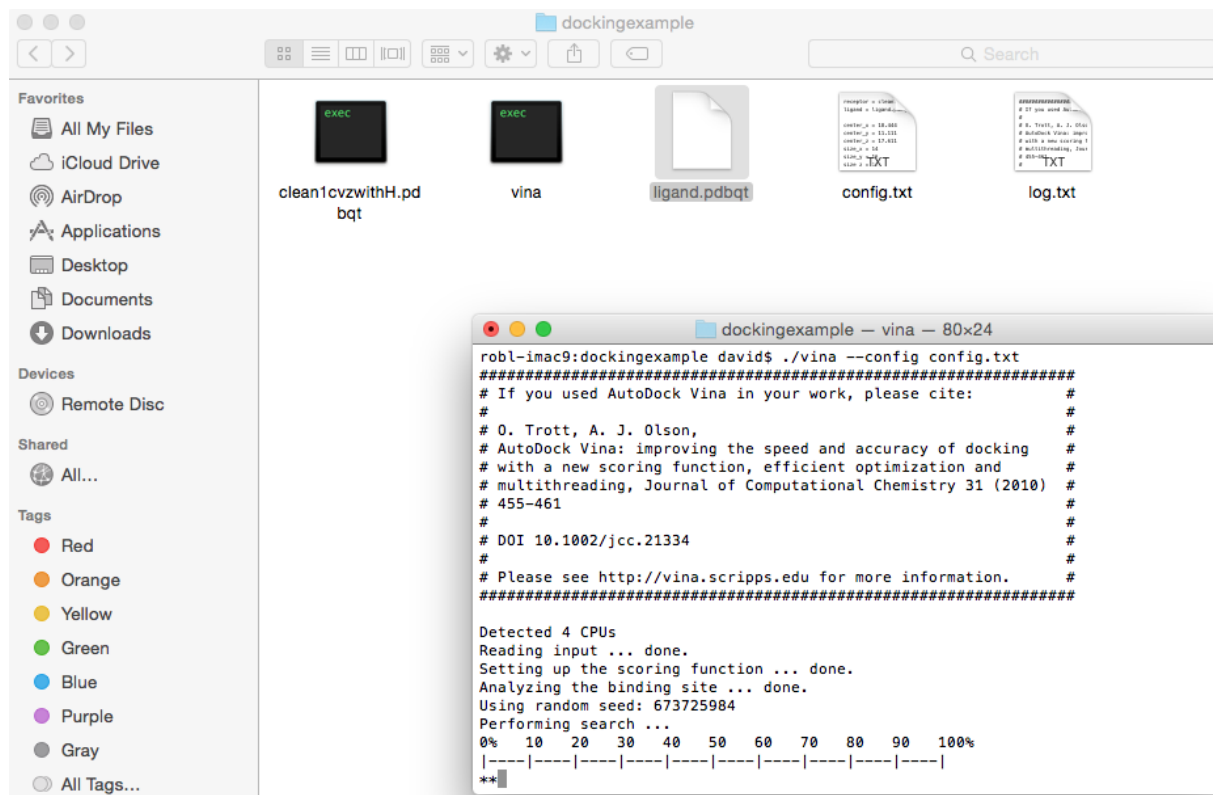


**Figure 2.6:** Defining the search space for docking with AutodockTools.<sup>176</sup> The search space is visualised with a 3D grid box and the corresponding coordinates, which define the search space, can be noted down for docking.

Obtaining the 3D coordinates of the macromolecule of interest was achieved by downloading a crystal structure from the protein data bank (PDB).<sup>166</sup> Subsequent preparation of the macromolecule was then conducted. This involved removing any unbound atoms such as counter ions and water molecules as well as any covalently bound inhibitors in the active site. This is a simple 'highlight and delete' procedure using Pymol,<sup>177</sup> although other molecular visualisation software, such as AutodockTools, would suffice.<sup>176</sup> Once the macromolecule had been prepared it then had to be saved as a .pdbqt file, again this is an AutoDock specific format and was therefore carried out through AutodockTools GUI.

AutoDock Vina operates by searching a defined space, which is defined by visual assessment in AutoDockTools as shown in figure 2.6. The corresponding coordinates were then saved in a configuration file (a simple text document defining the macromolecule file, coordinates to search, and ligand to dock). It is worth noting that this was a one time process for the macromolecule of interest and all subsequent docking simulations use the same macromolecule and configuration files, hence this is highly time efficient.

## Running Docking



**Figure 2.7:** Running AutoDock Vina, an example. Image displays directory containing the required files in the background: macromolecule (clean1cvzwithH.pdbqt), AutoDock Vina executable (Vina), Ligand (Ligand.pdbqt) and the configuration file (config.txt). A Log file (log.txt) is also generated upon running docking. The foreground shown AutoDock Vina running, giving live updates in the shell window.

Finally docking was conducted; the method for this depends on the chosen docking software. In this instance Autodock Vina was chosen which is always executed through shell scripting (it does not carry its own GUI). Direct execution through bash shell also made the later incorporation into a bash script intuitive and seamless. In short, the user must include all of the necessary files in one directory (ligand, macromolecule, configuration file and AutoDock Vina executable), navigate to this directory through their respective bash shell and execute AutoDockVina. The docking process will then take place with updates given through the shell window (Figure 2.7).

Upon completion a results file is generated which contains the top 8 docked conformations. A log file is also generated which defines the relative predicted binding affinities of each docked conformation serving as AutoDock Vinas 'scoring function'. Unless there is a considerable difference observed in the predicted binding energies all should be considered possible and user discretion is often required to determine the most likely binding pose. This process should be guided by known inhibitor binding modes where possible allowing the user to give greater weight to predicted binding modes by comparison with known binding modes. In the absence of such information, the lowest energy conformation should be considered the most likely binding mode. Of course, these are predictive models of a complex system so caution should be exercised by the user, especially to avoid over interpretation of computational results. It is for this reason that a combinatorial approach has been taken in the work to follow whereby docking studies are used to help guide rational modifications leading on to empirical results to solidify the findings.

### **The *DrawToDock* Script**

As discussed, all of the previous steps were carefully chosen in such a way to be easily executable through shell scripting. Hence, all steps have been incorporated into one convenient script forming the *DrawToDock* script, as shown in figure 2.8 (this is for illustration purposes only; a full text copy has also been provided in the appendix for ease of reading). Furthermore, *DrawToDock* has been written to execute iteratively for all provided ChemDraw files named Ligand\_ 'X', where 'X' is a variable name chosen by the user. As such, all that is required by the user is to include the relevant docking files (ChemDraw files of all desired ligands named with the convention Ligand\_ 'X', macromolecule file and configuration file) along with the *DrawToDock* script and *DrawToDock* will complete the whole process upon execution. Importantly, once complete for one ChemDraw file it automatically continues on to the next facilitating time efficient screening. The *DrawToDock* script has also been written to sort all of the various files generated at each step into folders dependent upon their file type to make accessing results, or intermediate files, easy. This reduces the overall user input time from approximately 30 minutes per ligand of interest when compared with the 'manual' options presented previously to the time taken for the user to draw the ligand in ChemDraw which, as a conservative estimate, is ca. 2 minutes. It is envisioned that *DrawToDock* will prove a valuable means for both time and cost efficient virtual screening to be utilised by medicinal chemists with little to no experience of computational techniques. As such, *DrawToDock* has been used extensively throughout the following work.



```

1 |#!/bin/bash
2 |#System requirements and notes: designed and tested on MacOS. Require install of Open Babel and install of Python 2.7. for script to operate.
3 |#Input parameters here:
4 |
5 |pH=6.5 # Input the pH for protonation of ligands.
6 |steps=3000 # Input the number of energy minimisation steps used for generation of ligand 3D coordinates (for conversion to .pdb)
7 |
8 |echo Running ligand conversion ... # all echo commands give a readout to the user in terminal to update on the process
9 |echo
10 |echo parameters:
11 |echo protonation state \(\pH\) = $pH
12 |echo energy minimisation steps = $steps
13 |echo Energy minimisation method: WMF94\, steepest descending #these parameters can be altered in line 19, this is assumed to be the standard
14 |. required.
15 |
16 |for f in ligand_*.cdx; do #this for loop converts all .cdx files starting with the name ligand_ to corresponding .pdb files
17 |    b=`basename $f .cdx`
18 |    echo Processing $b to .pdb format ...
19 |    echo
20 |    obabel $f -O ${b}.pdb --gen3d -p $pH --minimize --steps $steps --sd
21 |    echo
22 |    echo $b conversion complete.
23 |    echo
24 |done
25 |
26 |echo All ligands converted to .pdb
27 |
28 |for f in ligand_*.pdb; do #this for loop converts all .pdb files starting with the name ligand_ to corresponding .pdbqt files
29 |    b=`basename $f .pdb`
30 |    echo
31 |    echo Processing $b to .pdbqt format ...
32 |    echo
33 |    python ./prepare_ligand4.py -l ${f} # this line calls on the python script to prepare .pdbqt files from the corresponding .pdb files. python
34 |. version 2.7 is required for this.
35 |    echo
36 |    echo $b conversion to .pdbqt complete.
37 |    echo
38 |done
39 |echo All ligands converted to .pdbqt
40 |for f in ligand_*.pdbqt; do #this for loop rund the docking simulation for each of the previously prepared ligand .pdbqt files
41 |    b=`basename $f .pdbqt`
42 |    echo Docking ligand $b ...
43 |    mkdir -p $b
44 |    ./vina --config config.txt --ligand $f --out ${b}/${b}_out.pdbqt --log ${b}/${b}_log.txt #this is the vina executable to run docking. vina.exe
45 |. must be present in the directory for this to function along with a configuration file named config.txt
46 |done
47 |
48 |# The following code is to 'clean' the apperance of the results and files generated
49 |mkdir chemdraw\ files # makes a new directory named Chemdraw files
50 |mv *.cdx chemdraw\ files # moves all Chemdraw files to new directory
51 |mkdir pdb\ files # makes a new directory named pdb files
52 |mv ligand_*.pdb pdb\ files # moves all pdb files to new directory
53 |mkdir pdbqt\ files # makes a new directory named pdbqt files
54 |mv ligand_*.pdbqt pdbqt\ files # moves all .pdbqt files to new directory
55 |mkdir results # makes a new directory named results
56 |mv ligand_* results # moves all docking result files to new directory
57 |
58 |echo All ligands docked.
59 |echo Job Complete.
60 |cat DrawtoDock1.0.sh >> scriptlog.txt # saves a copy of the DrawtoDock1.0.sh script used to a log file for future reference
61 |mv scriptlog.txt results # Moves the DrawtoDock script log finle into the results folder.

```

**Figure 2.8:** *DrawtoDock* script. The script has been fully annotated to aid user understanding of the process and further development if desired. This is for illustration purposes only, a full text version can be found in the appendix.

## 2.2 Summary and Future Work

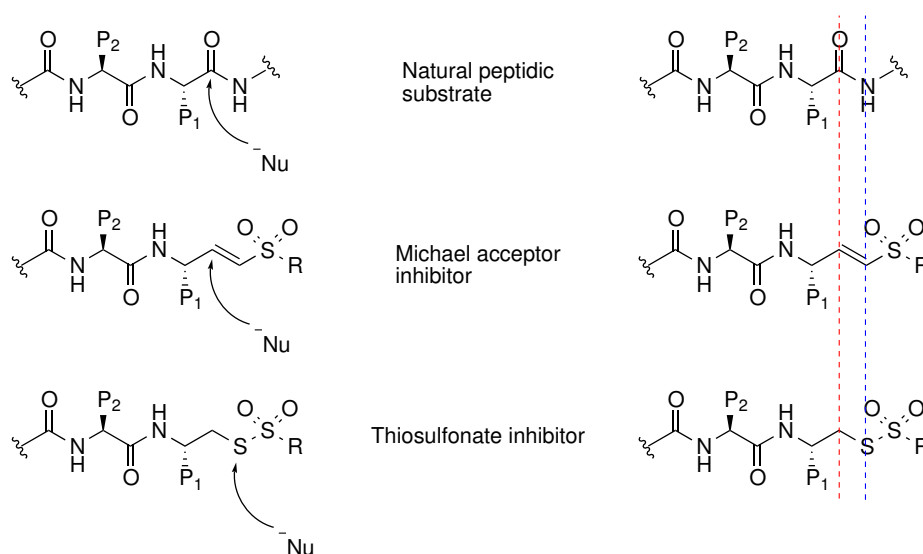
A virtual screening script has been developed, named *DrawToDock*, which offers a fantastic collection of free, openly available software to facilitate rapid virtual screening with minimal user input and importantly no upfront cost. As such, this represents a valuable tool for time and cost efficient virtual screening however, further steps are possible to improve the user experience. Namely, integrating all of the described script into a clean user interface. It was determined that building a GUI for *DrawToDock* was beyond the scope of this work however, in the current iteration of *DrawToDock*, the user must navigate through shell to execute the script. Although this is not a major challenge, it may be off-putting to users with little computational experience. Rather, developing a clean user

interface with user prompts such as ‘select ligand files’ and ‘select save location’ as well as the option to directly draw structures for docking is certainly possible and may prove useful. This is particularly true in realising the philosophy of overcoming computational barriers for synthetic chemists and would be a valuable addition in future work.

# Chapter 3

## First Generation Thiosulfonate Inhibitors

### 3.1 Docking Study: Are Thiosulfonates Viable Warheads?

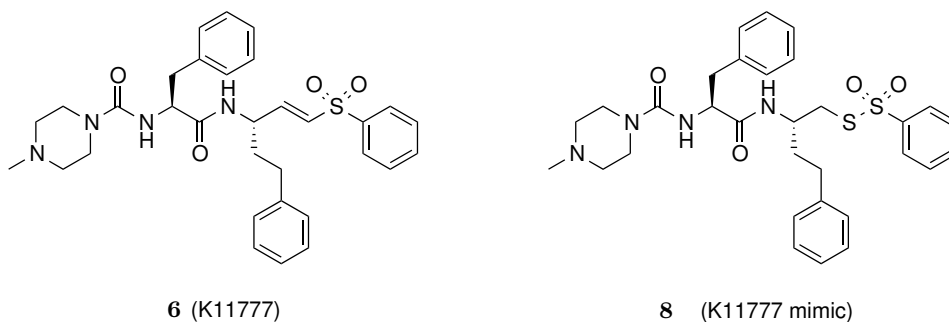


**Figure 3.1:** Comparison of the site of nucleophilic attack on a natural peptidic substrate (top) compared with a Michael acceptor (middle) and a thiosulfonate (bottom). Red dashed line highlights the site of attack on the carbonyl of the natural substrate and vinyl sulfones. Blue dashed line highlights site of attack on thiosulfonates which, as shown, is shifted towards the bivalent sulfur one position further along.

There are many factors to consider when designing a novel electrophilic trap. An important consideration for covalent inhibitors is the kinetics related to covalent bond formation. One kinetic implication which was immediately apparent for the thiosulfonates

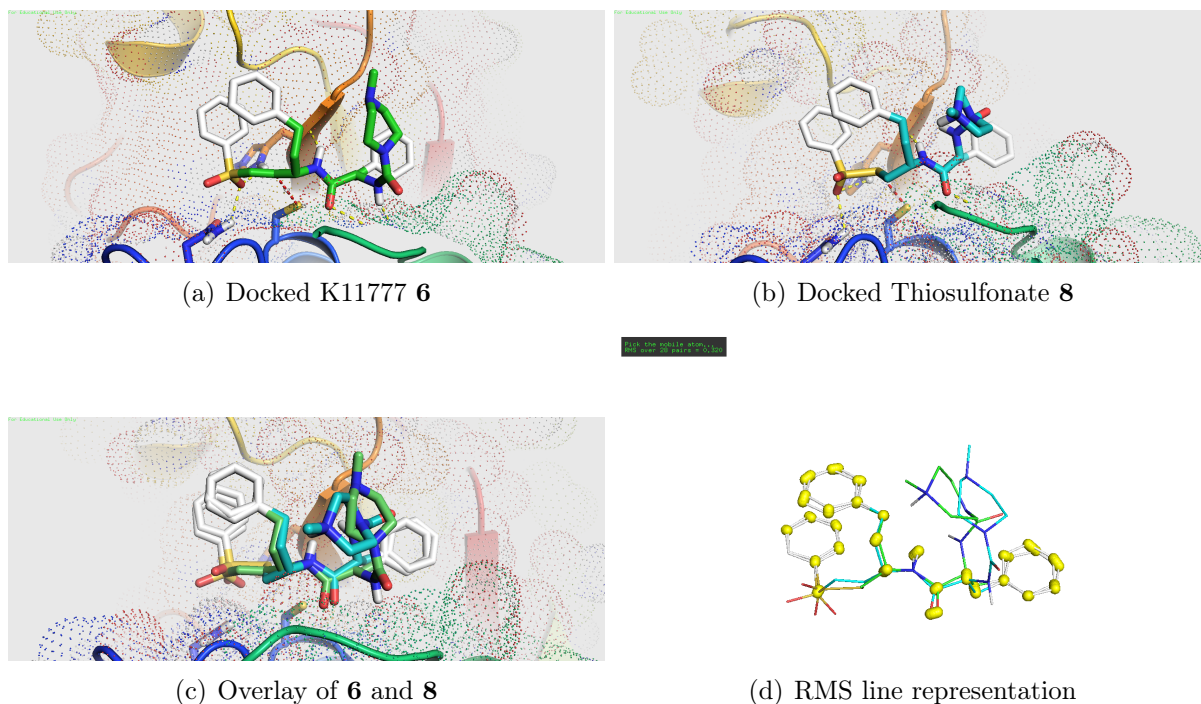
in peptidomimetic inhibitors was the site of nucleophilic attack, as shown in figure 3.1. Introducing thiosulfonates into peptidomimetic inhibitor constructs effectively translates the site of attack one position along the inhibitor backbone (towards the natural substrate's C terminus) when compared with the natural site of attack on the carbonyl of peptidic amide bonds. Furthermore, this natural site of attack is well mimicked by current Michael acceptors such as the vinyl sulfones, which offer well established covalent inhibitor constructs (figure 3.1). Of course, proteins are complex flexible structures and as such it may be reasonable to suggest that altering the site of attack by such a small margin may be well tolerated. Indeed, previous work within the Liskamp group towards sulfonyl fluorides as inhibitors of the proteasome,<sup>178-181</sup> serine proteases<sup>182</sup> and prolyl oligopeptidase inhibitors,<sup>182-184</sup> has demonstrated that such a shift in the site of attack can be well tolerated and should be of little concern. It is however prudent to consider that such a change may be kinetically detrimental towards covalent bond formation. An obvious solution would be to develop chemistry which accounts for this, in effect shifting the site of attack back to the 'natural' position. Although this may be possible, this would involve functionalising next to a stereogenic centre which would greatly increase the complexity of the chemistry required. In order to address this issue, a modelling study was conducted to better understand the possible binding mode of the thiosulfonate constructs and rationalise if altering the site of attack would be detrimental to the binding mode of thiosulfonate derived peptidomimetic inhibitors.

It was hypothesised that the inherent flexibility of the target proteins would be sufficient to account for this difference as has been displayed previously within the Liskamp group with the sulfonyl fluorides. However, if this could be further supported by modelling studies then efforts to develop the more complex chemistry, functionalising adjacent to a stereogenic centre, could be reasonably considered to be unnecessary. Importantly, it was believed that developing chemistry which was robust and readily accessible would be fundamental to the future success and uptake of thiosulfonate derived inhibitors. Hence, exploring the synthetic complexity *vs* inhibitor potential is a critical consideration which was further explored before synthesis was conducted.



**Figure 3.2:** Vinyl sulfone inhibitor K11777 (**6**, left) and the thiosulfonate mimic (**8**, right).

Papain was taken as a prototypical cysteine protease for this study as it is representative of Clan CA family C1 cysteine proteases and has been well studied in the literature. It was decided that modelling K11777 (**6**, figure 3.2) against its closest thiosulfonate mimic (compound **8**, figure 3.2) would allow investigation of the relative binding modes of the two inhibitor constructs, which was conducted with the use of *DrawToDock* (discussed in chapter 2).



**Figure 3.3:** Docking binding mode of (a) K11777 **6**, (b) Thiosulfonate mimic **8**, (c) overlay of both **6** and **8**, (d) RMS calculation between binding modes of **6** and **8**, line representation with paired atoms highlighted in yellow.

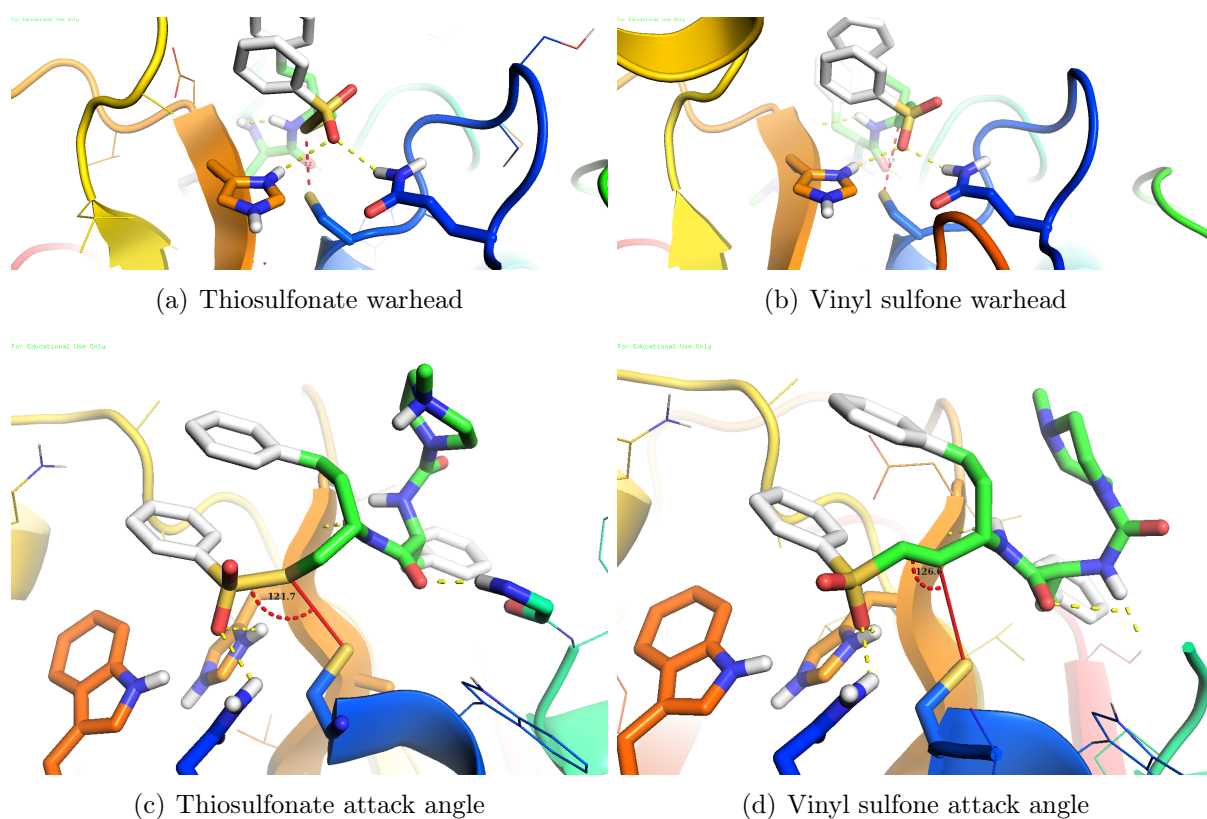
Firstly, the observed binding mode of K11777 (compound **6**) was evaluated to determine if this modelling study was producing reasonable, reliable data as shown in figure 3.3. K11777 was docked into papain (PDB code 1CVZ)<sup>185</sup> which gratifyingly gave an overall binding mode representative of the known binding mode for K11777. The known binding mode was determined by examination of the crystal structure of K11777 with two other clan CA cysteine proteases, Cruzain (PDB code 2OZ2)<sup>186</sup> and *S. mansoni* Cathepsin B1 (SmCB1, PDB code 3S3R).<sup>187</sup> In both of these crystal structures, the aromatic ring of the warhead occupies the  $S_1'$  pocket with the homophenylalanine residue occupying the  $S_1$  pocket, the phenylalanine residue occupies the  $S_2$  pocket and the N-methyl piperazine cap occupies the  $S_3$  region. The same trend was observed in docking K11777 with papain in this study, with the phenyl ring of the warhead occupying the  $S_1'$ , homophenylalanine side chain occupying the  $S_1$ , phenylalanine side chain occupying the  $S_2$  and N-methyl piperazine cap solvent facing in the  $S_3$  region. This gave further confirmation that a

reasonable prediction had been generated before further interpretation.

Confident that *DrawToDock* had generated a reasonable predicted binding mode, by comparison with literature crystal structures, the thiosulfonate warhead was then analysed. As shown in figure 3.3 the overall binding mode of the thiosulfonate mimic **8** was largely in agreement with that observed for the vinyl sulfone **6** (K11777). The greatest variability between **6** and **8** was observed for the *N*-methyl piperazine cap in the S<sub>3</sub> region, which was expected as this region is solvent facing and anticipated to be freely rotatable in the solvated environment. To further highlight the agreement in the binding modes between each inhibitor construct the RMS deviation, determined by Pymol align function, was 0.320 Å (overlay of 28 atom pairs as shown in figure 3.3). The *N*-methyl piperazine cap was not included in the RMS determination as it is considered to be freely rotating in the solvated environment thus does not contribute significantly towards the primary binding event.

Importantly, this modelling study also suggested that moving the position of the electrophilic trap further along the inhibitor backbone, as shown in figure 3.2 was not detrimental to the overall binding mode. It was found that the bivalent sulfur of the thiosulfonate was positioned 3.2 Å from the active site nucleophile (Cys-25) whereas the  $\beta$  sp<sub>2</sub> carbon of the vinyl sulfone was 3.7 Å from the active site thiolate. Furthermore, the angle of attack could also be assessed. As the thiosulfonates operate through an S<sub>N</sub> mechanism the optimal angle of attack by the active site nucleophile would be 180°. This modelling study suggests that in the bound conformation the angle of attack is approximately 128° as shown in figure 3.4. comparitively, the vinyl sulfone, being sp<sub>2</sub> hybridised, would be expected to have an optimal attack angle of ca. 107° in accordance with the Bürgi-Dunitz angle.<sup>188,189</sup> In this study an angle of ca. 127° was observed. Of course, considerations involving distance between nucleophile and electrophile as well as angle of attack can offer valuable insight into potential mechanistic advantages an inhibitor construct may have. On the other hand, these considerations tend to assume reactivity dependent on random collision and do not take into account the local, relatively constrained, enzymatic environment. For example, the Bürgi-Dunitz angle has been found to be as low as 88° in a fatty acid amide hydrolase.<sup>190</sup> Combining this with the limitations of modelling, one of which is particularly apparent in this system is treating the protein as a rigid structure, it is fair to conclude that the thiosulfonate appears no less suited to inhibition of cysteine proteases than the vinyl sulfones on the basis of attack angle and distance from the active site.

Another important finding from this docking study was the potential for *in situ* activation of the thiosulfonate warheads. Vinyl sulfones have long been accepted to undergo hydrogen bonding between active site H-bond donor residues and the H-bond acceptor of the



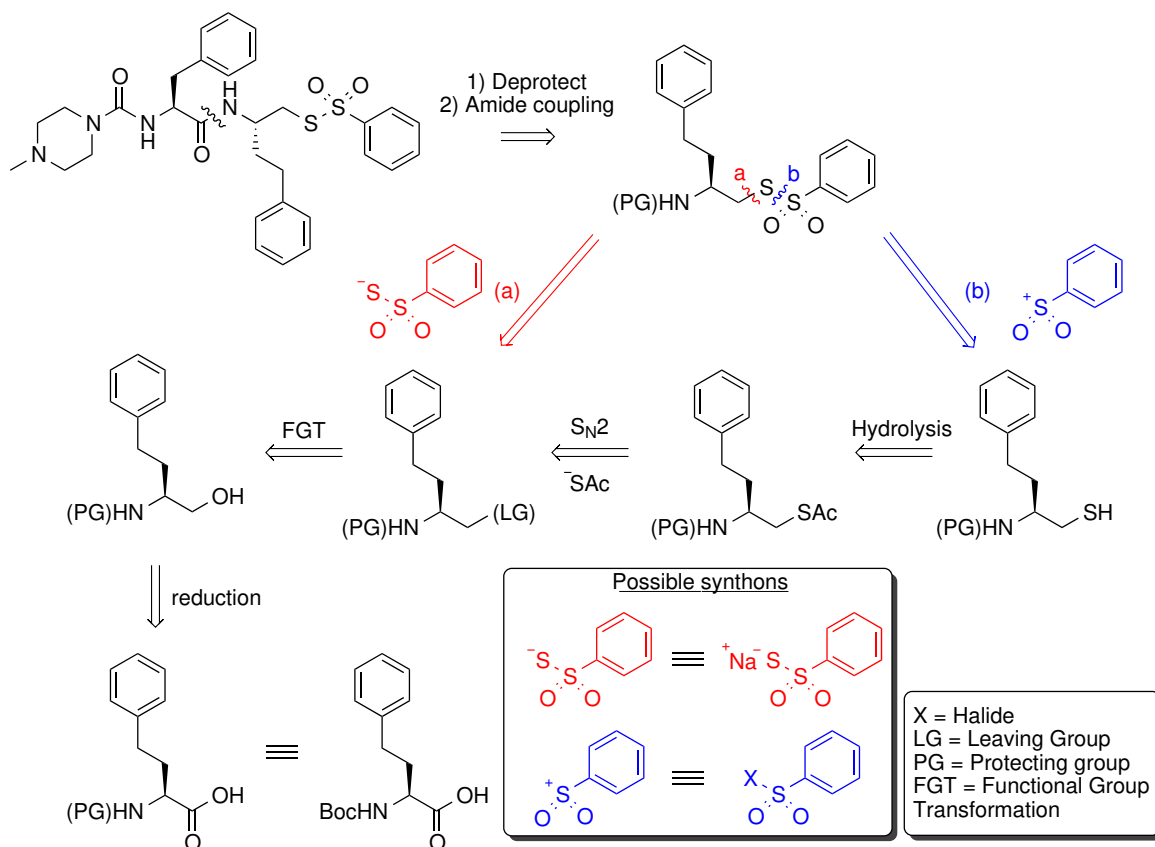
**Figure 3.4:** *In situ* activation of warheads through H-bonding. (a) Thiosulfonate warhead, (b) vinyl sulfone warhead, (c) Angle of attack on thiosulfonate warhead, (d) Angle of attack on vinyl sulfone.

hexavalent sulfur centres oxygen atoms.<sup>136</sup> Given the chemical similarity in this regard between the vinyl sulfone and thiosulfonates it is reasonable to suggest that a similar process may take place. In this modelling study it was found that hydrogen bonding between the sulfur oxygen atoms to His<sup>159</sup> may prove particularly important, as this residue forms the catalytic dyad with the active site Cys<sup>25</sup> (as discussed in introduction section 1.1.2, figure 1.8). The combination of this H-bond donor (His<sup>159</sup>) and H-bond acceptor character of the sulfonate oxygen atoms may in turn polarize the thiosulfonate S-S bond, thereby increasing the rate of cleavage. Importantly, cysteine proteases utilise the catalytic histidine residue as a general acid in their catalytic cycle, facilitating the H-bond donor specifically in cysteine proteases as the His<sup>159</sup> is in its protonated state. This is in contrast to other proteases, such as serine proteases, which utilise a histidine residue as a general base in their catalytic cycle.<sup>191</sup> As such, these differing catalytic mechanisms are favourable for the inhibitory specificity of thiosulfonates towards cysteine proteases.

## 3.2 Exploring Routes Towards an Efficient Warhead synthesis

Fundamental to the development and subsequent uptake of any newly emerging inhibitor is an efficient and reproducible synthesis. As thiosulfonates are likely to find their place amongst drug candidates targeted towards neglected tropical diseases, or as probes to better understand biological processes, an efficient synthesis from readily accessible materials was desired. It was envisioned that being able to access useful thiosulfonate constructs from amino acids by robust chemistry would not only meet these criteria, but encourage their use in future research. In the following chapter, efficient and reproducible routes towards thiosulfonates will be explored forming the foundation of their development as cysteine protease inhibitors.

### 3.2.1 A Retrosynthetic Approach

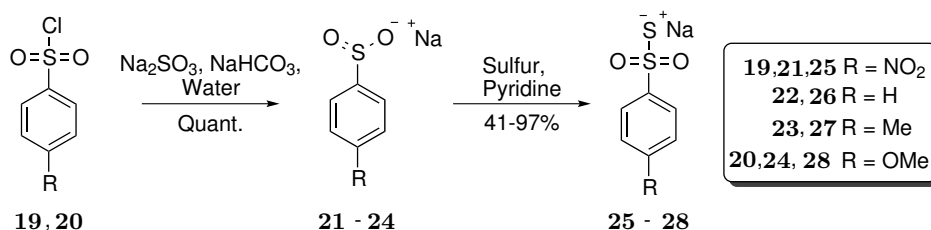


Beginning with the retrosynthesis of the desired thiosulfonate inhibitors, as shown in scheme 3.1, it was clear there were two main routes which could be explored. In each case, the final step would involve coupling the backbone to the warhead fragment by standard amide



coupling techniques. This leads back to the key amino acid derived thiosulfonate building block, with the potential for two clear disconnect options (scheme 3.1). Disconnection (a) is across the carbon-sulfur bond, offering a more direct substitution route with a thiosulfonate salt or (b) disconnecting across the sulfur-sulfur bond back to a nucleophilic bivalent sulfur, such as a thiol, and an electrophilic hexavalent sulfur, such as a sulfonyl halide. Either of these routes benefit from common synthetic intermediates earlier in the synthesis, providing redundancy in the key step of thiosulfonate formation, which can be readily accessed by a series of functional group transformations from the commercially available amino acid homophenylalanine. The common precursor to both of these intermediates is a protected amine functionalised with a leaving group. Again, there is apparent redundancy in this step as many possible leaving groups could be used, of which sulfonate esters or halides are the most common and readily accessible. Either of these leaving groups could be accessed by the corresponding alcohol which in turn could be accessed by reduction of the corresponding amino protected homophenylalanine. This retrosynthetic analysis also suggests that synthesis from other amino acids, not just homophenylalanine, should be well tolerated and easily accessible allowing for expansion of inhibitor constructs or structure-activity relationships to be explored with ease.

### 3.2.2 Warhead Synthesis: From Thiosulfonate Salts

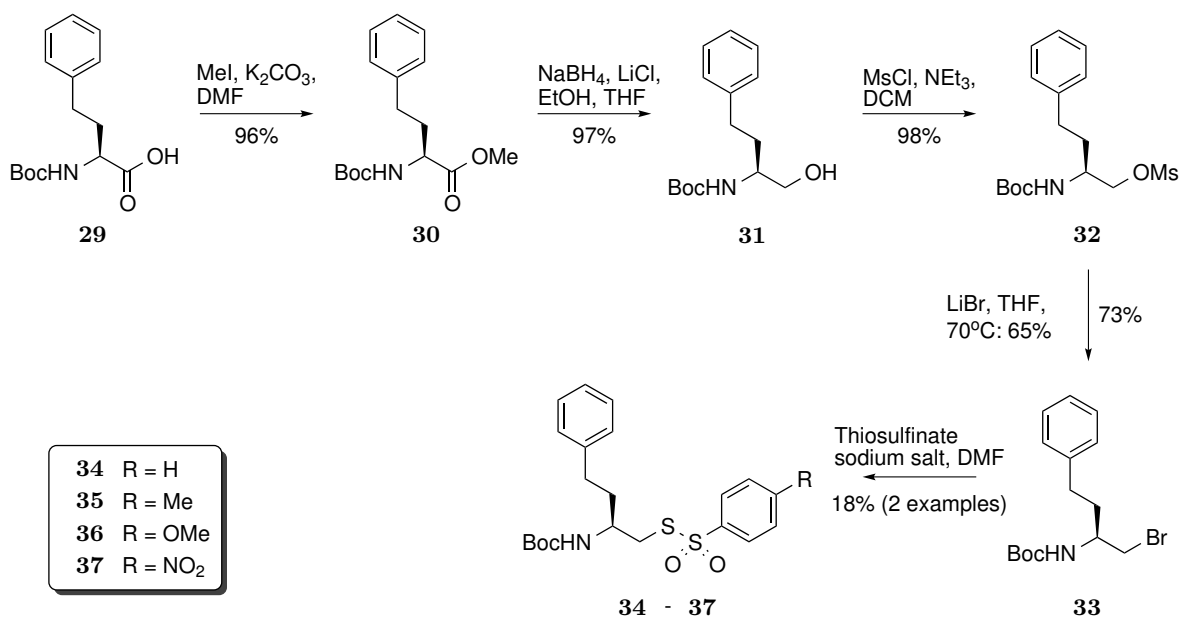


**Scheme 3.2:** Thiosulfonate salt synthesis from sulfinates.

Considering the retrosynthetic analysis in scheme 3.1, section 3.2.1, it was hypothesised that the synthesis of thiosulfonates would be most efficient by substitution with thiosulfonate salts (disconnect (a), scheme 3.1). This strategy was thought to be more convergent, allowing for a combination of two more complex fragments in fewer synthetic steps. Hence this route was the first to be explored.

The key step of this synthetic route was the introduction of the thiosulfonate warhead with thiosulfonate salts. As such, the formation of thiosulfonate salts was first assessed to ensure they were easily accessible as outlined in scheme 3.2 based on the work of Sato *et al.* from sulfinic acid salts and sulfur.<sup>192</sup> This method provided the corresponding thiosulfonate sodium salts **26** - **25** in good to excellent yields. Where sulfinic acid salts were not commercially

available, in the case of compounds **24** and **21**, they were accessed by reduction of the corresponding sulfonyl chlorides **20** and **19** in excellent yields based on the literature procedure by Lacour *et al.*<sup>193</sup> The ability to quickly access the required thiosulfonate salt from sulfonyl chlorides was thought to be of particular importance as the sulfonyl chloride moiety is a common synthetic intermediate with many commercially available, or easily accessible, potentially allowing rapid access to a wide range of thiosulfonate warheads.



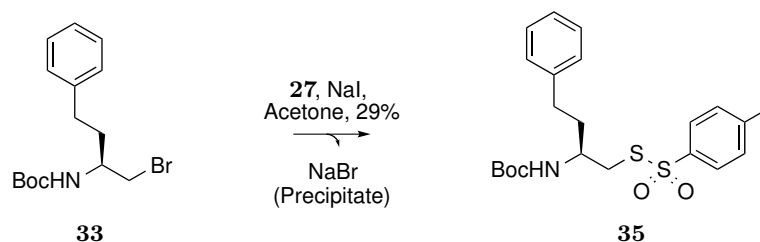
**Scheme 3.3:** Homophenylalanine derived aromatic thiosulfonates from thiosulfonate salts.

With the thiosulfonate salts to hand synthesis from the amino acid starting material could now be explored. Starting from commercially available Boc-hF-OH **29** the carboxylic acid was first converted to the corresponding methyl ester by alkylation under basic conditions. This method was chosen over the ubiquitous thionyl chloride / methanol approach to avoid in situ generation of HCl, which could potentially deprotect the acid labile Boc protecting group. This proved to be a convenient reaction as excess reagents could be removed by evaporation (MeI) and aqueous washing (K<sub>2</sub>CO<sub>3</sub>) giving **30** in excellent yields.

Next, the methyl ester was reduced by the *in situ* generation of LiBH<sub>4</sub> to the corresponding alcohol **31** in excellent yields. Previous work in the Liskamp group found the two step approach to amino alcohols to be higher yielding and more convenient than the direct conversion from carboxylic acids.<sup>181</sup> This also offers the advantage of milder reaction conditions rather than employing more powerful reducing agents such as LiAlH<sub>4</sub>, which may facilitate a broader substrate scope for use on more complex systems in future.

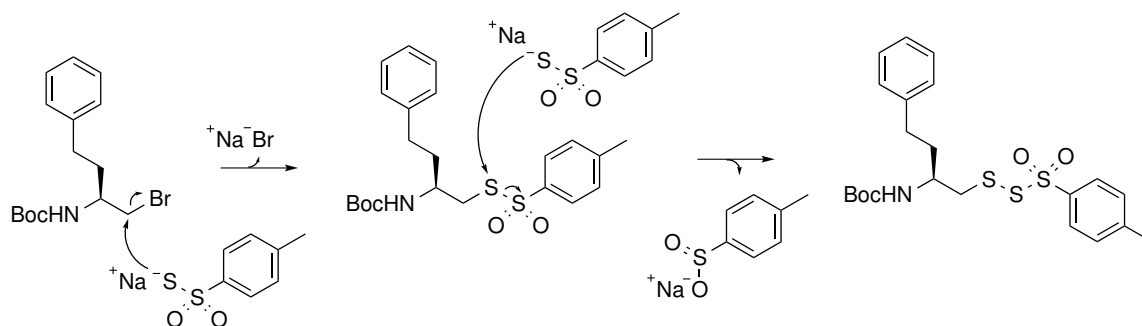
Ultimately, the alkyl bromide **33** was desired for the introduction of thiosulfonates by substitution of their corresponding thiosulfonate salts. This was built on the previous work

in the Liskamp group, which found the bromides offered better yields than other leaving groups, such as mesylates.<sup>194</sup> This could have been achieved directly by conversion of amino alcohol **31** to alkyl bromide **33** by the Appel reaction however, previous experience found the Appel reaction to be low yielding and the generation of triphenylphosphine oxide by-product complicated purifications. As such the Boc protected amino alcohol **31** was first converted to mesylate **32** in excellent yields. Subsequently, mesylate **32** was substituted for bromide **33** in good yields with the use of LiBr under reflux.



**Scheme 3.4:** Employing Finkelstein conditions to improve thiosulfonate formation from thiosulfonate salts

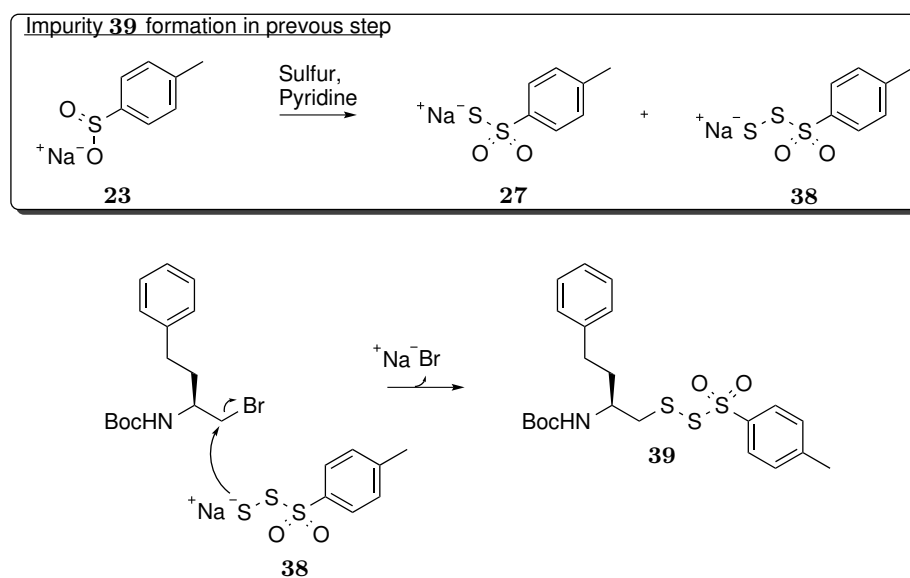
Finally the key transformation forming the thiosulfonate warhead could be explored. This involved direct substitution of bromide **33** to thiosulfonates **34** - **37**. It was intended to conduct this reaction for all four warheads presented in scheme 3.3, however when a relatively low yield of 18% was observed for the first two attempts (towards compounds **34** and **36**) it was decided to further investigate and optimise the reaction conditions before applying to all warhead constructs. Initial attempts at optimising the substitution took advantage of Finkelstein conditions (scheme 3.4) to generate the more reactive alkyl iodide *in situ*. Acetone was employed as the solvent to take advantage of the differing solubilities of both sodium bromide and sodium iodide, precipitating out the generated sodium bromide and driving the reaction towards product formation. This was met with a moderate improvement in the yield from 18% to 29%. Although this was an improvement, the yield was still low and not considered comparable with the search for an ‘efficient’ synthesis towards this novel warhead.



**Scheme 3.5:** Impurity formation by subsequent substitution of thiosulfonates.

Furthermore, when characterising the thiosulfonate product it was clear by LC-MS analysis

that an impurity was present, which could not be removed by standard chromatography and has a mass increase of +32 from the desired thiosulfonate. Furthermore, NMR analysis suggested that this impurity showed effectively identical NMR to the desired products **34** and **36**, with impurity peaks overlapping with product peaks. Given the high similarity by NMR and mass increase observed it was hypothesised that the impurity may contain an additional sulfur, generated through the potential side reaction outlined in scheme 3.5. It was reasoned that, upon formation of the desired warhead, an additional equivalent of the thiosulfonate salt was attacking back on the bivalent sulfur and generating the unwanted byproduct observed.

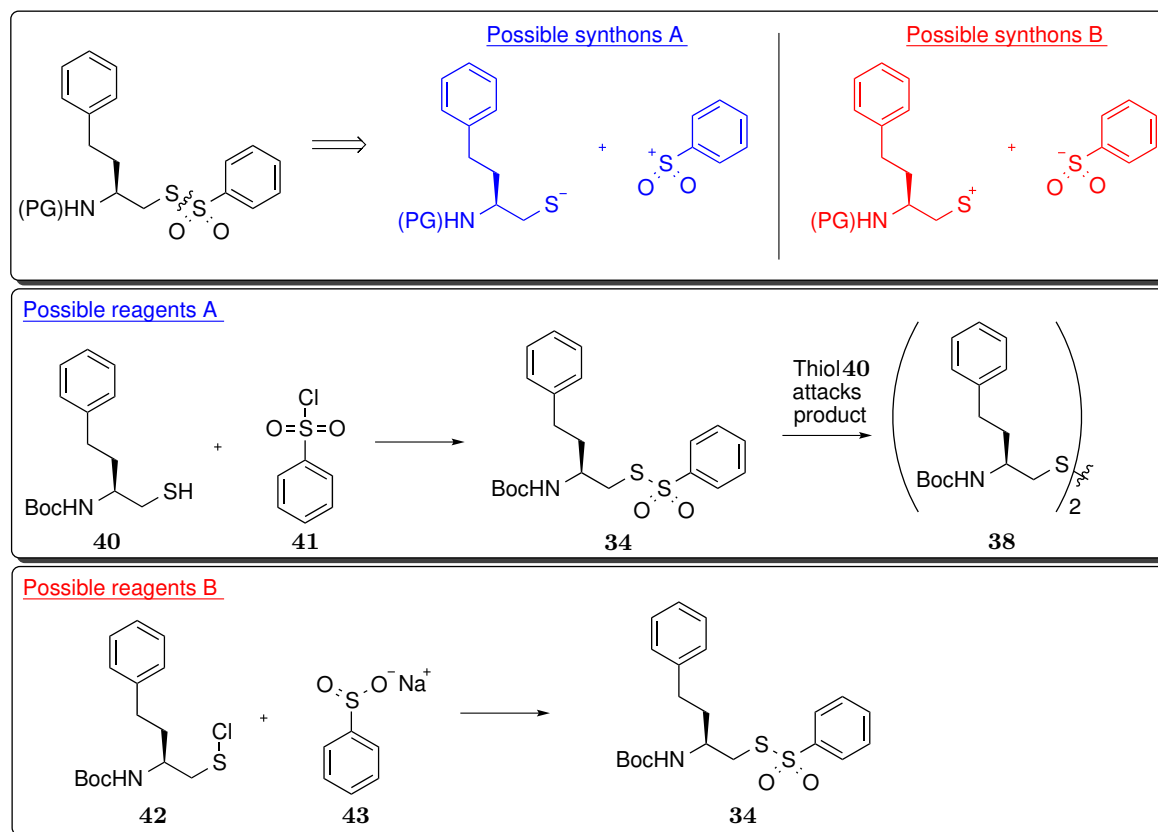


**Scheme 3.6:** Thiosulfonate impurity formation due to impure thiosulfonate salts.

Another potential route towards formation of the undesired product **39** is displayed in scheme 3.6. It is possible that the additional sulfur had been carried forward from the synthesis of the thiosulfinate salt which, in turn, was incorporated during the substitution of the alkyl halide. Either of these two mechanisms were considered to be plausible and inherent in this synthetic approach. It was hypothesised that attempting the synthesis of thiosulfonates by alternate methods that avoided these potential pitfalls would be preferable to optimising a synthesis with inherent competing reactivity. The reason for this was two fold: (1) the yield by direct substitution of alkyl halides by thiosulfonate salts was considered to be poor, even after attempted optimisation. Although further optimisation of this is possible to increase the yield, such as assessing other leaving groups, the results may still be hampered by the presence of an inseparable impurity. (2) Determining the exact source of the impurity, although possible, would be time consuming and challenging. The thiosulfonate salts were not found by mass spectrometry which removed the most obvious method for determining at which point the additional sulfur was being incorporated. Other standard characterisation techniques, such as NMR, were predicted to be insignif-

icantly different between the thiosulfonate salt and additional sulfur salt. Even if it was found that the additional sulfur was incorporated at the thiosulfonate salt formation step, the separation of the thiosulfonate salt from the impurity would prove difficult (separation of two salts). Furthermore, even if pure thiosulfonate salt could be obtained and proven to be pure, this would not prevent the potential for the same impurity to form in the following substitution step. It was determined that the impurity observed may well be through a combination of both routes outlined, meaning multiple reaction steps would have to be optimised to avoid an inherent weakness in the synthesis, which was thought to be an inefficient approach. As such, it was decided to explore an alternate approach to avoid these issues rather than undertake a lengthy optimisation process which may still have proven inefficient. Fortunately the earlier retrosynthetic analysis highlighted that this could be achieved by the common intermediate **32** which appeared to offer a synthetic route which could circumvent these issues.

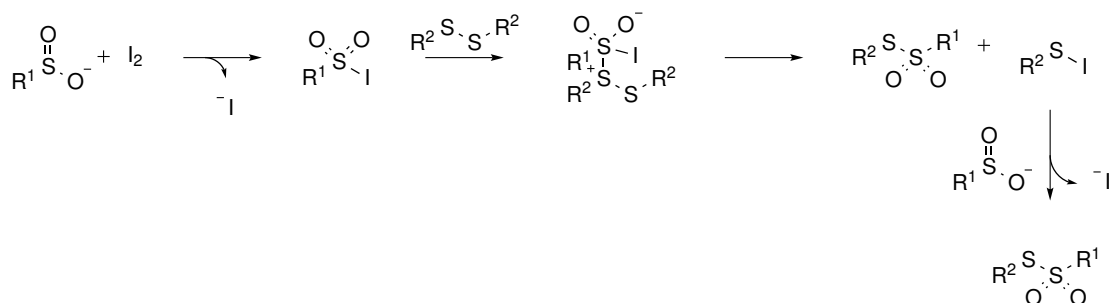
### 3.2.3 Warhead Synthesis: Oxidative Cleavage of Disulfides



**Scheme 3.7:** Thiosulfonate S-S disconnect with synthon options (top) and possible reagents (middle and bottom)

Considering the previously discussed impurity observed in section 3.2.2 a new route was desired. Based on the earlier retrosynthetic analysis (scheme 3.1) the most obvious route

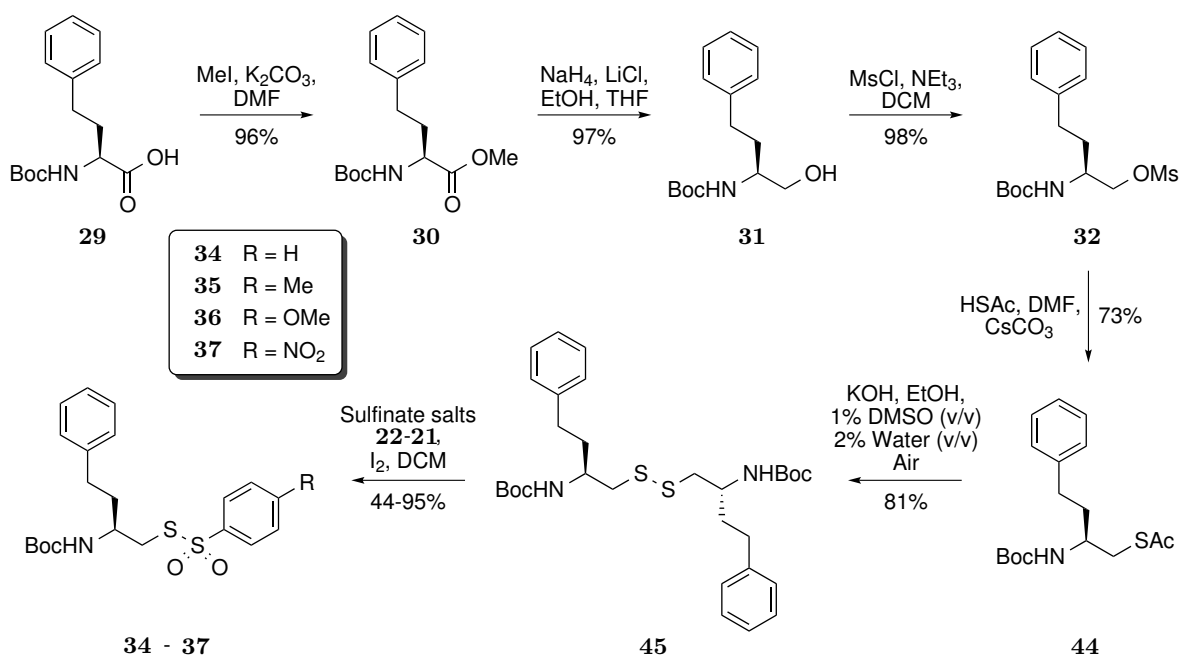
to attempt was thiosulfonate synthesis from sulfonyl chlorides and thiols. However, now armed with the information that a secondary attack on the thiosulfonate by sulfur centred nucleophiles may be possible, as shown previously in scheme 3.5, it was proposed that a similar side reaction may occur (scheme 3.7). This was because the synthesis of thiosulfonates from a thiol and sulfonyl chloride shared the same fundamental problem, generating an electrophile in the presence of a nucleophile. As shown in scheme 3.7 it was hypothesised that reversing the reactivity of the synthons such that the hexavalent sulfur acted as the nucleophile may overcome this issue, even if such a side reaction occurred (nucleophile attacking the desired product) this would only reform the desired product. As such, it was determined that this offered a more elegant and likely higher yielding approach, however the intended starting material (sulfuryl chloride **42**) would itself be difficult to access as a highly reactive intermediate. As such, a literature study was first conducted to assess the potential merits of each option. Firstly the method from sulfonyl chlorides was explored to assess if any previous literature had reported the potential disulfide forming side reaction. It was found that Mahieu *et al.* reported the potential of this side reaction suggesting that product formation could be controlled by the stoichiometry of the reaction.<sup>195</sup> More recently, Pham *et al.* have shown that when using a sulfonyl chloride, the major product tends to be the symmetrical disulfide rather than the thiosulfonate.<sup>196</sup> With these literature studies in mind it was considered that sulfonyl chlorides offered a somewhat inelegant but plausible synthetic route where slow addition and stoichiometric control may have been employed to yield the desired thiosulfonates.



**Scheme 3.8:** Mechanism: Thiosulfonate formation by oxidative cleavage of disulfides

Assessing the second option was mainly focused on a more stable, synthetically accessible electrophilic bivalent sulfur. Inspired by the concept of disulfide exchange, whereby a disulfide acts as a bivalent electrophilic sulfur centre for a sulfur centred nucleophile (thiol), the literature was searched for potential examples utilising a sulfinate salt as the nucleophile. This approach was in part inspired when considering how the originally intended sulfonyl chloride **42** could have been accessed from a disulfide and thionyl chloride.<sup>197</sup> Thus, it was thought that a disulfide itself may offer a shelf stable, easily accessible starting material for this transformation. Gratifyingly, a method reported by Fujiki *et al.* was

found which utilised the iodine mediated oxidation of disulfides and subsequent cleavage with a sulfinate salt to yield thiosulfonates in good yields, with the proposed mechanism shown in figure 3.8.<sup>198</sup> Although Fujiki *et al.* had only applied this method to simpler, non amino acid derived systems, it was considered an ideal and more elegant approach to access thiosulfonates for use as cysteine protease inhibitors. Importantly, it was considered that if a subsequent nucleophilic attack of the nucleophile utilised in this reaction (sulfinate salt) took place on the desired thiosulfonate product, this would simply reform the desired product. Based on the previous findings it was thought that this would offer a significant synthetic advantage and allow for improved efficiency in the key transformation of this work, which would prove fundamental to developing this novel warhead class for future research.



**Scheme 3.9:** Homophenylalanine derived aromatic thiosulfonates by oxidative cleavage of disulfides

This led to the new synthesis towards amino acid derived thiosulfonates as outlined in scheme 3.9. This takes advantage of the earlier recognised redundancy towards the synthesis of thiosulfonate warheads starting from the same synthetic intermediate, mesylate **32**. The sulfur functionality was then introduced by substitution with thioacetic acid yielding the thioacetate **44** in a good yield of 73%. Subsequent hydrolysis of the thioacetate under aerobic conditions allowed for direct conversion to disulfide **45** in a good yield of 81%. It was also considered that, if necessary, the thiol could be isolated through hydrolysis under anaerobic conditions which could provide a valuable building block if the oxidative cleavage of disulfides was found to be unsuccessful again, offering flexibility in the synthesis. Fortunately exploring thiol building blocks was not required as the subse-

quent oxidative cleavage to compounds **34** - **37** proceeded in moderate to excellent yields (44-95%).

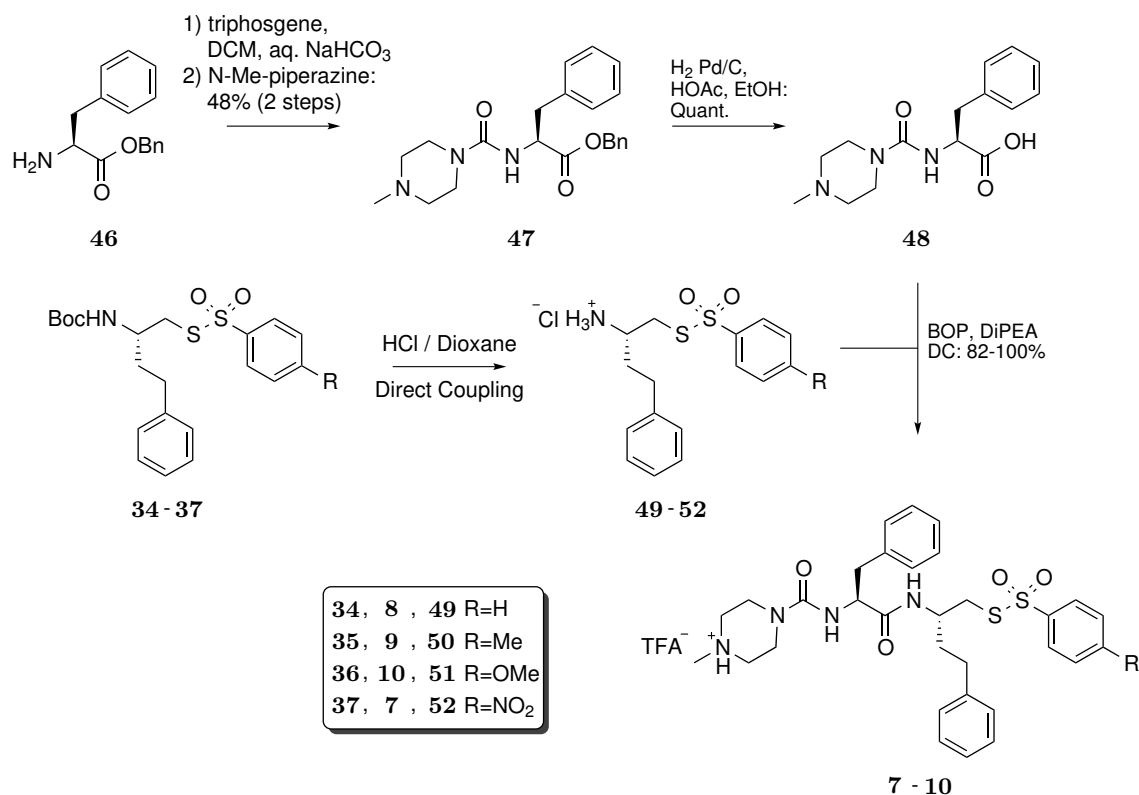
In summary, an alternate method towards the synthesis of the thiosulfonate warhead has been explored by the oxidative cleavage of disulfides with sulfinate salts. By considering appropriate retrosynthetic approaches, this overcame a fundamental issue with the synthesis of an electrophile in the presence of a nucleophile. Reversing the reactivity of the synthons in use led to a significant increase in yield, which was realised by applying the work of Fujiki *et al.* Furthermore, the overall synthesis shown in scheme 3.9 was believed to offer a high yielding (24-51% over 6 steps), robust synthesis from amino acid building blocks providing an excellent foundation for the work to follow.

### 3.3 Backbone Synthesis and Coupling to Warheads

With the thiosulfonate warhead building blocks to hand, focus turned to constructing the backbone. In general, the backbone can be altered to confer specificity towards different targets as it is responsible for the majority of the primary binding interactions. It was hypothesised that generating peptidomimetic inhibitor constructs would in principle facilitate a wide variety of peptidic backbones to be constructed, opening up a readily accessible route towards other macromolecule targets of interest. As such, focus was placed on generating a peptidic construct, which in this instance was derived from the K11777 backbone, to highlight the suitability of thiosulfonates for incorporation into such scaffolds by standard peptide coupling techniques. It was thought that confirming the thiosulfonates are compatible with standard peptide coupling techniques would be fundamental to the uptake of this novel warhead construct by the wider scientific community. Furthermore, this would also facilitate an ideal testing platform for comparison against a known cysteine protease inhibitor, K11777 (**6**), to assess changing only the warhead moiety. Maintaining the same backbone was considered important as the non-covalent interactions it provides can contribute significantly towards inhibitor specificity and potency. Thus, maintaining a consistent backbone motif allowed for assessment of the contribution from the warhead.

The backbone was made according to the literature procedure of Kiemele *et al.* (scheme 3.10) starting from phenylalanine benzyl ester **46**.<sup>199</sup> This was first converted to the isocyanate with the use of triphosgene before direct conversion to the *N*-methyl piperazine capped urea **47** in a moderate yield of 48%. Interestingly, following the literature procedure resulted in very low yields (8%) over these two steps. It was thought that this was due to an error in the reported procedure, which reports performing an acid wash during the workup, as this step would most likely extracting the tertiary amine into the aqueous phase reducing



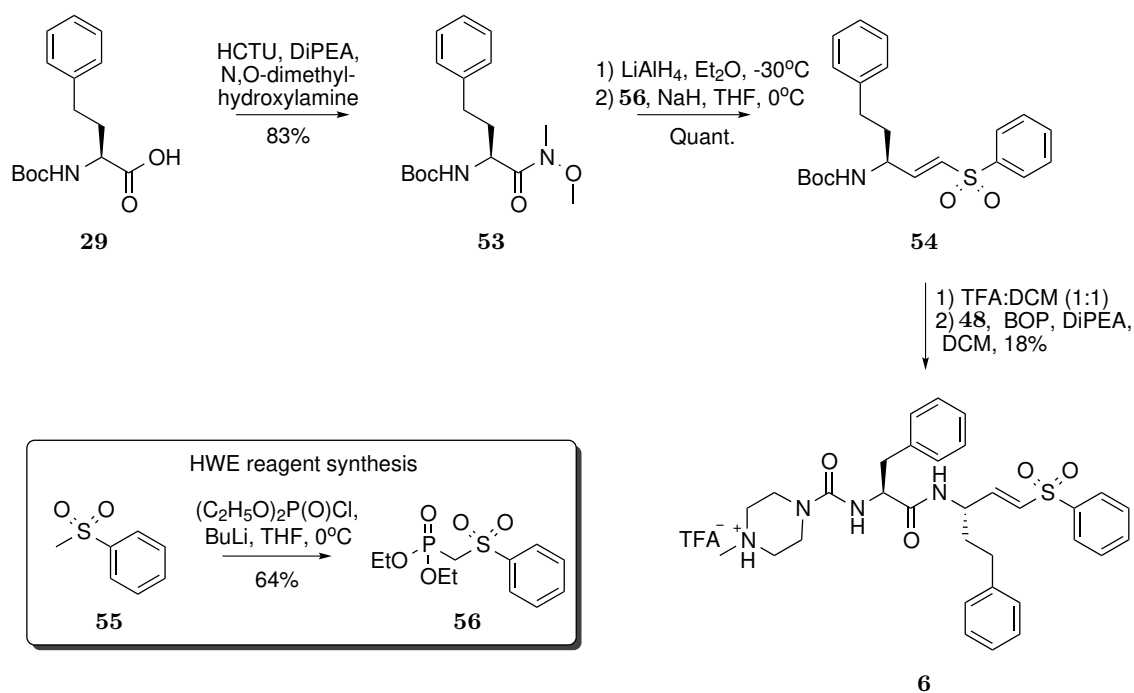


**Scheme 3.10:** Backbone synthesis and coupling to homophenylalanine derived aromatic warheads.

the yield. It was also noted that two solvents were being utilised, beginning with DCM, concentrating and then suspending in THF. This was deemed to be unnecessary thus an optimised approach, shown in scheme 3.10, was taken whereby the DCM layer was simply carried forward after separation of the aqueous NaHCO<sub>3</sub> phase and drying. Removing the reported acid wash and evaporation / resuspension in THF steps both improved the yield and convenience of the reaction. Following on from this the C terminus was deprotected by hydrogenolysis in preparation for the following peptide coupling. The amino acid derived thiosulfonate warheads **34-37** were deprotected with HCl before direct coupling to the backbone using the uronium coupling agent BOP as a representative, standard peptide coupling technique which proceeded in good to excellent yields (82-100%). It was thought that this represented a widely accessible, high yielding route towards peptidomimetic thiosulfonate inhibitor constructs, which could be reasonable expected to work with a wide variety of substrates.

### 3.4 Synthesis of K11777: A Reference Compound

In order to validate the newly synthesised thiosulfonates, K11777 (**6**) was synthesised as a reference compound for comparison. The synthesis, displayed in scheme 3.11, was based



**Scheme 3.11:** Synthesis of K11777 reference compound

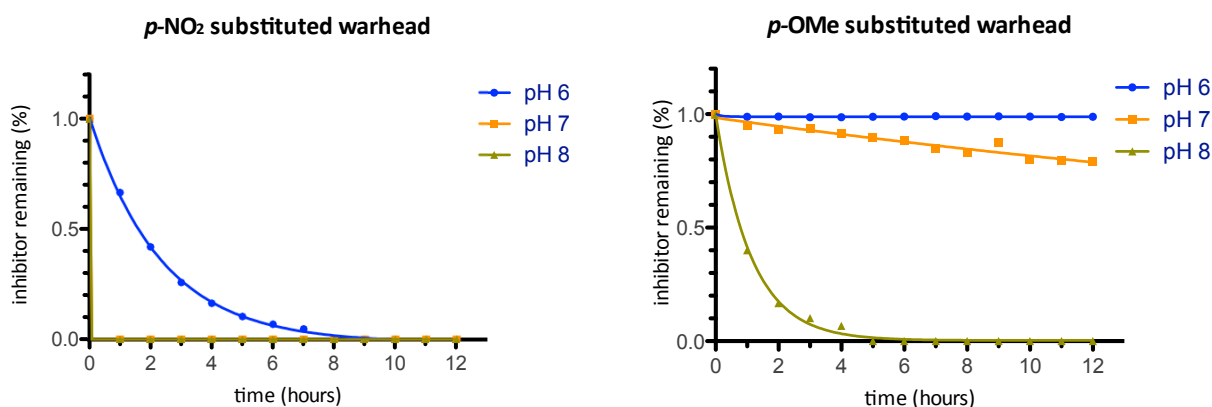
on previous work towards vinyl sulfone warheads which found first converting to the Weinreb amide, followed by reduction to the aldehyde and direct Horner Wadsworth Emmons (HWE) reaction was an efficient approach towards the synthesis of vinyl sulfones.<sup>200</sup> The Weinreb amide **53** offered a shelf stable and easily accessible aldehyde precursor whilst the HWE reagent **56** could be readily accessed by substitution of diethyl chlorophosphate with **55** yielding **56** in a good yield of 64%. The HWE reaction proceeded well giving vinyl sulfone **54** in a quantitative yield. The subsequent amide coupling to backbone **48**, synthesised previously, was low yielding at only 18%. It was thought that much of the product was lost during the purification as the tertiary amine became protonated and strongly adsorbed to the silica during chromatography before subsequent preparative HPLC yielding the TFA salt. This could be easily avoided in future synthesis by adding a base such as triethylamine to the eluent during chromatography, or perhaps direct purification by preparative HPLC. In this instance, further optimisation was not required as sufficient material was obtained for all required testing.

### 3.5 Stability Testing Towards Common Nucleophiles

Of any newly defined warhead class the stability and reactivity towards common organic nucleophiles should be considered. Not only does this prove useful for the design of new synthetic approaches to warheads of this type, by elucidating the conditions under which

the warhead is stable, but this also proves invaluable when considering biological activity by considering biologically relevant nucleophiles. It was hypothesised that testing the thiosulfonate warheads against a representative amine, representative thiol and pH aqueous stability would represent nitrogen, sulfur and oxygen centred nucleophiles respectively.

### 3.5.1 pH Dependent Aqueous Stability

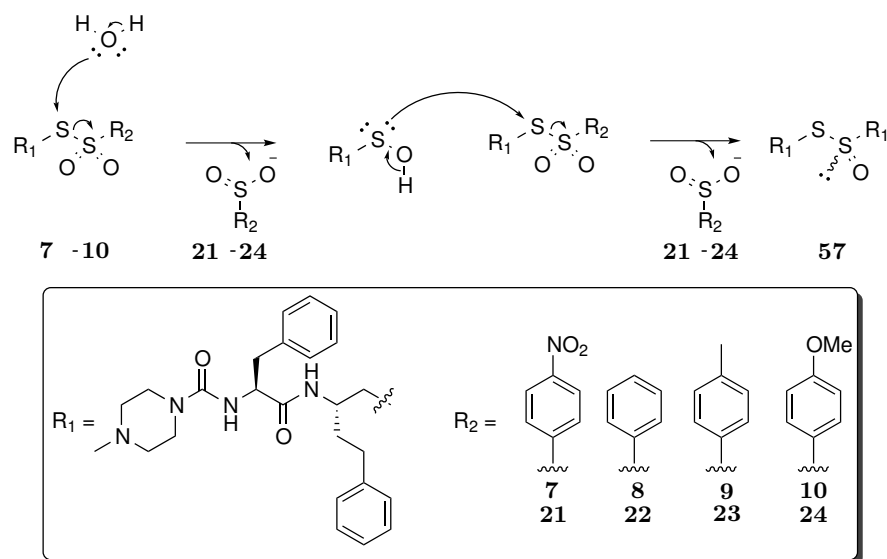


**Figure 3.5:** pH dependent aqueous hydrolysis of compounds **7** (left) and **10** (right) at pH 6 (blue), 7 (orange) and 8 (olive). Inhibitor degradation was determined by HPLC with peak areas integrated against an internal standard (Ac-Phe-OH), 13 measurements were taken 1 hour apart, representing time 0 to 12 hours. Sodium phosphate buffer (0.1 mM) of the relevant pH containing 5% DMSO was utilised, initial inhibitor concentration was 245  $\mu\text{M}$  and internal standard (Ac-Phe-OH) concentration was 338  $\mu\text{M}$ . Results were then normalised and plotted as scatter graphs for visualisation.

Starting with the pH dependent aqueous stability three pH's were examined (pH 6, 7 and 8) to represent a reasonable physiological pH range. Figure 3.5 represents the aqueous stability profiles of both the para-nitro substituted thiosulfonate inhibitor **7** and the paramethoxy substituted thiosulfonate inhibitor **10** over 12 hours as monitored by HPLC. It was anticipated that the warhead with an aromatic ring containing an electron withdrawing substituent (*p*-NO<sub>2</sub>, **7**) would lead to an increased reactivity of this warhead construct and, conversely, the electron donating substitution of compound **10** (*p*-OMe) would decrease the reactivity of the warhead. This was thought to be due to enhancing or reducing the leaving group ability of the sulfinate salt generated in each case. Gratifyingly, the aqueous stability profiles confirmed the highly tunable nature of this novel warhead construct showing a marked difference between the most electron rich and most electron deficient aromatic rings (**7** and **10**). As was anticipated, the stability was inversely proportional to pH, with increasing pH reducing the half life of the warheads.

Firstly, considering the stability profile of the *p*-NO<sub>2</sub> substituted compound **7**, it was found to be effectively unstable at pH 7 and 8, having completely decomposed within 1 hour, whilst at pH 6 a moderate half life of approximately 1.5 hours was observed. This would suggest that such a warhead would be incompatible for use in a more complex biological system, which is an important finding of this stability study. However, as a member of a first inhibitor series the true value of this result lies in comparison with the predicted most stable inhibitor construct **10** to demonstrate if the chemistry can be developed to control the reactivity of the thiosulfonate warheads.

Moving forward from the predicted most reactive warhead **7**, the *p*-OMe substituted warhead of compound **10** was found to be significantly more stable at across the tested pH range. This inhibitor construct was now effectively stable at pH 6 with a very reasonable half life of approximately 24 hours at pH 7 and even displayed a reasonable half life of approximately 1 hour at pH 8. This was a very encouraging finding, not only as this highlighted a potentially suitable warhead construct for use in more complex biological systems but the difference between the most reactive and least reactive warhead constructs could be confirmed and the highly tunable nature of the thiosulfonate displayed. Often, the ability to control the reactivity of a potential electrophilic trap is highly desirable in medicinal chemistry and chemical biology for the generation of specific molecular probes or inhibitors.<sup>143</sup> Thus, this early study highlights the suitability of thiosulfonates for further modifications and the ability to specifically tune their characteristics.

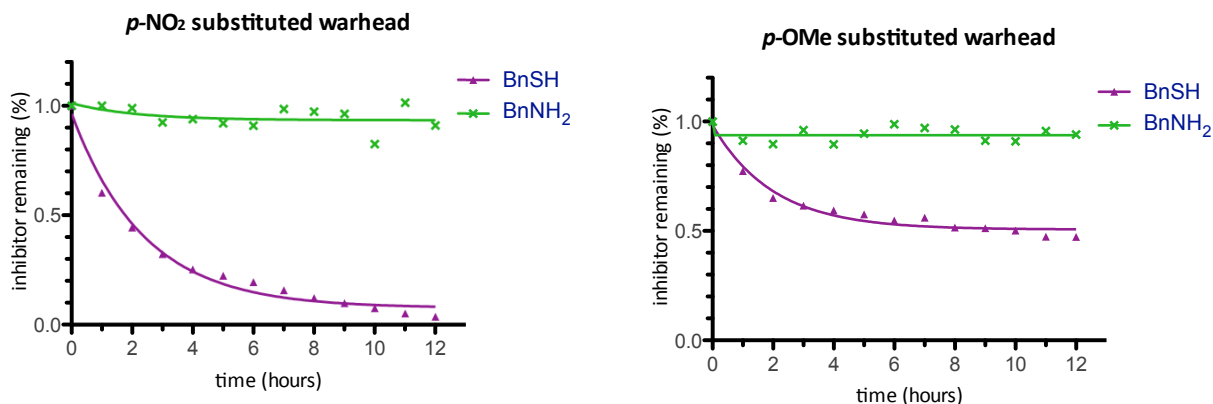


**Figure 3.6:** Chiral sulfoxides formed upon aqueous hydrolysis of thiosulfonate warheads *via* a sulfenic acid intermediate.

Furthermore this study was able to confirm the site of attack on the thiosulfonate warhead. As expected, the site of attack could be confirmed to be on the bivalent sulfur of the thiosulfonate. This was confirmed by the presence of two distinct peaks on the HPLC

chromatogram, found to be of the same mass by LC-MS and consistent with formation of chiral sulfoxide **57** (two peaks were observed due to separation of diastereomers). As shown in scheme 3.6 the mechanism of formation is proposed to occur through the sulfenic acid intermediate, generated by hydrolysis of the thiosulfonate warhead, leading to the chiral sulfoxide product shown. Interestingly, the secondary attack on the warhead by the generated sulfenic acid appears to outcompete further hydrolysis. This further confirms the highly selective nature of thiosulfonates towards sulfur centred nucleophiles.

### 3.5.2 Stability Towards Thiols and Amines - The Benzyl Thiol and Benzyl Amine Test System



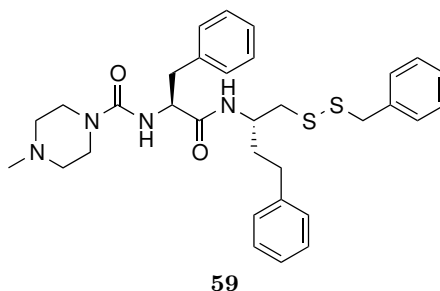
**Figure 3.7:** Stability towards BnSH and BnNH<sub>2</sub> of compounds **7** (*p*-NO<sub>2</sub> substituted, left) and **10** (*p*-OMe substituted, right). Inhibitor degradation was determined by HPLC in MeCN at room temperature with 13 measurements taken 1 hour apart, representing time 0 to 12 hours. Approximately 0.3 mg of inhibitor was accurately weighed and suspended in MeCN, to which 10 equivalents of BnSH or BnNH<sub>2</sub> was added immediately before measurement. Results were then normalised against time 0 and plotted as scatter graphs for visualisation.

Next the stability towards the common organic nucleophiles of interest, amines and thiols, was to be assessed. The main consideration here was to assess the chemical stability towards relevant nucleophiles in a chemical synthesis context. This was done for both the most reactive compound **7** bearing the *p*-NO<sub>2</sub> substituted warhead and the least reactive inhibitor **10** bearing the *p*-OMe substituted warhead. Although the findings could be extended towards biologically relevant nucleophiles, additional factors would have to be considered, such as the pH of the local environment, which was intentionally examined as a separate parameter in this work (section 3.5.1) to better understand the reactivity of the thiosulfonate warheads.

Firstly, benzyl amine was taken as a representative amine which was UV active to allow

progress to be monitored by UV absorption through HPLC analysis. 10 equivalents of benzyl amine were added to the thiosulfonate inhibitor construct in MeCN and the reaction progress monitored over 12 hours. Gratifyingly the thiosulfonates were found to be effectively stable toward benzyl amine over a 12 hour time period, even the more reactive *p*-NO<sub>2</sub> substituted compound. Of course, amines are common nucleophiles in organic synthesis so knowing that the thiosulfonate electrophilic trap displays little reactivity towards these is an encouraging finding for future work. As such, it should be possible to develop chemistry which requires a nucleophilic amine in the presence of the thiosulfonate, one example of which has already been displayed in the final amide coupling step of inhibitor constructs (section 3.3, scheme 3.10).

Moving on to the stability towards thiols, benzyl mercaptan was taken as a representative thiol. Again, the benzylic group was chosen as it allowed the time dependent concentration of the thiol to be monitored by UV absorption on HPLC. It was hypothesised that the thiosulfonates would display an inherent reactivity towards sulfur centred nucleophiles, as this is a fundamental feature believed to be beneficial towards specifically targeting biologically relevant sulfur centred nucleophiles, such as cysteine proteases. Encouragingly the thiosulfonates were found to react with a thiol in a manner consistent with their predicted reactivity profiles with the *p*-NO<sub>2</sub> substituted inhibitor **7** reacting almost entirely over 12 hours, while the less reactive *p*-OMe inhibitor **7** gradually decreased to 50% of its initial concentration. This mirrored the previously observed reactivity profiles during aqueous stability testing in section 3.5.1, confirming that the more electron deficient thiosulfonate warhead was more reactive towards incoming thiols and the more electron rich aromatic ring was less reactive towards incoming thiols.



**Figure 3.8:** Asymmetric disulfide formed by attack of benzyl mercaptan on thiosulfonate inhibitors

Again, it was observed by LC-MS that the site of attack on the thiosulfonate warhead was the bivalent sulfur, as the asymmetric disulfide (compound **59**, figure 3.8) was observed for both inhibitor constructs, giving further insight into the reactivity of this novel warhead construct. Furthermore the exceptional selectivity observed between amines and thiols was a highly encouraging result, suggesting that the fundamental premise of utilising a thiosulfonate as a warhead specific for sulfur centred nucleophiles was well founded. This

reactivity profile is also complementary to current warheads being developed within the Liskamp group, the sulfonyl fluorides, which are most reactive towards serine / threonine OH.<sup>178,181</sup> It is believed that developing such complementarity between warhead classes will help to expand the chemical toolbox available for probing specific biological problems or designing new inhibitors, such as those displayed in this work, with enhanced specificity.

Importantly, it was considered that this highly tunable nature could be pushed further, reducing the reactivity of the thiosulfonate warhead whilst maintaining specificity towards sulfur centred nucleophiles. In doing so, it was hoped that the thiosulfonates would be able to withstand biologically relevant thiols, such as circulating thiols, and upon binding to a target be able to illicit the desired effect. This was in part thought to be possible due to the specific environment invoked upon binding a target molecule which in many cases can enhance the reactivity of the warhead. Indeed, this very argument has been put forward previously during docking studies in section 3.1, which displayed the potential *in situ* activation of the thiosulfonate warheads. Thus the ultimate goal of this work became a two fold issue, which was successfully addressed; (1) to design an inhibitor backbone capable of delivering the warhead to a specific target and (2) to explore chemistry to downregulate the warhead reactivity such that circulating thiols, and indeed aqueous stability, were of little to no concern.

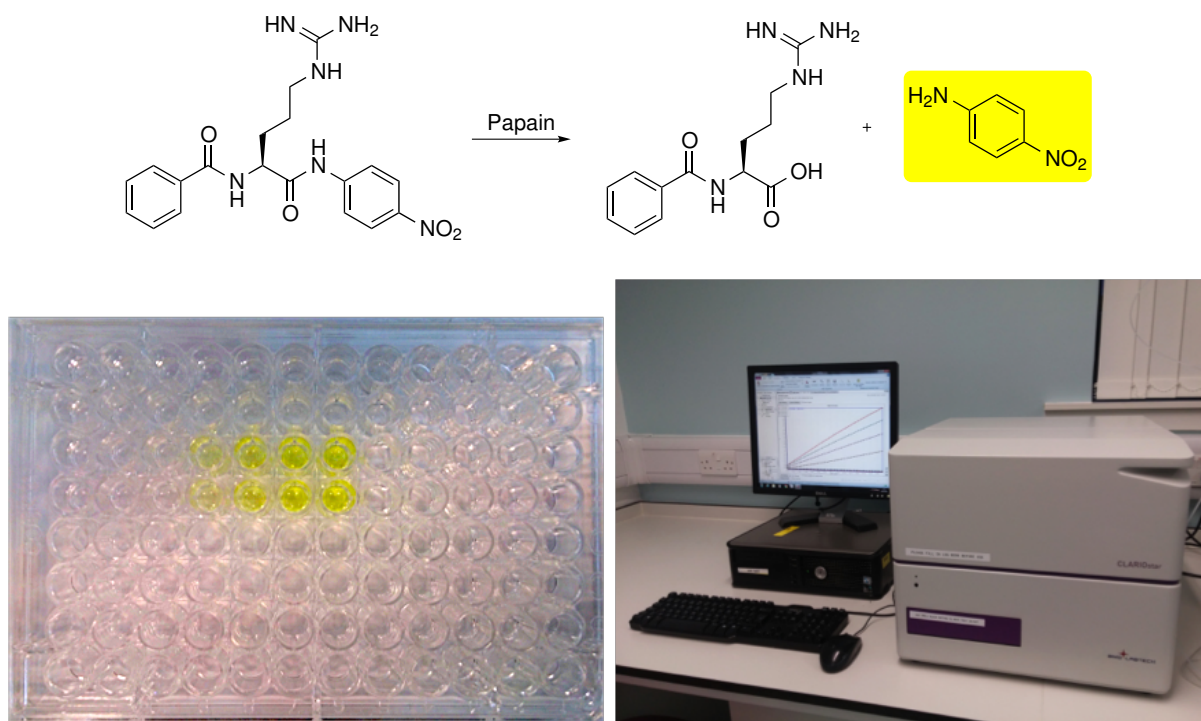
In summary, these stability studies suggested that the thiosulfonates displayed great potential, and that their highly tunable nature may be a desirable quality. Further work should be targeted towards increased stability combined with the ability to confer specificity through backbone design. Addressing these key issues would ensure the uptake, and ultimately success, of this newly emerging warhead construct. This may be of particular important in more complex biological systems where considerations, such as reaching the target macromolecule before degradation, become increasingly important. With this in mind, it was decided to move forward to simple biological testing to first assess the suitability of thiosulfonates as cysteine protease inhibitors in an enzymatic study. This would provide a baseline to work from and, if the thiosulfonates were deemed to offer suitable characteristics, the chemistry to further regulate the reactivity could be explored leading on to a potentially powerful new warhead moiety.

### 3.6 Biological Testing - Papain Enzyme Assay

Papain was selected as a prototypical cysteine protease being of the clan CA, family C1. Importantly, papain has been widely studied and is more economically favourable over other more specialised clan CA, family C1 proteases (such as parasitic analogues), offering

an ideal platform as an initial protease target. It was hypothesised that the high homology between these proteases would allow for rational conclusions to be drawn, which would still be applicable to more biologically relevant cysteine proteases.

### 3.6.1 Developing a Thiol Independent Assay

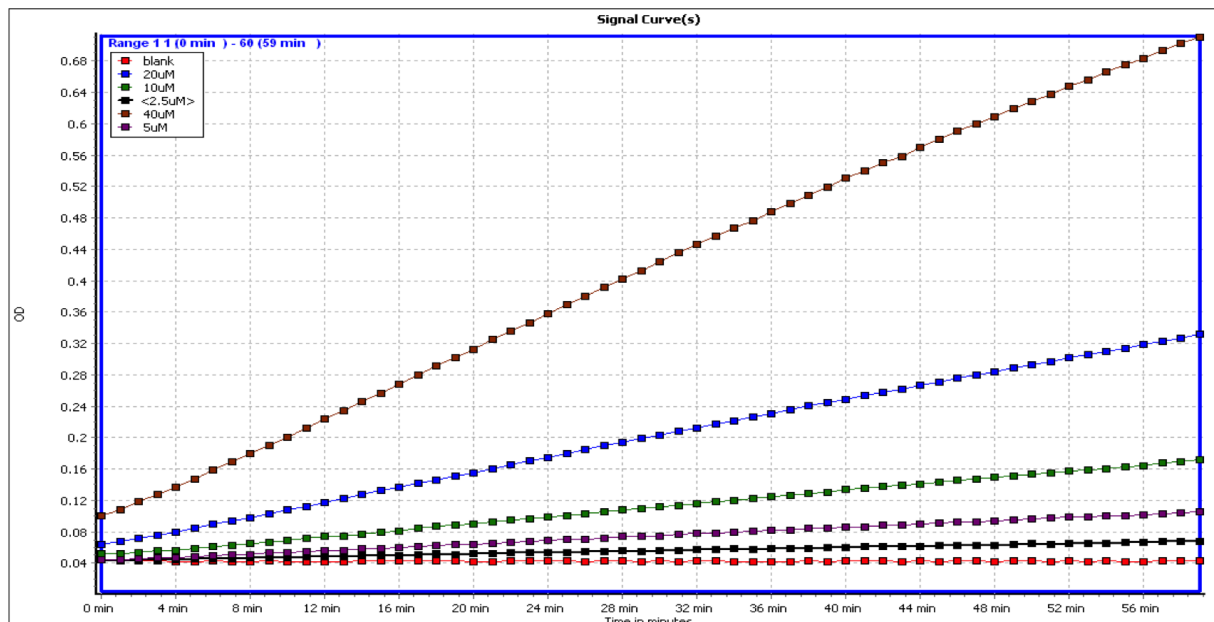


**Figure 3.9:** Papain assay overview: Hydrolysis of  $N_\alpha$ -benzoyl-L-arginine 4-nitroanilide substrate by papain yielding 4-nitro anilide chromophore (top) to be quantified by absorbance at 405 nm with a plate reader (bottom right) in a 96 well plate (bottom left).

Thiols are commonly used in cysteine protease enzymatic assays as activators to ensure the active site thiolate is in the reduced form. For preliminary biological testing it was thought to be beneficial to remove thiols from the system due to potential cross reactivity with the thiosulfonate warheads, as outlined in section 3.5.2. Previous work in the Liskamp group, conducted by Brouwer *et al.*<sup>183</sup> towards sulfonyl fluoride warheads developed a thiol independent papain assay for similar reasons, which served as a starting point for this work. The principle of the assay is outlined in figure 3.9 and is based on detecting the absorbance of the 4-nitroanilide chromophore produced upon hydrolysis of a substrate ( $N_\alpha$ -benzoyl-L-arginine 4-nitroanilide). As initial rate of substrate hydrolysis is directly proportional to the active enzyme concentration, this allows for quantification of active enzyme for assay setup and subsequent inhibition studies.

As the main purpose of thiols in cysteine protease enzymatic assays is to ensure the active site thiolate is in the active, reduced form it is possible to conduct the assay without



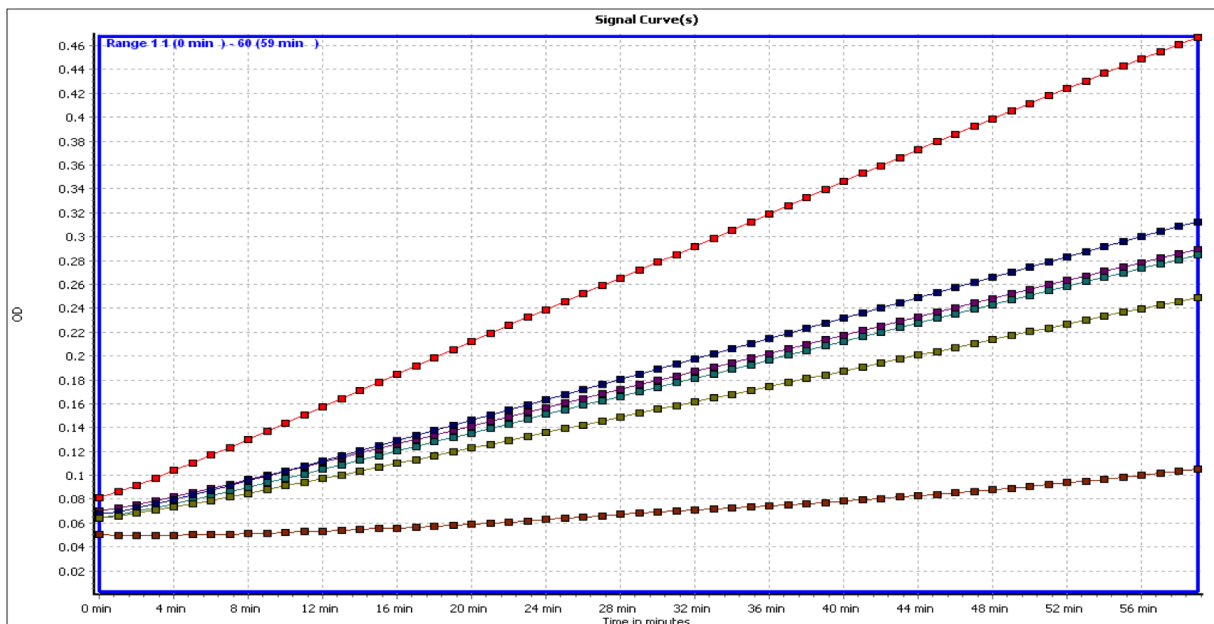


**Figure 3.10:** Determining papain concentration for thiol independent assay. Doubling dilution range from 40 – 2.5  $\mu\text{M}$ . From top to bottom: 40  $\mu\text{M}$  (brown); 20  $\mu\text{M}$  (blue); 10  $\mu\text{M}$  (green); 5  $\mu\text{M}$  (purple); 2.5  $\mu\text{M}$  (black); blank (red).

thiols present.<sup>183</sup> This results in a significant portion of the tested enzyme being in an inactive form thus higher enzyme concentrations can be used to counteract this. In this instance, approximately 10% of the enzyme was thought to be active and the first step was to define an adequate working concentration, as shown in figure 3.10. This was achieved by assessing a dilution range of enzyme with a colorimetric substrate to find the minimum working concentration which gave an adequate response, with an enzyme concentration of 20  $\mu\text{M}$  being found as optimal for further testing.

### 3.6.2 Testing Initial Inhibitor Series

With a thiol independent enzyme assay now defined, testing of the first thiosulfonate inhibitor series could take place (compounds **7** - **10** figure 3.11). Inhibitors were first screened at a broad concentration range to find the concentration at which differences in inhibition could be observed (between 100% inhibition and 0% inhibition), which was found to be 625 nM. As shown in figure 3.11 all inhibitors were tested at this concentration and a clear trend could be observed between the reactivity of the warheads and the potency of the inhibitor. More potent inhibitors reduced the enzyme activity, resulting in less observed hydrolysis of the substrate ( $N_{\alpha}$ -benzoyl-L-arginine 4-nitroanilide), which can be visualised as a less steep gradient of the linear response in figure 3.11. Encouragingly all inhibitors operated within the same order of magnitude as the reference compound K11777, although this was clearly the most potent inhibitor tested. Furthermore it was observed

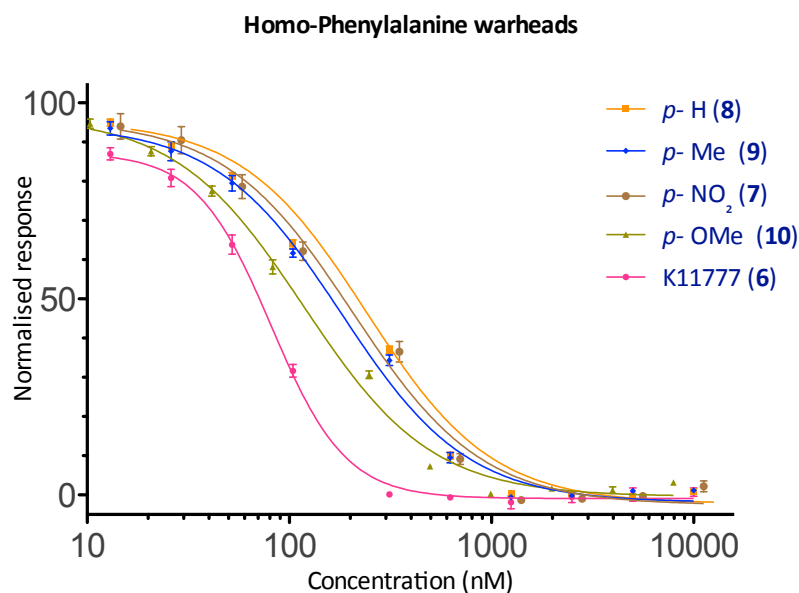


**Figure 3.11:** Papain inhibition curved for homophenylalanine derived aromatic thiosulfonates. From top to bottom the curves represent: No Inhibitor (red), **7** (dark blue), **8** (purple), **9** (cyan), **10** (olive), K11777 (**6**, maroon). All inhibitors were tested at 625 nM, 20  $\mu$ M papain.

that reducing the reactivity of the warheads resulted in increasingly potent inhibitors. This trend can be clearly observed with the most reactive *p*-NO<sub>2</sub> substituted inhibitor **7** showing the least inhibition (figure 3.11, dark blue curve). This trend was observed through the reactivity series and followed the predicted reactivity profiles with by the next most reactive compound **8** (no substitution, purple curve) being followed by **9** (*p*-Me substituted, cyan curve) and finally the least reactive inhibitor **10** (*p*-OMe substituted, olive curve) being the most potent in the thiosulfonate series.

The observed trend correlated well with the earlier stability studies, suggesting that the observed trend in potency may be due to partial decomposition of the thiosulfonate warheads under assay conditions. Upon comparison of the assay conditions (incubation time of 1Hr, pH 6.5) with the aqueous stability profiles in section 3.5.1, it was surprising that this effect was not more pronounced. It was previously observed that increased aqueous stability of the thiosulfonate warheads correlated with decreased pH, as expected. Compound **7** was previously observed to have completely decomposed within 1 hour at pH 6. As such, the fact that the inhibition falls within such a close range of other thiosulfonate warhead constructs, and indeed the vinyl sulfone reference compound, was remarkable. This suggests that the thiosulfonates are very fast acting, inhibiting the papain before decomposition can take place. This was a very encouraging finding as it suggests that the target binding and subsequent inhibition must be highly efficient for the thiosulfonate constructs enabling them to outcompete decomposition, even for relatively unstable warheads. Furthermore this may be in part attributed to the *in situ* activation of the thiosulfonates observed in

section 3.1, enhancing their reactivity in the binding pocket relative to the surrounding aqueous environment. Most importantly, this study suggested that the thiosulfonates were highly suitable warheads for cysteine protease inhibition and that further development to control their reactivity may be able to further improve their stability and potency simultaneously. Having demonstrated the highly tunable nature of the thiosulfonates it was believed that the chemistry surrounding the warhead stability could be further explored leading to a powerful new warhead class for specific inhibition of cysteine proteases.



Compound	Warhead	IC <sub>50</sub> nM
8		190 ± 6
9		182 ± 5
10		148 ± 6
7		186 ± 11
6 (K11777)	Vinyl sulfone reference compound	81 ± 2

**Figure 3.12:** IC<sub>50</sub> determination of homophenylalanine derived aromatic warheads. Assay buffer: sodium phosphate (100 mM, pH 6.5) containing EDTA (1.5 mM) and 2% DMSO. Final concentrations in the wells were; Papain: 4 μM; substrate: 1.0 mM; inhibitor: doubling dilution range starting at 10 μM giving: 10 μM, 5 μM, 2.5 μM, 1.25 μM, 0.625 μM, 0.3125 μM, 156.25 nM, 78.125 nM, 39.0625 nM. Error bars on graph display SEM across 3 independent repeats. Tabulated ± values represent the std error of the IC<sub>50</sub> values across 3 independent repeats.

The initial inhibition studies provided an excellent visual aid between reactivity and inhibition, to further extend this the IC<sub>50</sub> values were determined as shown in figure 3.12.

Initially, the  $IC_{50}$  values obtained correlated very well with the earlier observed inhibition studies when conducted with an enzyme concentration of 20  $\mu\text{M}$ . However, upon testing later inhibitor series it was found that there was inconsistency in the data due to batch variability in the active enzyme concentration of the papain. As such, a new batch of papain was purchased and the active papain concentration screened as it was earlier in figure 3.10, which was found to be more active than the previous batch. As such, a lower papain concentration of 4  $\mu\text{M}$  was used to correct for this resulting in an optical density (OD) of 0.2, consistent with previous experiments. This papain batch was then made into aliquots and stored at  $-80\text{ }^{\circ}\text{C}$  and all inhibitors retested with the same papain batch. The same papain batch was also used for all other inhibitors tested in this work to ensure consistency and avoid batch variability. Looking towards the results in figure 3.12 it can be seen that all inhibitors displayed similar  $IC_{50}$  values in the sub-micromolar range of 140-190 nM, similar to that of the reference compound K11777 at 80 nM. It was hypothesised that the relatively consistent  $IC_{50}$  values represented effective titration of the enzyme at the tested concentration range, with small difference observed in  $IC_{50}$  values reflecting competing pH dependent decomposition of inhibitor constructs correlating with earlier stability studies in section 3.5. As such, this assay offered a convenient testing platform to optimise the reactivity profiles of tested inhibitor constructs. Once stability arguments have been overcome, further optimisations should be made to this assay to reflect accurate  $IC_{50}$  determination with suggestions made in the following section (3.6.3, Future work on papain assay).

Moving forward there were two clear routes to be explored for the thiosulfonate derived cysteine protease inhibitors. Firstly, improved potency by improving the binding affinity with the target protease and secondly improved potency by improving the stability of the warhead. Improving the binding affinity is a common technique, often forming a large proportion of a structure-activity relationship; as such a convergent synthesis to incorporate this with the thiosulfonate warheads was to be explored. Following on from this, techniques to improve the inhibitor stability will be explored with the hypothesis that these will be synergistic with the noncovalent interactions of the backbone motif, further improving overall inhibitor design.

### 3.6.3 Future work on papain assay

The current papain assay offers a convenient platform to study the relationship between warhead stability (based on reactivity) and inhibition potential. There are however limitations which should be addressed once sufficiently stable warheads have been realised. Such assay optimisation should begin with detailed characterisation of the enzymatic and assay parameters, such as the  $K_m$  and  $K_{cat}$  of the enzyme and substrate combination used,

the amount of substrate turned over in the assay timeframe, the signal/background, and the  $Z'$  factor. Interested readers are referred towards the literature guidelines and reviews for assay optimisation.<sup>201–203</sup>

An extension of these optimisations is that the current assay appears to be titrating the enzyme with the inhibitor constructs, as such methods to overcome this should be considered. One area for optimisation would be the incubation time, as these inhibitors are proposed to be covalent the current incubation time of 1 hr may not reflect inhibitor binding in the  $IC_{50}$  value, but rather titrate the enzyme. Taking this into consideration, with the best compound (K11777) displaying an  $IC_{50}$  of 81 nM, there can not be more than 162 nM of active enzyme in the assay. This also suggests that the enzyme concentration is too high to characterise such potent inhibitors, and the current assay is not sensitive enough to accurately differentiate these inhibitors beyond comparing rates of decomposition vs rates of reaction with enzyme. A potential method to address these areas would be to reduce the incubation time, potentially to 0 by adding the inhibitors as the final assay component and observing the drop in substrate turnover. Methods to increase the sensitivity of the assay could focus on reduced enzyme concentration, which will in turn reduce the substrate turnover. With reduced substrate turnover increased sensitivity may be required in the assay readout, which could be achieved by exploring fluorescent substrates if required. As the inhibitors being tested are anticipated to be competitive in nature, substrate concentration dependence of inhibition may also be explored. This will then allow further kinetic measurements to be taken, such as determining  $K_i$  and  $K_{inact}$  for the inhibitors.

### 3.7 Summary

The first generation of peptidomimetic thiosulfonate inhibitors have been explored. Modelling studies suggested that the thiosulfonates were well placed to be explored as cysteine protease inhibitors which would be well accommodated within the active site and also offer potential *in situ* activation through the localised hydrogen bonding network. Furthermore this hydrogen bonding network is specific to cysteine proteases, further increasing the selectivity towards proteases of this class. An efficient synthesis towards thiosulfonate warheads was developed which allowed for incorporation of a broad reactivity series by differing the substitution on the aromatic ring of warheads. This broad reactivity series allowed a greater understanding of the relationship between reactivity and activity of this new warhead class. It was found that the thiosulfonate warhead was highly tunable, thought to be one of the key strengths of this warhead class, and that the predicted reactivity profiles correlated well with those observed in stability studies. A thiol independent enzyme assay

was developed to allow for testing of this initial inhibitor series which showed that the thiosulfonates were promising inhibitors of the cysteine protease papain. This assay also highlighted that future warhead development may be centred around warhead reactivity and stability. From this, a new inhibitor series can be developed to further improve on the potential of the thiosulfonates. This approach will be two fold, first targeting increased binding (led by modelling studies) and exploring new chemistry surrounding the warhead moiety to further improve the aqueous stability profiles of the thiosulfonates.

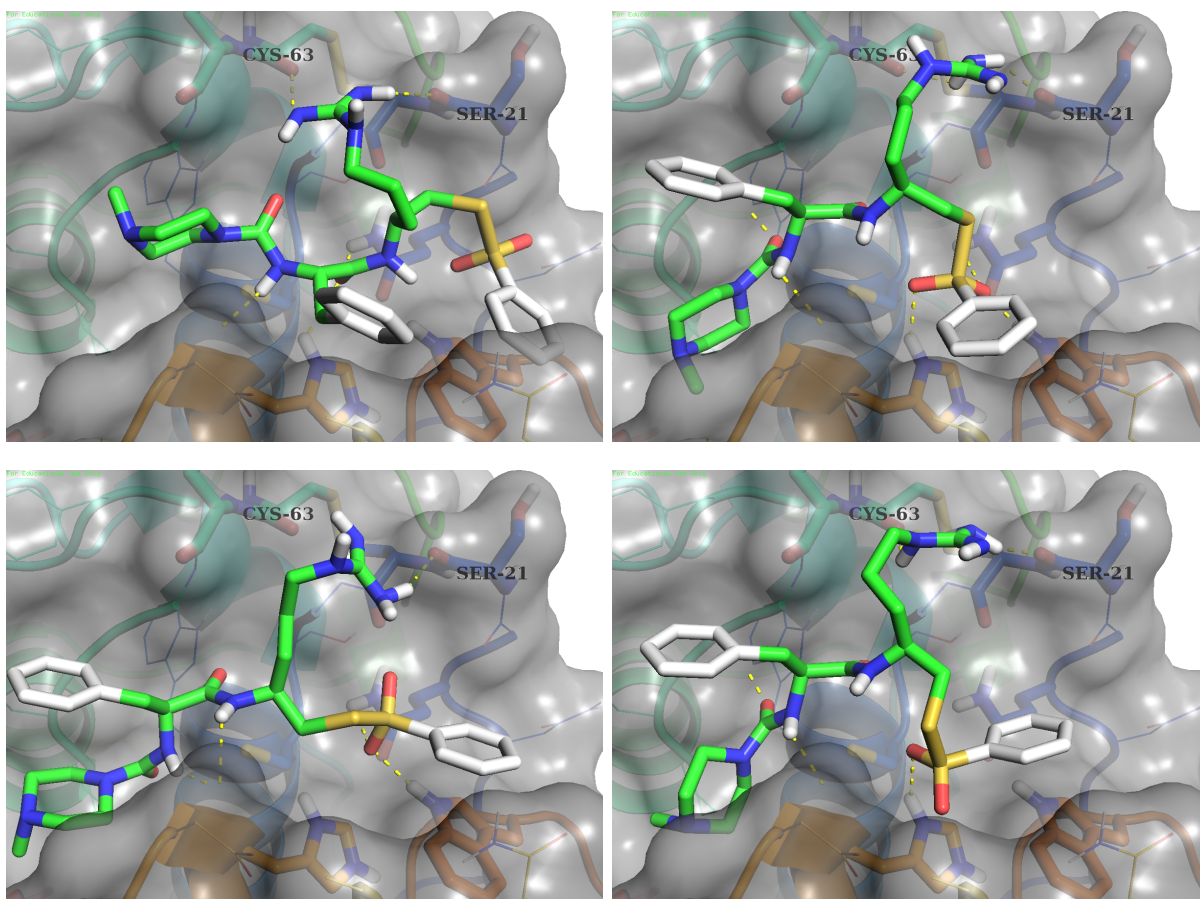
## Chapter 4

# Second Generation Inhibitors: A Convergent Strategy to Improved Binding

The ability to specifically target and differentiate between enzymes is of the utmost importance when designing a new inhibitor construct. As the thiosulfonates in this work are peptidomimetic in nature, and derived from amino acids, altering the amino acid used should have a direct impact on the binding of the final inhibitor constructs. This has been incorporated by design to facilitate an easily accessible method to improve the primary binding event of thiosulfonate inhibitors which in turn should correlate with an increased potency. To illustrate this, a modelling study was conducted to uncover a residue for improved binding of the thiosulfonates against papain. It was hypothesised that this improved binding would correlate with an improved potency, which would have an additive effect when combined with the more stable (and potent) thiosulfonate warheads observed previously.

### 4.1 Docking Study: Targeting Residues Introduced with Warhead Fragments

It has been well documented for papain like cysteine proteases that the P<sub>1</sub> residue plays an important role in the specificity of the inhibitor towards the target protease (see section 1.1.1. In the case of thiosulfonate inhibitors, the P<sub>1</sub> residue is incorporated along with the thiosulfonate moiety. Thus, it was thought to be important to highlight that alternate building blocks could be utilised to increase the binding affinity towards a specific



**Figure 4.1:** Docked pose of arginine-derived thiosulfonate inhibitor **14** with papain. 4 possible docked poses, all highlight coordination between backbone carbonyls of Cys<sup>63</sup>, Ser<sup>21</sup> and the arginine side chain.

protease.

In order to identify a suitable amino acid residue to incorporate, the crystal structure of papain (PDB code 1CVZ) was examined. As shown in figure 4.1, it was clear that the bottom of the P<sub>1</sub> binding pocket offered two potential H-bond acceptors (backbone carbonyls of Cys<sup>63</sup> and Ser<sup>21</sup>). The nature of having two H-bond acceptors in close proximity to each other suggested that a complementary residue, with two potential H-bond donors in close proximity, would offer an ideal residue. Considering the naturally occurring amino acids, arginine was the obvious candidate for this. The proposed suitability of arginine was further supported by a literature study whereby Turk *et al.*<sup>16</sup> proposed that the electrostatic environment at the bottom of the S<sub>1</sub> binding pocket, largely dominated by the backbone carbonyls, would be suitable for coordination to an arginine residue. To further explore this, and provide a visual cue to the suitability of the chosen residue in the P<sub>1</sub> position, *DrawToDock* was utilised to dock inhibitor **14** with Papain (PDB 1CVZ).

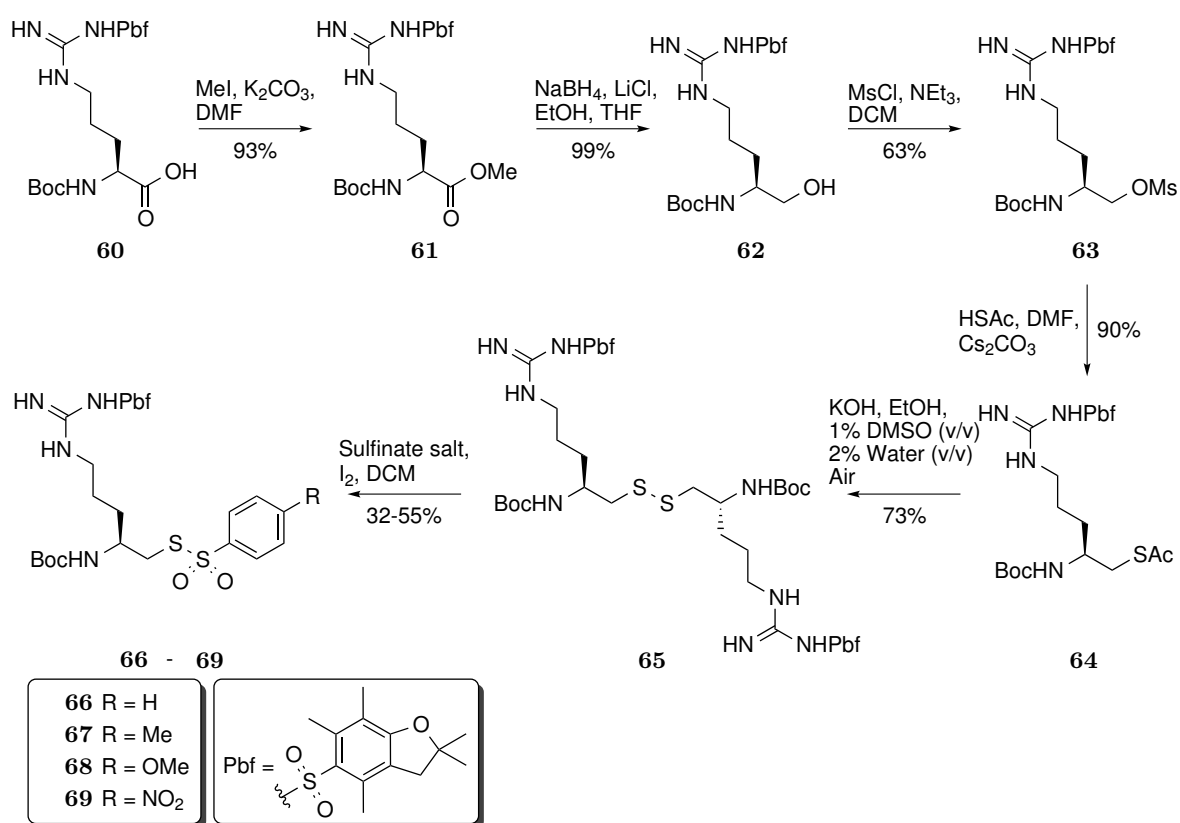
It was noted that the introduction of an arginine residue in the P<sub>1</sub> position placed the



arginine in close proximity to the backbone carbonyls of Cys<sub>63</sub> and Ser<sub>21</sub>, facilitating H-bonding (figure 4.1). Another distinct feature observed was the overall binding mode of the inhibitor had been altered. The P<sub>3</sub> *N*-methylpiperazine cap was no longer predicted to be solvent facing. This is likely due to the docking software placing a greater importance on the H-bonding which is now present in the P<sub>1</sub> binding site. In any case, it was hypothesised that enhancing the initial binding event would improve the potency of the inhibitors as the subsequent covalent bond formation is likely to be under kinetic control, thus any improvement in the initial binding (and bringing the warhead into closer proximity with the active site nucleophile) will likely result in an improved potency. Having proposed that an arginine residue is a suitable candidate to improve the non-covalent interactions of the warhead fragment, subsequent synthesis and testing could now take place.

## 4.2 Arginine Derived Thiosulfonates: Exploiting a Flexible Synthesis

### 4.2.1 Synthesis of Arginine Derived Warheads

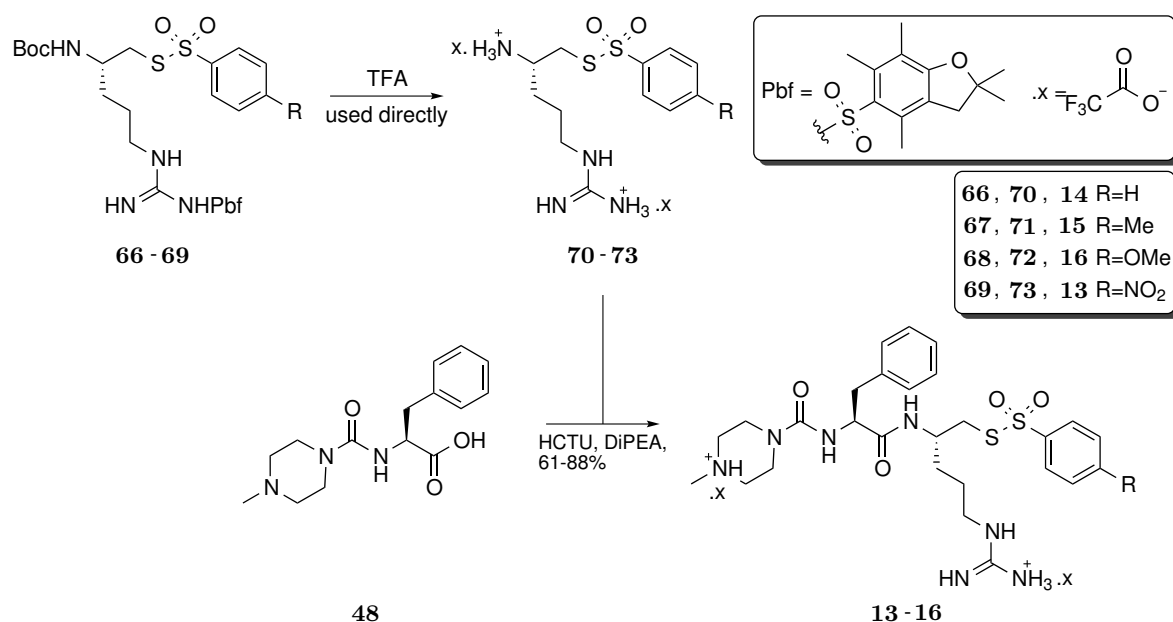


**Scheme 4.1:** Arginine derived aromatic thiosulfonates by oxidative cleavage of disulfides

Based on the earlier work towards thiosulfonate warheads, which defined a robust synthesis

towards amino acid derived thiosulfonates, a similar synthesis was conducted. This began with the protected arginine derivative Boc-Arg(Pbf)-OH (**60**). The Pbf group was selected as a suitable protecting group for this synthesis as the deprotection conditions (acidic) were known to be compatible with the thiosulfonate warhead from the previous synthesis. Following the same protocol as previously the carboxylic acid **60** was converted to the methyl ester **61** before reduction to the corresponding alcohol **62** in excellent yields of 93 and 99% respectively. Next, alcohol **62** was converted to mesylate **63** before introducing the thioacetate by substitution with thioacetic acid in a good to excellent yield of 63 and 90% respectively. Subsequent hydrolysis and oxidation to disulfide **65** proceeded in good yields (73%), although it was observed that the oxidation step from thiol to disulfide was less efficient than in the case of the homophenylalanine derivative. This was thought to be due to the increased sterics of the bulky Pbf protecting group on the side chain, as remaining thiol was observed by TLC and separated out during purification. Finally, the thiosulfonate warheads were isolated by the oxidative cleavage of disulfide **65** giving warhead building blocks **66 - 69** in moderate to good yields (32-55%).

#### 4.2.2 Final Inhibitors: Coupling Warheads to the Backbone

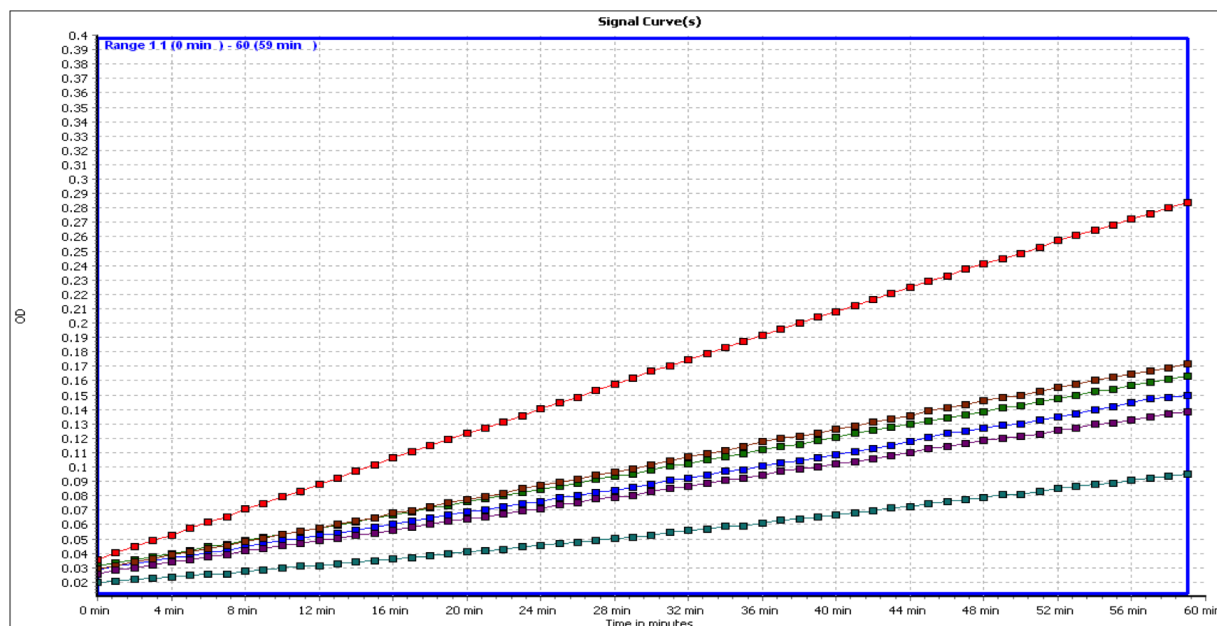


**Scheme 4.2:** Backbone synthesis and coupling to arginine derived aromatic warheads.

Moving towards final inhibitor constructs, the protected warhead fragments were first deprotected with TFA, which was required in the case of arginine constructs to effectively remove the Pbf protecting group as well as the Boc group. Previously, HCl was chosen to remove the Boc group as the subsequent coupling reaction could potentially take place between the carboxylic acid of the TFA salt rather than the desired backbone. Subsequent

coupling to the backbone **48** proceeded with good yields of 61-88% giving inhibitors **13** - **16** as TFA salts after preparative HPLC.

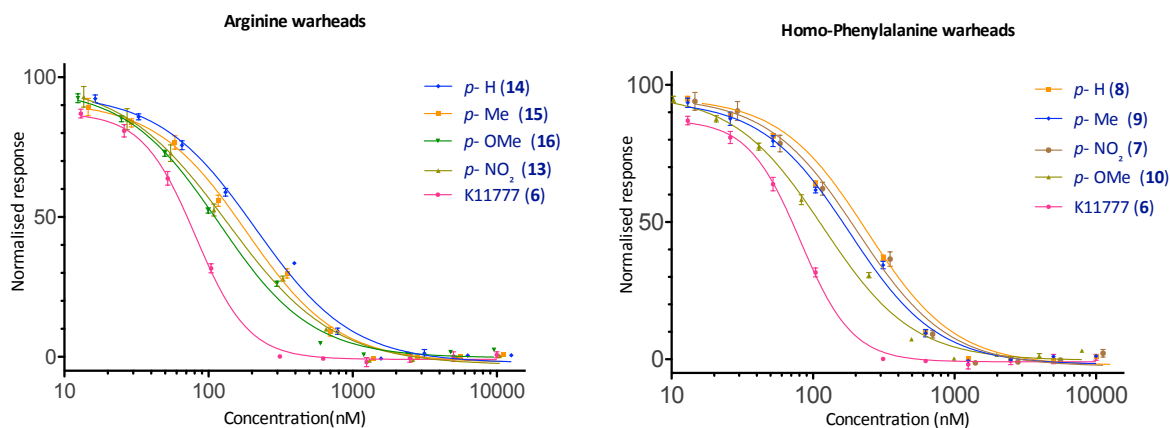
### 4.2.3 Biological Testing: Papain Enzyme Assay

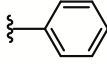
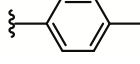
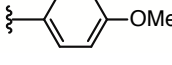
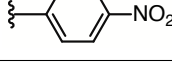
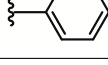


**Figure 4.2:** Papain inhibition curved for homophenylalanine derived aromatic thiosulfonates. From top to bottom the curves represent: No Inhibitor (red), **13** (maroon), **14** (green), **15** (blue), **16** (purple), K11777 (**6**, turquoise). All inhibitors were tested at 625 nM, 20  $\mu$ M papain.

The arginine derived inhibitors were first screened against papain as before for the homophenylalanine constructs at 625 nM to visualise any differences between each warhead. As can be seen in figure 4.2 the same trend was observed between warhead reactivity and potency, with the most reactive *p*-NO<sub>2</sub> substituted warhead being the least potent while the least reactive *p*-OMe substituted warhead was the most potent. Gratifyingly, all inhibitor constructs appeared to perform better than the previous homo-phenylalanine constructs being grouped closer to the reference compound K11777 (**6**, turquoise line), suggesting that more potent inhibitors have been generated. Although screening at the same concentration offers a great visual aid for comparison of warhead constructs, the IC<sub>50</sub> values were then determined to allow a more quantitative comparison between warhead constructs.

In order to determine the IC<sub>50</sub> values of the warhead arginine derived warheads each inhibitor was screened against papain at a doubling dilution range. This was conducted with the same batch of papain as used previously to remove any batch to batch variability in the active enzyme concentrations. All arginine-derived inhibitors performed better than the homophenylalanine-derived warheads with an approximate improvement of 20 nM



Warhead S <sub>1</sub> ' substituent	Arginine Compound number	Arg IC <sub>50</sub> nM	Homo-Phe Compound number	Homo-Phe IC <sub>50</sub> nM
	<b>14</b>	168 ± 3	<b>8</b>	190 ± 6
	<b>15</b>	162 ± 18	<b>9</b>	182 ± 5
	<b>16</b>	124 ± 1	<b>10</b>	148 ± 6
	<b>13</b>	129 ± 2	<b>7</b>	186 ± 11
	N/A	N/A	<b>6 (K11777)</b>	81 ± 2

**Figure 4.3:** IC<sub>50</sub> determination of arginine-derived aromatic warheads. Assay buffer: sodium phosphate (100 mM, pH 6.5) containing EDTA (1.5 mM) and 2% DMSO. Final concentrations in the wells were; Papain: 4 μM; substrate: 1.0 mM; inhibitor: doubling dilution range starting at 10 μM giving: 10 μM, 5 μM, 2.5 μM, 1.25 μM, 0.625 μM, 0.3125 μM, 156.25 nM, 78.125 nM, 39.0625 nM. Error bars on graph display SEM across 3 independent repeats. Tabulated ± values represent the std error of the IC<sub>50</sub> values across 3 independent repeats.

seen for all warhead construct when compared with their homophenylalanine counterparts (figure 4.3). This suggested that the modelling studies were correct and an improvement in potency has been achieved by increasing the initial binding event ( $K_i$ ), leading to inhibition before competing aqueous degradation, although further studies would be required to accurately determine  $K_i$  values. The hypothesis that inhibitor binding is out competing degradation to improve the  $IC_{50}$  values is also supported by the least reactive (most stable) warhead of compound **16** exhibiting a marginally improved potency compared with other warhead constructs. Interestingly, the *p*-NO<sub>2</sub> substituted warhead of inhibitor **13** was found to be equipotent to **16** in this study. This could, in part, be explained by the warheads contributing towards the  $K_i$  which would improve the initial binding. This supports the hypothesis that the inhibitor potency is determined by the initial binding event as well as the reactivity of the warheads. Furthermore, the *p*-NO<sub>2</sub> group of compound **13** offers an additional polar contact which may improve the binding of inhibitor constructs and explain the ability of the more reactive warheads to offer improved inhibition over the less reactive warheads. This leads on to the hypothesis that all warhead constructs must be fast acting, outcompeting aqueous degradation, which would be largely controlled by the initial binding event ( $K_i$ ). However further assay development and experiments would be required to fully interpret and discuss any kinetic effects, such as those outlined in section 3.6.3.

Although the variation observed between each warhead tested in this this assay is relatively small (1.3 fold), it is reasonable to suggest that this may become exaggerated in more complex biological systems, for example, a biological system whereby the time taken to reach the desired target will be significantly longer. Furthermore, the inhibitor constructs will have to safely negotiate numerous biologically relevant nucleophiles such as circulating thiols as well as the surrounding aqueous environment. Thus, future efforts focused on downregulating the reactivity of the thiosulfonate warhead as it was believed this would be favourable for selective target engagement in a more complex biological system where competing degradation pathways would be of greater significance.

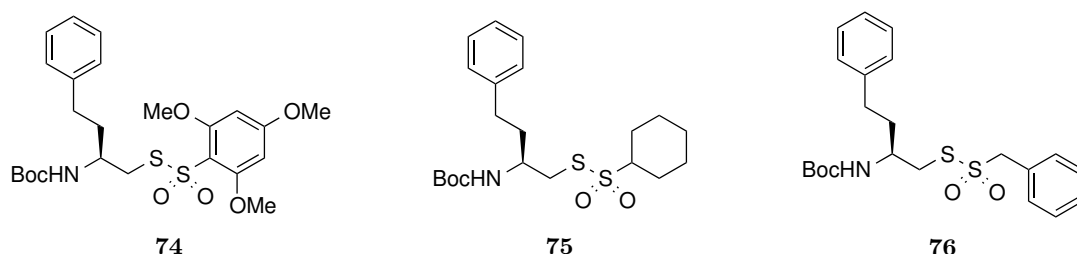
## 4.3 Summary

In summary, modelling studies suggested that introducing an arginine side chain in the P<sub>1</sub> position, simultaneously with the warhead fragment, would improve the binding of thiosulfonate derived inhibitors towards papain. This was hypothesised to result in an improved potency highlighting the suitability of the thiosulfonate warhead for rational modifications in a medicinal chemistry SAR like study. The previously described synthesis towards amino acid derived thiosulfonates has been expanded to encompass arginine

derivatives. Subsequent enzymatic studies suggested that the modelling studies hold true with a marginally improved potency observed for all arginine derived inhibitor constructs. This led on to the argument that the inhibitory potential of thiosulfonate inhibitors may be largely under kinetic control with the initial binding event ( $K_i$ ) being the major contributing factor, in combination with warhead reactivity. With this being said, warhead stability was observed to correlate with potency. This trend is expected to become exaggerated moving on to more complex biological systems whereby the influence of pH and other nucleophiles is anticipated to have greater impact. As such, future thiosulfonates will focus on both improving the initial binding event and downregulating reactivity which is expected to lead on to highly interesting inhibitors.

## Chapter 5

# Third Generation Inhibitors: Modulating Leaving Group Ability to Control Reactivity

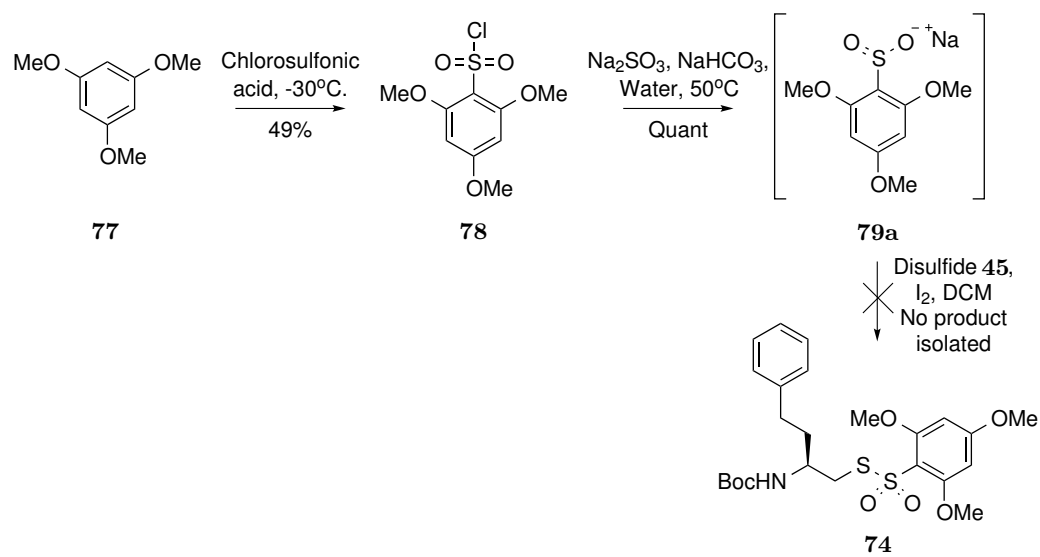


**Figure 5.1:** Target compounds for improved stability by reduced leaving group ability of the liberated sulfinate, two key strategies: (1) Increasing the electron density on the aromatic ring (compound **74**) and (2) loss of aromaticity (compound **75** and compound **76**)

From the results observed in the previous chapters it can be concluded that reduced reactivity of the thiosulfonate warhead leads improved stability with retention, or marginal improvement in potency. In this chapter, methods to reduce the reactivity of the thiosulfonate warhead by reducing the leaving group ability of the liberated sulfinate salt will be explored. It was hypothesised that modulating the leaving group ability could be achieved by tuning the electronic properties of the sulfinate liberated upon nucleophilic attack. As such, this chapter explores two key methods to influence the sulfinate leaving group ability. Firstly, generating a more electron rich aromatic ring by further substitution (as demonstrated by compound **74**, figure 5.1) will be considered. Then, expanding on this, loss of aromaticity to the aliphatic (cyclohexane) counterpart (compound **75**, figure 5.1) and introducing an aliphatic linker to an aromatic ring (**76**, figure 5.1) will be explored. Considering the wider implications of warhead reactivity, this may also offer improved pharmacodynamic properties of future thiosulfonate derived inhibitors when moving to-

wards more complex biological systems than the papain assay described earlier in section 3.6.1.

## 5.1 Improved Stability by Further Substitution: Tris-Methoxybenzene Analogues



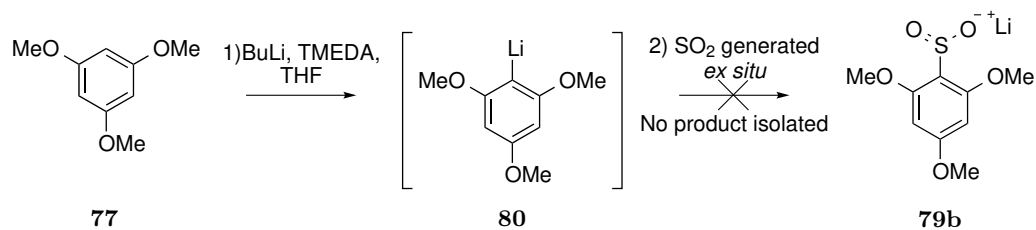
**Scheme 5.1:** 2,4,6-trimethoxybenzene sodium sulfinate synthesis by reduction of sulfonyl chloride.

The first target building block to be considered was the 2,4,6-tri-methoxybenzene derivative **74** (Figure 5.1). One noteworthy consideration of this building block is that the increased steric bulk may have a significant impact on initial inhibitor binding ( $K_i$ ). This could be either beneficial or detrimental, the outcome of which would be highly dependent on the specific protease being targeted, as it must be able to accommodate the bulkier substituent in the  $S_1'$  binding site. It was hypothesised that, from a chemical reactivity standpoint, this offers an interesting substrate which merits further investigation and future SAR considerations could follow if an improved stability was realised. Hence, the synthesis towards this building block was first explored.

The original intended synthesis towards building block **74** aimed to take advantage of the earlier developed oxidative cleavage of disulfides by sulfinate salts as described in section 3.2.3, scheme 3.9. As such, the first step was to form the desired sulfinate salt **79a** (scheme 5.1). Previously the sulfinate salts were found to be readily accessible from sulfonyl chlorides, therefore early attempts to access the sulfinate salt involved formation of the corresponding sulfonyl chloride **78** from 2,4,6-trimethoxybenzoic acid **77** with chlorosulfonic acid, which proceeded with a moderate yield of 49%. The subsequent reduction to a



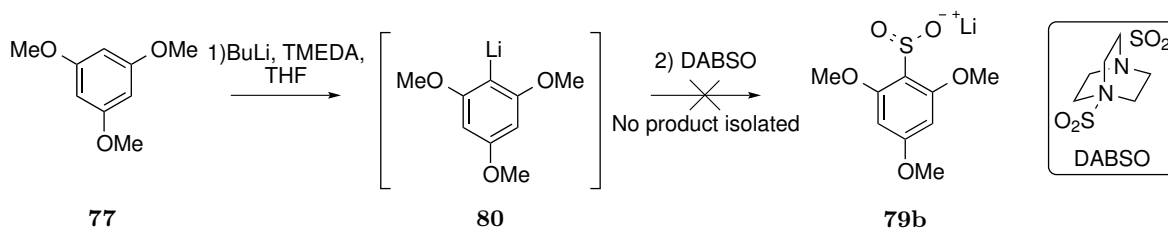
sulfinate salt appeared high yielding, however the characterisation was considered ambiguous by convenient characterisation techniques. The NMR of the product was consistent with the structure but no mass data could be obtained as the product did not ionise. It was considered that the NMR of potential by products, such as the sulfonic acid which may be produced by hydrolysis of the sulfonyl chloride, or even the starting material sulfonyl chloride were expected to be effectively identical with only slight differences in the chemical shift. As such, it was decided to subject what was believed to be the sodium sulfinate salt to the subsequent oxidative cleavage which, if successful, would confirm the production of the sulfinate salt. Unfortunately the following oxidative cleavage towards compound **74** did not yield any of the desired thiosulfonate, with the quality of sulfinate salt **79a** being considered the most likely cause for failure.



**Scheme 5.2:** 2,4,6-trimethoxybenzene sodium sulfinate synthesis by directed ortho lithation and SO<sub>2</sub>.

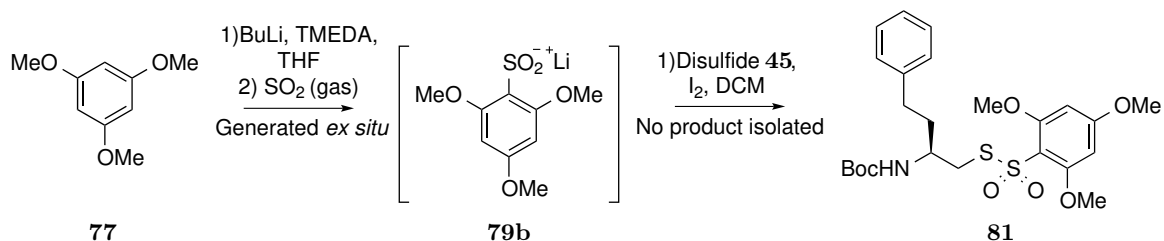
It was hypothesised that an alternative synthesis towards the sulfinate salt may circumvent the problems associated with potential side reactions and characterisation. As such, a new route was devised as outlined in scheme 5.2 towards the sulfinate salt by ortho lithation followed by nucleophilic addition of SO<sub>2</sub>. The main advantage of this route was that the formation of the sulfonic acid, which could occur by hydrolysis of the sulfonyl chloride previously, could be avoided. Furthermore the anticipated change in the characteristic NMR peaks between 2,4,6-trimethoxybenzene and the sulfinate salt was expected to be significantly easier to observe than the change from the sulfonyl chloride to the sulfinate salt (which may be so close they are difficult to differentiate). The ability to follow reaction progress by NMR would prove particularly useful if the product was again found to not ionise for mass spectrometry analysis. The first step, which involved ortho-lithation, was achieved with BuLi and *N,N,N',N'*-tetramethylethylenediamine (TMEDA). TMEDA was utilised as an additive which is widely reported to improve the efficiency of lithations with a mechanism which is likely to involve forming a coordination complex with the Li cations enhancing the reactivity of the BuLi. Next, SO<sub>2</sub> was generated *ex situ* from Na<sub>2</sub>SO<sub>3</sub> and conc. sulfuric acid, to be utilised as the electrophilic SO<sub>2</sub> source, which was bubbled through the reaction. Unfortunately only starting material **77** was isolated. Given the method of isolating the desired sulfinate salt involved extraction into water and washing with diethyl ether followed by freeze drying, which was intended to remove any excess

starting material, isolation of only the hydrophobic starting material **77** was unexpected. This observation led to the hypothesis that the sulfinate salt was potentially unstable, owing to the electron rich aromatic ring and possibility to liberate gaseous SO<sub>2</sub>. This would allow for extraction of the sulfinate salt into the aqueous phase before decomposition back to the starting material during the reduced pressure of the freeze drying process.



**Scheme 5.3:** 2,4,6-trimethoxybenzene sodium sulfinate synthesis by directed ortho lithiation and DABSO.

Another potential mechanism for the decomposition of sulfinate salt **79b** may first involve protonation of the salt to the less stable sulfinic acid.<sup>204</sup> To test the hypothesis of salt decomposition, the reaction was re-run using DABSO (scheme 5.3) as the SO<sub>2</sub> source and monitored at various stages. DABSO was chosen rather than generating SO<sub>2</sub> *ex situ* as a more convenient SO<sub>2</sub> source.<sup>205</sup> This would also prevent the potential for acid, used in *ex situ* SO<sub>2</sub> generation, from protonating the sulfinate salt yielding the less stable sulfinic acid. The reaction progress was followed by LC-MS at the crude reaction mixture stage, post extraction and finally post-freeze drying. The sulfinate salt was observed to form, be present in the aqueous layer and degrade after freeze drying. This was consistent with the hypothesis that sulfinate salt **79** was forming and decomposing upon freeze drying.



**Scheme 5.4:** Synthesis of 2,4,6-Trimethoxybenzene thiosulfonate warhead (compound **81**) by a one pot sulfinate salt formation followed by oxidative cleavage of disulfide **45**

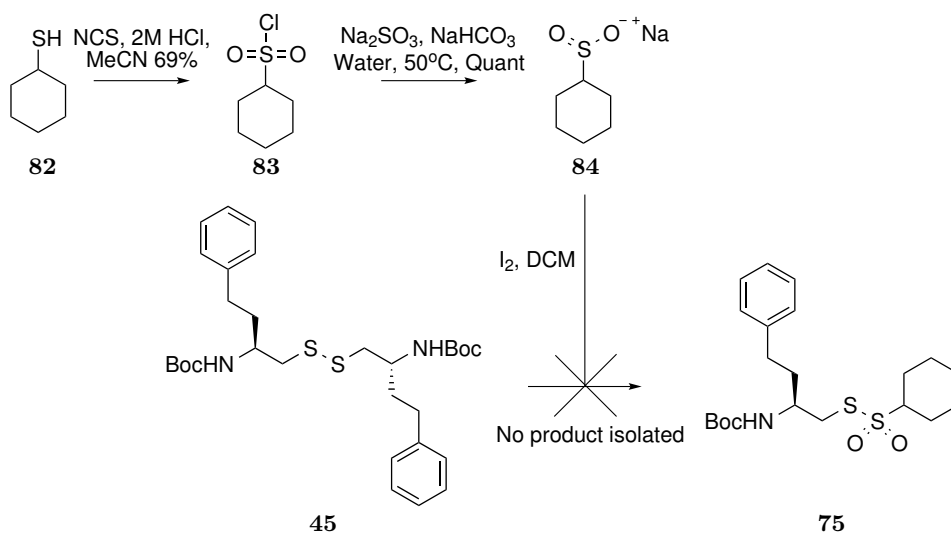
Having confirmed the apparent instability of the 2,4,6-trimethoxybenzene sulfinate **79b**, a one pot method was attempted. As shown in scheme 5.4 no product was isolated from this reaction. It was determined that the inherent instability of the intermediate sulfinate salt **79b** was unsuitable for further exploration at this stage. It was considered that, although the electron rich nature of 2,4,6-trimethoxybenzene analogue **81** was anticipated to improve stability and therefore potency, the bulky sterics may be detrimental to inhibitor binding. Combining the potential for decreased binding with the difficulty in the

synthesising compound **81**, synthesis efforts were ended here. It was decided to focus on what now appeared to be the more promising cyclohexane and benzylic analogues **75** and **76** respectively.

## 5.2 Improved Stability by Loss of Aromaticity

### 5.2.1 The Cyclohexane Warhead

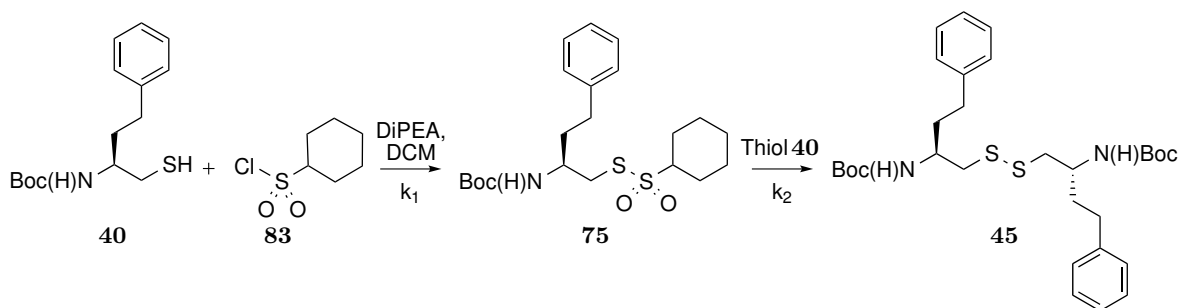
Combining the apparent instability of the 2,4,6-trimethoxybenzene sulfinate salt **79**, with the likelihood that the increased steric bulk would reduce its utility, new derivatives were desired in which the leaving group character of the sulfinate moiety was diminished. It was hypothesised that exploring the aliphatic analogue **75** (figure 5.1), bearing a cyclohexane ring rather than an aromatic ring, would reduce the leaving group ability of the liberated sulfinate salt. This, in turn, was proposed to offer a potential increase in stability, in a similar sense to the tri-methoxy analogues, with minimal increase in steric bulk at the S<sub>1</sub>' position.



**Scheme 5.5:** Synthesis of cyclohexane thiosulfonate warhead (compound **75**) by oxidative cleavage of disulfide **45**

Again, it was thought that this synthesis could build on the previous success of oxidative cleavage of disulfide **45** with sulfinate salts as shown in scheme 5.5. Starting from cyclohexane thiol **82** the sulfonyl chloride **83** was isolated in a good yield (69%) by oxidative chlorination with *N*-chlorosuccinimide / 2 M HCl, as described by Nishiguchi *et al.*<sup>206</sup> The following reduction of sulfonyl chloride **83** to sulfinate salt **84** was achieved with Na<sub>2</sub>SO<sub>3</sub> / NaHCO<sub>3</sub> and **84** was used directly in the oxidative cleavage of disulfide **45**. Unfortunately, this method did not yield any desired thiosulfonate **75**, only recovering unreacted disulfide

**45.** Interestingly, no aliphatic sulfinate salts were reported in the synthetic procedures by Fujiki *et al.* for the oxidative cleavage of disulfides.<sup>198</sup> This may well be due to intrinsic differences in reactivity which result in the aliphatic analogues being ineffective. One potential explanation would be that aliphatic sulfonyl iodides, generated as reactive intermediates, may be unstable due to competing  $\beta$ -elimination, which is not possible for the aromatic analogues. As such, an alternative method was desired for the efficient synthesis of aliphatic thiosulfonates.

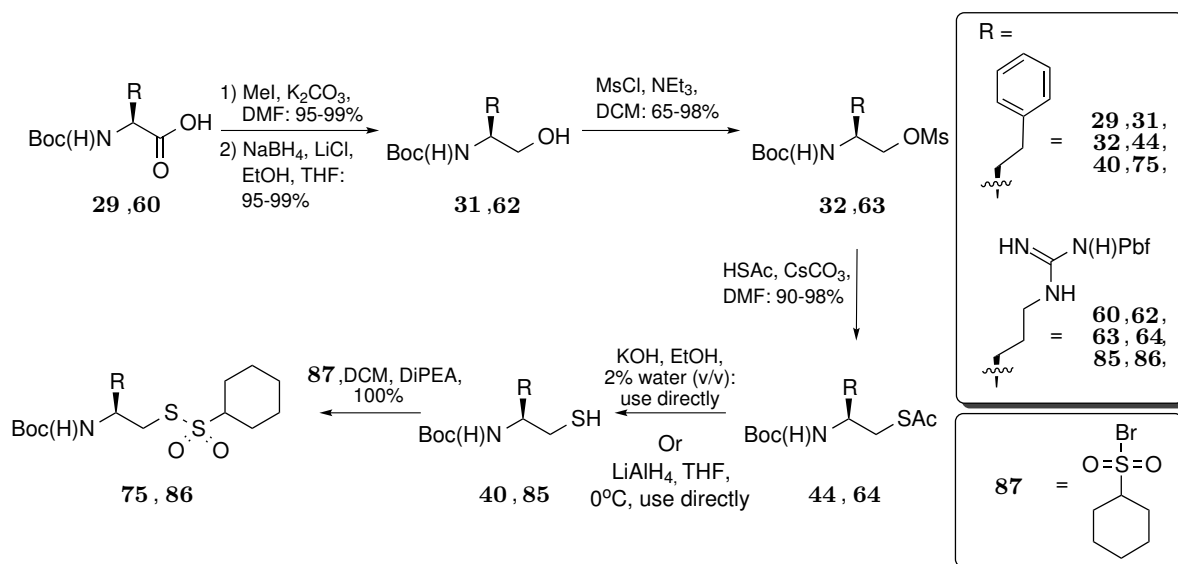


**Scheme 5.6:** Proposed synthesis of cyclohexane thiosulfonate **75** from a sulfonyl chloride **83**

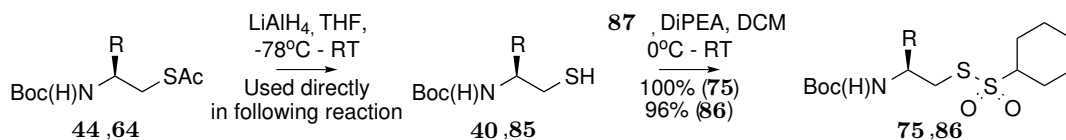
The earlier explored synthesis towards aromatic thiosulfonates avoided the use of sulfonyl chlorides as starting materials due to competing reactivity, as shown in scheme 5.6. One of the first thoughts was that moving from an aromatic to aliphatic system, the previously disregarded sulfonyl chlorides may now offer a sensible synthetic strategy. It was reasoned that the aliphatic thiosulfonate formed should have a reduced leaving group ability, enhancing its stability towards incoming nucleophile and in turn having a favourable impact on the ratio of  $k_1:k_2$  towards product formation (scheme 5.6). Hence, it was hypothesised that slow addition of a thiol to an excess of cyclohexanesulfonyl chloride **83** in the presence of DiPEA offered a plausible synthetic route, which was not previously possible for the aromatic analogues.

Having identified a potential method for the key transformation towards aliphatic thiosulfonates, the complete synthesis has been outlined in scheme 5.7. As previous work (chapter 4) had demonstrated the arginine derived warheads were highly potent, the arginine analogue was also synthesised along side the homophenylalanine derivative. Scheme 5.7 highlights the synthesis of the aliphatic thiosulfonate warheads starting from their respective amino acids. The key thioacetate building blocks **44** and **64** were isolated by the earlier reported synthesis in chapters 3 and 4 respectively.

Next, the corresponding thiols **40** and **85** were isolated by hydrolysis under an inert atmosphere. It was noted that, for the arginine derived construct, the reaction proceeded slower than that of the homophenylalanine counterpart. This was thought to be due to the increased sterics of the system, bearing a large Pbf protecting group on the arginine



**Scheme 5.7:** Synthesis of the cyclohexane thiosulfonate warhead.

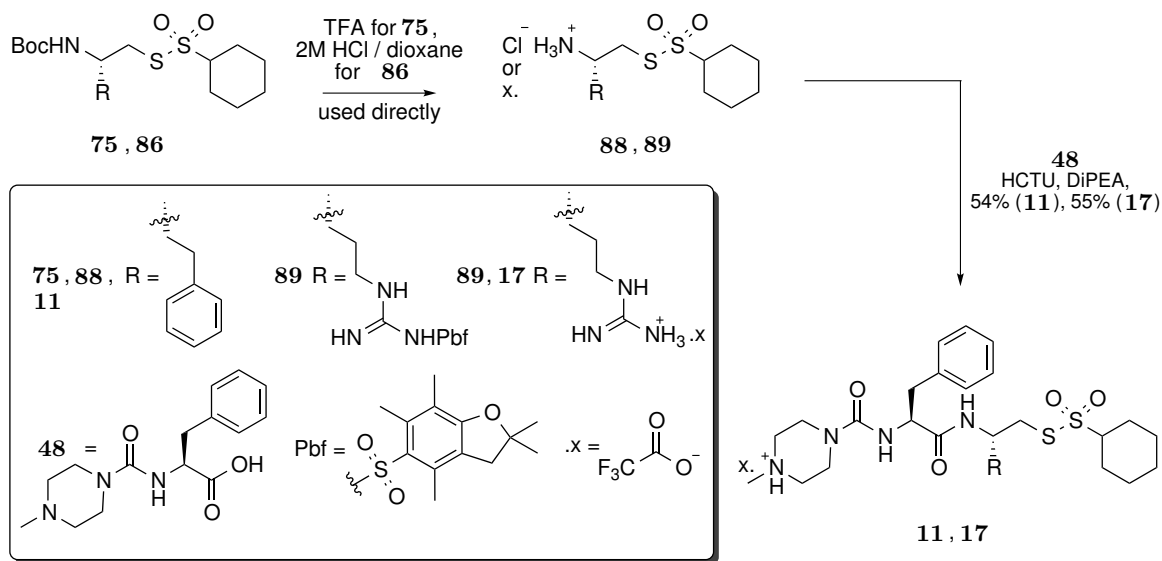


**Scheme 5.8:** Synthesis of thiols by reduction and subsequent thiosulfonate formation.

side chain. As a result, when hydrolysing thioacetate **64** to thiol **85**, an increased quantity of the undesired disulfide formed. This was considered to be inherent in the method used, as the extended reaction time exposed the desired thiol to basic conditions for longer, promoting disulfide formation. As such, an alternate method was applied (scheme 5.8) involving reduction with  $\text{LiAlH}_4$ . The reduction was found to be higher yielding and also more convenient owing to the reduced reaction time (<30 min). As such, reduction was also applied to the homophenylalanine constructs and became the method of choice moving forward.

With the desired thiols to hand these were then used directly in the subsequent thiosulfonate formation with cyclohexane sulfonyl chloride **83**. Unfortunately, when first trialed for the homophenylalanine analogue, this resulted in isolation of only the symmetrical disulfide **45** (scheme 5.6). This suggested that initial avoidance of sulfonyl chlorides for the earlier aromatic analogues was well founded. Fundamentally, the issue remained the relative rate of  $k_2$  to  $k_1$ , as shown in scheme 5.6, thus it was hypothesised that enhancing the reactivity of the sulfonyl halide would enhance  $k_1$  relative to  $k_2$  favouring product formation. As such, sulfonyl bromide **87** was isolated in good yields by adapting the oxidative chlorination method of Nishiguchi *et al.*<sup>206</sup> with brominated, rather than chlorinated reagents (scheme 5.8). Gratifyingly, the subsequent reaction with thiols **40**, **85**

(scheme 5.8) and sulfonyl bromide **87** was highly successful offering yields of 100% and 96% respectively for the homophenylalanine and arginine derivatives. This was ultimately achieved by a combination of slow addition, lowered temperature and enhanced reactivity all favouring product formation over competing product decomposition.



**Scheme 5.9:** Cyclohexane warhead coupling to backbone yielding final inhibitors **11** and **17**

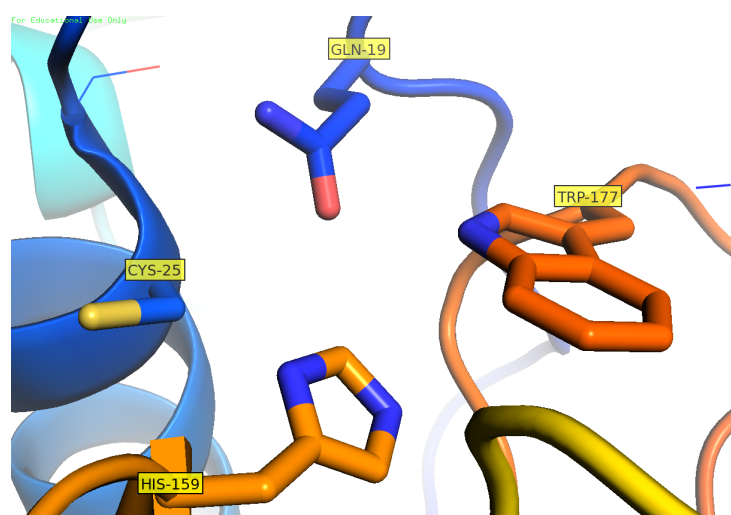
With the key cyclohexane derived warhead building blocks now to hand they were incorporated into a greater inhibitor construct as shown in scheme 5.9. The warheads were first deprotected and then directly coupled to the backbone yielding inhibitors **11** and **17** in good yields of 54% and 55% respectively. The only notable difference between the synthesis of each analogue was the conditions used for deprotecting warhead fragments **75** and **86**. In general, HCl was favoured for removal of the Boc protecting group as this generated the HCl salt. This offered the advantage over the TFA salt as it is possible, having been observed through previous experience, that the carboxylic acid of the TFA salt can couple to the amine, rather than the desired carboxylic acid of the backbone fragment during the amide coupling step. Hence, HCl was used where possible with TFA being utilised for the arginine constructs to also cleave the less labile Pbf protecting group.

### 5.2.2 Aliphatic Linker to Aromatic Ring: The Best of Both Worlds?

In the previous section, reducing the leaving group ability by loss of aromaticity was discussed. In theory, this is a valid method to improve stability of the thiosulfonate warheads. However, such structural changes not only alter the chemical reactivity of the warhead constructs but also their target engagement and binding characteristics. As such, it was considered that the loss of aromaticity may also be met with an altered binding affinity ( $K_i$ ), which is highly dependent on the target protease. In the case of papain

like cysteine proteases (Clan CA, Family C1) there is a conserved aromatic residue in the  $S_1'$  region (Trp<sup>177</sup>, papain numbering). As the thiosulfonates are being developed to selectively target cysteine proteases, which display this conserved residue, modifications which have the potential to target this may increase inhibitor binding and the resultant potency. As such, this section aims to explore introducing an aliphatic linker, to alter the reactivity of the thiosulfonates to be closer to the aliphatic analogues, joining on to an aromatic ring. It was hypothesised that this may enable targeting of a potential pi stacking interaction, which would be specific to cysteine proteases of this class. This aims to build on the previously observed stability trends whilst accommodating rational SAR for enhanced target engagement. Ultimately, it is envisioned that combining rational SAR alongside tuning of reactivity would reinforce the suitability of this novel warhead class in specific engagement of cysteine proteases.

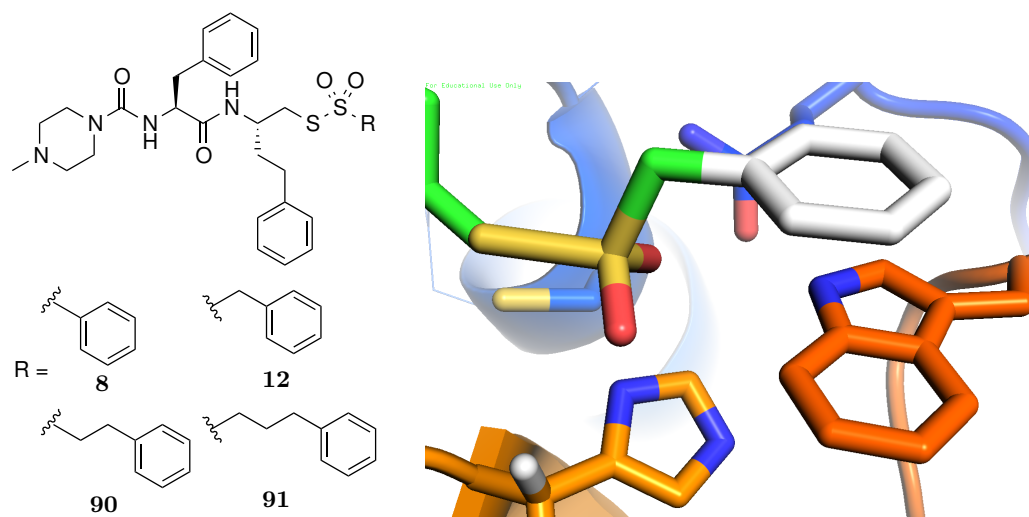
### Modelling Study: Targeting Conserved Active Site Residues in the $S_1'$ Pocket



**Figure 5.2:** Active site residues of papain (PDB code 1CVZ) highlighting the conserved tryptophan residue (Trp-177, papain numbering) of papain like cysteine proteases.

Upon examination of the active site of papain like cysteine proteases, it was noted that the primed side of the active site contains a conserved tryptophan residue (figure 5.2). It was hypothesised that this may offer a good pi-stacking interaction with potential inhibitor constructs which may in turn further increase the specificity of the thiosulfonate warheads towards cysteine proteases. With the previous hypothesis that an aliphatic substituent at the primed side of the warhead may offer a favourable reactivity profile, it was proposed that an aliphatic linker to an aromatic ring may offer an ideal combination of chemical reactivity and pi-stacking ability.

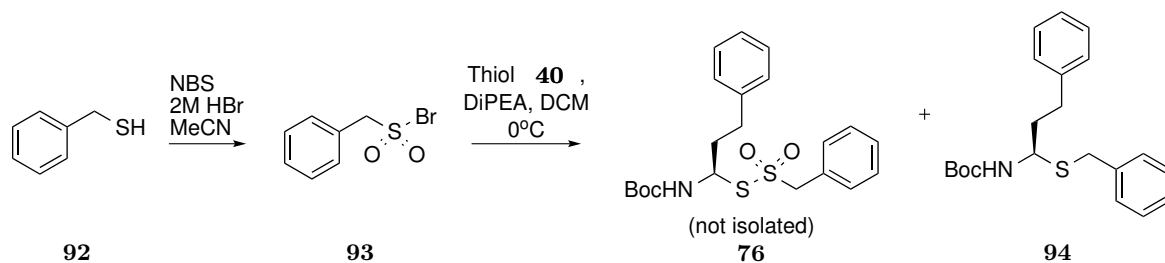
In order to direct the chemistry to undertake, *DrawToDock* was used to screen for suitable warhead constructs. As shown in figure 5.3, docking warheads with additional methylene



**Figure 5.3:** Varying the number of methylene units between the thiosulfonate warhead and aromatic ring in the  $S_1'$  region, screened compounds (left) with docked model of compound **12** shown on right, highlighting pi stacking ability.

units between the hexavalent sulfur and aromatic moiety (**8**, **12**, **90**, **91**) suggested that a single methylene unit (compound **12**) would be optimal. This was based on the fact that compound **12** allowed a suitable proximity of the aromatic ring to Trp-177 to facilitate pi-stacking whilst conserving the key hydrogen bonding network surrounding the active site thiolate and inhibitor sulfone oxygen atoms, as discussed earlier in section 3.1.

### Synthesis of Methylene Spaced Aromatic Warhead

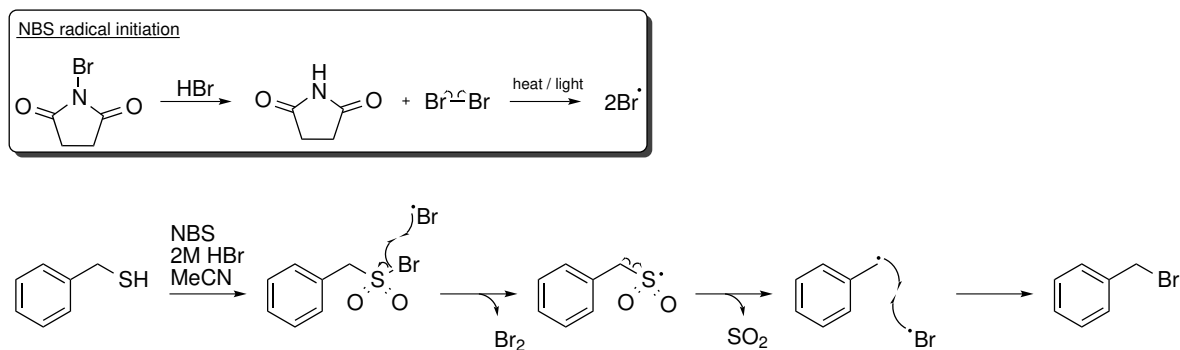


**Scheme 5.10:** Original intended route towards benzyl thiosulfonate warhead building block **76**, from sulfonyl bromide **93**

As the reactivity of the benzylic thiosulfonate warhead **76** was anticipated to be closer to the aliphatic than aromatic analogues, the initial synthesis was based on the previous success towards aliphatic thiosulfonates. It was hypothesised that warhead synthesis from benzyl sulfonyl bromide (**93**) would be more suitable than oxidative cleavage of disulfides with a sulfinate salt, which was found to be most applicable to aromatic analogues (see chapter 3). As such, scheme 5.10 outlines the intended route towards benzylic thiosulfonate

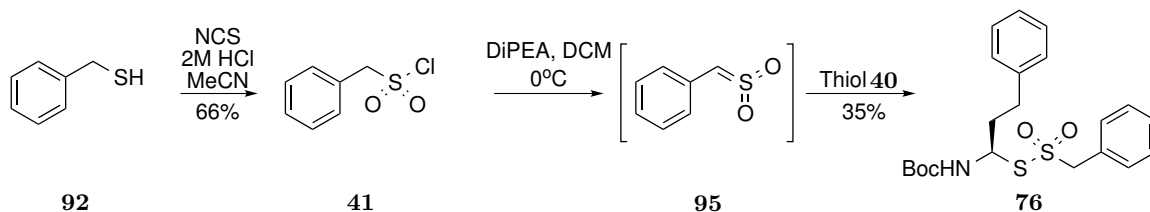


building block **76**. Starting from benzyl mercaptan **92**, synthesis towards the corresponding sulfonyl bromide was attempted by oxidative bromination with NBS / 2 M HBr based on the previous success towards cyclohexane derivatives (see section 5.2.1). The freshly prepared sulfonyl bromide was then used directly in the subsequent reaction with thiol **40**, which was isolated by reduction of the corresponding thioacetate discussed previously in scheme 5.8. Direct use of sulfonyl bromide was chosen as it was considered likely to be a unstable synthetic intermediate. Unfortunately, the reaction did not yield the desired thiosulfonate but rather thioether **94** was identified as the major product.



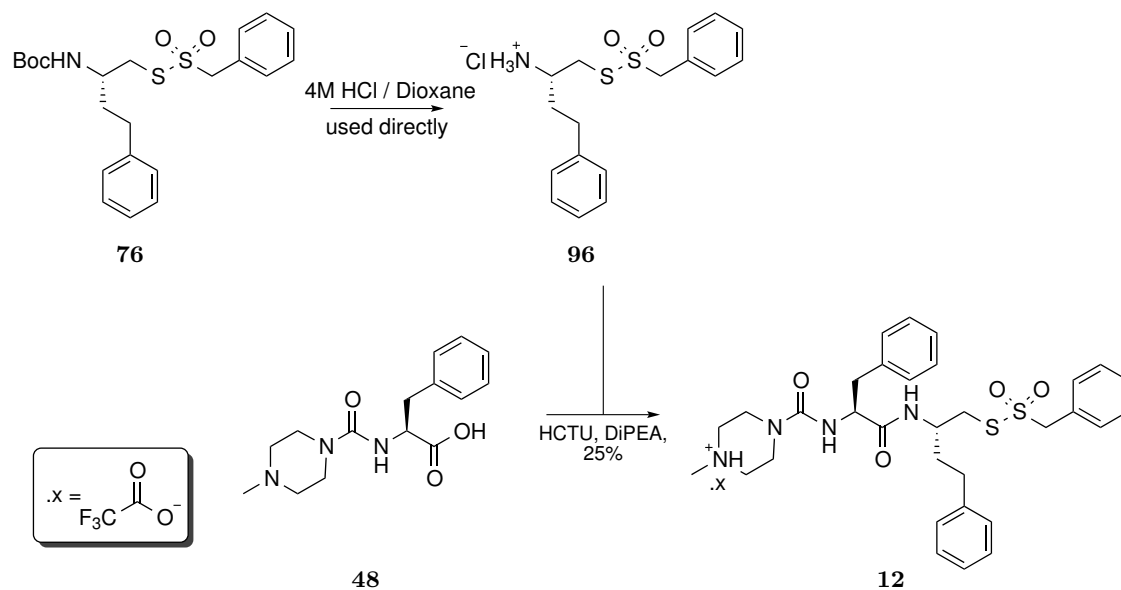
**Scheme 5.11:** Proposed radical decomposition pathway of benzylsulfonyl bromide.

Suspecting that the previously observed thioether **94** arose from the quality of the sulfonyl bromide, the sulfonyl bromide was isolated and characterised. The characterisation was consistent with benzyl bromide, rather than sulfonyl bromide **93**. This was also consistent with the isolation of thioether **94**, which would form upon alkylation of thiol **40** with benzyl bromide. Benzyl bromide formation could occur via desulfuration of the sulfonyl bromide **93**, however King *et al.*<sup>207</sup> previously reported this desulfuration required heating at 80 °C for 5 days, resulting in only 50% conversion. Given the short reaction time (<10 min.), reagents used and no observable sulfonyl bromide it has been proposed that decomposition to benzyl bromide was likely accelerated through a radical mechanism as shown in scheme 5.11. NBS is a known radical initiator, hence it has been incorporated into the potential mechanism shown in scheme 5.11, with stabilisation of the radical product (benzylic) as well as liberation of SO<sub>2</sub> being recognised as potential driving forces.



**Scheme 5.12:** Benzyl thiosulfonate warhead synthesis from benzylsulfonyl chloride *via* sulfene intermediate

Moving forward it was decided that the differing reactivity profile of the benzylic analogues, which facilitated the proposed radical mechanism, could be utilised rather than problematic towards the synthesis of thiosulfonates. As shown in scheme 5.12, it was proposed that the benzylic position, having more acidic protons, would favour sulfene formation under basic conditions which may be accessible through the more stable sulfonyl chlorides. The sulfene formed, being a highly reactive intermediate, allowed for the required increase in reaction rate to out compete the side reaction involving attack of thiol **40** on the warhead fragment (shown previously in scheme 5.6). Gratifyingly this yielded the benzylic thiosulfonate warhead derivative **76** with a moderate yield of 35%. The reduced yield was due to the competing symmetrical disulfide formation, occurring by attack of thiol **40** on the generated thiosulfonate warhead **76**, however the quantity isolated was sufficient to carry forward for further testing.



**Scheme 5.13:** Final inhibitor coupling: Coupling the benzyl derived thiosulfonate warhead building block to backbone.

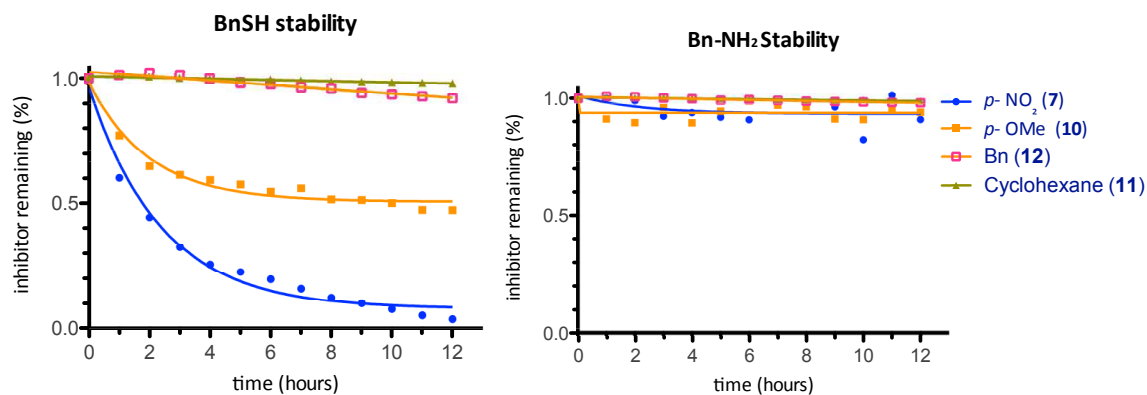
With the benzylic thiosulfonate warhead building block (compound **76**) now to hand it was incorporated into the overall inhibitor construct. As shown in scheme 5.13 the warhead fragment was first deprotected with HCl followed by direct coupling to the backbone **48** yielding final inhibitor construct **12** in 25% over two steps, which was isolated as a TFA salt after purification by preparative HPLC before biological testing.

### 5.2.3 Stability Testing Towards Common Nucleophiles

It was hypothesised that the inhibitors bearing the cyclohexane warhead (**11**) and the benzylic warhead (**12**) would show increased stability towards incoming nucleophiles when

compared with the previous aromatic warheads. To test this, each inhibitor construct was assessed for its stability towards amines, thiols and pH dependent hydrolysis. This followed the same testing undertaken for the earlier warhead constructs in section 3.5, allowing for a direct comparison.

### Stability Towards Thiols and Amines

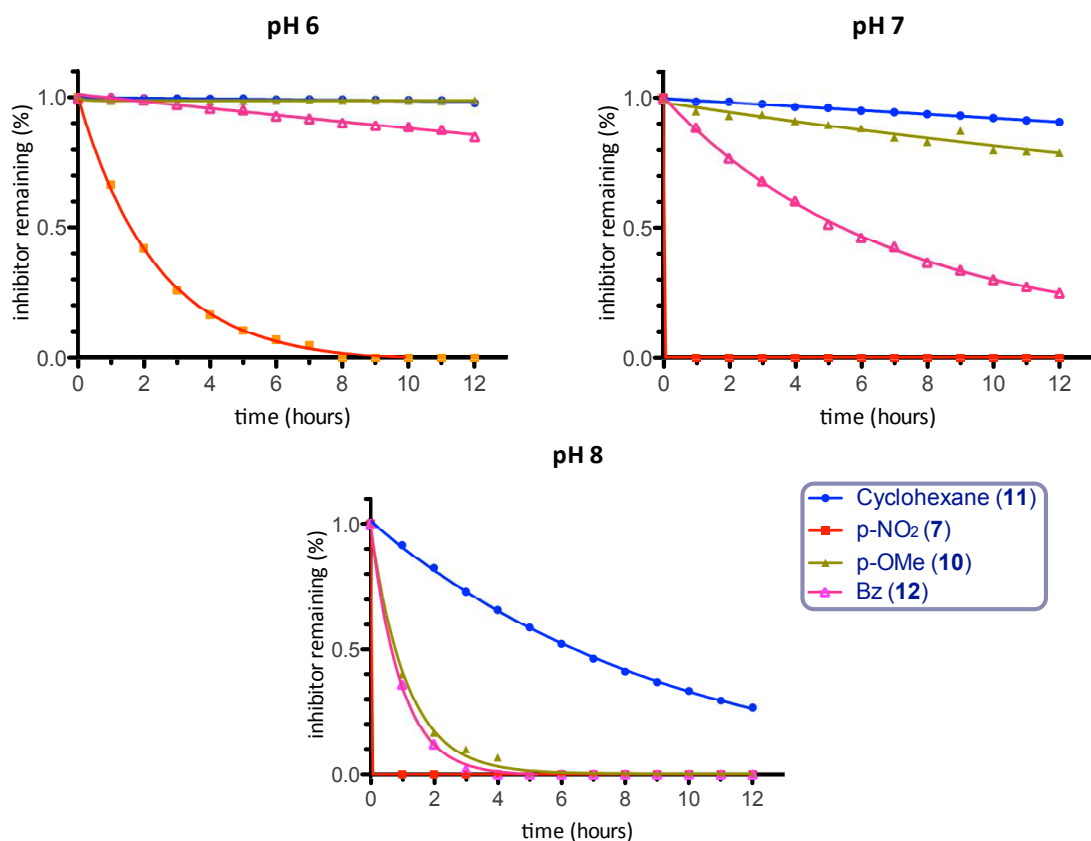


**Figure 5.4:** Cyclohexane warhead stability towards benzyl thiol (left) and benzyl amine (right). Inhibitor degradation was determined by HPLC in MeCN at room temperature with 13 measurements taken 1 hour apart, representing time 0 to 12 hours. Approximately 0.3 mg of inhibitor was accurately weighed and suspended in MeCN, to which 10 equivalents of BnSH or BnNH<sub>2</sub> was added immediately before measurement. Results were then normalised against time 0 and plotted as scatter graphs for visualisation.

Gratifyingly, both new inhibitor designs were significantly more stable than the previously synthesised aromatic analogues. In each case, stability towards amines (represented by benzylamine in this study, figure 5.4) was maintained, reinforcing the selective nature of the thiosulfonates towards sulfur centred nucleophiles. A remarkable improvement in stability towards thiols (benzyl mercaptan in this study, figure 5.4) was observed for both of the new inhibitor constructs when compared with the previous aromatic analogues. Cyclohexane derivative **11** displayed no significant thiol dependent degradation, whilst the benzylic derivative **12** displayed <10% degradation over the 12 hour time period. This was a significant improvement from the previously prepared aromatic warheads which displayed close to complete degradation (for the *p*-NO<sub>2</sub> substituted compound **7**) and 50% degradation (for the *p*-OMe substituted compound **10**) over 12 hours. The observed increase in stability towards thiols was particularly encouraging as thiols represent an abundant, biologically relevant nucleophile with the potential to halt the progress of thiosulfonates as cysteine protease inhibitors. This owes to the fact that, thiosulfonates are inherently more selective towards sulfur centred nucleophiles. Indeed, this is one of the core concepts

for their development into selective cysteine protease inhibitors. This does however carry the risk of cross reactivity with circulating thiols, thus observing a significant increase in stability towards thiols was considered to be a considerable advance in their development as cysteine protease inhibitors. Of course, the nucleophilicity of thiols will correlate with the local pH, which has not been accounted for in this study, and would be a significant contributing factor towards the stability of thiosulfonates in a biological system. This does however demonstrate a move in the right direction for further exploration, an encouraging finding in the development of this novel warhead class.

### pH Dependent Aqueous Stability



**Figure 5.5:** pH dependent aqueous stability of aliphatic thiosulfonate warheads compared to the previous aromatic thiosulfonate warheads. Inhibitor degradation was determined by HPLC with peak areas integrated against an internal standard (Ac-Phe-OH), 13 measurements were taken 1 hour apart, representing time 0 to 12 hours. Sodium phosphate buffer (0.1 mM) of the relevant pH containing 5% DMSO was utilised, initial inhibitor concentration was 245  $\mu\text{M}$  and internal standard (Ac-Phe-OH) concentration was 338  $\mu\text{M}$ . Results were then normalised and plotted as scatter graphs for visualisation.

Moving on to the aqueous stability profiles, shown in figure 5.5, the hydrolytic stability

correlated well with the predicted leaving group ability in each case. The leaving group ability of the liberated sulfinate was anticipated to be higher for the benzylic derivative **12** than that of the cyclohexane derivative **11**. Indeed, the benzylic derivative was found to be more prone to hydrolysis at all pH's tested when compared with the cyclohexane analogue, falling between the stability of the previously most stable *p*-OMe substituted compound **10** and the improved stability of cyclohexane analogue **11**. Previously, the most stable aromatic analogue (**10**) was effectively stable at pH 6 and had a half life of approximately 30 hours at pH 7, and 45 minutes at pH 8. This has been significantly improved upon with cyclohexane analogue **11** being stable at pH 6 and displaying a half life of approximately 60 hours at pH 7 and 6 hours at pH 8. This was a very encouraging finding as hydrolytic stability across a reasonable physiological pH range, such as that represented here, is a significant consideration in the development of a new warhead class.

Importantly, the above stability studies also reinforce the relative ease with which the chemical reactivity of the thiosulfonate warheads can be controlled. Moving forward, this may prove to be an important characteristic for controlling the reactivity of the thiosulfonate warheads.

#### 5.2.4 Biological Testing: Papain Enzyme Assay

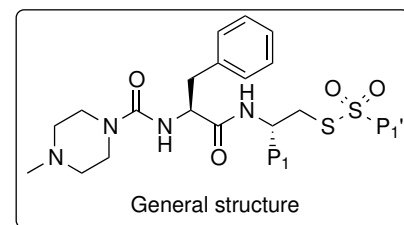
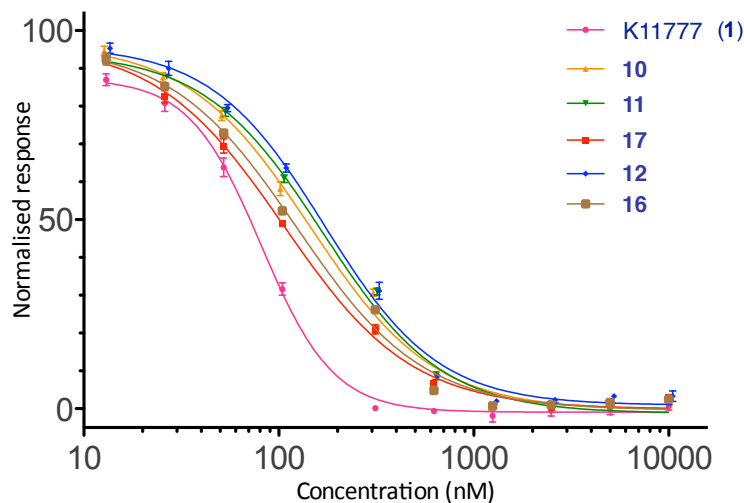
Having established the aliphatic thiosulfonate construct **11** offered a significant increase in stability over their aromatic counterparts, further biological evaluation was required. It was hypothesised that, based on the previously observed correlation between reactivity and potency of the aromatic warheads, this new construct would offer an enhanced potency. Furthermore it was hypothesised that compound **12** (bearing the benzylic warhead), which displayed a reactivity profile close to that of the most stable aromatic construct **10** (bearing the *p*-OMe substituted warhead), would be equipotent on the basis of chemical reactivity. However, it was hoped that the methylene linker to an aromatic ring may offer enhanced target binding ( $K_i$ ) by pi stacking to the conserved tryptophan residue as discussed in section 5.2.2 resulting in an improvement in potency. To test this, the aliphatic inhibitor constructs **11**, **17** and **12** were subjected to the papain assay developed earlier in section 3.6.1.

Gratifyingly, these new aliphatic warhead constructs performed as well as or better than the previous most promising compound (**10**). As shown in figure 5.6, cyclohexane derived warhead **11** displayed an  $IC_{50}$  of 168 nM, effectively equipotent to the previous best inhibitor (*p*-OMe benzene derived warhead of inhibitor **10** at 148 nM). Combining the marked increase in stability towards thiols and pH dependent hydrolysis with retained potency, the cyclohexane warhead bearing inhibitor **11** offers a significantly improved

inhibitor construct. Furthermore, when the sidechain of the P<sub>1</sub> position was altered to arginine, rather than homophenylalanine, a further reduction in the IC<sub>50</sub> value to 106 nM was observed. This enhanced potency was consistent with the previously observed trend in chapter 4 when installing an arginine rather than homophenylalanine side chain. Arginine derived inhibitor **17**, bearing a cyclohexane warhead, was also slightly more potent than the *p*-OMe aromatic warhead counterpart (inhibitor **16** which had an IC<sub>50</sub> of 124 nM), again suggesting this warhead offers an overall improvement in inhibitor design.

Interestingly, inhibitor **12**, bearing a benzylic group in the P<sub>1</sub>' position, was found to be equipotent with its cyclohexane counterpart (inhibitor **11**) at 164 nM. This is contrary to the observed correlation between stability and potency observed for all other inhibitor constructs, which would suggest the benzylic derivative (**12**) should be less potent rather than equipotent with the cyclohexane derivative (**11**). This may suggest that the anticipated reduction in potency (based on stability arguments) is being offset by an improvement in potency due to improved target binding (K<sub>i</sub>), potentially through pi stacking of the warhead aromatic ring. The above findings suggest that an additive effect is being observed between both backbone and warhead design. This is a significant observation, highlighting the potential benefits which can be derived from new electrophilic traps rather than relying on the ubiquitous 1,4-Michael acceptors (as discussed in section 1.2).

Papain IC<sub>50</sub> Determination and Comparison with best aromatic inhibitor



Compound	P <sub>1</sub>	P <sub>1</sub> '	IC <sub>50</sub> (nM)
11			168 ± 4
17			106 ± 2
12			164 ± 7
10			148 ± 6
16			124 ± 1
6 (K11777)			81 ± 2

**Figure 5.6:** IC<sub>50</sub> determination of aliphatic warheads compared with the most potent aromatic warhead. K11777 **6** reference compound (magenta curve), Homophenylalanine derived (p-)OMe aromatic warhead **10** (orange curve), Arginine derived (p-)OMe aromatic warhead **16** (brown curve) Homophenylalanine derived cyclohexane warhead **11** (green curve), Arginine derived cyclohexane warhead **17** (red curve), Homophenylalanine derived benzylic warhead **12** (blue curve). Assay buffer: sodium phosphate (100 mM, pH 6.5) containing EDTA (1.5 mM) and 2% DMSO. Final concentrations in the wells were; Papain: 4 μM; substrate: 1.0 mM; inhibitor: doubling dilution range starting at 10 μM giving: 10 μM, 5 μM, 2.5 μM, 1.25 μM, 0.625 μM, 0.3125 μM, 156.25 nM, 78.125 nM, 39.0625 nM. Error bars on graph display SEM across 3 independent repeats. Tabulated ± values represent the std error of the IC<sub>50</sub> values across 3 independent repeats.

## 5.3 Summary

In summary, electron rich aromatic thiosulfonates, synthesised by further substitution the aromatic ring, were not isolated for further development. The combination of their difficult synthesis alongside the likelihood that the additional steric bulk would be detrimental to inhibitor design led to new methods of controlling reactivity, by modulating leaving group ability, being explored. This lead on to the successful synthesis and testing of a new class of aliphatic thiosulfonate (**11**, **12** and **17**) bearing either a cyclohexane or benzylic warhead. Early attempts found that the differing reactivity of the aliphatic thiosulfonates required an alternate synthesis to be developed, culminating in their synthesis from sulfonyl halides, rather than the earlier described synthesis by oxidative cleavage of disulfides. Gratifyingly, the aliphatic thiosulfonates were found to maintain, or improve upon, the potency of the best in class aromatic thiosulfonate **10** against the cysteine protease papain. Furthermore, the aliphatic thiosulfonate **11** offered a considerable increase in stability towards thiols and also towards pH dependent hydrolysis. Moving forward, it was clear that improving the stability of the thiosulfonates was favourable for their development as cysteine protease inhibitors. With this in mind, new strategies to further improve the reactivity profile of the thiosulfonate warhead will be explored.

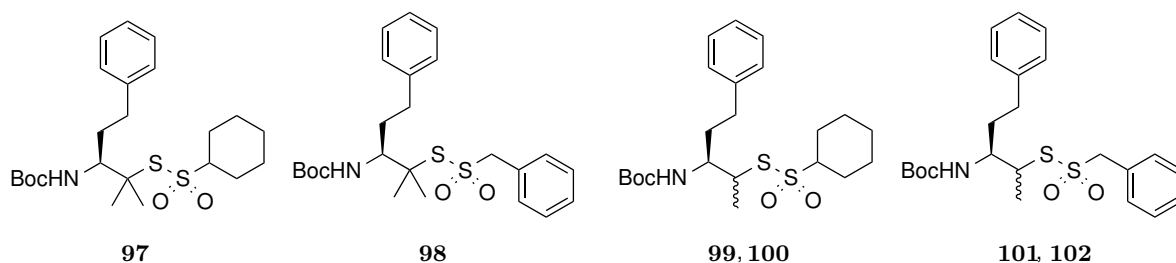


# Chapter 6

## Fourth Generation inhibitors: A Steric Approach for Increased Warhead Stability

### 6.1 Reasoning and Target Warheads

Previously, it has been demonstrated that downregulating the reactivity of the thiosulfonate warheads, based on reducing the leaving group ability, has led to an improvement in inhibitor design. Expanding on this trend, a new approach will be taken to regulate the reactivity of the thiosulfonates based on reducing the rate of nucleophilic attack through steric hindrance. It was hypothesised that the optimal reactivity profile of the thiosulfonates would be such that they could react with active site thiolates with minimal degradation towards other biologically relevant nucleophiles, such as water and thiols. This was believed to be possible due to the enhanced reactivity of the local enzymatic environment (such as the local hydrogen bonding network demonstrated in section 3.1). Taken in combination with the highly tunable nature of the thiosulfonates, which has been observed thus far, it was believed this trend could be further improved through steric arguments discussed herein.

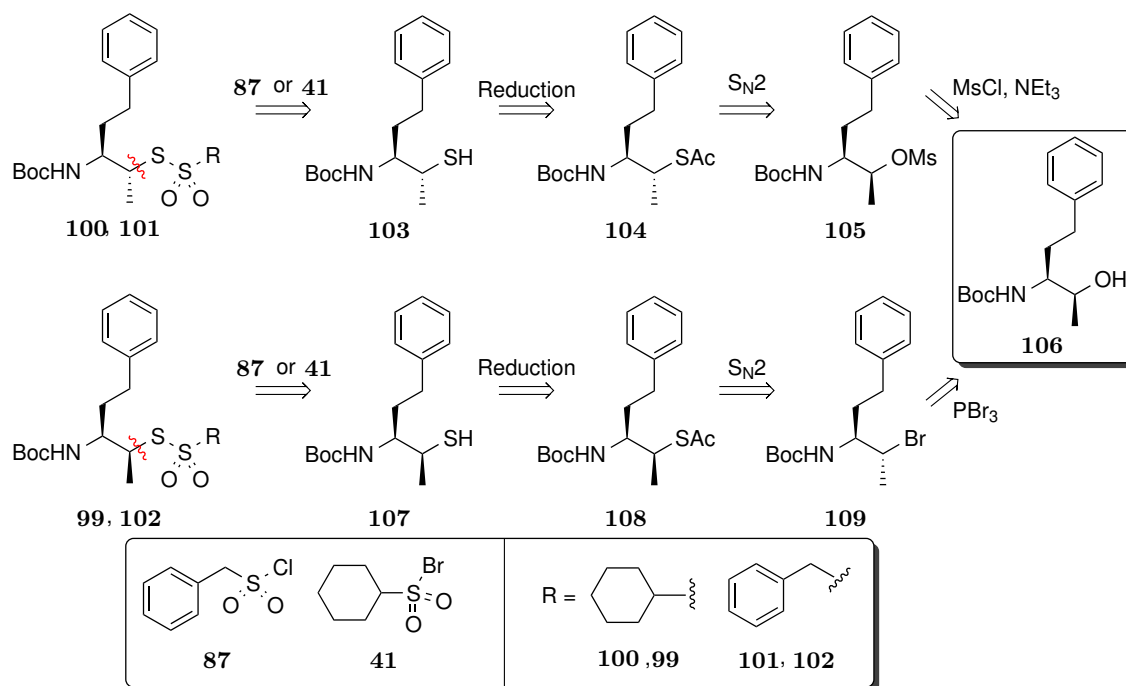


**Figure 6.1:** key building blocks to introduce steric arguments for decreased reactivity

As such, this section will explore the synthesis and evaluation of two key modifications, shown in figure 6.1, by introducing either one methyl or two methyl substituents to the  $\beta$ -amino position of the thiosulfonate warhead building blocks. These aimed to build on the previous success of the aliphatic warheads, thus functionalising both the cyclohexane and benzylic warheads to their mono-methyl and di-methyl counterparts was proposed, which shall be referred to as secondary and tertiary thiosulfonates respectively. It was hypothesised that the order of reactivity would be in line with that of all  $S_N2$  reactions, hence the tertiary thiosulfonates **97** and **98** were expected to display the slowest reaction rate, corresponding to the most stable warhead construct. Following on from this the secondary thiosulfonates **99-102** were anticipated to display the next slowest reaction rate and corresponding stability, with the previously generated primary thiosulfonates thought to be least stable in this series. Based on the previously observed trends between reactivity, stability and inhibitor potency, it was hypothesised the order of potency from most potent to least potent would be tertiary, secondary and then primary thiosulfonates.

## 6.2 Exploring a Route Towards $\beta$ -Methyl analogues

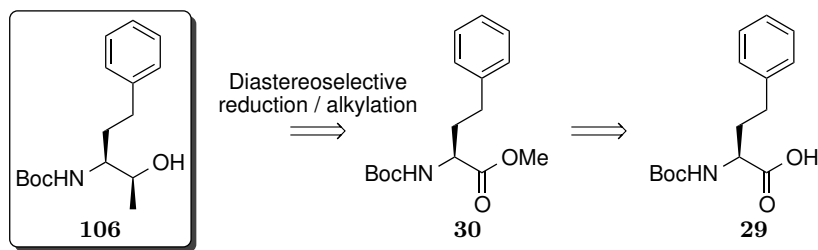
### 6.2.1 Retrosynthetic Analysis - Two Key Disconnections



**Scheme 6.1:** Retrosynthetic analysis: Towards the key secondary alcohol intermediate **106**

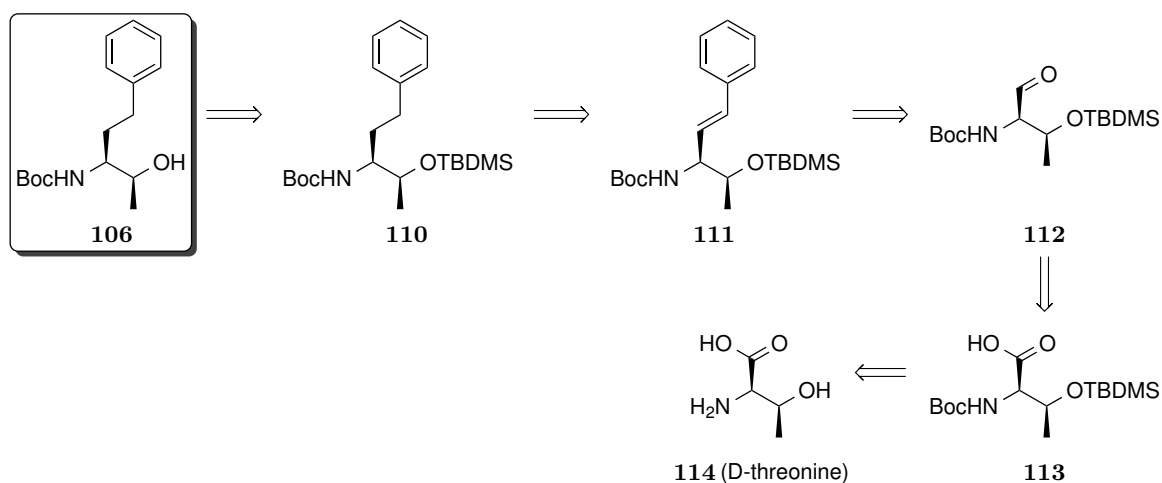
Two key strategies were identified towards the secondary thiosulfonates **99-102** which both went via the key synthetic intermediate **106** shown in scheme 6.1. These synthetic steps

aimed to build on the previous success towards aliphatic thiosulfonates with warheads generated from the corresponding sulfonyl halide and thiol. The thiol, in turn, could be accessed by the reduction of the corresponding thioacetate which was thought to be accessible by substitution of a suitable leaving group. Interestingly, altering the leaving group used offered a divergent synthesis to each diastereomer, which could be accessed by functionalising the corresponding alcohol (**106**) with retention of the stereocentre when generating the mesylate **105**, or inversion of the stereocentre when generating the bromide **109** with  $\text{PBr}_3$ .



**Scheme 6.2:** Retrosynthetic analysis: Towards secondary thiosulfonate warheads by reduction / diastereoselective addition

In retrosynthetic scheme 6.2 it has been proposed that alcohol **106** could be accessed by a diastereoselective nucleophilic attack and reduction of methyl ester **30**. The key advantage to this method was that, in principle, it could be conducted from any amino acid allowing easy access to various side chains for future SAR studies. Furthermore, this method would allow for broader substrate scope by selection of different alkylating reagents. In principle, this could prove interesting when invoking steric hindrance to alter reactivity as bulkier substituents could be introduced with ease.



**Scheme 6.3:** Retrosynthetic analysis: Towards secondary thiosulfonate warheads by sidechain functionalisation

Moving on to retrosynthesis shown in scheme 6.3, this also started from the common sec-

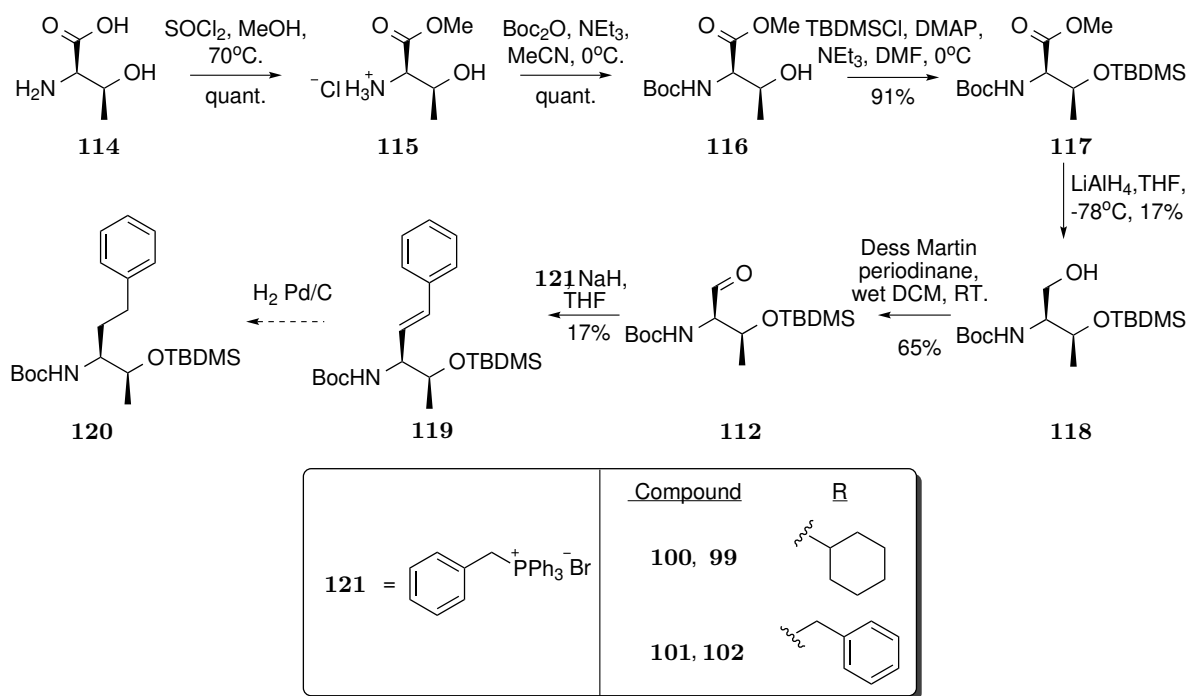
ondary alcohol intermediate **106**. It was hypothesised that a sidechain functionalisation approach could be employed allowing the methyl group of interest to be incorporated from the chiral pool. Retrosynthetically this would involve a final deprotection step of the protected alcohol **110**, with TBDMS proposed as a suitable protecting group. The sidechain functionalisation was proposed to be possible by reduction on the alkene **111** which could be accessed by either a Horner Wadsworth Emmons or Wittig olefination as a convenient carbon-carbon bond forming reaction. The corresponding aldehyde **112**, required for the olefination, could be accessed by reduction of carboxylic acid **113**. This carboxylic acid represents the commercially available, and relatively cheap, D-threonine starting material (**114**). Although this approach involved more synthetic steps, it offers the advantage of drawing from the chiral pool, potentially avoiding the need for diastereoselective transformations and / or separations. It was also thought that as the side chain is functionalised from a carboxylic acid, an excellent handle for further functional group interconversions, multiple non natural amino acids could be accessed through this route. As each method had apparent strengths for accessing new thiosulfonate warhead building blocks, both will be explored.

## 6.2.2 Drawing From the Chiral Pool: A Sidechain Functionalisation Approach

The experimental work presented in this section (section 6.2.2) was conducted by two project students, started by Alin Pirvan and continued by Damon Allan as part of their work towards their masters projects. All work was conducted under the supervision of the author.

One of the first strategies to be explored towards the synthesis of secondary thiosulfonate warheads involved drawing from the chiral pool, as shown in scheme 6.4. Starting from D-threonine (**114**) the carboxylic acid was converted to methyl ester **115** in quantitative yields. Next, the free N terminus was protected with a Boc group followed by masking the secondary alcohol with a TBDMS group, with both steps proceeding in excellent yields (quantitative and 91% respectively). Following on from this, the methyl ester **117** was reduced to alcohol **118** with  $\text{LiAlH}_4$  in a low yield of 17%.

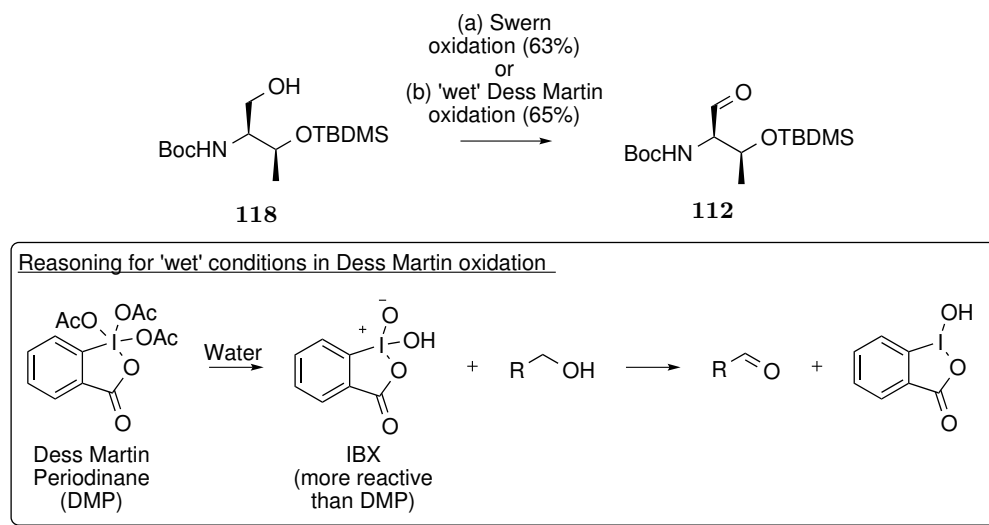
The subsequent oxidation to aldehyde **112** was originally trialled with a Swern oxidation (6.5), which was met with a good yield of 63%. In an attempt to improve upon this yield the Dess Martin oxidation was employed, using wet conditions to generate the more reactive IBX in situ (scheme 6.5) as has been reported by Meyer *et al.*, resulting in an equally good yield of 65%.<sup>208</sup> As the yields were similar, the Dess Martin was selected as the main route moving forward on the basis of being a more convenient reaction, with less



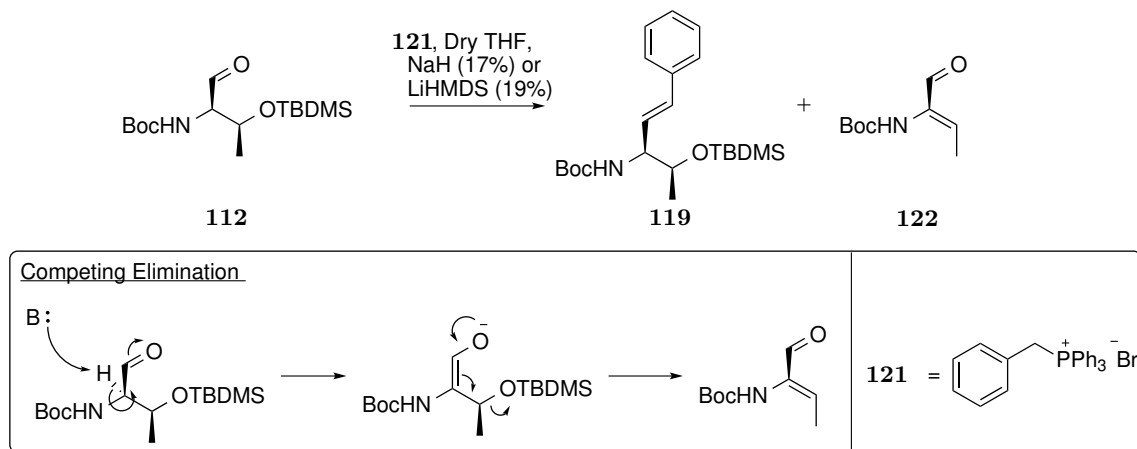
**Scheme 6.4:** Synthesis towards secondary thiosulfonates from D-threonine, dashed arrows indicate proposed synthetic steps which were not conducted.

noxious byproducts that the DMS, CO and CO<sub>2</sub> produced by the Swern oxidation.

The following Wittig reaction was anticipated to be the most synthetically challenging step of this synthesis, owing to the predicted instability of the  $\alpha$ -aldehyde intermediate **112**. Upon first attempts (scheme 6.6) the phosphonium ylide was generated from the phosphonium salt with NaH before addition of the aldehyde, resulting in a yield of only 17% of the desired product **119**, with the elimination product **122** being identified as shown in scheme 6.6. This was proposed to form through deprotonation of the highly acidic alpha proton, leading to an anion, which was resonance stabilised by the neighbouring aldehyde, before subsequent elimination as shown in scheme 6.6. As such, a weaker base, LiHMDS, was employed to form the ylide which was met with a marginal increase to 19% yield. When combined with the low yielding reduction of methyl ester **117** to alcohol **118** it was decided this synthesis was inefficient and would not be taken further. The choice to end the synthesis here was largely based on recognising the alternate synthetic strategy, highlighted by retrosynthesis in scheme 6.2. The alternate synthetic strategy offered increased simplicity, and circumvented the unstable aldehyde intermediate.



**Scheme 6.5:** Methods of oxidation to aldehyde **112**. (a) Swern oxidation: Oxalyl chloride, DMSO,  $\text{NEt}_3$ , dry DCM,  $-78^\circ\text{C}$ -RT. (b) Dess Martin oxidation: Dess Martin Periodinane, wet DCM (washed with water before use).

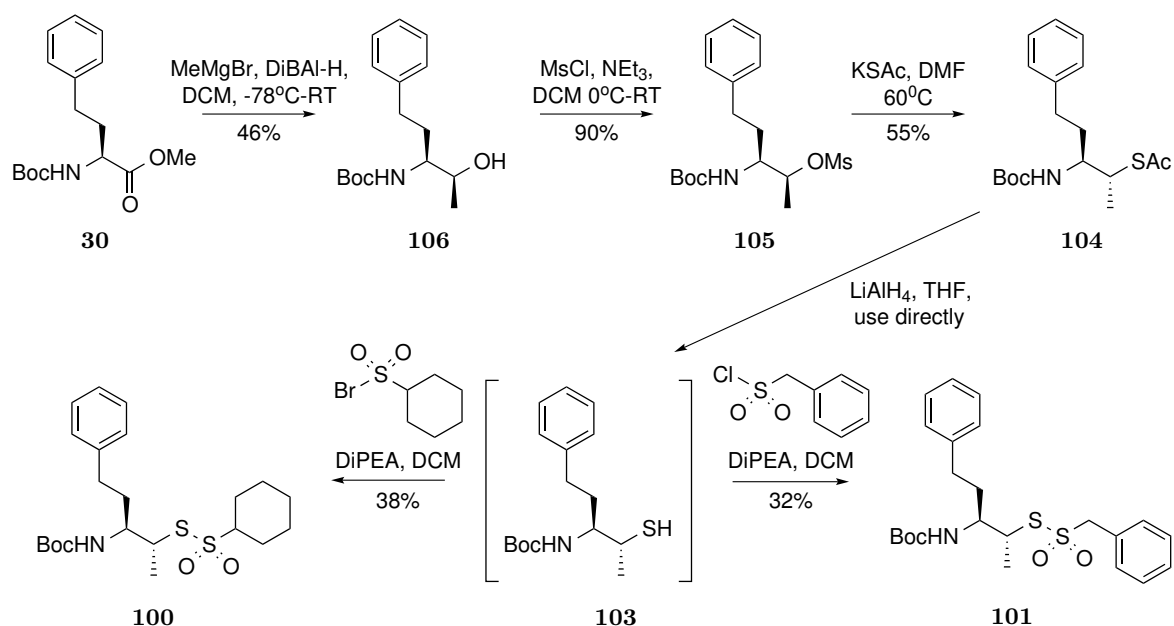


**Scheme 6.6:** Wittig reaction and competing elimination

### 6.2.3 Reduction & Diastereoselective Addition: Introduction of A Chiral Secondary Alcohol

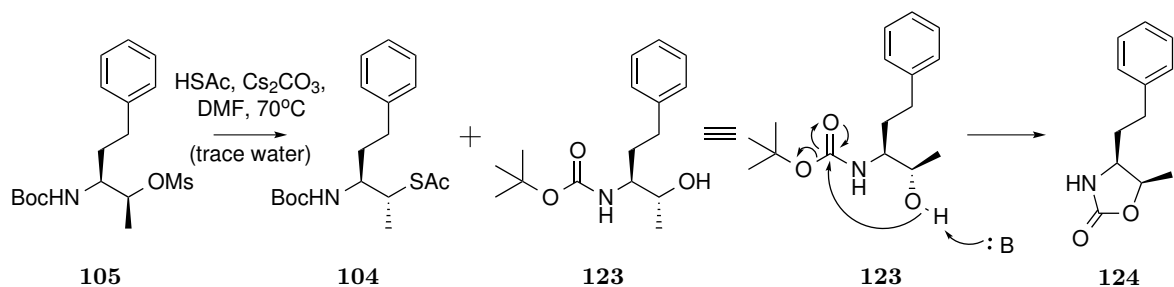
Having found the previous side chain functionalisation approach (section 6.2.2) to be limited by unstable synthetic intermediates and low yields, a new approach was desired. This was based on the retrosynthetic analysis shown in scheme 6.2 to include a reduction / diastereoselective addition to introduce the methyl group into amino acid starting materials.

As shown in scheme 6.7 this approach started from methyl ester **30**. The diastereoselective reduction / alkylation was based on the earlier work of Nyong *et al.*<sup>209</sup> and achieved a reasonable yield of 46%. The main route for loss of product was overreduction to the



**Scheme 6.7:** Secondary thiosulfonates by diastereoselective alkylation

corresponding primary alcohol **31** rather than the desired alkylation step which it may be possible to further optimise through addition times. The following mesylation proceeded in excellent yields (90%) to mesylate **105**.

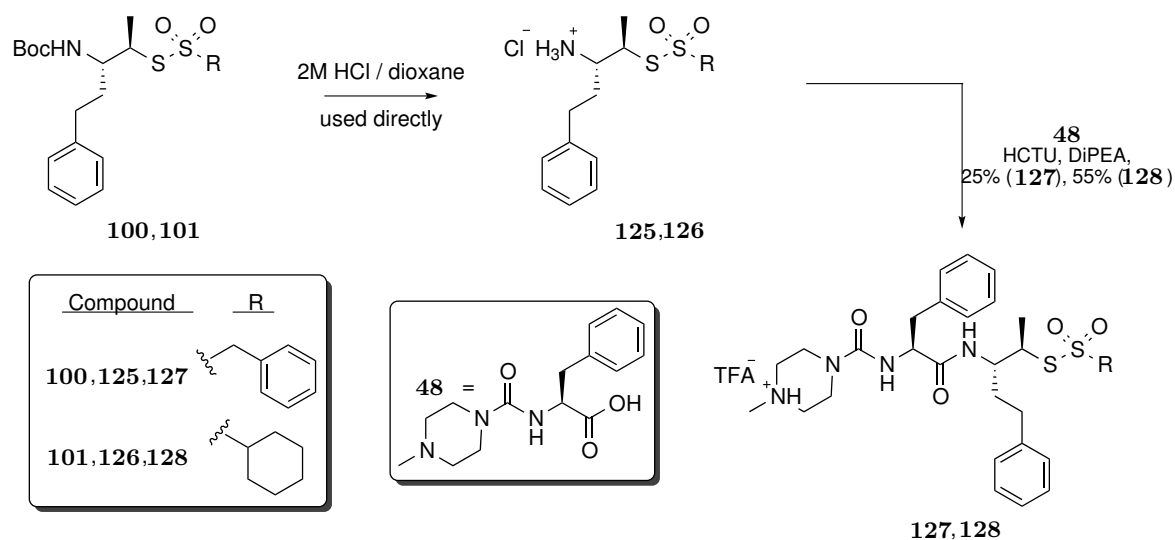


**Scheme 6.8:** Oxazolidinone formation through alcohol by-product

Following on from mesylate formation was substitution to thioacetate **104**, as shown in scheme 6.8, early attempts led to the undesired oxazolidinone formation. The conditions used were based on the earlier transformation of primary mesylates to thioacetates with thioacetic acid and caesium carbonate. Initially the reaction did not proceed, yielding only the mesylate starting material when conducted at room temperature. It was concluded that the rate of substitution for the secondary mesylate was significantly slower than the primary mesylates used in the earlier synthesis of thiosulfonate warheads. As such, the reaction was conducted at an elevated temperature of 70 °C which gave rise to oxazolidinone **124** as the major product. It was proposed that the increased sterics of the system, which led to increased reaction times, placed greater significance on the competing substitution with water (which had not been observed previously). As this reaction was

conducted with hygroscopic reagents, which had not been dried, the water content was likely significant leading to the proposed mechanism of oxazolidinone formation in scheme 6.8. The secondary alcohol, which formed by substitution of mesylate **105** with water, was able to perform an intramolecular cyclisation to generate oxazolidinone **124** under the basic conditions present. This was easily overcome by using dry solvents and KSAc, rather than the hygroscopic  $\text{Cs}_2\text{CO}_3$  and HSAc method employed previously, resulting in a reasonable yield of 55% shown in scheme 6.7.

With thioacetate **104** in hand it was reduced to thiol **103** with  $\text{LiAlH}_4$ , which was used directly in the following thiosulfonate warhead formation. Thiol **103** was diversified out into thiosulfonates **100** and **101** using the earlier developed methods with cyclohexane sulfonyl bromide and benzylsulfonyl chloride methods in moderate yields of 38% and 50% respectively. Unfortunately, time constraints prevented the corresponding diastereomer synthesis from being explored, which would involve isolating the corresponding bromide **109** from alcohol **106** followed by applying the same conditions above for all subsequent transformations from thioacetate to the thiosulfonate warheads. In terms of the development of the thiosulfonate warheads, this was not considered to be detrimental as the chemical reactivity, and subsequent stability, was considered to be the same for each diastereomer. However, for future work, it would be interesting to generate the corresponding diastereomer for biological testing as it may offer an increase, or decrease in binding to the target protease with the corresponding change observed in the inhibitor potency.



**Scheme 6.9:** Coupling the Secondary Thiosulfonate Warheads to their Backbone

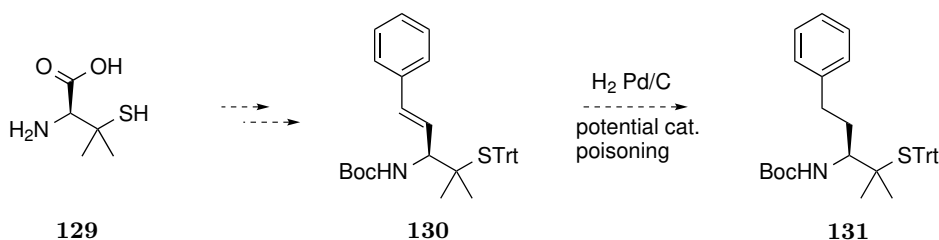
At present these newly developed secondary thiosulfonate warheads, with a methyl substituent anti to the amino acid side chain, were sufficient to test the hypothesis of increased stability leading on to improved potency. With each warhead fragment now to hand they were incorporated into the peptidic backbone as shown in scheme 6.9 in low to moder-



ate yields of 25% (**128**) and 55% (**127**). The inhibitors were isolated as the TFA salts after purification by preparative HPLC before further stability and biological testing was conducted.

## 6.3 Exploring a Route Towards $\beta$ -Gem-DiMethyl analogues

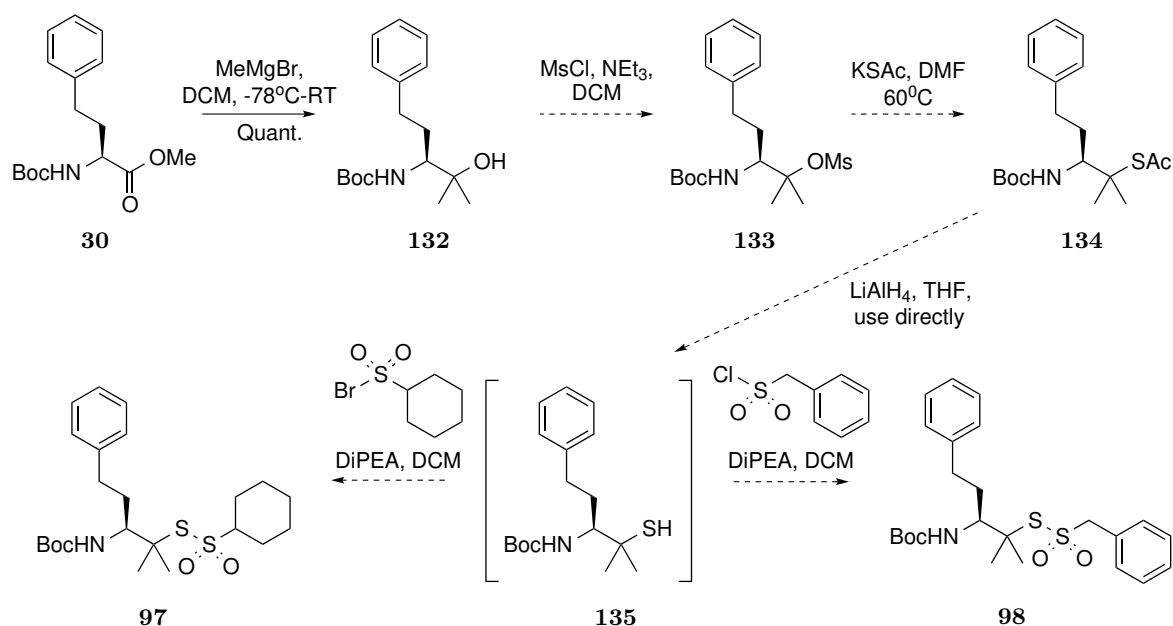
### 6.3.1 Substitution of a Hindered Tertiary Centre



**Scheme 6.10:** Synthesis towards tertiary thiosulfonates starting from D-penicillamine (**129**).

In an effort to further expand upon steric arguments to alter the reactivity profile of the thiosulfonate warheads, tertiary thiosulfonate warheads were desired to complement the previously discussed secondary thiosulfonate warheads. Similar to the previous synthesis towards the secondary thiosulfonates, discussed in section 6.2.2, two potential synthetic routes were identified. The first involved a ‘sidechain functionalisation’ approach starting from penicillamine and the second involved starting from an amino acid and methylating. The merits of each route were considered to be analogous to that of the mono-methyl derivatives with one key difference for the ‘sidechain functionalisation’ approach. It was thought that the presence of a bivalent sulfur may poison the Pd catalyst during the reduction of the alkene, as shown in scheme 6.10. Considering the potential for catalyst poisoning, combined with the previously observed instability of the aldehyde intermediate, this route was considered to be unsuitable.

As such, it was hypothesised that the tertiary thiosulfonate warheads could be obtained by di-methylation of an ester (scheme 6.11). Starting from methyl ester **30** tertiary alcohol **132** was obtained in good yields (72%) by alkylation with the Grignard reagent MeMgBr. Based on the previous success of the primary and secondary thiosulfonates, it was intended to functionalise the tertiary alcohol to mesylate **137** before substitution to introduce the sulfur functionality. Unfortunately, mesylate **137** was not found to form under these conditions, thus optimisation to find a suitable leaving group began.

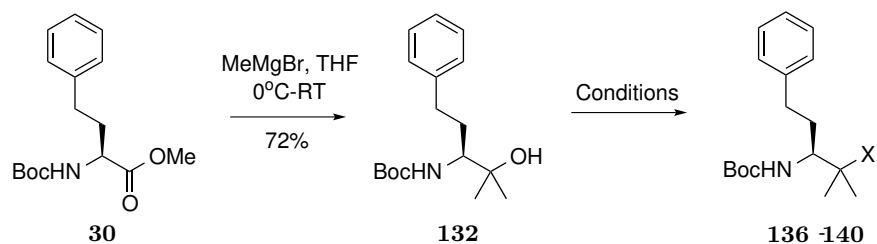


**Scheme 6.11:** Synthesis towards tertiary thiosulfonates by a tertiary alcohol intermediate

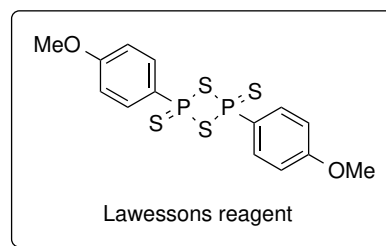
As shown in scheme 6.12 functionalisation of alcohol **132** to tosylate **133** was then attempted. Unfortunately, this led to no reaction, recovering only starting material. This was not entirely unexpected given the similarity with the previously attempted mesylation. Based on this it was decided to move away from sulfonate esters.

As generating the leaving group appeared to be the limiting step, it was hypothesised that phosphorous reagents may offer a viable alternative, as the phosphorous-oxygen bond is known to be one of the strongest and an excellent driving force in synthesis.<sup>197</sup> As such, functionalising with diphenylphosphine chloride was attempted based on the work of Pluempunapat *et al.* who reported using this method to functionalise tertiary alcohols before substitution with sulfur based nucleophiles.<sup>210</sup> Unfortunately a very low yield of only 8% was achieved.

Encouraged by the finding that phosphorous reagents were able to functionalise the tertiary alcohol, albeit in very low yields, this method was explored further. Lawesson's reagent (shown in scheme 6.12), which also derives its driving force through phosphorous-oxygen bond formation, has been reported to convert tertiary alcohols directly to tertiary thiols.<sup>211</sup> As such, this method was initially attempted in THF under reflux (65 °C), with no product identified as forming. It was thought that the reaction may not proceed due to the sterics of the tertiary alcohol, thus the reaction was attempted at an elevated temperature by reflux in toluene (110 °C). This led to a complex mixture of by-products, with none of the desired thiol isolated or observed. Encouraged by the consumption of the alcohol starting material, the reaction was then attempted at a lower temperature (50 °C) in an effort to reduce the formation of by-products. Unfortunately the outcome of this was the same, only



Compound	X	Conditions	Yield
133	OTs	TsCl, NEt <sub>3</sub> , DCM	No product isolated
137	OMs	MsCl, NEt <sub>3</sub> , DCM	No product isolated
138	OPPh <sub>2</sub>	CIPPh <sub>2</sub> , NEt <sub>3</sub> , DMAP, THF	8%
139	SH	Lawessons reagent, 65°C (reflux), THF	No product isolated
139	SH	Lawessons reagent, 110°C (reflux), Toluene	No product isolated
139	SH	Lawessons reagent, 50°C, toluene	No product isolated
TMP10	OTFA	TsCl, NEt <sub>3</sub> , DCM	No product isolated

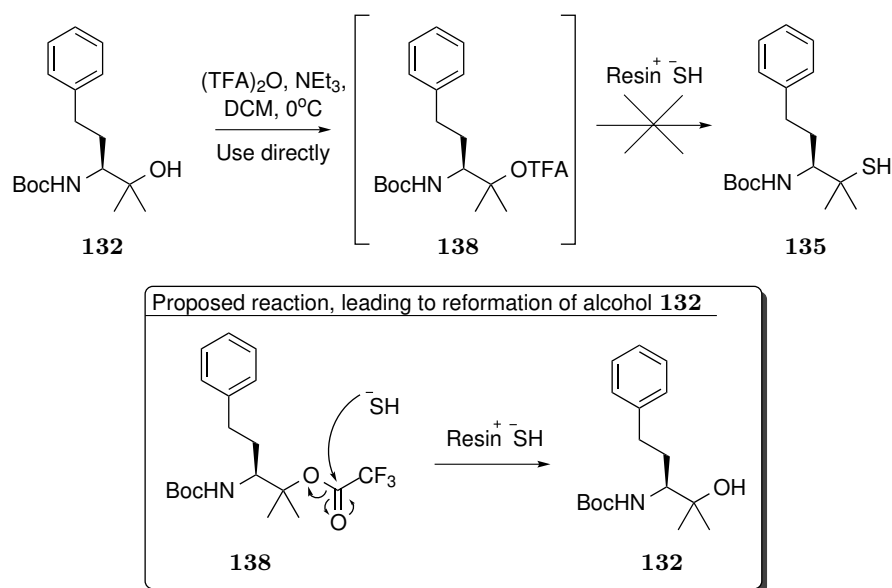


**Scheme 6.12:** Synthesis of tertiary thiosulfonates by di-methylation

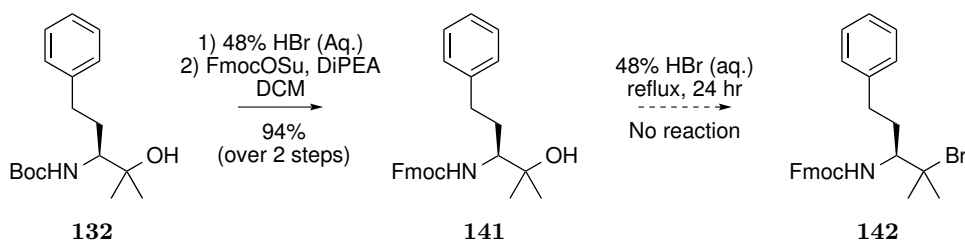
observing a complex mixture of products with none of the desired thiol observed.

As shown in scheme 6.13, it was attempted to isolate the desired tertiary thiol *via* substitution of a trifluoroacetate with resin bound thiol, based on the work of Bandgar *et al.*<sup>212</sup> Bandgar *et al.* reported using this method to generate *tert*-butyl thiol directly from *tert*-butanol in a 75% yield. The reported high yield combined with the direct isolation of the desired thiol made this method appear highly attractive. It was reported that the enhanced nucleofugal character of the trifluoroacetate moiety, in combination with the enhanced nucleophilicity of the resin bound thiol were favourable for this reaction. This argument was considered reasonable, with the resin bound thiol thought to be comparable to techniques such as employing crown ethers to enhance nucleophilicity (when compared with NaSH). As shown in scheme 6.13 the intermediate trifluoroacetate was observed to form by TLC however upon addition of resin bound thiol only the alcohol starting material was found to form and be isolated. It is proposed that competing addition to the carbonyl of the TFA moiety led to reforming of the starting material and trifluorothioacetic acid.

Finally, It was then hypothesised that functionalising the tertiary centre may be possible through an S<sub>N</sub>1 mechanism. As such, conc. HBr (Aq.) was thought to offer a viable option to isolate the corresponding bromide. This method was initially avoided as it was



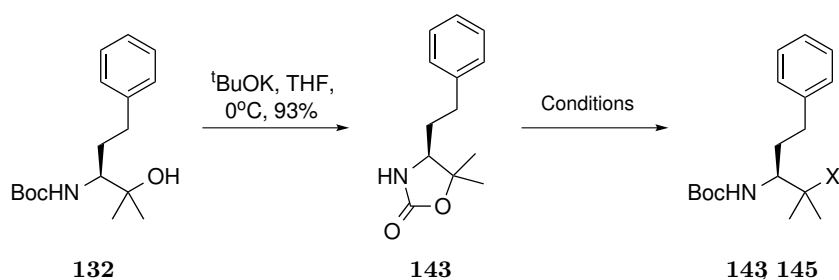
**Scheme 6.13:** Attempted synthesis of tertiary thiol by direct substitution of a trifluoroacetate with resin bound thiol.<sup>212</sup> Proposed side reaction leading to reformation of alcohol **132** shown (bottom).



**Scheme 6.14:** Converting tertiary alcohol to tertiary bromide under acidic conditions

incompatible with the Boc protection group, thus the Boc group was first deprotected and exchanged for an Fmoc protecting group (scheme 6.14). Then, Fmoc protected compound **141** was subjected to 48% HBr however no product was isolated, yielding only unreacted starting material. In hindsight, this reaction may have been more efficient if conducted in non-aqueous solvents such as HBr / AcOH, which would offer a suitable polar protic solvent for the  $S_N1$  reaction whilst taking advantage of Le Chatelier's principle to drive the reaction to completion (as water should be liberated).

As functionalising the alcohol appeared to be the limiting step in the synthesis, rather than the following substitution, a new approach was trailed. Based on the previously observed oxazolidinone formation, which occurred accidentally during the synthesis towards secondary thiosulfonates (section 6.2.3, scheme 6.8), the method in scheme 6.15 was proposed. It was hypothesised that oxazolidinone **143** could be generated under basic conditions, which occurred in an excellent yield of 93% with the use of potassium tert-butoxide. Following on from this, the reactivity of the oxazolidinone, being a cyclic carbamate, was compared with that of a Boc group, being a linear carbamate. It was



Compound	X	Conditions	Yield
144	H	TFA:DCM (1:1), TIPS (2% v/v), 50°C.	No reaction
145	SBn	4M HCl/dioxane, BnSH	No reaction
145	SBn	33% HBr in AcOH, reflux	No reaction
145	SBn	BF <sub>3</sub> ·OEt <sub>2</sub> , BnSH, DCM	No reaction

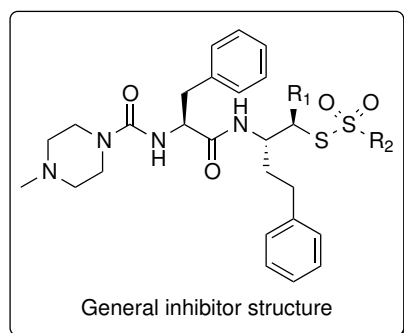
**Scheme 6.15:** Attempted ring opening of oxazolidinone **143** under acidic conditions. TIPS and BnSH were trialed as carbocation scavengers to determine if ring opening under acidic conditions was possible.

thought that the same driving forces for Boc deprotection under acidic conditions also existed for oxazolidinone **143**. Namely, stabilisation of the generated carbocation by a tertiary centre, as well as liberation of CO<sub>2</sub>, which could drive the reaction forward. Hence, it was hypothesised that under acidic conditions the oxazolidinone **143** could be ring opened, generating a carbocation that could be captured with a suitable carbocation scavenger (ideally sulfur based, leading to a thiol). To test this hypothesis oxazolidinone **143** was subjected to increasingly strong acidic conditions, eventually leading to reflux in conc. HBr / AcOH. Either TIPS, or BnSH, were introduced as a carbocation scavengers, with the intention of optimising the scavenger used if the reaction was seen to progress. In each case only unreacted oxazolidinone **143** was recovered, with no other products observed. It was thought that acetic acid, being a polar protic solvent, would be optimal for this transformation and the generated carbocation would be scavenged by either the bromide, leading to the corresponding tertiary bromide or BnSH, leading to the thioether **145**. Surprisingly, oxazolidinone **143** was found to be very stable under acidic conditions, offering no detectable decomposition. A final attempt was made with the use of a Lewis acid (BF<sub>3</sub>·OEt<sub>2</sub>) under dry conditions, with the main reasoning being exclusion of water from the reaction, however this was also unsuccessful in opening the ring with only unreacted oxazolidinone **143** being observed. It was reasoned that, being a cyclic system, the reactivity of **143** was significantly different to the linear (Boc) counterpart. This may be explained by poor orbital overlap, which is required to push electrons into

the  $\sigma^*$  antibonding orbital, breaking the bond and open the ring system.

Due to time constraints, attempts towards the synthesis of tertiary thiosulfonates ended here. It is clear that further optimisation can be done and it is believed that the tertiary thiosulfonates will be accessible by one of the methods, or a close analogue, presented above. A next obvious step would be to attempt the synthesis with a triflate, which also offers enhanced nucleofugal character similar to the trifluoroacetate presented, however would not suffer from the corresponding side reaction (attack on the carbonyl). This, in combination with resin supported thiol, may offer a suitable synthesis towards the required tertiary thiol intermediate to continue towards tertiary thiosulfonates.

### 6.3.2 Stability Testing Towards Common nucleophiles



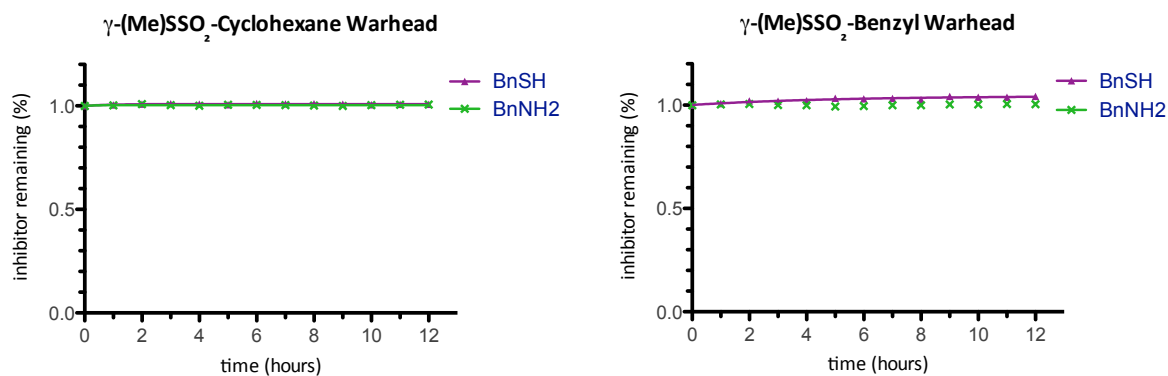
Compound	R <sub>1</sub>	R <sub>2</sub>
11	H	
127	Me	
12	H	
128	Me	
7	H	
10	H	

**Figure 6.2:** Structure of primary thiosulfonate inhibitors and their secondary thiosulfonate counterparts. Also included are the first generation aromatic inhibitors, to highlight the progression of the thiosulfonate inhibitors.

As the core reasoning for introducing steric arguments was to improve the stability of the thiosulfonate warheads, the stability of the secondary thiosulfonates towards common nucleophiles was assessed. It was hypothesised that these warheads would mimic the trend seen earlier with the cyclohexane derivative being most stable, followed by the benzylic derivative. It was also expected that introduction of a methyl group, to generate the secondary thiosulfonates in each instance, would further improve the warhead stability. As such, comparisons will be drawn with the respective primary thiosulfonate warheads

to deduce the effect steric arguments have upon thiosulfonate warhead stability in this section. Figure 6.2 has been provided for reference to the structure of each compound discussed in the following section.

### Stability Towards Thiols and Amines - The benzyl Thiol and Benzyl Amine test system

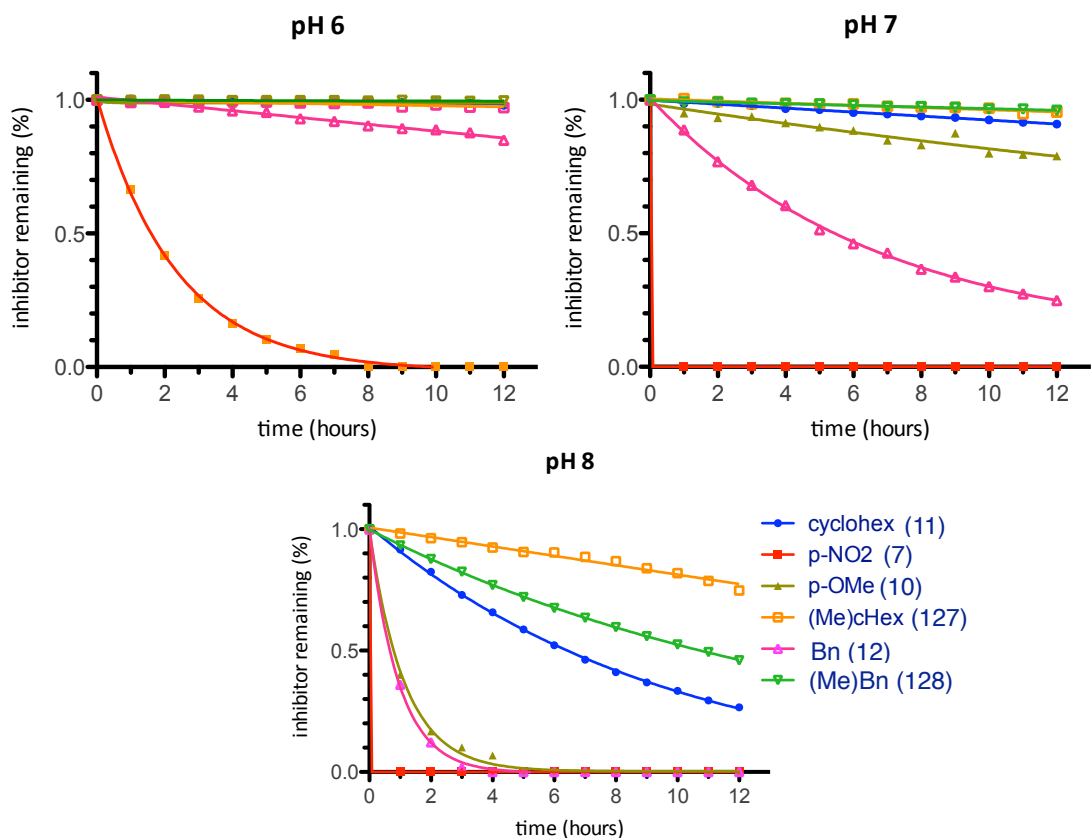


**Figure 6.3:** Stability of inhibitor **127** (left) and **128** (right) towards benzyl amine and benzyl mercaptan. Inhibitor degradation was determined by HPLC in MeCN at room temperature with 13 measurements taken 1 hour apart, representing time 0 to 12 hours. Approximately 0.3 mg of inhibitor was accurately weighed and suspended in MeCN, to which 10 equivalents of BnSH or BnNH<sub>2</sub> was added immediately before measurement. Results were then normalised against time 0 and plotted as scatter graphs for visualisation.

To begin with, the stability of the secondary thiosulfonates **127** and **128** towards amines and thiols was assessed. Based on the previous success of the primary thiosulfonate bearing an aliphatic warhead (compounds **11** and **12**, section 5.2.3) it was expected that the secondary thiosulfonates would display very high stability towards both amines and thiols. Testing was conducted by the earlier described method with benzyl amine and benzyl thiol, used for the primary thiosulfonates, for comparison (section 5.2.3). As can be seen in figure 6.3 the secondary thiosulfonates were stable towards both amines and thiols over the 12 hour test period. When compared with the earlier aromatic thiosulfonate warheads, this was a significant improvement. However, when compared against the primary, aliphatic thiosulfonates this was expected, hence the main issue to be addressed was the aqueous stability profiles.

### pH Dependent Aqueous Hydrolysis

Moving on to the aqueous stability profiles, comparison of the newly developed secondary thiosulfonates (**127** and **128**) with their primary thiosulfonate counterparts (**11** and **12**)



**Figure 6.4:** pH dependent aqueous stability of secondary thiosulfonate warheads compared to the previous primary aromatic and aliphatic thiosulfonate warheads. Inhibitor degradation was determined by HPLC with peak areas integrated against an internal standard (Ac-Phe-OH), 13 measurements were taken 1 hour apart, representing time 0 to 12 hours. Sodium phosphate buffer (0.1 mM) of the relevant pH containing 5% DMSO was utilised, initial inhibitor concentration was 245  $\mu\text{M}$  and internal standard (Ac-Phe-OH) concentration was 338  $\mu\text{M}$ . Results were then normalised and plotted as scatter graphs for visualisation.



there was a clear improvement, as shown in figure 6.4.

Looking towards the cyclohexane warhead analogues, the primary thiosulfonate warhead of compound **11** was known to be stable at pH 6, display minimal degradation at pH 7 (half life approximately 60 hours) and had a half life of approximately 6 hours at pH 8. Now, with the secondary thiosulfonate warhead of compound **127** the inhibitor was stable at both pH 6 and pH 7, with a greatly increased half life of approximately 24 hours at pH 8.

A similar trend was observed when comparing the benzylic warheads of compounds **12** and **128**. Previously, the primary thiosulfonate warhead of compound **12** showed a half life of approximately 40 hours at pH 6, 5 hours at pH 7 and 45 minutes at pH 8. Now, with the incorporation of a methyl group to form the corresponding secondary thiosulfonate warhead of inhibitor **128** the warhead was effectively stable at pH 6 and 7, with a greatly increased half life of approximately 11 hours at pH 8.

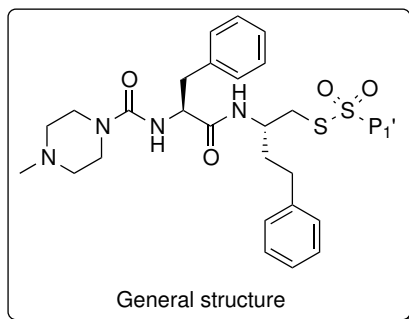
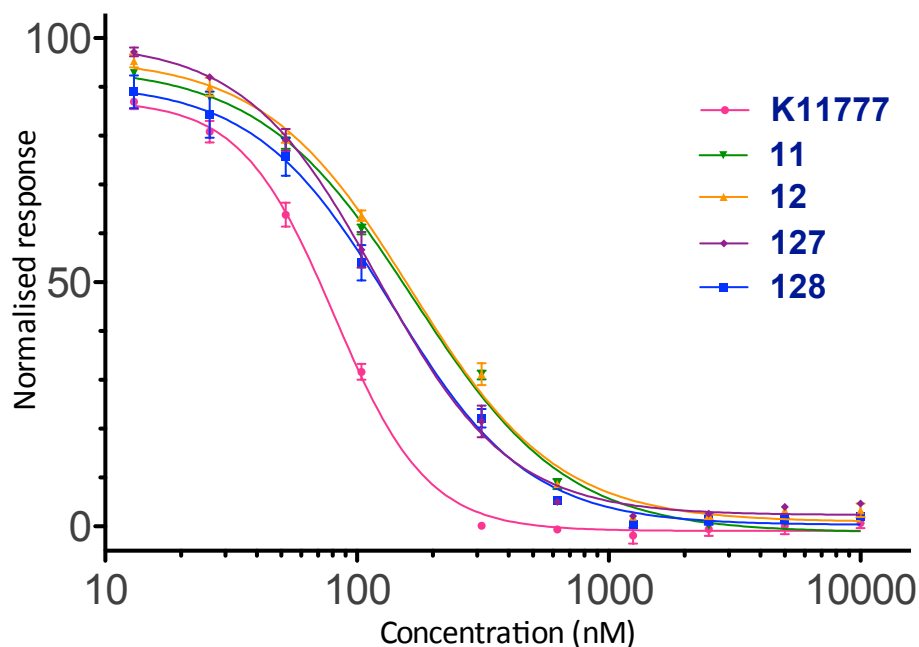
As expected, the cyclohexane derived warhead (**11**) was more stable than the benzylic counterpart (**12**), as had been observed previously for the primary thiosulfonates (section 5.2.3). However, in both cases the thiosulfonates were now found to be significantly more stable towards aqueous hydrolysis, especially at the relatively basic pH of 8 (in terms of physiological pH). This was considered a significant step forward, as a half life in the order of hours at higher physiological pH levels would offer a wider potential therapeutic window. The earlier aromatic warheads **7** and **10** have also been displayed in figure 6.4 to highlight the progression of the thiosulfonate warheads from the early aromatic constructs.

## 6.4 Biological evaluation - Papain assay

Moving on from the success of the stability testing towards common nucleophiles, the inhibition of papain was to be assessed. It was hypothesised that, based on the trends observed in the work thus far, the increased stability of the new secondary thiosulfonate warheads would offer improved potency in the papain assay. Figure 6.5 displays the newly made secondary thiosulfonates (**127** and **127**) compared with their primary thiosulfonate counterparts (**11** and **12**). In accordance with the hypothesis, both secondary thiosulfonate warheads offered an improved potency when compared with their primary thiosulfonate counterparts (approximately 40 nM, 1.3 fold increase). Furthermore, the most stable warhead of compound **127** (secondary thiosulfonate with a cyclohexane warhead) was again observed to be the most potent at 121 nM. The inhibitor bearing a secondary thiosulfonate, benzylic warhead (**12**) was close to follow, with an  $IC_{50}$  value of 137 nM).

It can be concluded that secondary thiosulfonate warheads of **127** and **128** retained, or

### Papain IC<sub>50</sub> of Secondary Thiosulfonate Warheads



Compound	P <sub>1</sub> '	IC <sub>50</sub> (nM)
<b>11</b>		168 ± 4
<b>12</b>		164 ± 7
<b>127</b>		121 ± 11
<b>128</b>		137 ± 8
<b>6 (K11777)</b>		81 ± 2

**Figure 6.5:** Papain IC<sub>50</sub> values of the secondary thiosulfonate warheads **127** and **128** compared to their primary thiosulfonate counterparts (inhibitors **11** and **12**). Assay buffer: sodium phosphate (100 mM, pH 6.5) containing EDTA (1.5 mM) and 2% DMSO. Final concentrations in the wells were; Papain: 4 μM; substrate: 1.0 mM; inhibitor: doubling dilution range starting at 10 μM giving: 10 μM, 5 μM, 2.5 μM, 1.25 μM, 0.625 μM, 0.3125 μM, 156.25 nM, 78.125 nM, 39.0625 nM. Error bars on graph display SEM across 3 independent repeats. Tabulated ± values represent the std error of the IC<sub>50</sub> values across 3 independent repeats.

slightly improved upon the previous  $IC_{50}$  values of their primary thiosulfonate counterparts (**11** and **12**). When combined with the significant increase in stability towards common nucleophiles displayed in section 6.3.2, this displays a significant advance in the development of thiosulfonates as cysteine protease inhibitors. Of particular significance was the increased stability towards aqueous hydrolysis, with a substantial increase in half life even at higher pH values. As it is believed the aqueous stability will become more significant in increasingly complex biological settings, the findings of improved potency in the papain assay are anticipated to become exaggerated with increasing biological complexity.

## 6.5 Summary and Future Work

In summary, controlling the reactivity by introducing steric arguments to reduce the rate of nucleophilic attack on the thiosulfonate warhead has been explored. This led to the successful synthesis, stability and biological testing of the secondary thiosulfonates, bearing an additional methyl group adjacent to the site of attack. It was found that increasing the sterics surrounding the thiosulfonate warhead successfully down regulated their reactivity, which was met with an improved  $IC_{50}$  against papain. Efforts towards the synthesis of tertiary thiosulfonates were made to further expand upon this trend, however they were unsuccessful. Based on the observed results, future work should include accessing the tertiary thiosulfonates and assessing their reactivity profiles. Furthermore, the opposite diastereomer of the secondary thiosulfonates presented could be made to assess the potential impact of chirality on the function of the thiosulfonates. With the thiosulfonates having a substantially improved reactivity profile, compared with the early aromatic primary thiosulfonate warheads, it was decided to move on to more complex and relevant biological testing targeting the neglected tropical disease Schistosomiasis.

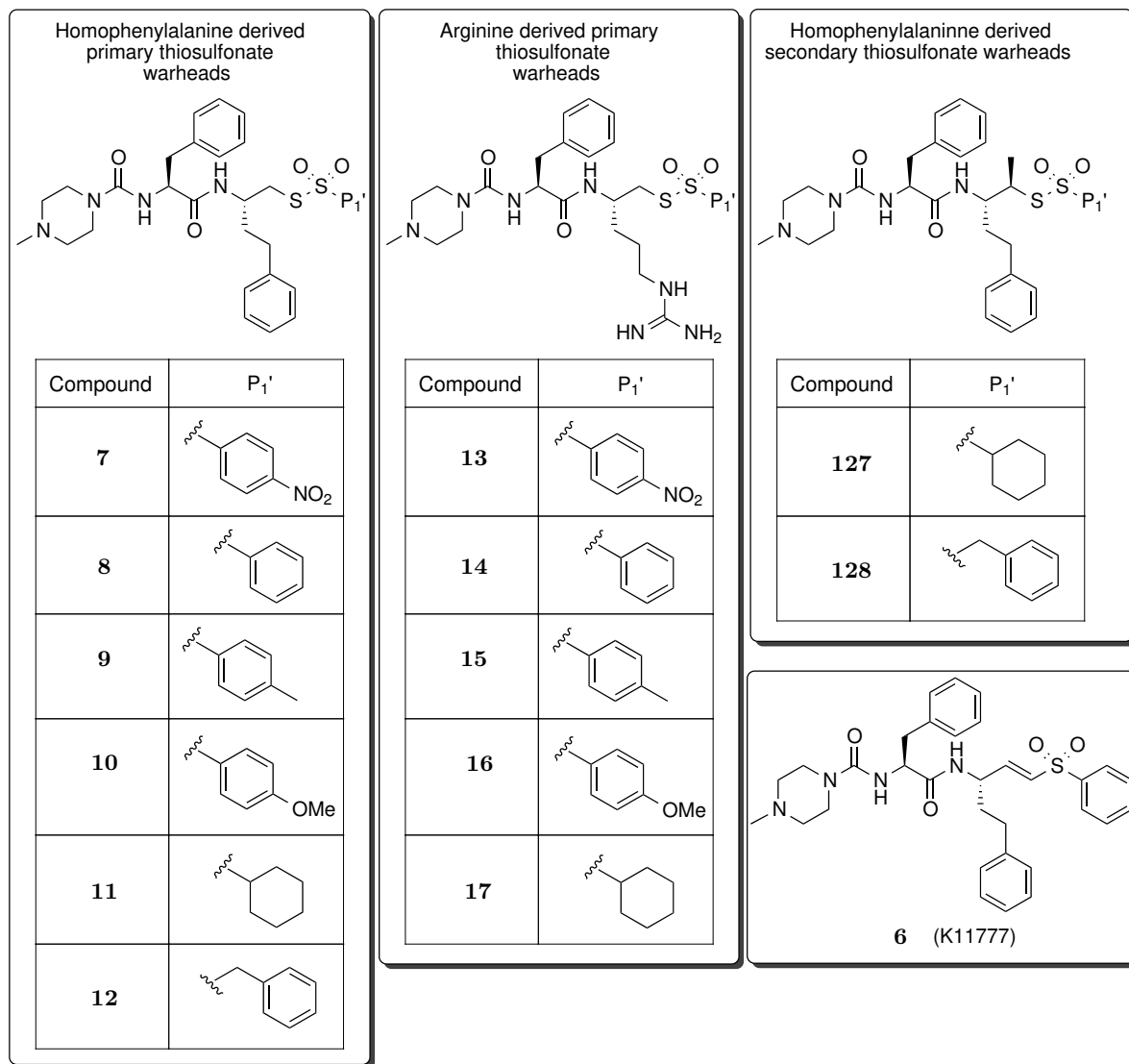
## Chapter 7

# Applying Thiosulfonates to target a Neglected Tropical Disease: Schistosomiasis

The work presented in this chapter was conducted in collaboration with the research group of Dr. Andrea Kriedenweiss at The Institute of Neglected Tropical Diseases, Tubingen University, Germany. All viability testing presented herein was conducted by Mr Erik Koehne with the author providing compounds for testing.

Parts of this chapter have been published: D.J. Ward, H. Van de Langemheen, E. Koehne, A. Kriedenweiss, and R.M.J. Liskamp. Highly tunable thiosulfonates as a novel class of cysteine protease inhibitors with anti-parasitic activity against *Schistosoma mansoni*. *Bioorganic & Medicinal Chemistry*, 27(13):2857 - 2870, 2019.

Having explored the reactivity profiles of the thiosulfonate warheads and proven their suitability as cysteine protease inhibitors in a model (papain) system, it was decided to expand testing towards a more biologically relevant target. It was hypothesised that moving towards increasingly complex biological systems, such as *S. mansoni* (the parasite responsible for schistosomiasis) the earlier observed stability and potency trends would become exaggerated. *S. mansoni* is a trematode blood fluke causing acute and chronic intestinal schistosomiasis (bilharzia) in humans. Schistosomiasis is widely prevalent in Africa, Middle East, and South America affecting >200 million people. Large-scale elimination activities are ongoing and praziquantel, the only schistosomiasis medication, is widely deployed by mass drug administration programs. New anti-schistosomal drug candidates are urgently needed since reduced susceptibility to praziquantel has been reported for many decades.<sup>91,92</sup> Research on candidate antischistosomal targets has focused on digestive enzymes in *S. mansoni* since the parasite feeds on the haemoglobin of red blood



**Figure 7.1:** Structure of all thiosulfonate inhibitors tested in this chapter. Homophenylalanine derived primary thiosulfonates (left), arginine derived primary thiosulfonates (middle) and homophenylalanine derived secondary thiosulfonates (right)

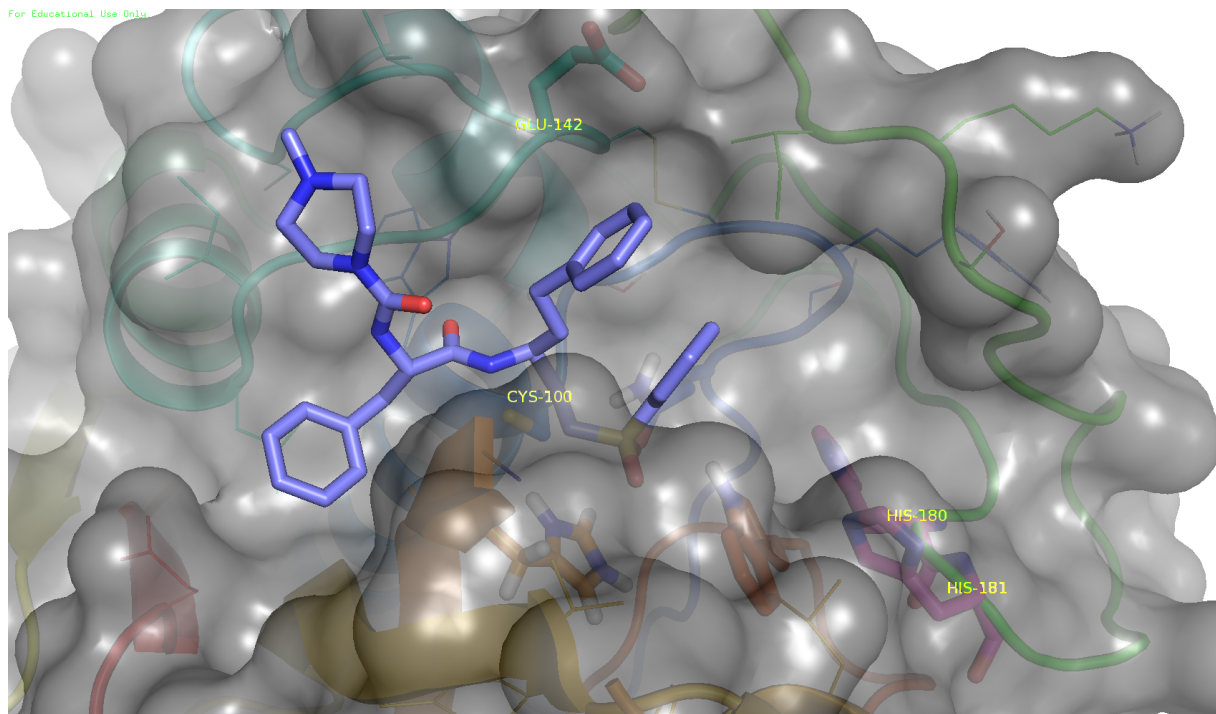
cells, essential for the parasite's growth, development, and reproduction.<sup>93</sup> Papain like cysteine proteases play a major role in haemoglobin degradation and have become an important drug target.<sup>44,213,214</sup> Namely, the cathepsin B like cysteine protease SmCB1. The reference compound in this study (**6**, K11777) has shown a high efficacy against the *S. mansoni* parasite in the murine model.<sup>101</sup> Furthermore, K11777 (**6**) has been utilised to probe the structural basis for inhibition of SmCB1, providing a foundation for the thiosulfonate derivatives presented.<sup>187</sup>

First, the structural basis for targeting *S. mansoni* was explored with *DrawToDock* before viability testing was conducted. Then, conclusions will be drawn which relate back to earlier observed trends in the papain assay test system between reactivity and potency, as well as new considerations such as permeability. A total of 13 thiosulfonate derivatives (figure 7.1) were tested *in vitro* against *S. mansoni*, of which 8 compounds showed high activity. The first 10 compounds were tested at the same point in time with the final 3 compounds being tested at a later date. The most potent thiosulfonate from the first 10 was tested alongside the final 3, to account for any variation between biological assays.

## 7.1 Modelling Study: Applying Thiosulfonates Inhibitors to SmCB1

As mentioned previously, SmCB1 is a major cysteine protease in *S. mansoni*, validating it as a suitable target. SmCB1 is a papain like cysteine protease, being of the Clan CA family C1, the active site shares many common features with papain used in the earlier test system. There are however a few key differences surrounding the active site, most notable a feature named the occluding loop which can be seen in figure 7.2. The occluding loop is a feature of cathepsin B and cathepsin B like proteases that is defined as a loop extending beyond the C-terminus of the peptidic substrate in the active site. A notable feature of the occluding loop is the presence of two histidine residues, His-180 and His-181 shown in figure 7.2, which allow Cathepsin B like proteases to anchor the C-terminus through H-bonding to these key His residues (see section 1.1.1). As such, cathepsin B like proteases can function as di-peptidyl carboxypeptidases in addition to their endopeptidase activity, with the occluding loop anchoring the C-terminus in such a way that the scissile bond is placed two amino acid residues back from the C terminus. This di-peptidyl carboxypeptidase activity has been reported as being highly pH dependent. The occluding loop of mammalian cathepsin B was found to alter its position and thus C-terminal anchoring ability in a pH dependent manner, whilst the parasitic variant SmCB1 was reported to

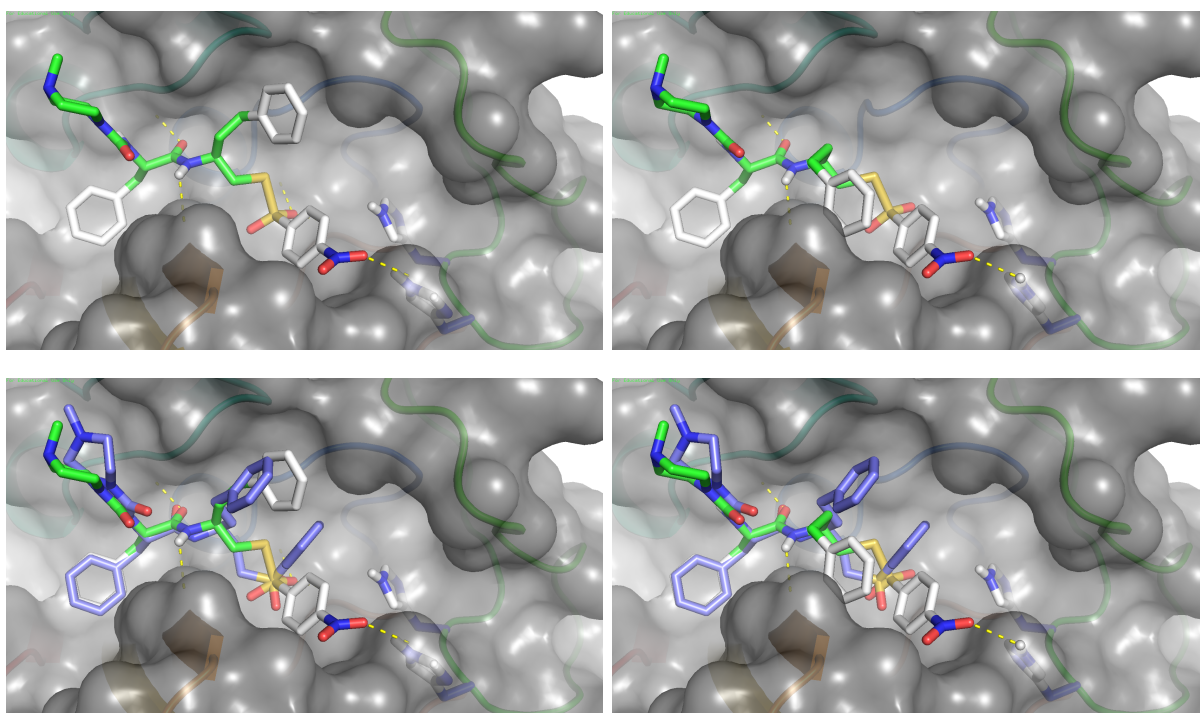
be less flexible through stabilisation by two salt bridges.<sup>28,215</sup> It was hypothesised that the occluding offered a unique H-bonding partner amongst the cysteine proteases, which could be exploited to improve binding of the thiosulfonate inhibitors.



**Figure 7.2:** Crystal Structure of *S. mansoni* cathepsin B like Cysteine Protease SmCB1 with K11777 bound in active site; PDB code: 3S3R.

A further key difference between papain and SmCB1 is found at the bottom of the  $S_1$  binding pocket. SmCB1 has a glutamic acid residue at the bottom of the  $S_1$  binding pocket, again shown in figure 7.2, which offers an excellent H-bond partner for basic residues. As such, it was expected that the basic arginine residue of inhibitors **13-17** would offer an ideal partner for optimal binding. Indeed, the ability for Cathepsin B as well as SmCB1 to accept basic residues in this position has been reported.<sup>216</sup> Gratifyingly, it was anticipated that this would correlate well with the earlier observed trends in the papain test system that arginine derived warheads would offer improved inhibition (section 4.2.3). In this case, the preference for arginine is due to coordination towards a slightly different residue (Glu-142 in this case) rather than coordinating backbone carbonyls of Cys-63 and Ser-21 (in the case of Papain, see chapter 4 section 4.1). It was hypothesised that Glu-142 would offer a more favourable H-bonding partner with arginine than the electrostatic environment created by backbone carbonyls of Cys-63 and Ser-21 in the papain test system previously, thus the arginine constructs were anticipated to enhance target binding to SmCB1.

Initially all inhibitor constructs were screened against SmCB1 with the use of *DrawToDock* to determine if these inhibitors would offer reasonable binding towards this new target. Gratifyingly, the overall binding mode observed mirrored that of K11777, based on the

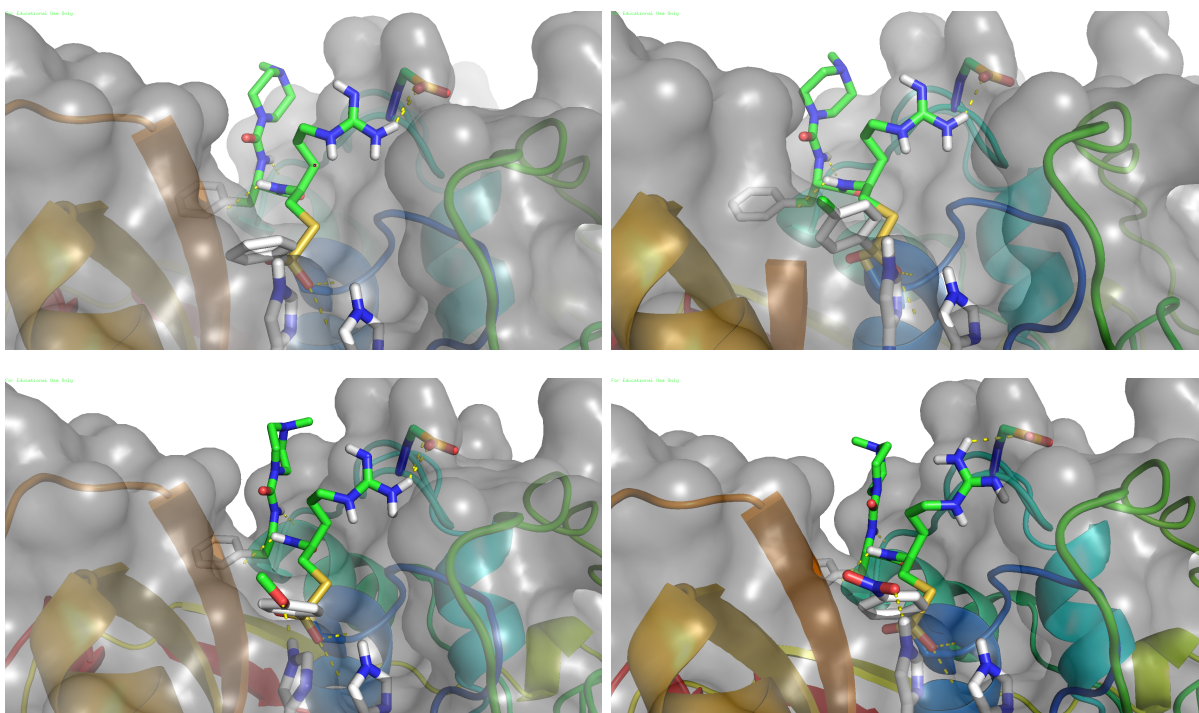


**Figure 7.3:** Docked pose of thiosulfonate inhibitor **7** with SmCB1. 2 possible docked poses (top left and top right) and the overlay with K11777 from the crystal structure (bottom left and bottom right). K11777 is shown with blue carbon backbone, thiosulfonate **7** is shown with green carbon backbone.

crystal structure of K11777 bound to SmCB1 (PDB code 3S3R). Two noteworthy observations were made during this screening. Firstly, as shown in figure 7.3, it was found that the *p*-NO<sub>2</sub> substituted warheads offered a favourable H-bonding pair with His-180 and His-181 of the occluding loop. When compared with K11777 (bound in the original crystal structure) the overall binding mode correlated well with the homophenylalanine residue occupying the same S<sub>1</sub> region, the phenylalanine residue occupying the S<sub>2</sub> region and the N-methylpiperazine cap occupying the S<sub>3</sub> region. The greatest difference was observed in the S<sub>1</sub>' region where the *p*-NO<sub>2</sub> substituted warhead of compound **7** coordinated His-181 of the occluding loop. This was a very encouraging find, suggesting that the *p*-NO<sub>2</sub> moiety may serve as a suitable mimic of the C-terminus of natural substrates. Thus, it was expected that the *p*-NO<sub>2</sub> substituted warheads may perform better than others on the basis of the initial binding event, K<sub>i</sub>. Earlier studies correlated the decreased stability of the *p*-NO<sub>2</sub> warheads with reduced potency (section 3.6), which may counteract the potential contribution to enhanced binding.

Secondly, it was found that arginine constructs **14-13** all offered H-bonding between the arginine side chain and Glu-142 at the base of the S<sub>1</sub> binding pocket. Furthermore, it was observed that compound **13** bearing the *p*-NO<sub>2</sub> substitution displayed H-bonding between the His-181 of the occluding loop and the nitro group. It was hypothesised that



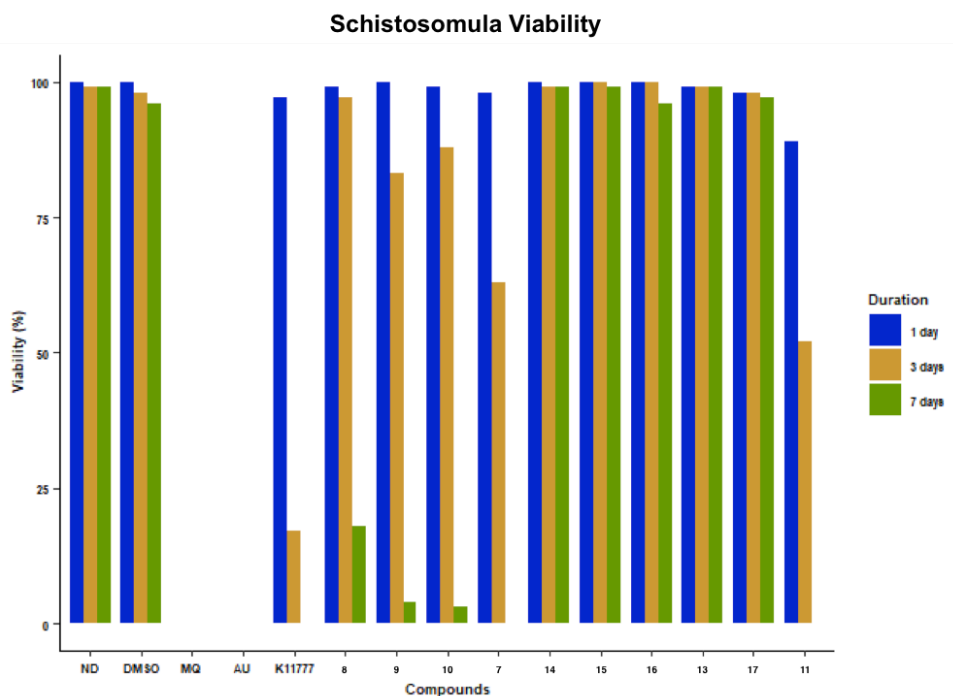


**Figure 7.4:** Docked pose of thiosulfonate inhibitors **14** (top left), **15** (top right), **16** (bottom left) and **13** (bottom right) with SmCB1. All inhibitor constructs show coordination of the arginine side chain towards Glu-142. Compound **13** (bottom right) also shows coordination of the *p*-NO<sub>2</sub> group towards His of the occluding loop.

the additional H-bonding partners presented here would enhance the initial binding of inhibitor constructs and translating through to more potent inhibitors, as was previously demonstrated in the papain assay.

## 7.2 Viability Testing Against Schistosomula: The Larval Form of Schistosomiasis

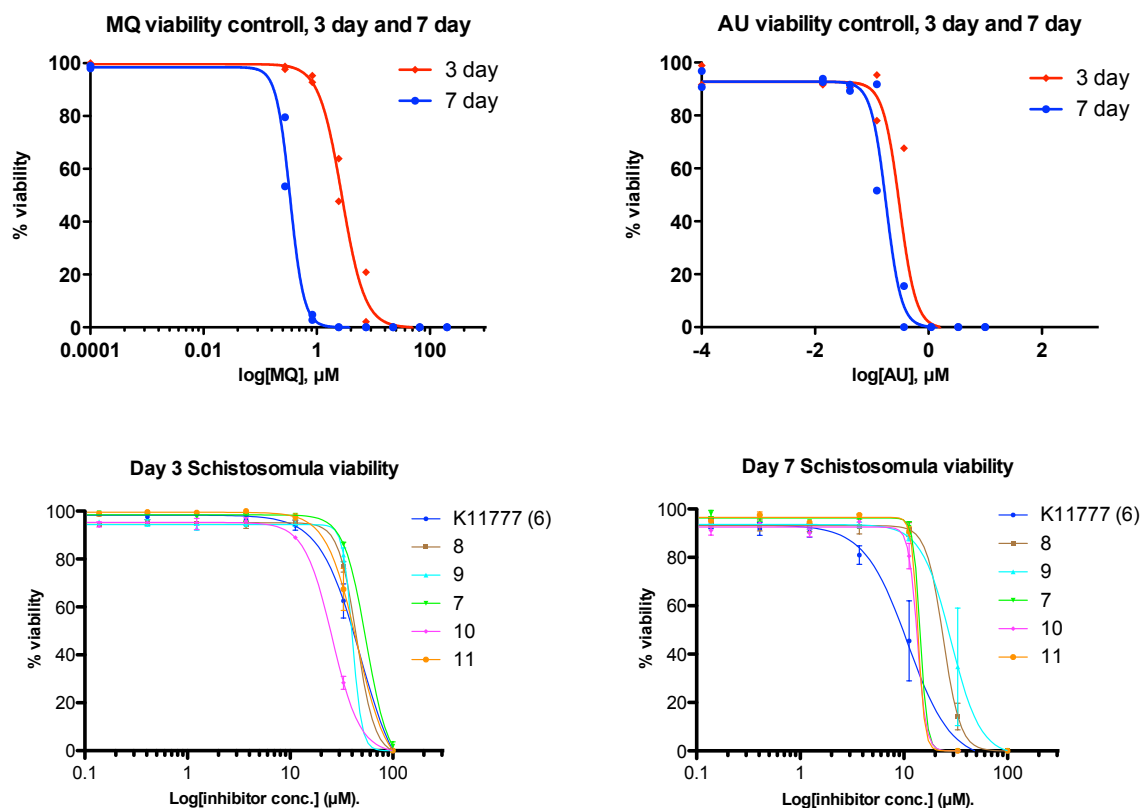
For the first 10 thiosulfonate inhibitors tested (**7-11** and **13-17**), viability testing was conducted at a set concentration of 30  $\mu$ M to screen for active compounds, as shown in figure 7.5. Mefloquine and auranofin were used as control compounds and K11777 (**6** included as a reference compound for comparison. Interestingly all arginine derived constructs (**13-17**), which were anticipated to be more potent based on earlier papain studies and modelling towards SmCB1, exhibited effectively no activity. This was in contrast to the homophenylalanine derivatives **7-11**), which were found to be highly active. Based on this, it was believed that the lack of activity of the arginine derived inhibitors was likely due to poor permeability of the parasitic tegument. The parasitic tegument is the outer body covering the parasite, which is highly hydrophobic potentially reducing the permeability



**Figure 7.5:** In vitro drug susceptibility screen of schistosomula. Mean % viability of schistosomula after drug exposure for 1 day, 3 days and 7 days. Derivatives (**7-11** and **13-17**) and K11777 (**6**) control were tested at 30  $\mu\text{M}$ , MQ (mefloquine) at 100  $\mu\text{M}$ , and AU (auranofin) at 1  $\mu\text{M}$ . ND (no drug, medium only) and DMSO control for schistosomula viability. Assay was done in triplicate.

of the more polar arginine derivatives, carrying a positive charge at physiological pH.<sup>217</sup> There are many other factors which may account for the selective uptake, or increased permeability, of the homophenylalanine derivatives tested here. For example, it has been reported that lipids are of particular interest in the research of *S.mansoni* with some extraordinary lipids, such as  $\Delta^5$ -octadecanoic acid, and large amounts of lysophospholipids being found in the tegument.<sup>218</sup> Interestingly, *S.mansoni* worms are unable to synthesise sterols and fatty acids *de novo*.<sup>219</sup> Recent reports on the lipid topography of the oral and ventral suckers (involved in the uptake of nutrients) suggesting an enhanced lipid concentration in this region.<sup>217</sup> These characteristics of the parasite biology may be significant in the design of inhibitors, resulting in the loss of potency moving from homophenylalanine (**7-11**) to arginine constructs (**13-17**).

Following on from this, the trend observed for all active compounds (**7-11**) broadly followed the earlier observed trends between warhead reactivity and potency in the papain test system. To further confirm this all active constructs (**7-11**) were carried forward to determine their  $\text{EC}_{50}$  values, as shown in figure 7.6. Most notably, the cyclohexane analogue **11** was found to be the most potent in this early screen with the *p*-NO<sub>2</sub> substituted warhead **7** appearing almost equipotent. Interestingly, this suggests that the potency mirrors

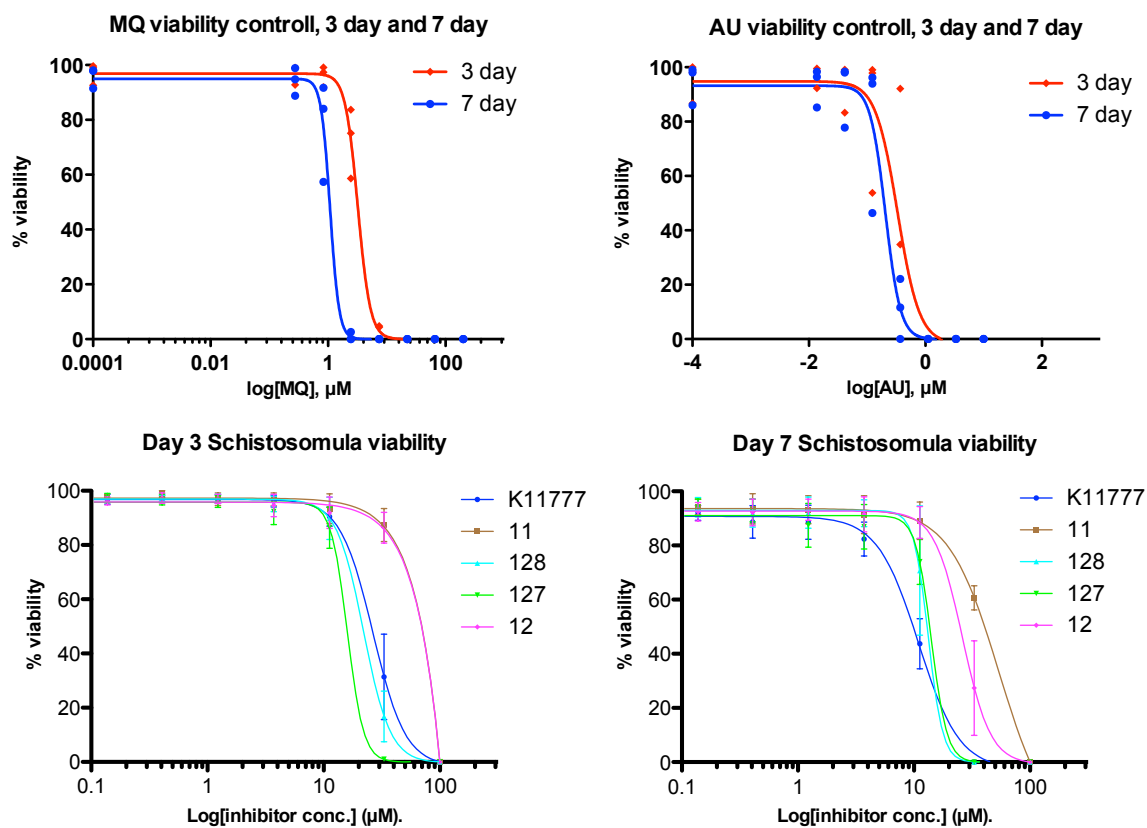


Compound number	Concentration range ( $\mu\text{M}$ ) tested	3 Day Mean $\text{IC}_{50}$ (SD) ( $\mu\text{M}$ )	7 Day Mean $\text{IC}_{50}$ (SD) ( $\mu\text{M}$ )
MQ	200-0.27	2.4 (0.6)	0.4 (0.1)
AU	10-0.014	0.2 (0.1)	0.1 (0.1)
K11777 <b>6</b>	100-0.14	59.9 (23.5)	7.3 (3.3)
<b>8</b>	300-0.14	64.4 (19.3)	21.4 (3.5)
<b>9</b>	300-0.14	41.1 (7.6)	26.9 (4.7)
<b>10</b>	300-0.14	44.9 (27)	13.6 (1.0)
<b>7</b>	300-0.14	29.3 (12.6)	14.7 (1.1)
<b>11</b>	300-0.14	25.4 (16.7)	13.5 (1.2)

**Figure 7.6:** Schistosomula viability dose-response curves for compounds **7-11**. Viability (%) is displayed per compound concentration ( $\log_{10}$ ,  $\mu\text{M}$ ). Drug exposure: blue: 1 day, red: 7 days. Dots and triangles represent individual data points per drug concentration (triplicates) from one exemplary assay. Lines represent the modelled curve fit. MQ (mefloquine), AU (auranofin).

stability, with the least reactive cyclohexane analogue **11** being the most potent inhibitor at 25.4 and 13.5  $\mu\text{M}$  after 3 and 7 days respectively. An exception to this trend is the most reactive inhibitor (*p*-NO<sub>2</sub>) substituted warhead, compound **7**, which was equipotent with the least reactive inhibitor (cyclohexane analogue **11**) at 29.3 and 14.7  $\mu\text{M}$  after 3 and 7 days respectively. Considering the anticipated stability under these assay conditions (pH 7.3), when combined with earlier stability tests (section 3.5), compound **7** was expected to decompose rapidly under these conditions. This led to the belief that the thiosulfonate analogues must be fast acting, with a rate capable of overcoming the competing pH dependent hydrolysis of compound **7**. Combining this with the earlier modelling studies, which suggested the *p*-NO<sub>2</sub> may be able to anchor the inhibitors by coordinating His-180 and His-181 of the occluding loop, may explain this unanticipated increase in the observed potency, despite its low stability.

With the previous finding that arginine analogues were not effective in targeting *S. mansoni* in this system, future inhibitor constructs focused only on the homophenylalanine analogues. These represent the benzylic warhead **12** and the secondary thiosulfonates **127** and **128**. It was hypothesised that the correlation between increased warhead stability and inhibitor potency would continue, as such the cyclohexane derived warhead (inhibitor **11**) was anticipated to be more potent than the benzylic warhead (inhibitor **12**) and the secondary thiosulfonates (**127** and **128**) were anticipated to be more potent than their primary thiosulfonate counterparts (**11** and **12**). As testing was taking place on a different batch of *S. mansoni*, a number of months apart, the previous most potent thiosulfonate inhibitor **11** was run alongside K11777 (**6**) as a control to account for any potential biological variability in the assay. In agreement with the hypothesis based on stability arguments, the secondary thiosulfonate warheads were found to be more potent than their primary thiosulfonate counterparts, with approximately a 2 fold increase observed in this assay (figure 7.7, comparing compound **11** and **12** with **127** and **128** respectively). A marked improvement from the previous most potent thiosulfonate inhibitors was observed. One interesting finding was that, in this assay, the previous most potent thiosulfonate (**11**) offered reduced potency of 43.4 and 38.9  $\mu\text{M}$  over 3 and 7 days respectively when compared with the previous assay (at 29.3 and 14.7  $\mu\text{M}$  respectively). This was considered to be consistent with biological variation, hence the importance of inclusion of this as a control compound for comparison between assays. Comparing this with the secondary thiosulfonate (**127**), derived from the same cyclohexane warhead, showed a considerable improvement to 15.9 and 14.1  $\mu\text{M}$  over 3 and 7 days respectively. A similar trend was observed for the benzylic warhead (**12**) with an EC<sub>50</sub> of 43.4 and 26.3  $\mu\text{M}$  over 3 and 7 days respectively, which was equipotent with the previous best primary thiosulfonate (cyclohexane derivative **11**). When moving to the secondary benzylic thiosulfonate derivative **128** there was again a significant improvement in EC<sub>50</sub> to 22.0 and 13.2  $\mu\text{M}$  over 3 and 7



Compound number	Concentration range ( $\mu\text{M}$ ) tested	3 Day Mean $\text{IC}_{50}$ (SD) ( $\mu\text{M}$ )	7 Day Mean $\text{IC}_{50}$ (SD) ( $\mu\text{M}$ )
MQ	200-0.27	3.3 (0.5)	1.1 (0.2)
AU	10-0.014	0.3 (0.1)	0.2 (0.1)
K11777 <b>6</b>	100-0.14	26.7 (10.1)	10.8 (1.8)
<b>11</b>	100-0.14	43.4 (1.3)	38.9 (3.8)
<b>127</b>	100-0.14	15.9 (2.6)	14.1 (0.4)
<b>12</b>	100-0.14	43.4 (1.2)	26.3 (10.4)
<b>128</b>	100-0.14	22.0 (6.7)	13.2 (2.6)

**Figure 7.7:** Schistosomula viability dose-response curves for compounds **11** and **12** compared to **127** and **128**. Viability (%) is displayed per compound concentration ( $\log_{10}$ ,  $\mu\text{M}$ ). Drug exposure: blue: 1 day, red: 7 days. Dots and triangles represent individual data points per drug concentration (triplicates) from one exemplary assay. Lines represent the modelled curve fit. MQ (mefloquine), AU (auranofin).

days respectively. Possibly the most interesting find was that, when moving to the secondary thiosulfonate warheads, these were now approximately equipotent with the known vinyl sulfone reference compound K11777 **6**, suggesting the thiosulfonates are well placed as cysteine protease inhibitors for further investigation and application towards treatment of schistosomiasis.

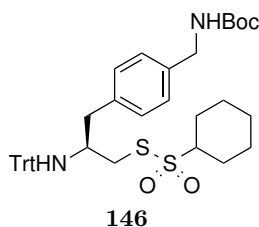
### 7.3 Summary

Modelling studies were used to understand the suitability of thiosulfonate constructs for application towards *S. mansoni* cathepsin B like cysteine protease SmCB1. Thiosulfonate derived inhibitors (**7-17**, **127** and **128**) were tested against larval stage *S. mansoni* and compared against the known cysteine protease inhibitor K11777 (**6**). Gratifyingly, thiosulfonates **7-12**, **127** and **128** were found to be highly active and followed a similar trend between reactivity and potency observed in the previous papain assay test system. It is believed that the increased charge of the arginine derived constructs lead to reduced permeability, correlating with the composition of the parasitic tegument, which in turn greatly decreased the potency of these constructs. Upon considering the assay conditions, namely pH and time, with the earlier explored stability profiles, it was determined that the thiosulfonates must be fast acting inhibitors with long lasting effects. Namely, It was suggested that inhibitor **7** bearing *p*-NO<sub>2</sub> substitution may offer a suitable mimic of the C-terminus, enhancing its target binding and resultant potency to outcompete hydrolysis based on the warheads earlier observed aqueous stability profile. A significant increase in potency, which again correlated with stability, was realised for the secondary thiosulfonate warheads **127** and **128** which were comparable with the activity of the vinyl sulfone reference compound K11777 (**6**).

Moving forwards, these results suggest that the thiosulfonates are well suited for use as cysteine protease inhibitors in targeting schistosomiasis. One potential area for improvement would be to generate arginine mimics with increased permeability, as modelling suggested basic residues would be well accepted in the P<sub>1</sub> position. This may facilitate improved target engagement (K<sub>i</sub>) through rational SAR, whilst accounting for the required improvement in permeability.

## Chapter 8

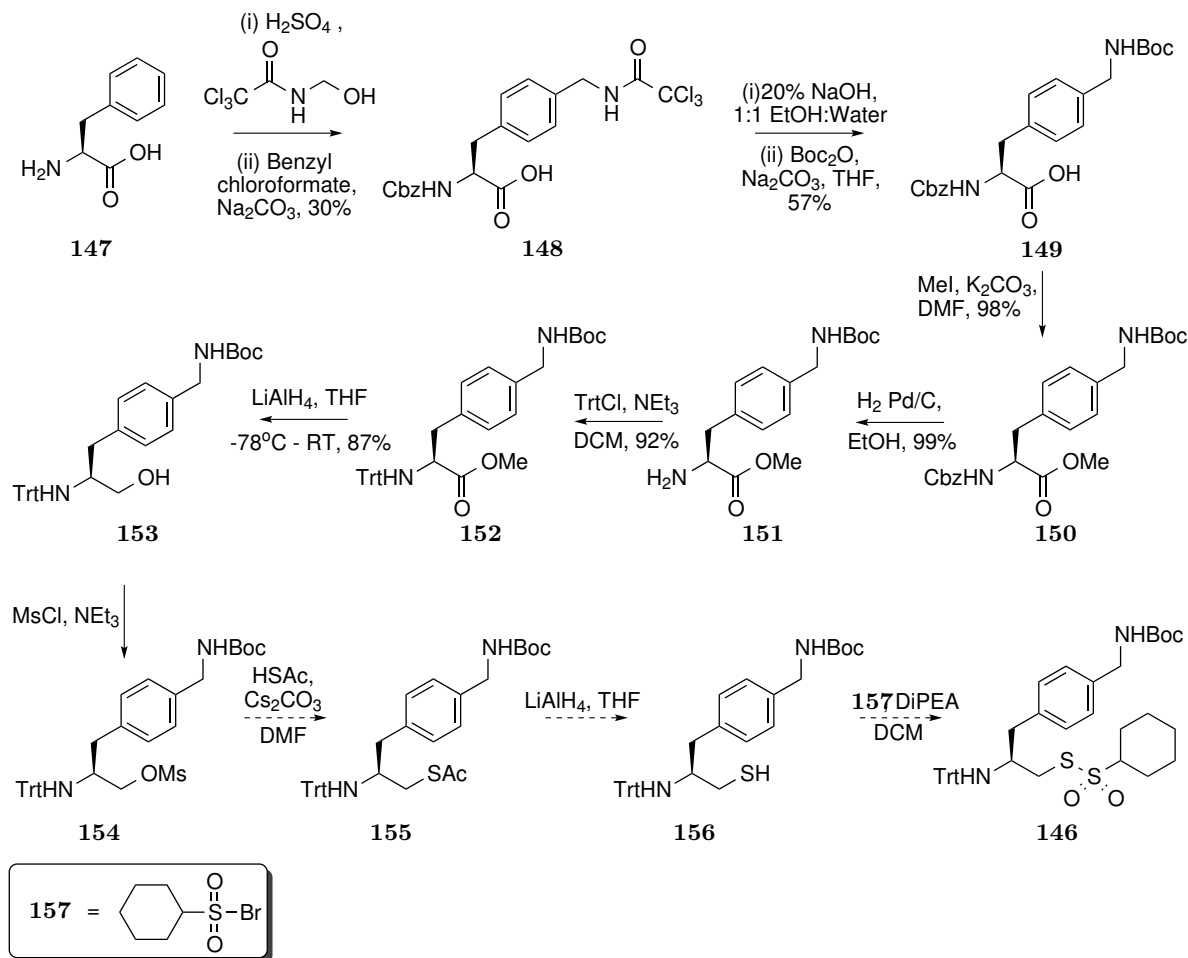
# Strategy to Improve Permeability: Towards a Benzyl Amine-Derived Arginine Mimic



**Figure 8.1:** Warhead building block with a benzyl amine arginine mimic side chain.

Previous work has shown that inhibitors bearing an arginine residue in the P<sub>1</sub> position are highly potent in the papain test system. Both modelling and literature suggested that a similar response could be expected when targeting the cathepsin B like cysteine protease of SmCB1 in schistosomiasis. Interestingly, testing of arginine derived constructs on *S.mansoni* found them to be highly ineffective leading to on to the hypothesis that permeability, rather than selectivity and binding, was reducing the inhibitor potency. As such, this chapter will explore the synthesis towards an arginine mimic with improved permeability. Figure 8.1 shows the key building block for this work, the non-natural amino acid derivative bearing a benzylamine side chain. this building block was chosen based on the work of Geurink *et al.* who successfully exploited the benzylamine side chain as an arginine mimic with improved cell permeability.<sup>220</sup> The amine of the side chain offers a suitable H-bond donor to mimic that of arginine, while the benzylic amine offers a reduced pK<sub>a</sub>, reducing the charge at physiological pH and improving the permeability.

## 8.1 Attempted Synthesis of Benzyl Amine-Derived Warheads

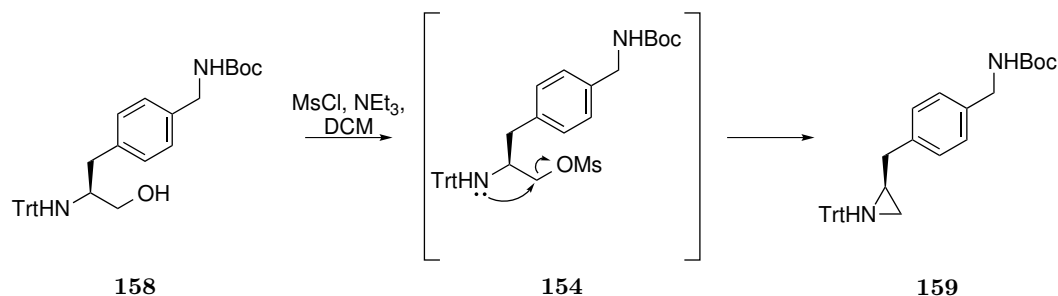


**Scheme 8.1:** Synthesis towards the arginine mimic benzyl amine derived non natural amino acid building block **146**.

The intended synthesis towards the benzylamine-derived non natural amino acid was based largely on the earlier work of Geurink *et al.*, which incorporated this non-natural amino acid into inhibitor constructs for improved permeability.<sup>220</sup> Starting from phenylalanine **147** the para substituted *N*-chloroacetamide moiety was introduced followed by sequential Cbz protection of the N terminal amine to compound **148**. This first step was met with low yields of 30%, mainly due to the difficult separation of the di-substituted by-product (ortho and para substitution) on the aromatic ring. The following hydrolysis of the *N*-trichloroacetamide was followed by direct Boc protection yielding the free carboxylic acid **149** in a moderate yield of 57%. The carboxylic acid was then converted to the corresponding methyl ester for two reasons, (1) to aid purifications as the polar carboxylic acid tends to be poorly resolved by normal phase column chromatography and (2) to allow for a mild reduction to the corresponding alcohol later in the synthesis (as employed previ-



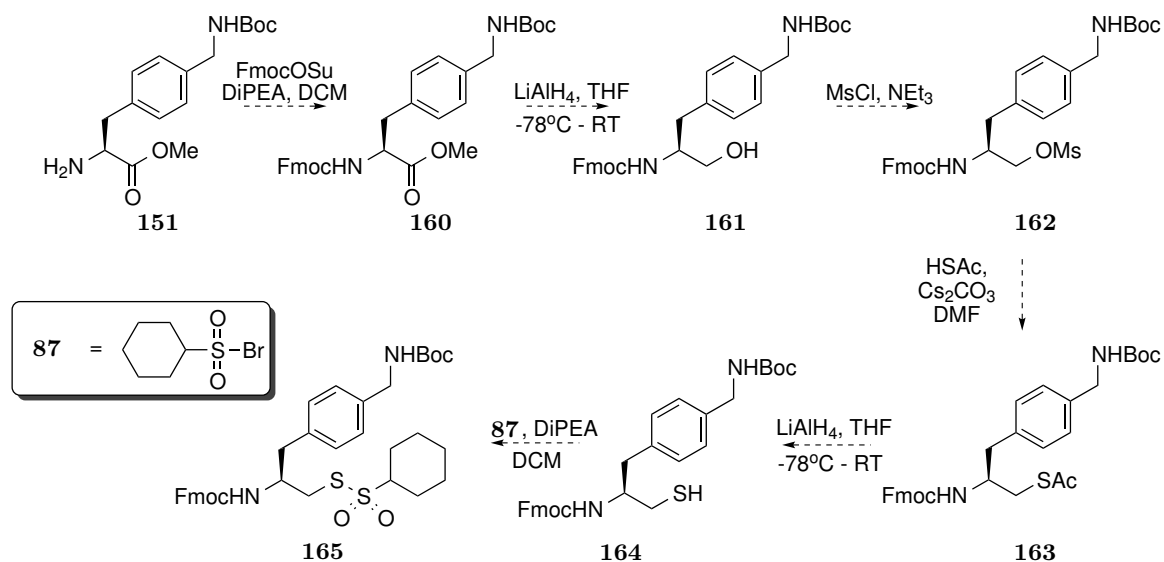
ously in section 3.2.2, scheme 3.3). At this stage the Cbz protecting group was exchanged for the trityl protecting group, as the reducing conditions required for its removal were considered to be incompatible with the final thiosulfonate warhead. This was achieved over two steps, first by hydrogenolysis with H<sub>2</sub> Pd/C (99% yield) followed by protection with TrtCl under basic conditions (92% yield).



**Scheme 8.2:** Intramolecular cyclisation and aziridine formation.

The route forward from the primary alcohol aimed to take advantage of the previously explored synthetic routes towards thiosulfonate warheads. This involved the conversion of the primary alcohol to a mesylate for subsequent substitution with thioacetic acid. Once thioacetate **155** was to hand it was intended to access thiol **156** by reduction rather than hydrolysis as this was found to be the more efficient method, especially when preparing analogues with increasingly bulky side chains such as that of arginine attempted previously. Once the thiol was to hand it was intended to incorporate the cyclohexane thiosulfonate warhead, as this was found to be the most promising warhead construct previously, with sulfonyl bromide **157**. However, as shown in scheme 8.2 this route was not successful, it was found that upon generating mesylate **154** a rapid intramolecular substitution occurred generating aziridine **159**. This is likely due to the available lone pair on the nitrogen accelerated by the aziridine formation placing the steric bulk of the trityl group away from other substituents. A subsequent literature search concluded that the combination of a bulky trityl group and neighbouring mesylate is common for aziridine formation.

With the above aziridine formation in mind, along with the suggestion that the sterics and reactivity of the trityl protected amine accelerated the reaction, a new synthetic route has been proposed in scheme 8.3. The key difference with this approach is to replace the trityl protecting group with a carbamate, Fmoc, which will still allow for orthogonal deprotection whilst delocalising the lone pair of the nitrogen into the carbamate and preventing the intramolecular cyclisation. One potential weakness of this route is that Fmoc deprotection may occur during the final step to form thiosulfonate warhead **165** generating the free amine. This was not considered a significant issue as previous studies have shown the thiosulfonate to be stable in the presence of amines and the subsequent reaction would involve Fmoc removal before coupling to a peptidic backbone. Unfortunately



**Scheme 8.3:** Revised synthesis towards the arginine mimic benzyl amine derived non natural amino acid building block **165**

time constraints prevented this route from being fully explored as it was felt improving cell permeability with the use of non natural amino acids was interesting but not a key milestone towards developing the thiosulfonates as cysteine protease inhibitor. As such, work focused on the other key milestone which has been realised through this work, further controlling the reactivity of the thiosulfonate warhead. However, the above synthesis should be considered and the final compound is likely to be highly interesting for future work.

## 8.2 Summary

The synthesis towards an arginine mimic non natural amino acid building block to be incorporated as a thiosulfonate inhibitor was attempted. This synthetic approach was unsuccessful due to aziridine formation when attempting to generate the key mesylate synthetic intermediate. A new synthesis has been proposed which may overcome this side reaction however time constraints prevented the full exploration of this route. It was determined that increasing the permeability of with arginine mimics may provide interesting future inhibitor constructs but this work was not deemed to be fundamental to the development of the thiosulfonate warhead itself. Rather, focus turned to further controlling the reactivity of the thiosulfonate warheads, which was believed to be of greater significance for the overall development of this new warhead class.

# Chapter 9

## Conclusions

In conclusion, the thiosulfonate moiety has been developed as a novel electrophilic trap to be utilised as cysteine protease inhibitors. A convenient platform for running virtual screening direct from ChemDraw, named *DrawToDock*, was also developed and applied to the thiosulfonate inhibitors. This highlighted the suitability of thiosulfonates as cysteine protease inhibitors, displaying their ability to take advantage of the local hydrogen bonding network of cysteine proteases through hydrogen bonding through the sulfoxide oxygens. This resulted in rational SAR, leading on to the incorporating of arginine in the P<sub>1</sub> position which improved inhibitor potency towards papain, a representative cysteine protease. Stability studies with common nucleophiles, such as amines, thiols and pH dependent hydrolysis, displayed the highly tunable nature of the thiosulfonates. Although some of the first generation inhibitors, namely **7** with an electron withdrawing *p*-NO<sub>2</sub> substituent on the aromatic ring, were found to be unstable towards aqueous hydrolysis. There was a marked difference with inhibitor **10** (*p*-OMe substituted) highlighted the ease with which the reactivity could be controlled. This led on to a second generation of aliphatic inhibitors, with significantly improved stability through reduced leaving group ability of the liberated sulfinate salt. This was achieved by replacing the previous aromatic rings with aliphatic (cyclohexane) or benzylic motifs. Again, it was found that increasing the stability of the thiosulfonate warhead improved the potency.

Encouraged by the highly tunable nature of the thiosulfonates, steric arguments were then introduced, to improve stability by reducing the rate of nucleophilic attack. It was found that generating secondary thiosulfonates, by introduction of a methyl group, significantly improved the warhead stability (even at the relatively high pH 8). It is believed that the thiosulfonates now offer highly suitable warhead constructs to be used as cysteine protease inhibitors, which may find potential use as chemotherapeutics in human diseases such as cancer and osteoporosis. Another potential application is towards neglected tropical dis-

eases, with 8 of the final inhibitor constructs being found to be active towards *S.mansoni*, the parasite responsible for schistosomiasis. At the time of writing, interest in the thiosulfonates is growing and testing (by request) is ongoing against the malaria parasite *P.falciparum*.

From the early stages, where the thiosulfonates were a concept derived from the simple methyl methane thiosulfonate found in plants such as cauliflower and garlic, significant advances have been made. It is believed that the chemoselectivity of the thiosulfonate motif offers a valuable addition to the currently available warheads, moving away from the over-represented Michael acceptors. Future work to further improve upon this may involve the synthesis of tertiary thiosulfonates, as an extension of the steric stability arguments put forward, which may prove to be a powerful new warhead class.

# Chapter 10

## Experimental

### 10.0.1 Materials and Methods

Reagents: All reagents and starting materials (where applicable) were obtained from either Sigma Aldrich<sup>®</sup> or Fluorochem Ltd. of the highest available quality, and utilised without further purification.

Solvents: All solvents were obtained from Fisher Scientific. Where necessary (under strict anhydrous conditions) solvents were obtained from a dry source. All deuterated solvents for NMR were obtained from Cambridge Isotope Laboratories, Inc.

Reactions: Air and moisture sensitive reactions were performed under a nitrogen atmosphere. All glassware used was dried under reduced pressure with heating (250 °C) from a heat gun. All reactions were conducted in septa sealed vessels with equipped with a nitrogen balloon, unless otherwise stated, and stirred with the use of a magnetic stirrer bead. A glass stopper replaced the rubber septum in strongly acidic reaction conditions, such as Boc removal with TFA.

Thin Layer Chromatography (TLC): All TLCs were conducted on aluminium backed TLC plates coated (0.25 mm), with silica gel 60 F<sub>254</sub> obtained from Merck. Compounds were then detected by fluorescent quenching at 254 nm, by UV light from a UV Minerallight<sup>®</sup> lamp. Non-fluorescent quenching substances were visualised by oxidative staining with KMnO<sub>4</sub> (2.00 g in 400 mL water), or Ninhydrin (1.5 g in 100 mL n-Butanol, 3 mL AcOH) for staining amines. Stained TLC plates were developed by heating with a heat gun (250 °C), on a hot plate (250 °C) or a combination of both as required.

Automated column chromatography: was conducted with the Biotage<sup>®</sup> Isolera One<sup>®</sup> automated chromatograph. Products were purified on Biotage<sup>®</sup> SNAP Ultra cartridges pre packed with Biotage<sup>®</sup> HP-sphere<sup>™</sup> Spherical silica. UV absorption was detected

with Biotage<sup>®</sup> Isolera<sup>™</sup> Spektra UV detector at both UV<sub>1</sub> (254 nm) and UV<sub>2</sub> (280 nm) to identify fractions for collection in combination with TLC analysis.

Preparative HPLC: was conducted with the Angilent Technologies 1260 Infinity Preparative-scale Purification system. Separation was achieved on a Phenomenex Gemini<sup>®</sup>, 10  $\mu$  C18 110 Å AXIA, 250 x 21.2 mm. A linear gradient of 5  $\rightarrow$  95% MeCN in Ultra pure water with 0.1% TFA was utilised. All runs were conducted over 80 minutes with a flow rate of 12.5 mL/Min. Fraction collection was based on UV absorption detected at both 214 and 254 nm. LC-MS: was conducted with a Thermo Scientific, Dionex UltiMate 3000 standard LC system coupled to a Thermo Scientific LCQ Fleet<sup>™</sup> Ion trap mass spectrometer. Separations were achieved with a Dr. Maisch GmbH Reprosil Gold 120 C18, 3  $\mu$ m, 150 x 4 mm column with a flow rate of 1 mL/Min. A linear gradient of 5  $\rightarrow$  95% MeCN in Ultra pure water with 0.1% TFA over either 40 min. or 10 min. was utilised and retention times given. UV absorption was detected at 214 nm.

Nuclear Magnetic Resonance (NMR): characterization employed a Bruker<sup>®</sup> 400 MHz and a Bruker<sup>®</sup> 500 MHz spectrometer measuring <sup>1</sup>H, <sup>13</sup>C, COSY and HSQC as required. Chemical shifts are given in parts per million (ppm). Shifts were downfield of a TMS reference ( $\delta$ TMS = 0 ppm) and the resonances of the rest protons of the deuterated solvents served as internal standard.<sup>221</sup> CDCl<sub>3</sub> 7.24 ppm (<sup>1</sup>H-NMR), 77.0 ppm (<sup>13</sup>C-NMR); Methanol-d<sub>4</sub> 3.31 ppm (<sup>1</sup>H-NMR), 49.0 ppm (<sup>13</sup>C-NMR); DMSO-d<sub>6</sub>: 2.49 ppm (<sup>1</sup>H-NMR), 39.5 ppm (<sup>13</sup>C-NMR). <sup>1</sup>H-<sup>1</sup>H COSY experiments and <sup>1</sup>H-<sup>13</sup>C HSQC experiments were also conducted for correct signal assignment as required.

## 10.0.2 General Procedures

(1) General procedure for amino acid derived methyl esters: the Boc protected amino acid (1.00 equiv.) and K<sub>2</sub>CO<sub>3</sub> (3.00 equiv.) were slurried in DMF (3.5 mL/mmol) for 1.5 hrs. under a nitrogen atmosphere. Methyl iodide (3.00 equiv.) was added dropwise and the solution stirred overnight. TLC (0.1% AcOH in a suitable mixture of EtOAc and Pet-ether 40-60) confirmed reaction completion and the solution was evaporated to dryness. The residue was taken up in EtOAc (10 mL/mmol), washed with water (10 mL/mmol), aqueous back extracted with EtOAc (3 $\times$ 10 mL/mmol) The combined organic layers were washed with 5% aq. Na<sub>2</sub>SO<sub>3</sub> (3 $\times$ 20 mL/mmol), brine (50 mL/mmol) and dried over MgSO<sub>4</sub>. Concentration in vacuo yielded the title compound with no further purification required unless otherwise stated.

(2) General procedure for obtaining amino alcohols by methyl ester reduction: The previously formed methyl ester (1.00 equiv.) and LiCl (2.50 equiv.) were stirred in dry THF (3.00 mL/mmol) and cooled to 0°C for 15 min. under a nitrogen atmosphere. NaBH<sub>4</sub>

(2.50 equiv.) was added and stirred for 15 min. followed by the addition of EtOH (4 mL/mmol). The solution was stirred for 1 hr. at 0 °C before removal of the ice bath. TLC (suitable eluent mixture of EtOAc and pet-ether 40-60) confirmed reaction completion after 3 hrs. Water (1.5 mL/mmol) was added followed by addition of sat. NH<sub>4</sub>Cl (2.5 mL/mmol) and EtOAc (10 mL/mmol). The aqueous and organic layers were separated and the aqueous layer was back extracted with EtOAc (3×10 mL/mmol). The combined organic layers were washed with brine (50 mL/mmol) and dried over MgSO<sub>4</sub>. Concentration *in vacuo* yielded the title compound without further purification unless otherwise stated.

(3) General procedure for preparation of mesylates: The previously formed alcohol (1.00 equiv.) was dissolved in DCM (10 mL/mmol) and cooled to 0 °C. Triethylamine (5.00 equiv.) was added followed by MsCl (3.50 equiv.) and stirred overnight. Completion of the reaction was confirmed by TLC (suitable concentration of EtOAc, pet-ether) and the solution washed with 1 M KHSO<sub>4</sub> (2×5 mL/mmol), water (2×15 mL/mmol), brine (30 mL/mmol) and dried over MgSO<sub>4</sub>. Concentration *in vacuo* yielded the title compound with no further purification required unless otherwise stated.

(4) General procedure for preparation of thioacetates by mesylate substitution: CsCO<sub>3</sub> (1.00 equiv.) was suspended in DMF (10mL/mmol) and thioacetic acid (2.00 equiv.) was added under an N<sub>2</sub> atmosphere before stirring for 15 min. In a separate flask the previously formed mesylate (1.00 equiv.) was dissolved in DMF (2 mL/mmol) and added dropwise to the CsCO<sub>3</sub>/HSAc solution before DMF (2 mL/mmol) was used to rinse any remaining mesylate into the CsCO<sub>3</sub>/HSAc solution. The reaction vessel was then covered with aluminium foil and stirred overnight, TLC confirmed completion (suitable mixture of EtOAc and pet-ether 40-60). The reaction mixture was evaporated to dryness and taken up in EtOAc (10 mL/mmol), washed with water (3×mL/mmol), back extracted with EtOAc (4×5 mL/mmol). The combined organic layers were washed with brine (7.5 mL/mmol) and dried over MgSO<sub>4</sub>. Purification by column chromatography (using a suitable gradient of EtOAc and pet-ether 40-60) yielded the title compound.

(5) General Procedure for the preparation of a disulfide from a thioacetate: The previously prepared thioacetate (1.00 equiv.) was dissolved in EtOH (25 mL/mmol) and KOH (3.00 equiv.) was added followed by DMSO (1% by volume) and water (2% by volume). Air was bubbled through the solution for 30 min. before stirring vigorously in an open topped flask overnight. TLC (suitable eluent mixture of EtOAc, pet-ether 40-60) confirmed reaction completion. The thiol formation was always observed as a spot on TLC with a higher R<sub>f</sub> value than the disulfide, which tends to have an R<sub>f</sub> value close to that of the starting material thioacetate. The solution was then evaporated to dryness, taken up in EtOAc (20 mL/mmol), washed with water (3×20 mL/mmol), brine (20 mL/mmol), dried over

MgSO<sub>4</sub> and concentrated in vacuo to yield the title compound which was used without further purification unless otherwise stated.

(6)(a) General procedure for preparation of a thiol from a thioacetate: All solvents used in this procedure were first degassed by bubbling N<sub>2</sub> gas through for 30 min. Under a N<sub>2</sub> atmosphere the previously prepared thioacetate (1.00 equiv.) was dissolved in EtOH (25 mL/mmol) and KOH (3.00 equiv.) was added and stirred overnight. TLC (suitable eluent mixture of EtOAc, pet-ether 40-60) confirmed reaction completion. The solution was then evaporated to dryness on a rotary evaporator fitted with a N<sub>2</sub> balloon to ensure a N<sub>2</sub> atmosphere was maintained upon removal from the rotary evaporator. The resultant residue was taken up in EtOAc (20 mL/mmol), washed with water (3×20 mL/mmol), brine (20 mL/mmol), dried over MgSO<sub>4</sub> and concentrated in vacuo to yield the title compound which was used without further purification.

(6)(b) Under a N<sub>2</sub> atmosphere the previously prepared thioacetate (1.00 equiv.) was dissolved in dry THF (10 mL/mmol), cooled to -78 °C and LiAlH<sub>4</sub> (2 M in THF, 1.00 equiv.) was added. The reaction was stirred for 10 min. before allowing to warm to 0 °C. TLC (suitable eluent mixture of EtOAc, pet-ether 40-60) confirmed reaction completion within 30 min. to the reaction was added EtOAc (2 mL/mmol), water (2 mL/mmol) and sat. NH<sub>4</sub>Cl (2 mL/mmol). The solution was then diluted with EtOAc (20 mL/mmol), washed with 1 M KHSO<sub>4</sub> (20 mL/mmol), brine (20 mL/mmol), dried over MgSO<sub>4</sub> and concentrated *in vacuo* to yield the title compound which was used without further purification.

(7) General procedure for thiosulfonate formation from the disulfide: The previously formed disulfide (1.00 equiv.) was dissolved in DCM (10 mL/mmol) and the respective sodium sulfinate salt (3.20 equiv.) was added followed by iodine (2.00 equiv.). Completion was confirmed by TLC (relevant concentration of EtOAc, pet-ether) after 1.5 hrs. The reaction mixture was diluted with DCM (10 mL/mmol), washed with 1 M Na<sub>2</sub>SO<sub>3</sub> (2×10 mL/mmol), water (20 mL/mmol), back extract aqueous with DCM (10 mL/mmol), combine organic and wash with brine (10 mL/mmol) and dry over MgSO<sub>4</sub>. Purification by column chromatography (relevant gradient of EtOAc, pet-ether) yielded the title the title compound.

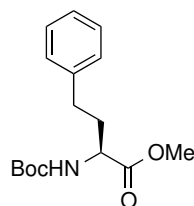
(8) Thiosulfonate formation from Sulfonyl bromide: Cyclohexane sulfonyl bromide, which was prepared by modification of the procedure of Nishiguchi *et al.*<sup>206</sup> (1.2 equiv.) was stirred in DCM (10 mL/mmol) and DiPEA was added (1.2 equiv.). The solution was cooled to 0 °C in an ice bath and the relevant thiol dissolved in DCM (10 mL/mmol) was added dropwise. Reaction completion was confirmed by TLC after 30 min., the solution was diluted with 10 mL/mmol DMF and evaporated to dryness before being taken up in EtOAc (10 mL/mmol) and washed with (3×10 mL/mmol) 1 M KHSO<sub>4</sub>, Brine (1×10 mL/mmol), dried over MgSO<sub>4</sub> and concentrated in vacuo to yield the title compound.



(9) Peptide coupling of the warhead containing derivative to backbone peptides - Method 1: The free amine hydrochloride salt was prepared from the relevant Boc protected amine by stirring in 2 M HCl/dioxane (10 mL/mmol) for 1 hr. and completion confirmed by TLC (15% EtOAc, *n*-Hexane). The solution was then evaporated to dryness and co-evaporated with toluene (2×10 mL/mmol) yielding the unprotected amine as the hydrochloride salt. The amine was then taken up in DCM (10 mL/mmol) and BOP (1.20 equiv.) was added followed by DiPEA (initially 2.00 equiv. with more added as necessary to maintain basic conditions) and reaction progress followed by TLC (10% MeOH in DCM). Coupling was usually complete within 3 hrs. at which point the solution was diluted with EtOAc (100 mL/mmol), washed with water (2×50 mL/mmol), brine (2×50 mL/mmol), dried over MgSO<sub>4</sub> and purified by column chromatography (0→10% MeOH, DCM) yielding the title compounds. Purification by preparative HPLC yielded the title compounds as fluffy white TFA salts after lyophilisation.

(10) Peptide coupling of the warhead containing derivative to backbone peptides - Method 2: The free amine TFA-salt was prepared from the relevant Boc-Arg(Pbf)-warhead by stirring in TFA (10 mL/mmol) with 2% water and 1% TIPS for 2 hr. and completion confirmed by LCMS to ensure both Boc and Pbf removal. The solution was then evaporated to dryness and co-evaporated with toluene (2×10 mL/mmol) yielding the unprotected amine as the bis-TFA salt. The amine was then taken up in DCM (10 mL/mmol) and HCTU (1.20 equiv.) was added followed by DiPEA (initially 2.00 equiv. with more added as necessary to maintain basic conditions) and reaction progress followed by TLC (10% MeOH, DCM). Coupling was usually complete within 3 hrs. at which point the solution was diluted with DMF (10mL/mmol) and evaporated to dryness. The residue was taken up in 0.1% TFA in water and freeze dried before purification by reverse phase column chromatography (5→40% MeCN, water, 0.1% TFA) yielding the title compounds. Further purification by preparative HPLC yielded compounds as fluffy white TFA salts after lyophilisation.

### 10.0.3 Chemistry

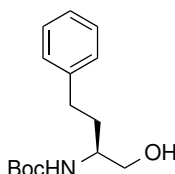


Boc-HomoPhe-OMe (**30**): Literature compound.<sup>222</sup> Methyl ester was prepared according to general procedure 1 on a 1 mmol (279 mg) scale. Yield: 280 mg, 0.955 mmol, 96%.

$^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ )  $\delta$  7.24-7.09 (m, 5H), 5.00 (d,  $J = 8.4$  Hz, 1H), 4.29 (dd,  $J = 12.5, 7.5$  Hz, 1H), 3.64 (s, 3H), 2.65-2.55 (m, 2H), 2.14-2.03 (m, 1H), 1.93-1.82 (m, 1H), 1.38 (s, 9H).

$^{13}\text{C}$  NMR (101 MHz,  $\text{CDCl}_3$ )  $\delta$  173.16, 155.35, 140.77, 128.48, 128.40, 126.15, 79.94, 53.25, 52.27, 34.40, 31.64, 28.33.

LC-MS (Linear gradient 5 $\rightarrow$ 95% MeCN, 0.1% TFA, 10 min): Rt (min): 6.91 (ESI-MS (m/z): 293.66 ( $\text{M}^+$ )).



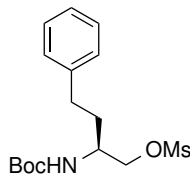
Boc-HomoPhe-CH<sub>2</sub>OH (**31**): Literature compound.<sup>223</sup> Reduction to the alcohol was achieved according to general procedure 2 on a 0.933 mmol (272 mg) scale. Yield: 239 mg, 0.901 mmol, 97%.

$^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ )  $\delta$  7.23-7.07 (m, 5H), 4.73 (d,  $J = 8.0$  Hz, 1H), 3.57 (d,  $J = 10.2$  Hz, 2H), 3.48 (br s, 1H), 2.75 (s, 1H), 2.68-2.53 (m, 2H), 1.82-1.72 (m, 1H), 1.72-1.60 (m, 1H), 1.38 (s, 9H).

$^{13}\text{C}$  NMR (101 MHz,  $\text{CDCl}_3$ )  $\delta$  156.47, 141.55, 128.45, 128.36, 125.98, 79.65, 65.71, 52.49, 33.32, 32.40, 28.42.

LC-MS (Linear gradient 5 $\rightarrow$ 95% MeCN, 0.1% TFA, 10 min): Rt (min): 6.21 (ESI-MS (m/z): 265.75 ( $\text{M}^+$ )).

HRMS: calcd. for  $\text{C}_{15}\text{H}_{23}\text{NO}_3\text{Na}$ , 288.1570 [ $\text{M}+\text{Na}^+$ ]; found 288.1559.



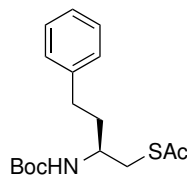
Boc-HomoPhe-CH<sub>2</sub>OMs (**32**): Literature compound.<sup>224</sup> Prepared according to general procedure 3 on a 1 mmol (263 mg) scale. Yield: 335 mg, 0.976 mmol, 98%.

$^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ )  $\delta$  7.35-7.18 (m, 5H), 4.66 (d,  $J = 8.7$  Hz, 1H), 4.29 (dd,  $J = 9.9, 4.0$  Hz, 1H), 4.23 (dd,  $J = 10.1, 4.1$  Hz, 1H), 3.90 (br s, 1H), 3.03 (s, 3H), 2.83-2.64 (m, 2H), 1.99-1.81 (m, 2H), 1.48 (s, 9H).

$^{13}\text{C}$  NMR (101 MHz,  $\text{CDCl}_3$ )  $\delta$  155.30, 140.78, 128.57, 128.37, 126.22, 71.14, 49.42, 37.56, 32.98, 32.09, 28.34.

LC-MS (Linear gradient 5 →95% MeCN, 0.1% TFA, 10 min): Rt (min): 6.59 (ESI-MS (m/z): 366.08 (M+Na<sup>+</sup>)).

HRMS: calcd. for C<sub>16</sub>H<sub>25</sub>NO<sub>5</sub>Na, 366.1346 [M+Na<sup>+</sup>]; found 366.1343.



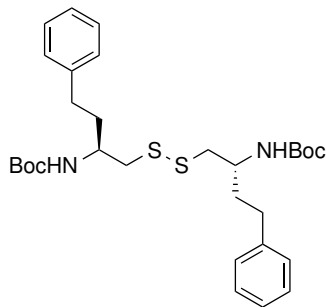
Boc-HomoPhe-CH<sub>2</sub>SAC (**44**): Literature compound.<sup>225</sup> Prepared according to general procedure 4 on a 23.5 mmol (8.07 g) scale. Purified by column chromatography 0 →10% EtOAc, Hexane. Yield: 5.55 g, 17.2 mmol, 73%.

<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 7.24-7.08 (m, 5H), 4.46 (d, *J* = 9.1 Hz, 1H), 3.73 (br s, 1H), 3.05 (dd, *J* = 13.9, 4.8 Hz, 1H), 2.97 (dd, *J* = 13.9, 7.0 Hz, 1H), 2.69-2.53 (m, 2H), 2.28 (s, 3H), 1.81-1.70 (m, 1H), 1.70-1.61 (m, 1H), 1.38 (s, 9H).

<sup>13</sup>C NMR (101 MHz, CDCl<sub>3</sub>) δ 155.58, 141.52, 128.58, 128.49, 126.11, 79.54, 50.45, 36.33, 34.06, 32.49, 30.72, 28.51.

LC-MS (Linear gradient 5 →95% MeCN, 0.1% TFA, 40 min): Rt (min): 22.14 (ESI-MS (m/z): 323.67 (M<sup>+</sup>)).

HRMS: calcd. for C<sub>17</sub>H<sub>25</sub>NO<sub>3</sub>SNa, 346.1447 [M+Na<sup>+</sup>]; found 346.1437.



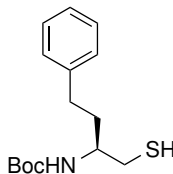
Boc-HomoPhe-Disulfide (**45**): Prepared according to general procedure 5 on a 2.67 mmol (864 mg) scale. Yield: 605 mg, 1.08 mmol, 81%.

<sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>) δ 7.23-7.08 (m, 10H), 4.90 (br s, 2H), 3.81 (br s, 2H), 2.92 (d, *J* = 11.9 Hz, 2H), 2.72 (dd, *J* = 14.2, 6.3 Hz, 2H), 2.65 (dd, *J* = 9.8, 5.0 Hz, 2H), 2.57 (ddd, *J* = 13.8, 9.8, 6.6 Hz, 2H), 1.86 (br s, 2H), 1.70 (br s, 2H), 1.39 (s, 18H).

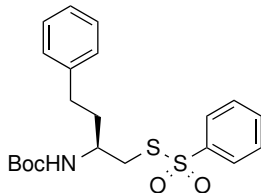
<sup>13</sup>C NMR (101 MHz, CDCl<sub>3</sub>) δ 155.50, 141.46, 128.44, 125.98, 49.96, 45.05, 35.10, 32.44, 28.49.

LC-MS (Linear gradient 5→95% MeCN, 0.1% TFA, 10 min): Rt (min): 8.36 (ESI-MS (m/z): 860.89 (M<sup>+</sup>), 883.20 (M+Na<sup>+</sup>)).

HRMS: calcd. for  $C_{30}H_{44}N_2O_4S_2Na$ , 583.2635  $[M+Na^+]$ ; found 583.2610.



Boc-HomoPhe- $CH_2SH$  (**40**): Prepared according to general procedure (6) on a 1.20 mmol (684 mg) scale. Yield: Quant. Due to air sensitivity the thiol was not characterized and used directly in following reaction.



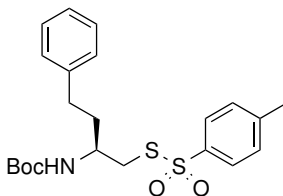
Boc-Homo- $CH_2SSO_2-Ph$  (**34**): Prepared according to general procedure 7 on a 0.196 mmol (110 mg) scale. Purified by automated column chromatography 5→25% EtOAc, pet-ether 40-60. Yield: 90 mg, 0.213 mmol, 54%.

$^1H$  NMR (400 MHz,  $CDCl_3$ )  $\delta$  7.90 (dd,  $J = 8.3, 1.4$  Hz, 2H), 7.60 (tt,  $J = 7.3, 1.3$  Hz, 1H), 7.52 (tt,  $J = 7.5, 1.5$  Hz, 2H), 7.29-7.23 (m, 2H), 7.18 (tt,  $J = 7.4, 1.3$  Hz, 1H), 7.11 (dd,  $J = 8.3, 1.5$  Hz, 2H), 4.67 (d,  $J = 8.7$  Hz, 1H), 3.86-3.74 (m, 1H), 3.22 (dd,  $J = 13.4, 4.9$  Hz, 1H), 3.15 (dd,  $J = 13.1, 6.3$  Hz, 1H), 2.70-2.60 (m, 1H), 2.56 (ddd,  $J = 13.9, 9.2, 7.0$  Hz, 1H), 1.83-1.67 (m, 2H), 1.43 (s, 9H).

$^{13}C$  NMR (101 MHz,  $CDCl_3$ )  $\delta$  155.24, 144.64, 140.91, 133.83, 129.39, 128.56, 128.37, 127.01, 126.18, 79.79, 49.42, 40.91, 35.62, 32.26, 28.41.

LC-MS (Linear gradient 5→95% MeCN, 0.1% TFA, 40 min): Rt (min): 23.52 (ESI-MS (m/z): 421.57 ( $M^+$ ), 444.07( $M+Na^+$ )).

HRMS: calcd. for  $C_{21}H_{27}NO_4S_2Na$ , 444.1274  $[M+Na^+]$ ; found 444.1260.



Boc-HomoPhe- $CH_2SSO_2-(p-Me)Ph$ ] (**35**): Prepared according to general procedure 7 on a 0.212 mmol (119 mg) scale. Purified by automated column chromatography 5→25% EtOAc, pet-ether 40-60. Yield: 157 mg, 0.266 mmol, 63%.

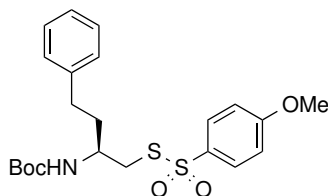
$^1H$  NMR (400 MHz,  $CDCl_3$ )  $\delta$  7.78 (d,  $J = 8.4$  Hz, 2H), 7.32-7.28 (m, 2H), 7.28-7.23 (m,

2H), 7.21-7.14 (m, 1H), 7.13-7.09 (m, 2H), 4.71 (d,  $J = 8.7$  Hz, 1H), 3.87-3.76 (m, 1H), 3.20 (dd,  $J = 13.1, 4.2$  Hz, 1H), 3.14 (dd,  $J = 13.2, 6.2$  Hz, 1H), 2.70-2.60 (m, 1H), 2.56 (ddd,  $J = 13.9, 9.2, 7.0$  Hz, 1H), 2.41 (s, 3H), 1.81-1.68 (m, 2H), 1.43 (s, 9H).

$^{13}\text{C}$  NMR (101 MHz,  $\text{CDCl}_3$ )  $\delta$  155.25, 144.92, 141.80, 140.95, 129.94, 128.51, 128.35, 127.08, 126.13, 79.70, 49.41, 40.79, 35.59, 32.25, 28.39, 21.69.

LC-MS (Linear gradient 5→95% MeCN, 0.1% TFA, 10 min): Rt (min): 7.56 (ESI-MS (m/z): 435.65 ( $\text{M}^+$ ), 458.07( $\text{M}+\text{Na}^+$ )).

HRMS: calcd. for  $\text{C}_{22}\text{H}_{29}\text{NO}_4\text{S}_2\text{Na}$ , 458.1430 [ $\text{M}+\text{Na}^+$ ]; found 458.1417



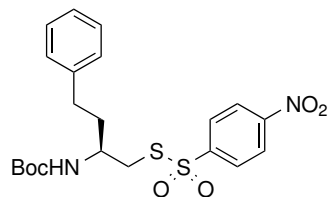
Boc-HomoPhe-SSO<sub>2</sub>-(*p*-OMe)Ph (**36**): Prepared according to general procedure 7 on a 0.212 mmol (119 mg) scale. Purified by automated column chromatography 5→25% EtOAc, pet-ether. Yield: 158 mg, 0.350 mmol, 83%.

$^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ )  $\delta$  7.88-7.80 (m, 2H), 7.31-7.24 (m, 2H), 7.22-7.15 (m, 1H), 7.15-7.10 (m, 2H), 7.01 -6.93 (m, 2H), 4.62 (d,  $J = 8.7$  Hz, 1H), 3.87 (s, 3H), 3.86-3.77 (m, 1H), 3.20 (dd,  $J = 13.6, 4.2$  Hz, 1H), 3.15 (dd,  $J = 13.3, 6.0$  Hz, 1H), 2.72-2.63 (m, 1H), 2.58 (ddd,  $J = 13.9, 9.1, 7.1$  Hz, 1H), 1.86-1.70 (m, 2H), 1.44 (s, 9H).

$^{13}\text{C}$  NMR (101 MHz,  $\text{CDCl}_3$ )  $\delta$  163.75, 155.25, 140.96, 136.28, 129.43, 128.52, 128.35, 126.13, 114.43, 79.69, 55.80, 49.40, 40.75, 35.61, 32.26, 28.39.

LC-MS (Linear gradient 5→95% MeCN, 0.1% TFA, 10 min): Rt (min): 7.39 (ESI-MS (m/z): 451.62 ( $\text{M}^+$ ), 474.06 ( $\text{M}+\text{Na}^+$ )).

HRMS: calcd. for  $\text{C}_{22}\text{H}_{29}\text{NO}_5\text{S}_2\text{Na}$ , 474.1379 [ $\text{M}+\text{Na}^+$ ]; found 474.1366.



Boc-HomoPhe-CH<sub>2</sub>SSO<sub>2</sub>-(*p*-NO<sub>2</sub>)Ph (**37**): Prepared according to general procedure 7 on a 1 mmol (560 mg) scale. Purified by automated column chromatography 5→25% EtOAc, pet-ether. Yield: 520 mg, 1.11 mmol, 56%.

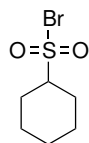
$^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ )  $\delta$  8.34 (d,  $J = 8.9$  Hz, 2H), 8.08 (d,  $J = 8.9$  Hz, 2H), 7.31-7.23 (m, 2H), 7.22-7.16 (m, 1H), 7.14 -7.08 (m, 2H), 4.52 (d,  $J = 8.4$  Hz, 1H), 3.82-3.69 (m,

1H), 3.27 (dd,  $J = 13.6, 5.4$  Hz, 1H), 3.21 (dd,  $J = 13.9, 6.2$  Hz, 1H), 2.68 (ddd,  $J = 14.6, 9.1, 5.9$  Hz, 1H), 2.59 (ddd,  $J = 13.9, 9.0, 7.1$  Hz, 1H), 1.86-1.76 (m, 1H), 1.76-1.66 (m, 1H), 1.43 (s, 9H).

$^{13}\text{C}$  NMR (101 MHz,  $\text{CDCl}_3$ )  $\delta$  155.18, 150.46, 149.72, 140.59, 128.62, 128.35, 128.27, 126.34, 124.72, 80.04, 60.47, 49.38, 41.52, 35.54, 32.18, 28.39, 21.12, 14.28.

LC-MS (Linear gradient 5 $\rightarrow$ 95% MeCN, 0.1% TFA, 10 min): Rt (min): 7.32 (ESI-MS (m/z): 466.31 ( $\text{M}^+$ )).

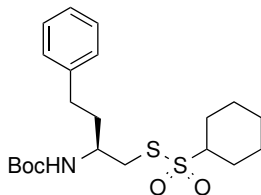
HRMS: calcd. for  $\text{C}_{21}\text{H}_{26}\text{N}_2\text{O}_6\text{S}_2\text{Na}$ , 489.1124 [ $\text{M}+\text{Na}^+$ ]; found 489.1119.



Cyclohexane sulfonyl bromide (**87**) was prepared by adapting the method of Nishiguchi *et al.* for the synthesis of sulfonyl chlorides:<sup>206</sup> N-Bromosuccinimide (7.12g, 40 mmol, 4 equiv) was added to MeCN/2 M HBr (1:1, v/v) at 0 °C. To this was added cyclohexane thiol (1.225 mL, 10 mmol, 1.00 equiv.) and the reaction allowed to stir for 10 minutes. Completion was confirmed by TLC (100% pet ether, 40-60), the reaction was filtered over a silica plug which was rinsed with DCM (10 mL). The title compound was isolated as a slightly yellow oil after evaporation *in vacuo* (2.828 g, quant.)

$^1\text{H}$  NMR (400 MHz, Chloroform- $d$ )  $\delta$  3.50 (tt,  $J = 11.9, 3.6$  Hz, 1H), 2.41 (dddt,  $J = 11.8, 3.8, 2.6, 1.3$  Hz, 2H), 1.99 (ddt,  $J = 12.1, 4.2, 2.4$  Hz, 2H), 1.80-1.62 (m, 3H), 1.47-1.33 (m, 2H), 1.28 (tt,  $J = 12.7, 3.3$  Hz, 1H).

$^{13}\text{C}$  NMR (101 MHz,  $\text{CDCl}_3$ )  $\delta$  78.92, 27.60, 25.01, 24.84.



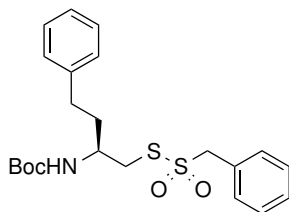
Boc-HomoPhe- $\text{CH}_2\text{SSO}_2$ -cHex] (**75**): Prepared according to general procedure 8 on a 0.500 mmol (141 mg) scale. Purified by column chromatography 3 $\rightarrow$ 30% EtOAc, pet-ether 40-60. Yield: 214 mg, Quant.

$^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ )  $\delta$  7.31-7.25 (m, 2H), 7.22-7.15 (m, 3H), 4.68 (d,  $J = 8.8$  Hz, 1H), 3.90-3.82 (m, 1H), 3.39 (dd,  $J = 13.4, 4.1$  Hz, 1H), 3.18 (dd,  $J = 13.8, 7.0$  Hz, 1H), 3.17-3.11 (m, 1H), 2.77-2.61 (m, 2H), 2.31-2.22 (m, 2H), 1.95-1.87 (m, 2H), 1.87-1.76 (m, 2H), 1.75-1.67 (m, 1H), 1.63-1.49 (m, 2H), 1.46 (s, 9H), 1.37-1.26 (m, 2H), 1.26-1.15 (m,

1H). <sup>13</sup>C NMR (101 MHz, CDCl<sub>3</sub>) δ 155.36, 140.96, 128.53, 128.38, 126.14, 71.37, 50.07, 41.17, 35.88, 32.26, 28.38, 26.36, 26.21, 25.17, 25.11, 25.04.

LC-MS (Linear gradient 5→95% MeCN, 0.1% TFA, 10 min): Rt (min): 7.39 (ESI-MS (m/z): 427.75 (M<sup>+</sup>)).

HRMS: calcd. for C<sub>21</sub>H<sub>33</sub>NO<sub>4</sub>S<sub>2</sub>Na, 450.1743 [M+Na<sup>+</sup>]; found 450.1746.

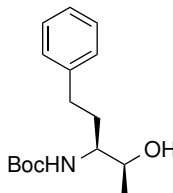


Boc-HomoPhe-CH<sub>2</sub>SSO<sub>2</sub>-MePhe] (**76**): Benzyl sulfonyl chloride (216 mg, 1.13 mmol, 2.26 equiv.) was dissolved in dry DCM (5 mL) and cooled to 0 °C followed by the addition of DiPEA (87 μL, 0.5 mmol, 1.00 equiv.). The solution was stirred for 5 min. before Thiol **40** (0.5 mmol, 1.00 equiv) was taken up in dry DCM (5 mL) and added dropwise. The reaction was allowed to stir for 30 min before a second addition of DiPEA (87 μL, 0.5 mmol, 1.00 equiv.). TLC confirmed reaction completion before diluting with EtOAc (15 mL), washing with 1 M KHSO<sub>4</sub> (2×10 mL), NaHCO<sub>3</sub> (2×10 mL), water (1×10 mL), brine (1×10 mL) and drying over MgSO<sub>4</sub>. The crude product was concentrated *in vacuo* before purification by column chromatography (3→30% EtOAc / pet-ether 40-60) yielding the title compound after concentration *in vacuo* (76 mg, 0.17 mmol, 35%).

<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 7.47-7.43 (m, 2H), 7.41-7.36 (m, 3H), 7.27 (t, *J* = 7.5 Hz, 2H), 7.22-7.16 (m, 1H), 7.14 (d, *J* = 7.6 Hz, 2H), 4.59 (d, *J* = 9.0 Hz, 1H), 4.53 (s, 2H), 3.74 (br s, 1H), 3.04-2.95 (m, 1H), 2.84 (dd, *J* = 13.8, 7.3 Hz, 1H), 2.69-2.52 (m, 2H), 1.80-1.65 (m, 2H), 1.45 (s, 9H).

<sup>13</sup> NMR (101 MHz, CDCl<sub>3</sub>) δ 155.31, 140.91, 131.45, 130.65, 129.44, 129.13, 128.97, 128.85, 128.55, 128.45, 128.37, 127.78, 126.18, 79.86, 68.98, 56.31, 49.90, 41.45, 35.91, 32.19, 28.40.

HRMS: calcd. for C<sub>22</sub>H<sub>29</sub>NO<sub>4</sub>S<sub>2</sub>Na, 458.1430 [M+Na<sup>+</sup>]; found 458.1428.



Boc-HomoPhe-CH(Me)OH (**106**): Methyl ester **30** (5.866 g, 20 mmol, 1.00 equiv.) was dissolved in dry DCM (200 mL) under a N<sub>2</sub> atmosphere and cooled to -78 °C. MeMgBr

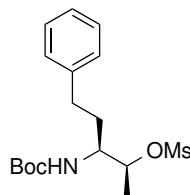
(3 M in diethyl ether, 13.2 mL, 40 mmol, 2.00 equiv.) was added dropwise over 15 minutes followed by the slow addition of DiBAIH (1 M in THF, 40 mL, 40 mmol, 2 equiv.) over 2 hours. A further portion of MeMgBr (3 M in diethyl ether, 20 mL, 60 mmol, 3.00 equiv.) was added dropwise over 10 minutes and the reaction allowed to stir for a further 1 hour before removal from cooling and stirring overnight. The reaction mixture was slowly added to sat. NaHCO<sub>3</sub>, diluted with DCM (200 mL) and washed with Rochelles salt (sat. 400 mL) before separating the layers and back extracting the aqueous with DCM (4×100 mL). The organic layer was dried over MgSO<sub>4</sub>, concentrated *in vacuo* and purified by column chromatography (5→60% EtOAc / pet ether 40-60) yielding the title compound after concentration *in vacuo* (2.558 g, 9.156 mmol, 46%) Product contained 12% anti diastereomer by NMR.

<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 7.31-7.23 (m, 2H), 7.21-7.14 (m, 3H), 4.77 (d, *J* = 9.5 Hz, 1H), 3.79 (br s, 1H), 3.54-3.42 (m, 1H), 2.81-2.56 (m, 2H), 2.32-2.22 (m, 1H), 1.88 -1.76 (m, 2H), 1.46 (s, 9H), 1.17 (d, *J* = 6.3 Hz, 3H).

<sup>13</sup>C NMR (101 MHz, CDCl<sub>3</sub>) δ 156.57, 141.85, 128.48, 128.40, 126.02, 125.91, 79.37, 70.64, 69.70, 58.95, 55.59, 34.39, 32.99, 32.12, 31.82, 28.40, 20.49.

HRMS: calcd. for C<sub>16</sub>H<sub>25</sub>NO<sub>3</sub>Na, 302.1727 [M+Na<sup>+</sup>]; found 302.1719.

[α]<sub>D</sub><sup>20</sup> = -8.3 (*c* 1.00, CHCl<sub>3</sub>)



Boc-HomoPhe-CH(Me)OMs (**105**): Prepared according to general procedure (3) on a 1 mmol (279 mg) scale. Purified by automated column chromatography 10→50% EtOAc, pet-ether 40-60. Yield: 278 mg, 0.778 mmol, 90%.

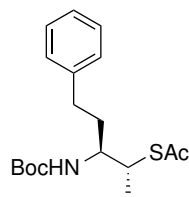
<sup>1</sup>H NMR (400 MHz, cdcl<sub>3</sub>) δ 7.32-7.24 (m, 2H), 7.22-7.13 (m, 3H), 4.81 (qd, *J* = 6.4, 2.5 Hz, 1H), 4.61 (d, *J* = 9.8 Hz, 1H), 3.79-3.68 (m, 1H), 2.99 (s, 3H), 2.78-2.63 (m, 2H), 1.91-1.73 (m, 2H), 1.46 (d, *J* = 1.4 Hz, 9H), 1.41 (d, *J* = 6.4 Hz, 3H).

<sup>13</sup>C NMR (101 MHz, CDCl<sub>3</sub>) δ 155.79, 141.09, 128.53, 128.50, 128.47, 128.37, 126.11, 80.89, 79.75, 53.72, 38.77, 34.10, 32.17, 28.36, 18.10.

HRMS: calcd. for C<sub>17</sub>H<sub>27</sub>NO<sub>5</sub>SNa, 380.1502 [M+Na<sup>+</sup>]; found 380.1499

[α]<sub>D</sub><sup>20</sup> = -14.0 (*c* 1.02, CHCl<sub>3</sub>)





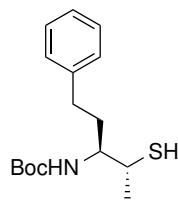
Boc-HomoPhe-CH(Me)SAc (**104**): Mesylate **105** (272 mg, 0.761 mmol, 1 equiv.) was dissolved in dry DMF under a N<sub>2</sub> atmosphere. KSAc (104 mg, 0.913 mmol, 1.2equiv.) was added and the reaction heated to 60 °C. after 1 hour, a further portion of KSAc (156 mg, 1.37 mmol, 1.8 equiv) was added and TLC (20% EtOAc, pet ether 40-60) confirmed reaction completion after a further 2 hours. The reaction mix was evaporated to dryness *in vacuo*, the residue taken up in EtOAc (10 mL), washed with brine (1×10 mL), dried over MgSO<sub>4</sub> and concentrated *in vacuo*. Purification by column chromatography (3→25% EtOAc/ pet ether 40-60) gave yielded the title compound (62 mg, 0.421 mmol, 55%). Note: at this stage chiral resolution was achieved during chromatographicseperation from the other diastereomer (which eluted sooner, with a higher R<sub>f</sub> value).

<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 7.27 (d, *J* = 8.5 Hz, 2H), 7.22-7.15 (m, 3H), 4.52 (d, *J* = 9.7 Hz, 1H), 3.90 – 3.72 (m, 1H), 2.73 (td, *J* = 12.4, 10.4, 5.1 Hz, 1H), 2.68-2.56 (m, 1H), 2.32 (s, 3H), 1.88 (dddd, *J* = 13.9, 10.1, 6.5, 3.4 Hz, 1H), 1.56 (s, 2H), 1.47 (s, 9H), 1.27 (d, *J* = 7.1 Hz, 3H).

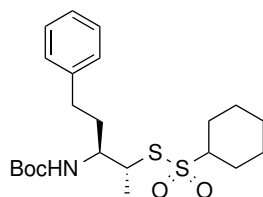
<sup>13</sup>C NMR (101 MHz, CDCl<sub>3</sub>) δ 155.57, 128.45, 128.36, 125.99, 53.93, 44.06, 34.18, 32.67, 30.80, 28.39, 17.68.

HRMS: calcd. for C<sub>18</sub>H<sub>27</sub>NO<sub>3</sub>SNa, 360.1604 [M+Na<sup>+</sup>]; found 360.1602

[α]<sub>D</sub><sup>20</sup> = +17.9 (*c* 1.2, CHCl<sub>3</sub>)



Boc-HomoPhe-CH(Me)SH (**103**): Prepared according to general procedure (6.a) on a 0.174 mmol (59 mg) scale. Yield: Quant. Due to air sensitivity the thiol was not characterized and used directly in following reaction.



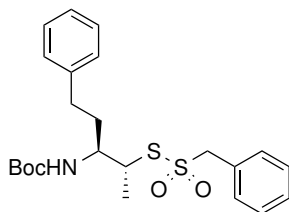
Boc-HomoPhe-CH(Me)SSO<sub>2</sub>cHex (**100**): Prepared according to general procedure 8 on a 0.154 mmol (52 mg) scale. Purified by automated column chromatography 3→30% EtOAc, pet-ether 40-60. Yield: 26 mg, 0.059 mmol, 38%.

<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 7.32-7.25 (m, 2H), 7.23-7.14 (m, 3H), 4.70 (d, *J* = 9.6 Hz, 1H), 3.89 (dq, *J* = 10.0, 5.2, 3.5 Hz, 1H), 3.73 (qd, *J* = 7.2, 3.2 Hz, 1H), 3.14 (tt, *J* = 12.2, 3.5 Hz, 1H), 2.76 (ddd, *J* = 14.7, 9.9, 5.2 Hz, 1H), 2.62 (ddd, *J* = 13.9, 9.5, 6.8 Hz, 1H), 2.34-2.20 (m, 2H), 1.92 (dd, *J* = 12.0, 4.5 Hz, 2H), 1.88-1.80 (m, 1H), 1.75-1.67 (m, 1H), 1.65-1.51 (m, 3H), 1.47 (s, 9H), 1.43 (d, *J* = 7.3 Hz, 3H), 1.38-1.13 (m, 4H).

<sup>13</sup>C NMR (101 MHz, CDCl<sub>3</sub>) δ 155.49, 141.04, 128.53, 128.37, 126.16, 79.81, 71.25, 54.28, 52.41, 33.34, 32.58, 28.40, 26.44, 26.09, 25.24, 25.14, 25.08, 18.94.

HRMS: calcd. for C<sub>22</sub>H<sub>35</sub>NO<sub>4</sub>S<sub>2</sub>Na, 464.1900 [M+Na<sup>+</sup>]; found 464.1894

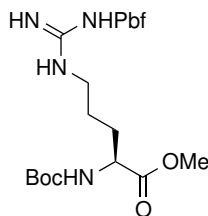
[α]<sub>D</sub><sup>20</sup> = +3.3 (*c* 0.575, CHCl<sub>3</sub>)



Boc-HomoPhe-CH(Me)SSO<sub>2</sub>MePh (**101**): Phenylmethane sulfonyl chloride (67 mg, 0.350 mmol, 2.00 equiv.) was dissolved in dry DCM (2 mL) and cooled to 0 °C. DiPEA was then added (61 μL, 0.350 mmol, 2.00 equiv.) before the addition of freshly prepared thiol **103** (0.175 mmol, 1.00 equiv.). The reaction was allowed to stir for 1 hour before addition of a second portion of phenylmethane sulfonyl chloride (34 mg, 0.175 mmol, 1.00 equiv.) followed by DiPEA (61 μL, 0.350 mmol, 2.00 equiv.). The reaction mix was then diluted with EtOAc (10 mL) before washing with 1 M KHSO<sub>4</sub> (2×10 mL), brine (1×10 mL) and drying over MgSO<sub>4</sub>. Purification by column chromatography (3→30% EtOAc, Pet ether 40-60) yielded the title compound (39 mg, 0.087 mmol, 50%).

HRMS: calcd. for C<sub>23</sub>H<sub>31</sub>NO<sub>4</sub>S<sub>2</sub>Na, 472.1587 [M+Na<sup>+</sup>]; found 472.1589

[α]<sub>D</sub><sup>20</sup> = +9.0 (*c* 0.74, CHCl<sub>3</sub>)



Boc-Arg(Pbf)-OMe (**61**): Prepared according to general procedure 1 on a 10 mmol (5.27 g) scale. Purified by automated column chromatography 20→100% EtOAc, pet-ether 40-60.

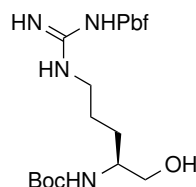
Yield: 5.02 g, 9.29 mmol, 93%.

$^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ )  $\delta$  8.02 (s, 1H), 6.18-6.04 (m, 2H), 5.26 (d,  $J = 8.3$  Hz, 1H), 4.35-4.19 (m, 1H), 3.73 (s, 3H), 3.35-3.25 (m, 1H), 3.23-3.13 (m, 1H), 2.96 (s, 2H), 2.58 (s, 3H), 2.52 (s, 3H), 2.09 (s, 3H), 1.87-1.75 (m, 1H), 1.70-1.55 (m, 3H), 1.46 (s, 6H), 1.42 (s, 9H).

$^{13}\text{C}$  NMR (101 MHz,  $\text{CDCl}_3$ )  $\delta$  173.07, 158.83, 156.29, 156.08, 138.47, 133.05, 132.39, 124.70, 117.59, 86.49, 80.46, 52.81, 52.60, 43.37, 40.85, 30.71, 28.74, 28.45, 25.30, 19.40, 18.03, 12.61.

LC-MS (Linear gradient 5 $\rightarrow$  95% MeCN, 0.1% TFA, 10 min): Rt (min): 6.70 (ESI-MS (m/z): 541.12 ( $\text{M}+\text{H}^+$ )).

HRMS: calcd. for  $\text{C}_{25}\text{H}_{40}\text{N}_4\text{O}_7\text{SNa}$ , 563.2510 [ $\text{M}+\text{Na}^+$ ]; found 563.2501.



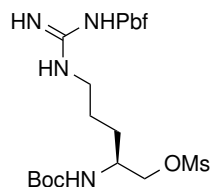
Boc-Arg(Pbf)- $\text{CH}_2\text{OH}$  (**62**): Prepared was achieved according to general procedure 2 on a 9.29 mmol (5.02 g) scale. Yield: 4.72 g, 9.21 mmol, 99%.

$^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ )  $\delta$  6.39-6.12 (m, 3H), 5.16 (d,  $J = 7.9$  Hz, 1H), 3.64-3.51 (m, 3H), 3.31-3.15 (m, 3H), 2.95 (s, 2H), 2.56 (s, 3H), 2.50 (s, 3H), 2.09 (s, 3H), 1.64-1.51 (m, 4H), 1.46 (s, 6H), 1.41 (s, 9H).

$^{13}\text{C}$  NMR (101 MHz,  $\text{CDCl}_3$ )  $\delta$  158.92, 156.80, 156.47, 138.46, 132.88, 132.37, 124.77, 117.67, 86.56, 79.74, 65.12, 43.37, 41.20, 29.03, 28.74, 28.55, 25.75, 19.43, 18.09, 12.62.

LC-MS (Linear gradient 5 $\rightarrow$ 95% MeCN, 0.1% TFA, 10 min): Rt (min): 6.15 (ESI-MS (m/z): 513.08 ( $\text{M}+\text{H}^+$ )).

HRMS: calcd. for  $\text{C}_{24}\text{H}_{40}\text{N}_4\text{O}_6\text{SNa}$ , 535.2561 [ $\text{M}+\text{Na}^+$ ]; found 535.2543.



Boc-Arg(Pbf)- $\text{CH}_2\text{OMs}$  (**63**): Prepared according to general procedure 3 on a 9.21 mmol (4.72 g) scale. Purified by automated column chromatography 20 $\rightarrow$ 100% EtOAc, pet-ether 40-60. Yield: 3.42 g, 5.79 mmol, 63%.

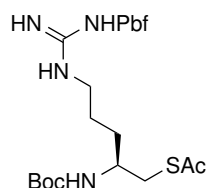
$^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ )  $\delta$  6.20-5.98 (m, 3H), 5.03 (d,  $J = 9.0$  Hz, 1H), 4.23 (dd,  $J =$

10.2, 4.2 Hz, 1H), 4.16 (dd,  $J = 10.2, 4.4$  Hz, 1H), 3.90-3.79 (m, 1H), 3.32-3.15 (m, 2H), 3.04 (s, 3H), 2.96 (s, 2H), 2.57 (s, 3H), 2.51 (s, 3H), 2.10 (s, 3H), 1.70-1.53 (m, 4H), 1.46 (s, 6H), 1.42 (s, 9H).

$^{13}\text{C}$  NMR (101 MHz,  $\text{CDCl}_3$ )  $\delta$  159.07, 156.19, 155.96, 138.53, 132.48, 124.90, 117.74, 86.65, 79.97, 71.52, 49.69, 43.29, 40.95, 37.39, 28.71, 28.67, 28.47, 25.50, 19.42, 18.06, 14.30, 12.59.

LC-MS (Linear gradient 5 $\rightarrow$ 95% MeCN, 0.1% TFA, 10 min): Rt (min): 6.40 (ESI-MS (m/z): 591.08 ( $\text{M}+\text{H}^+$ )).

HRMS: calcd. for  $\text{C}_{25}\text{H}_{42}\text{N}_4\text{O}_8\text{S}_2\text{Na}$ , 613.2336 [ $\text{M}+\text{Na}^+$ ]; found 613.2312.



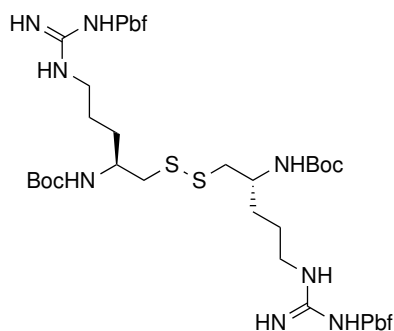
Boc-Arg(Pbf)- $\text{CH}_2\text{SAC}$  (**64**): Prepared according to general procedure 5 on a 5.79 mmol (3.42 g) scale. Purified by automated column chromatography 20 $\rightarrow$ 100% EtOAc, pet-ether 40-60. Yield: 2.98 g, 5.23 mmol, 90%.

$^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ )  $\delta$  6.19-6.02 (m, 3H), 4.71 (d,  $J = 9.2$  Hz, 1H), 3.78-3.64 (m, 1H), 3.32-3.20 (m, 1H), 3.20-3.09 (m, 1H), 2.95 (s, 2H), 3.03-2.86 (m, 2H), 2.58 (s, 3H), 2.52 (s, 3H), 2.35 (s, 3H), 2.09 (s, 3H), 1.63-1.53 (m, 2H), 1.53-1.46 (m, 2H), 1.46 (s, 6H), 1.41 (s, 9H).

$^{13}\text{C}$  NMR (101 MHz,  $\text{CDCl}_3$ )  $\delta$  158.81, 156.52, 156.21, 138.50, 132.44, 124.69, 117.58, 86.47, 80.04, 50.19, 43.39, 41.12, 34.15, 32.44, 30.72, 28.75, 28.52, 25.60, 19.41, 18.04, 12.63.

LC-MS (Linear gradient 5 $\rightarrow$ 95% MeCN, 0.1% TFA, 10 min): Rt (min): 6.74 (ESI-MS (m/z): 571.08 ( $\text{M}+\text{H}^+$ )).

HRMS: calcd. for  $\text{C}_{26}\text{H}_{42}\text{N}_4\text{O}_6\text{S}_2\text{Na}$ , 593.2438 [ $\text{M}+\text{Na}^+$ ]; found 593.2408.



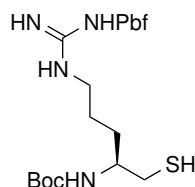
Boc-Arg(Pbf)-Disulfide (**65**): Prepared according to general procedure 5 on a 4.21 mmol (2.40 g) scale. Purified by automated column chromatography 20→100% EtOAc, pet-ether 40-60. Yield: 1.62 g, 1.54 mmol, 73%.

$^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ )  $\delta$  6.47-6.18 (m, 6H), 3.91-3.77 (m, 2H), 3.40-3.13 (m, 4H), 2.96 (s, 4H), 3.04-2.81 (m, 4H), 2.57 (s, 6H), 2.51 (s, 6H), 2.10 (s, 6H), 1.76-1.50 (m, 8H), 1.46 (s, 12H), 1.41 (s, 18H).

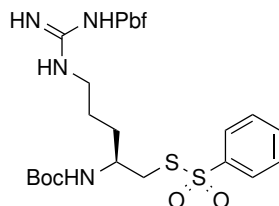
$^{13}\text{C}$  NMR (101 MHz,  $\text{CDCl}_3$ )  $\delta$  158.90, 156.76, 156.69, 138.45, 132.98, 132.42, 124.74, 117.65, 86.52, 79.65, 50.68, 43.42, 41.01, 31.53, 28.76, 28.64, 26.04, 19.49, 18.11, 12.65.

LC-MS (Linear gradient 5→95% MeCN, 0.1% TFA, 10 min): Rt (min): 7.62 (ESI-MS (m/z): 1055.25 ( $\text{M}+\text{H}^+$ )).

HRMS: calcd. for  $\text{C}_{48}\text{H}_{78}\text{N}_8\text{O}_{10}\text{S}_4\text{Na}$ , 1077.4616 [ $\text{M}+\text{Na}^+$ ]; found 1077.4598.



Boc-Arg(Pbf)- $\text{CH}_2\text{SH}$  (**85**): Prepared according to general procedure 7 on a 3.285mmol scale. Yield: 94%. Thiol was used directly in following reaction due to air sensitivity.



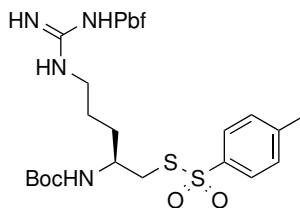
Boc-Arg(Pbf)- $\text{CH}_2\text{SSO}_2\text{-Ph}$ ] (**66**): Prepared according to general procedure 7 0.331 mmol (349 mg) scale. Purified by automated column chromatography 20→80% EtOAc, pet-ether 40-60. Yield: 142 mg, 0.212 mmol, 32%.

$^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ )  $\delta$  7.94 -7.89 (m, 2H), 7.65-7.59 (m, 1H), 7.58-7.51 (m, 2H), 6.30-6.02 (m, 3H), 5.01 (d,  $J = 8.8$  Hz, 1H), 3.86-3.66 (m, 1H), 3.23-3.10 (m, 3H), 3.06 (dd,  $J = 13.7, 6.8$  Hz, 1H), 2.95 (s, 2H), 2.56 (s, 3H), 2.50 (s, 3H), 2.09 (s, 3H), 1.63-1.50 (m, 2H), 1.45 (s, 6H), 1.50 - 1.37 (m, 2H), 1.40 (s, 9H).

$^{13}\text{C}$  NMR (101 MHz,  $\text{CDCl}_3$ )  $\delta$  158.79, 156.16, 155.81, 144.37, 138.33, 133.90, 132.80, 132.28, 129.42, 127.02, 124.67, 117.52, 86.43, 79.92, 49.37, 43.23, 40.92, 40.82, 31.27, 28.61, 28.36, 25.60, 19.33, 17.97, 12.50.

LC-MS (Linear gradient 5→95% MeCN, 0.1% TFA, 10 min): Rt (min): 7.18 (ESI-MS (m/z): 669.02 ( $\text{M}+\text{H}^+$ )).

HRMS: calcd. for C<sub>30</sub>H<sub>44</sub>N<sub>4</sub>O<sub>7</sub>S<sub>3</sub>Na, 691.2246 [M+Na<sup>+</sup>]; found 691.2235.



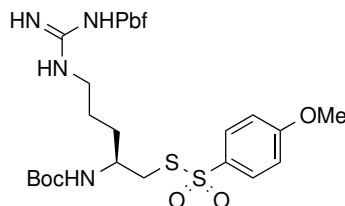
Boc-Arg(Pbf)-CH<sub>2</sub>SSO<sub>2</sub>-(*p*-Me)Ph] (**67**): Prepared according to general procedure 7 on a 0.383 mmol (404 mg) scale. Purified by automated column chromatography 20→80% EtOAc, pet-ether 40-60. Yield: 423 mg, 0.423 mmol, 55%.

<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 7.79 (d, *J* = 8.4 Hz, 2H), 7.33 (d, *J* = 8.1 Hz, 2H), 6.20-6.01 (m, 3H), 4.94 (d, *J* = 9.0 Hz, 1H), 3.83-3.72 (m, 1H), 3.25-3.08 (m, 3H), 3.04 (dd, *J* = 13.7, 6.8 Hz, 1H), 2.95 (s, 2H), 2.57 (s, 3H), 2.51 (s, 3H), 2.44 (s, 3H), 2.09 (s, 3H), 1.63-1.47 (m, 4H), 1.45 (s, 6H), 1.40 (s, 9H).

<sup>13</sup>C NMR (101 MHz, CDCl<sub>3</sub>) δ 158.92, 156.21, 145.25, 141.64, 138.51, 132.47, 130.14, 127.25, 124.78, 117.63, 86.54, 80.16, 49.33, 43.38, 40.98, 40.85, 31.61, 31.60, 28.75, 28.50, 25.65, 21.81, 19.44, 18.08, 12.63.

LC-MS (Linear gradient 5→95% MeCN, 0.1% TFA, 10 min): Rt (min): 7.33 (ESI-MS (*m/z*): 683.01 (M+H<sup>+</sup>)).

HRMS: calcd. for C<sub>31</sub>H<sub>46</sub>N<sub>4</sub>O<sub>7</sub>S<sub>3</sub>Na, 705.2421 [M+Na<sup>+</sup>]; found 705.2402.



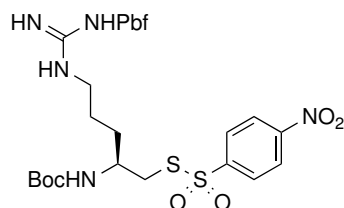
Boc-Arg(Pbf)-CH<sub>2</sub>SSO<sub>2</sub>-(*p*-OMe)Ph] (**68**): Prepared according to general procedure 7 on a 0.319 mmol (336 mg) scale. Purified by automated column chromatography 20→80% EtOAc, pet-ether 40-60. Yield: 142 mg, 0.212 mmol, 33%.

<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 7.85 (d, *J* = 9.0 Hz, 2H), 7.00 (d, *J* = 9.0 Hz, 2H), 6.16-5.97 (m, 3H), 4.92 (d, *J* = 9.0 Hz, 1H), 3.88 (s, 3H), 3.85-3.73 (m, 1H), 3.29-3.08 (m, 3H), 3.04 (dd, *J* = 13.8, 6.7 Hz, 1H), 2.95 (s, 2H), 2.57 (s, 3H), 2.51 (s, 3H), 2.09 (s, 3H), 1.64-1.48 (m, 4H), 1.45 (s, 6H), 1.42 (s, 9H).

<sup>13</sup>C NMR (101 MHz, CDCl<sub>3</sub>) δ 163.87, 158.77, 156.02, 138.38, 132.35, 129.49, 124.63, 117.49, 114.50, 86.40, 80.05, 55.81, 43.24, 40.86, 40.64, 31.52, 28.61, 28.36, 25.47, 19.29, 17.93, 12.48.

LC-MS (Linear gradient 5→95% MeCN, 0.1% TFA, 10 min): Rt (min): 7.18 (ESI-MS (m/z): 699.01 (M+H<sup>+</sup>)).

HRMS: calcd. for C<sub>31</sub>H<sub>46</sub>N<sub>4</sub>O<sub>8</sub>S<sub>3</sub>Na, 721.2370 [M+Na<sup>+</sup>]; found 721.2341.



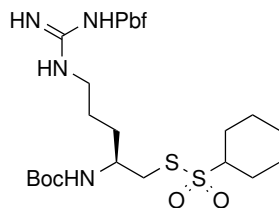
Boc-Arg(Pbf)-CH<sub>2</sub>SSO<sub>2</sub>-(*p*-NO<sub>2</sub>)Ph] (**69**): Prepared according to general procedure 7 0.337 mmol (355 mg) scale. Purified by automated column chromatography 20→80% EtOAc, pet-ether 40-60. Yield: 172 mg, 0.241 mmol, 36%.

<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 8.34 (d, *J* = 8.9 Hz, 2H), 8.11 (d, *J* = 8.5 Hz, 2H), 6.32-6.04 (m, 3H), 5.10 (d, *J* = 8.9 Hz, 1H), 3.83-3.71 (m, 1H), 3.24-3.16 (m, 3H), 3.12 (dd, *J* = 13.6, 7.1 Hz, 1H), 2.96 (s, 2H), 2.55 (s, 3H), 2.49 (s, 3H), 2.09 (s, 3H), 1.63-1.48 (m, 4H), 1.46 (s, 6H), 1.38 (s, 9H).

<sup>13</sup>C NMR (101 MHz, CDCl<sub>3</sub>) δ 158.96, 156.15, 155.77, 150.46, 149.41, 138.32, 132.44, 132.28, 128.34, 124.82, 124.71, 117.65, 86.57, 80.02, 49.51, 43.19, 41.28, 40.80, 31.21, 28.60, 28.34, 25.69, 19.32, 17.97, 12.49.

LC-MS (Linear gradient 5→95% MeCN, 0.1% TFA, 10 min): Rt (min): 7.21 (ESI-MS (m/z): 713.99 (M<sup>+</sup>)).

HRMS: calcd. for C<sub>30</sub>H<sub>43</sub>N<sub>5</sub>O<sub>9</sub>S<sub>3</sub>Na, 736.2115 [M+Na<sup>+</sup>]; found 736.2095.



Boc-Arg(Pbf)-CH<sub>2</sub>SSO<sub>2</sub>-cHex] (**86**): Prepared according to general procedure 8 0.500 mmol (264 mg) scale. Purified by automated column chromatography 20→100% EtOAc, pet-ether 40-60. Yield: 96%.

<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 6.28-6.06 (m, 3H), 5.12 (d, *J* = 8.9 Hz, 1H), 3.83-3.73 (m, 1H), 3.36-3.05 (m, 5H), 2.96 (s, 2H), 2.57 (s, 3H), 2.51 (s, 3H), 2.31-2.22 (m, 2H), 2.09 (s, 3H), 1.97-1.86 (m, 2H), 1.76-1.66 (m, 1H), 1.64-1.49 (m, 6H), 1.46 (s, 6H), 1.42 (s, 9H), 1.37-1.16 (m, 3H).

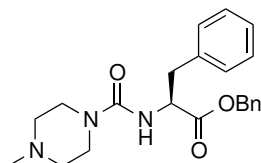
<sup>13</sup>C NMR (101 MHz, CDCl<sub>3</sub>) δ 158.75, 156.23, 156.00, 138.30, 132.87, 132.25, 124.64, 117.49, 86.41, 79.80, 71.27, 50.15, 43.24, 41.09, 40.84, 31.56, 28.62, 28.39, 26.32, 26.17,

25.71, 25.12, 25.06, 25.02, 19.33, 17.98, 12.50.

LC-MS (Linear gradient 5→95% MeCN, 0.1% TFA, 10 min): Rt (min): 7.47 (ESI-MS (m/z): 675.17 (M+H<sup>+</sup>)).

HRMS: calcd. for C<sub>30</sub>H<sub>50</sub>N<sub>4</sub>O<sub>7</sub>S<sub>3</sub>Na, 697.2734 [M+Na<sup>+</sup>]; found 697.2709.

## Backbone Synthesis

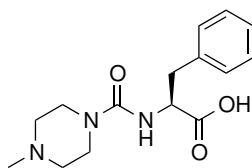


Synthesis of phenyl alanine derived Urea **47** according to the literature procedure.<sup>199</sup> step 1, Isocyanate formation: was achieved from the TFA salt TFA.H<sub>2</sub>N-Phe-OBn (17.2 g, 48.7 mmol, 1.00 equiv.) and reaction progress was followed by TLC (5% MeOH/DCM). Step 2: Urea formation: according to the literature procedure.<sup>199</sup> Reaction progress was followed by TLC (5% MeOH, DCM) and the title compound purified by column chromatography (0→6% MeOH, DCM). Yield: (8.84 g, 23.2 mmol, 48%).

<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 7.35-7.30 (m, 3H), 7.30-7.26 (m, 2H), 7.21-7.17 (m, 3H), 7.02-6.97 (m, 2H), 5.16 (d, *J* = 12.2 Hz, 1H), 5.07 (d, *J* = 12.2 Hz, 1H), 5.00 (d, *J* = 7.6 Hz, 1H), 4.82 (dt, *J* = 7.6, 5.9 Hz, 1H), 3.37-3.31 (m, 4H), 3.09 (dd, *J* = 5.9, 3.1 Hz, 2H), 2.36-2.31 (m, 4H), 2.27 (s, 3H).

<sup>13</sup>C NMR (101 MHz, CDCl<sub>3</sub>) δ 172.51, 156.47, 136.13, 135.23, 129.34, 128.55, 128.47, 128.43, 128.42, 126.91, 67.06, 54.48, 54.38, 45.98, 43.56, 38.19.

LC-MS (Linear gradient 5→95% MeCN, 0.1% TFA, 40 min): Rt (min): 14.63 (ESI-MS (m/z): 382.04 (M+H<sup>+</sup>)).



Carboxylic acid **48** was made according to the literature procedure.<sup>199</sup> Benzyl ester **47** (8.84 g, 23.2 mmol, 1.00 equiv.) was dissolved in 1% AcOH / EtOH (250 mL) and 10wt. % Pd/C (2.3 g) used. TLC (5% MeOH in DCM) confirmed reaction completion. Yield: 7.36 g, 25.3 mmol, quant.

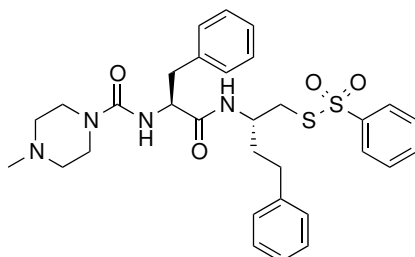
<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 7.18-7.08 (m, 5H), 5.88 (s, 1H), 4.48 (dd, *J* = 12.3, 6.4 Hz, 1H), 3.59-3.42 (m, 2H), 3.42-3.26 (m, 2H), 3.14 (dd, *J* = 13.8, 5.0 Hz, 1H), 2.99 (dd, *J* =



13.8, 7.1 Hz, 1H), 2.73-2.49 (m, 4H), 2.39 (s, 3H).

$^{13}\text{C}$  NMR (101 MHz,  $\text{CDCl}_3$ )  $\delta$  177.12, 156.91, 138.26, 129.71, 128.24, 126.60, 56.12, 53.09, 43.96, 42.02, 38.18.

### Coupling warheads to backbone



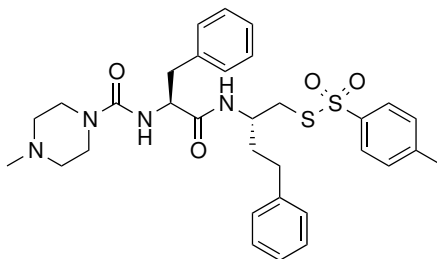
Inhibitor [HomoPhe- $\text{CH}_2\text{SSO}_2\text{-Ph}$ ] (**8**): Prepared according to general procedure 9 on a 0.577 mmol (243 mg) scale. Purified by automated column chromatography 0 $\rightarrow$ 10% MeOH, DCM. Yield: 281 mg, 0.472 mmol, 82%. To ensure high purity for biological testing an 82 mg portion was further purified by preparative HPLC, returning 55 mg of the title compound as a TFA salt after lyophilisation which was > 98% pure by HPLC.

$^1\text{H}$  NMR (400 MHz, Methanol- $\text{d}_4$ )  $\delta$  7.91 (d,  $J = 8.4$  Hz, 2H), 7.70-7.63 (m, 1H), 7.62-7.55 (m, 2H), 7.26 (d,  $J = 4.3$  Hz, 4H), 7.23-7.18 (m, 3H), 7.16-7.13 (m, 1H), 7.12-7.07 (m, 2H), 4.44 (dd,  $J = 9.0, 6.6$  Hz, 1H), 4.21-4.06 (m, 2H), 4.01-3.91 (m, 1H), 3.49-3.36 (m, 2H), 3.19-3.10 (m, 2H), 3.07 (dd,  $J = 13.7, 6.6$  Hz, 1H), 3.00 (dd,  $J = 6.4, 4.2$  Hz, 2H), 2.92 (dd,  $J = 13.7, 9.0$  Hz, 1H), 2.95-2.80 (m, 2H), 2.84 (s, 3H), 2.65-2.56 (m, 1H), 2.50-2.40 (m, 1H), 1.82-1.71 (m, 1H), 1.71-1.60 (m, 1H).

$^{13}\text{C}$  NMR (101 MHz, Methanol- $\text{d}_4$ )  $\delta$  174.63, 158.51, 145.98, 142.54, 138.70, 135.16, 130.66, 130.38, 129.53, 129.46, 129.40, 128.05, 127.82, 126.98, 57.98, 54.10, 49.38, 43.61, 42.38, 41.11, 39.12, 36.22, 33.00.

LC-MS (Linear gradient 5 $\rightarrow$ 95% MeCN, 0.1% TFA, 40 min): Rt (min): 17.06 (ESI-MS (m/z): 595.13 ( $\text{M}+\text{H}^+$ )).

HRMS: calcd. for  $\text{C}_{31}\text{H}_{39}\text{N}_4\text{O}_4\text{S}_2$ , 595.2407 [ $\text{M}+\text{H}^+$ ]; found 595.2397 and calcd. for  $\text{C}_{31}\text{H}_{39}\text{N}_4\text{O}_4\text{S}_2\text{Na}$ , 617.2227 [ $\text{M}+\text{Na}^+$ ]; found 617.2214.



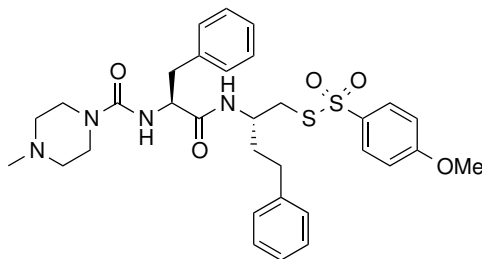
Inhibitor HomoPhe-CH<sub>2</sub>SSO<sub>2</sub>-(*p*-Me)Ph] (**9**): Prepared according to general procedure 9 on a 0.485 mmol (211 mg) scale. Purified by automated column chromatography 0→10% MeOH, DCM. Yield: 297 mg, 0.488 mmol, Quant. To ensure high purity for biological testing a 33 mg portion was further purified by preparative HPLC, returning 13 mg of the title compound as a TFA salt after lyophilisation which was > 99% pure by HPLC.

<sup>1</sup>H NMR (500 MHz, Methanol-d<sub>4</sub>) δ 8.09 (d, *J* = 8.4 Hz, 1H), 7.79 (d, *J* = 8.4 Hz, 2H), 7.40 (d, *J* = 8.1 Hz, 2H), 7.31-7.18 (m, 6H), 7.15 (d, *J* = 7.5 Hz, 1H), 7.13-7.08 (m, 2H), 4.42 (dd, *J* = 9.0, 6.6 Hz, 1H), 4.25-4.04 (m, 2H), 4.01-3.89 (m, 1H), 3.52-3.35 (m, 2H), 3.21-3.07 (m, 2H), 3.03 (dd, *J* = 13.8, 6.7 Hz, 1H), 2.98 (t, *J* = 6.6 Hz, 2H), 3.04-2.92 (m, 2H), 2.90 (dd, *J* = 13.8, 9.0 Hz, 1H), 2.87 (s, 3H), 2.65-2.57 (m, 1H), 2.48-2.43 (m, 1H), 2.42 (s, 3H), 1.81-1.73 (m, 1H), 1.69-1.60 (m, 1H).

<sup>13</sup>C NMR (126 MHz, Methanol-d<sub>4</sub>) δ 174.69, 158.57, 146.63, 143.29, 142.61, 138.71, 131.14, 130.38, 129.55, 129.45, 129.40, 128.20, 127.84, 126.98, 57.95, 54.20, 43.65, 42.44, 41.10, 39.17, 36.25, 33.05, 21.59.

LC-MS (Linear gradient 5→95% MeCN, 0.1% TFA, 40 min): Rt (min): 18.29 (ESI-MS (m/z): 609.10 (M+H<sup>+</sup>)).

HRMS: calcd. for C<sub>32</sub>H<sub>41</sub>N<sub>4</sub>O<sub>4</sub>S<sub>2</sub>, 609.2564 [M+H<sup>+</sup>]; found 609.2544.



Inhibitor HomoPhe-CH<sub>2</sub>SSO<sub>2</sub>-(*p*-OMe)Ph] (**10**): Prepared according to general procedure 9 on a 0.441 mmol (199 mg) scale. Purified by automated column chromatography 0→10% MeOH, DCM. Yield: 315 mg, 0.504 mmol, Quant. To ensure high purity for biological testing an 83 mg portion was further purified by preparative HPLC, returning 60 mg of the title compound as a TFA salt after lyophilisation which was > 99% pure by HPLC.

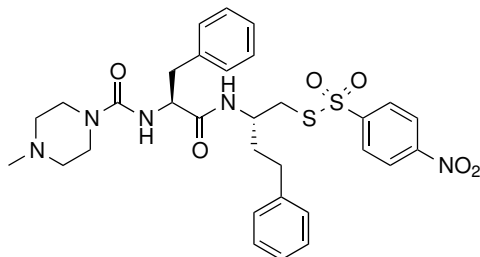
<sup>1</sup>H NMR (400 MHz, Methanol-d<sub>4</sub>) δ 7.84 (d, *J* = 9.0 Hz, 2H), 7.28-7.18 (m, 7H), 7.16-7.05 (m, 5H), 4.43 (dd, *J* = 9.0, 6.5 Hz, 1H), 4.23-4.04 (m, 2H), 4.00-3.91 (m, 1H), 3.86 (s, 3H), 3.50-3.34 (m, 2H), 3.22-3.08 (m, 2H), 3.05 (dd, *J* = 13.7, 6.5 Hz, 1H), 2.98 (t, *J* = 6.2 Hz, 2H), 3.05-2.90 (m, 2H), 2.91 (dd, *J* = 13.6, 9.0 Hz, 1H), 2.85 (s, 3H), 2.65-2.56 (m, 1H), 2.50-2.40 (m, 1H), 1.83-1.72 (m, 1H), 1.71-1.59 (m, 1H).

<sup>13</sup>C NMR (101 MHz, Methanol-d<sub>4</sub>) δ 174.61, 165.47, 158.54, 142.58, 138.72, 137.55, 130.56,

130.38, 129.53, 129.44, 129.39, 127.82, 126.98, 115.65, 57.96, 56.42, 54.13, 43.62, 42.40, 41.02, 39.15, 36.22, 33.04.

LC-MS (Linear gradient 5→95% MeCN, 0.1% TFA, 40 min): Rt (min): 17.33 (ESI-MS (m/z): 625.11 (M+H<sup>+</sup>)).

HRMS: calcd. for C<sub>32</sub>H<sub>41</sub>N<sub>4</sub>O<sub>5</sub>S<sub>2</sub>, 625.2513 [M+H<sup>+</sup>]; found 625.2482 and calcd. for C<sub>32</sub>H<sub>40</sub>N<sub>4</sub>O<sub>4</sub>S<sub>2</sub>Na, 647.2332 [M+Na<sup>+</sup>]; found 647.2305.



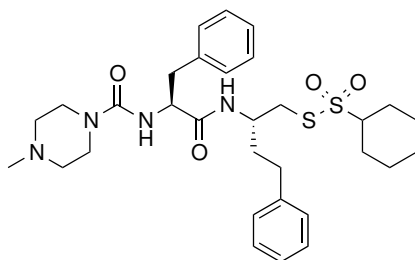
Inhibitor HomoPhe-CH<sub>2</sub>SSO<sub>2</sub>-(*p*-NO<sub>2</sub>)Ph (**7**): Prepared according to general procedure 9 on a 0.204 mmol (95 mg) scale. Purified by automated column chromatography 0→10% MeOH, DCM. Yield: 108 mg, 0.169 mmol, 83%. To ensure high purity for biological testing all 108 mg was further purified by preparative HPLC, returning 36 mg of the title compound as a TFA salt after lyophilisation which was > 98% pure by HPLC.

<sup>1</sup>H NMR (400 MHz, Methanol-d<sub>4</sub>) δ 8.36 (d, *J* = 9.0 Hz, 2H), 8.11 (d, *J* = 8.9 Hz, 2H), 7.32-7.15 (m, 7H), 7.13-7.04 (m, 3H), 4.42 (dd, *J* = 8.8, 6.8 Hz, 1H), 4.23-4.07 (m, 2H), 3.87-3.77 (m, 1H), 3.51-3.36 (m, 2H), 3.22-3.09 (m, 2H), 3.09-3.01 (m, 3H), 3.17-2.86 (m, 2H), 2.92 (dd, *J* = 13.7, 8.8 Hz, 1H), 2.86 (s, 3H), 2.66-2.57 (m, 1H), 2.50-2.39 (m, 1H), 1.80-1.70 (m, 1H), 1.68-1.57 (m, 1H).

<sup>13</sup>C NMR (101 MHz, Methanol-d<sub>4</sub>) δ 174.72, 158.55, 151.94, 151.01, 142.37, 138.65, 130.37, 129.58, 129.47, 129.46, 129.35, 127.87, 127.00, 125.89, 57.93, 54.15, 43.63, 42.42, 41.73, 39.10, 36.14, 32.84.

LC-MS (Linear gradient 5→95% MeCN, 0.1% TFA, 40 min): Rt (min): 17.55 (ESI-MS (m/z): 640.11 (M+H<sup>+</sup>)).

HRMS: calcd. for C<sub>31</sub>H<sub>38</sub>N<sub>5</sub>O<sub>6</sub>S<sub>2</sub>, 640.2258 [M+H<sup>+</sup>]; found 640.2232.



Inhibitor HomoPhe-CH<sub>2</sub>SSO<sub>2</sub>-cHex (**11**): Prepared according to general procedure 9 on

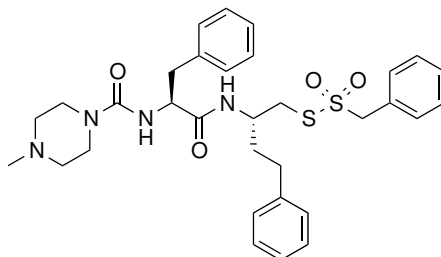
a 0.391 mmol (167 mg) scale. Purified by automated column chromatography 0→10% MeOH, DCM. Yield: 163 mg, 0.271 mmol, 69%. To ensure high purity for biological testing all 163 mg was further purified by preparative HPLC, returning 130 mg of the title compound as a TFA salt after lyophilisation which was > 99% pure by HPLC.

$^1\text{H}$  NMR (400 MHz, Methanol- $d_4$ )  $\delta$  8.15 (d,  $J = 8.5$  Hz, 1H), 7.34-7.10 (m, 10H), 4.48 (dd,  $J = 9.2, 6.2$  Hz, 1H), 4.27-4.08 (m, 2H), 4.08-3.96 (m, 1H), 3.52-3.34 (m, 2H), 3.26 (tt,  $J = 11.7, 3.4$  Hz, 1H), 3.23-3.05 (m, 5H), 3.07-2.86 (m, 2H), 2.96 (dd,  $J = 13.8, 9.3$  Hz, 1H), 2.86 (s, 3H), 2.71 (ddd,  $J = 14.4, 9.3, 5.4$  Hz, 1H), 2.59 (ddd,  $J = 13.7, 9.1, 7.0$  Hz, 1H), 2.26-2.14 (m, 2H), 1.94-1.84 (m, 3H), 1.84-1.74 (m, 1H), 1.73-1.66 (m, 1H), 1.55-1.43 (m, 2H), 1.42-1.27 (m, 2H), 1.22 (tt,  $J = 12.6, 3.2$  Hz, 1H).

$^{13}\text{C}$  NMR (101 MHz, Methanol- $d_4$ )  $\delta$  174.76, 158.59, 142.66, 138.84, 130.37, 129.57, 129.55, 129.43, 127.80, 127.03, 71.85, 58.00, 54.17, 50.14, 43.63, 42.44, 41.37, 39.17, 36.65, 33.03, 27.51, 27.45, 26.23, 26.11, 26.07.

LC-MS (Linear gradient 5→95% MeCN, 0.1% TFA, 40 min): Rt (min): 17.97 (ESI-MS (m/z): 601.08 ( $\text{M}+\text{H}^+$ )).

HRMS: calcd. for  $\text{C}_{31}\text{H}_{45}\text{N}_4\text{O}_4\text{S}_2$ , 601.2877 [ $\text{M}+\text{H}^+$ ]; found 601.2859 and calcd. for  $\text{C}_{31}\text{H}_{44}\text{N}_4\text{O}_4\text{S}_2\text{Na}$ , 623.2696 [ $\text{M}+\text{Na}^+$ ]; found 623.2672.



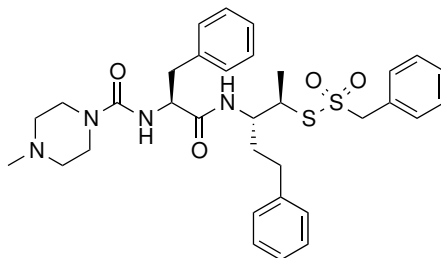
Inhibitor HomoPhe- $\text{CH}_2\text{SSO}_2\text{-CH}_2\text{Ph}$  (**12**): Prepared according to general procedure 9 on a 0.174 mmol (76 mg) scale. Purified by preparative HPLC 5→95% MeCN/Water, 0.1% TFA. Yield: 32 mg, 0.044 mmol, 25%.

$^1\text{H}$  NMR (500 MHz, Methanol- $d_4$ )  $\delta$  8.08 (d,  $J = 9.1$  Hz, 1H), 7.46 (dd,  $J = 8.0, 1.6$  Hz, 2H), 7.36 (tdd,  $J = 8.7, 6.9, 3.7$  Hz, 3H), 7.32-7.23 (m, 6H), 7.23-7.13 (m, 4H), 4.68 (s, 2H), 4.53 (dd,  $J = 9.4, 6.1$  Hz, 1H), 4.17-4.10 (m, 3H), 3.41 (s, 2H), 3.29-3.22 (m, 1H), 3.14 (dd,  $J = 13.9, 6.1$  Hz, 2H), 3.10 (br s, 2H), 2.96 (dd,  $J = 13.9, 9.4$  Hz, 2H), 2.85 (s, 3H), 2.86-2.81 (br s, 2H), 2.72-2.64 (m, 1H), 2.49 (dt,  $J = 13.7, 8.2$  Hz, 1H), 1.74 -1.63 (m, 2H), 1.10 (dd,  $J = 7.2, 1.1$  Hz, 3H).

$^{13}\text{C}$  NMR (126 MHz, MeOD)  $\delta$  173.53, 173.44, 160.77, 160.41, 157.17, 141.26, 137.45, 131.32, 128.98, 128.88, 128.38, 128.33, 128.19, 128.07, 126.45, 125.67, 67.84, 60.68, 56.69, 48.65, 48.56, 42.24, 41.01, 40.19, 37.76, 35.31, 31.63

LC-MS (Linear gradient 5→95% MeCN, 0.1% TFA, 40 min): Rt (min): 17.29 (ESI-MS (m/z): 609.33 (M+H<sup>+</sup>)).

HRMS: calcd. for C<sub>32</sub>H<sub>41</sub>N<sub>4</sub>O<sub>4</sub>S<sub>2</sub>, 609.2564 [M+H<sup>+</sup>]; found 609.2251 and calcd. for C<sub>32</sub>H<sub>40</sub>N<sub>4</sub>O<sub>4</sub>S<sub>2</sub>Na, 631.2383 [M+Na<sup>+</sup>]; found 631.2362



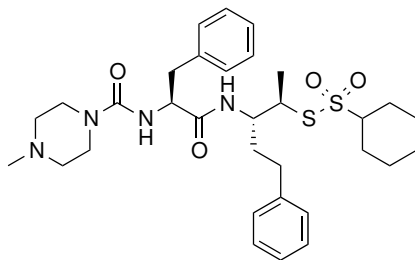
Inhibitor HomoPhe-CH(Me)SSO<sub>2</sub>-CH<sub>2</sub>Ph (**128**): Prepared according to general procedure 9 on a 80 μmol (36 mg) scale. Purified by preparative HPLC 5→95% MeCN/Water, 0.1% TFA. Yield: 25 mg, 34 μmol, 25%.

<sup>1</sup>H NMR (500 MHz, Methanol-d<sub>4</sub>) δ 8.08 (d, *J* = 9.1 Hz, 1H), 7.46 (dd, *J* = 8.0, 1.6 Hz, 2H), 7.36 (tdd, *J* = 8.7, 6.9, 3.7 Hz, 3H), 7.32-7.23 (m, 6H), 7.23-7.13 (m, 4H), 4.68 (s, 2H), 4.53 (dd, *J* = 9.4, 6.1 Hz, 1H), 4.17-4.10 (m, 3H), 3.41 (s, 2H), 3.29-3.22 (m, 1H), 3.14 (dd, *J* = 13.9, 6.1 Hz, 2H), 3.10 (br s, 2H), 2.96 (dd, *J* = 13.9, 9.4 Hz, 2H), 2.85 (s, 3H), 2.86-2.81 (br s, 2H), 2.72-2.64 (m, 1H), 2.49 (dt, *J* = 13.7, 8.2 Hz, 1H), 1.74 -1.63 (m, 2H), 1.10 (dd, *J* = 7.2, 1.1 Hz, 3H).

<sup>13</sup>C NMR (126 MHz, MeOD) δ 173.80, 157.28, 141.29, 137.45, 131.34, 129.01, 128.84, 128.39, 128.28, 128.25, 128.18, 128.04, 126.38, 125.66, 68.37, 56.62, 52.75, 52.63, 51.79, 42.22, 41.05, 37.80, 33.62, 31.86, 15.98.

LC-MS (Linear gradient 5→95% MeCN, 0.1% TFA, 40 min): Rt (min): 17.29 (ESI-MS (m/z): 609.33 (M+H<sup>+</sup>)).

HRMS: calcd. for C<sub>32</sub>H<sub>41</sub>N<sub>4</sub>O<sub>4</sub>S<sub>2</sub>, 609.2564 [M+H<sup>+</sup>]; found 609.2251 and calcd. for C<sub>32</sub>H<sub>40</sub>N<sub>4</sub>O<sub>4</sub>S<sub>2</sub>Na, 631.2383 [M+Na<sup>+</sup>]; found 631.2362

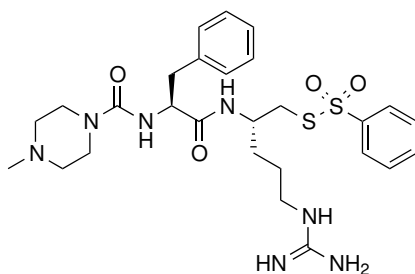


Inhibitor HomoPhe-CH(Me)SSO<sub>2</sub>cHex (**127**): Prepared according to general procedure 9 on a 59 μmol (26 mg) scale. Purified by preparative HPLC 5→95% MeCN/Water, 0.1% TFA. Yield: 25 mg, 25 μmol, 42%.

$^1\text{H}$  NMR (400 MHz, Methanol- $d_4$ )  $\delta$  8.17 (d,  $J = 9.0$  Hz, 1H), 7.39-7.08 (m, 10H), 4.56 (dd,  $J = 9.5, 5.9$  Hz, 1H), 4.29-3.96 (m, 3H), 3.61 (qd,  $J = 7.2, 3.8$  Hz, 1H), 3.42 (br s, 2H), 3.27-3.10 (m, 2H), 3.09 (s, 2H), 3.00-2.88 (m, 2H), 2.86 (s, 3H), 2.74 (ddd,  $J = 13.9, 9.0, 5.0$  Hz, 1H), 2.62-2.51 (m, 1H), 2.28-2.16 (m, 2H), 1.95-1.83 (m, 3H), 1.83-1.73 (m, 1H), 1.49 (tdt,  $J = 12.1, 7.4, 3.7$  Hz, 2H), 1.41-1.31 (m, 2H), 1.29 (dd,  $J = 7.3, 0.9$  Hz, 3H), 1.26-1.11 (m, 1H).

$^{13}\text{C}$  NMR (101 MHz, MeOD)  $\delta$  173.65, 157.31, 141.23, 137.52, 129.00, 128.23, 126.37, 125.65, 70.71, 56.51, 52.83, 52.77, 51.12, 42.23, 41.07, 37.76, 33.30, 31.82, 26.09, 25.96, 24.84, 24.71, 24.66, 16.75.

HRMS: calcd. for  $\text{C}_{32}\text{H}_{46}\text{N}_4\text{O}_4\text{S}_2\text{Na}$ , 637.2853 [ $\text{M}+\text{Na}^+$ ]; found 637.2852.



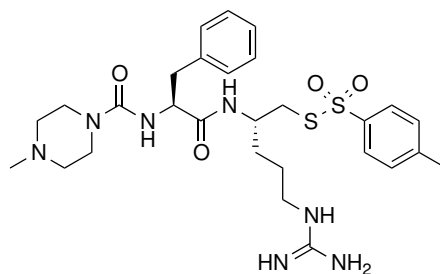
Inhibitor Arg- $\text{CH}_2\text{SSO}_2\text{-Ph}$  (**14**): Prepared according to general procedure 10 on a 0.190 mmol (127 mg) scale. Purification directly by preparative HPLC (Linear gradient 5 $\rightarrow$ 40% MeCN, 0.1% TFA, 80 min) returned the title compound as a TFA salt after lyophilisation which was >99% pure by HPLC: Yield: 131 mg, 0.167 mmol, 88%.

$^1\text{H}$  NMR (400 MHz, Methanol- $d_4$ )  $\delta$  8.05 (d,  $J = 8.7$  Hz, 1H), 7.98-7.93 (m, 2H), 7.77-7.70 (m, 1H), 7.69-7.61 (m, 2H), 7.27-7.15 (m, 5H), 4.33 (dd,  $J = 9.0, 6.6$  Hz, 1H), 4.20-4.03 (m, 2H), 4.06-3.94 (m, 2H), 3.55-3.32 (m, 2H), 3.15-3.01 (m, 5H), 2.97-2.92 (m, 3H), 2.88 (s, 3H), 1.69-1.34 (m, 4H).

$^{13}\text{C}$  NMR (101 MHz, Methanol- $d_4$ )  $\delta$  174.99, 158.64, 158.56, 145.98, 138.64, 135.26, 130.75, 130.34, 129.56, 128.16, 127.88, 58.21, 54.13, 43.63, 42.40, 41.94, 40.93, 39.03, 31.39, 26.23.

LC-MS (Linear gradient 5 $\rightarrow$ 95% MeCN, 0.1% TFA, 40 min): Rt (min): 12.24 (ESI-MS ( $m/z$ ): 590.33 ( $\text{M}+\text{H}^+$ )).

HRMS: calcd. for  $\text{C}_{27}\text{H}_{40}\text{N}_7\text{O}_4\text{S}_2$ , 590.2578 [ $\text{M}+\text{H}^+$ ]; found 590.2557.



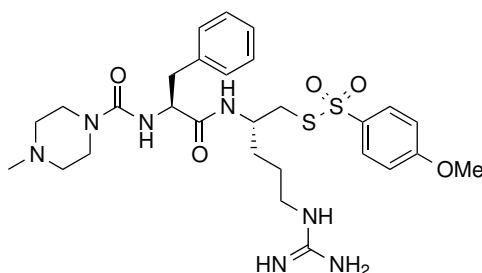
Inhibitor Arg-CH<sub>2</sub>SSO<sub>2</sub>-(*p*-Me)Ph] (**15**): Prepared according to general procedure 10 on a 0.180 mmol (124 mg) scale yielding 254 mg of crude material. Purification of a 44 mg portion of crude material directly by preparative HPLC (Linear gradient 5→40% MeCN, 0.1% TFA, 80 min) gave the title compound as a TFA salt after lyophilisation which was >99% pure by HPLC: Yield: 20 mg, (0.152 mmol, 84%, assuming all crude was to be purified).

<sup>1</sup>H NMR (500 MHz, Methanol-d<sub>4</sub>) δ 8.03 (d, *J* = 8.7 Hz, 1H), 7.82 (d, *J* = 8.4 Hz, 2H), 7.46 (d, *J* = 8.4 Hz, 2H), 7.28-7.17 (m, 5H), 4.33 (dd, *J* = 9.1, 6.6 Hz, 1H), 4.29-3.83 (m, 2H), 4.06-3.94 (m, 1H), 3.59-3.21 (m, 2H), 3.15-3.00 (m, 3H), 2.98-2.89 (m, 3H), 2.88 (s, 3H), 2.45 (s, 3H), 1.69-1.60 (m, 1H), 1.61-1.49 (m, 2H), 1.46-1.38 (m, 1H).

<sup>13</sup>C NMR (126 MHz, Methanol-d<sub>4</sub>) δ 174.98, 158.64, 158.57, 146.77, 143.16, 138.66, 131.20, 130.34, 129.56, 128.27, 127.87, 58.21, 54.13, 43.64, 42.40, 41.95, 40.90, 39.04, 31.43, 26.24, 21.59.

LC-MS (Linear gradient 5→95% MeCN, 0.1% TFA, 40 min): Rt (min): 13.70 (ESI-MS (*m/z*): 604.33 (M+H<sup>+</sup>)).

HRMS: calcd. for C<sub>28</sub>H<sub>42</sub>N<sub>7</sub>O<sub>4</sub>S<sub>2</sub>, 604.2734 [M+H<sup>+</sup>]; found 604.2717.



Inhibitor Arg-CH<sub>2</sub>SSO<sub>2</sub>-(*p*-OMe)Ph] (**16**): Prepared according to general procedure 10 on a 0.094 mmol (66 mg) scale. Purification by preparative HPLC (Linear gradient 5→40% MeCN, 0.1% TFA, 80 min) gave the title compound as a TFA salt after lyophilisation which was >99% pure by HPLC: Yield: 50 mg, 0.61 mmol, 65%.

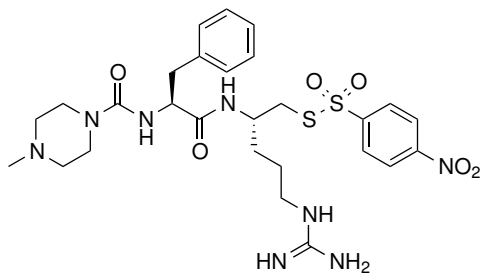
<sup>1</sup>H NMR (400 MHz, Methanol-d<sub>4</sub>) δ 8.05 (d, *J* = 8.7 Hz, 1H), 7.88 (d, *J* = 8.9 Hz, 2H), 7.45-7.18 (m, 5H), 7.13 (d, *J* = 8.9 Hz, 2H), 4.33 (dd, *J* = 9.1, 6.5 Hz, 1H), 4.20-4.05 (m, 2H), 4.03-3.94 (m, 1H), 3.89 (s, 3H), 3.54-3.35 (m, 2H), 3.24-3.00 (m, 5H), 2.99-2.89 (m,

3H), 2.88 (s, 3H), 1.69-1.48 (m, 3H), 1.47-1.37 (m, 1H).

$^{13}\text{C}$  NMR (101 MHz, Methanol- $d_4$ )  $\delta$  175.00, 165.56, 158.61, 158.54, 138.68, 137.43, 130.66, 130.34, 129.54, 127.85, 115.70, 58.27, 56.45, 54.09, 49.02, 43.62, 42.37, 41.93, 40.87, 39.00, 31.46, 26.23.

LC-MS (Linear gradient 5 $\rightarrow$ 95% MeCN, 0.1% TFA, 40 min): Rt (min): 12.88 (ESI-MS (m/z): 620.25 (M+H $^+$ )).

HRMS: calcd. for  $\text{C}_{28}\text{H}_{42}\text{N}_7\text{O}_5\text{S}_2$ , 620.2683 [M+H $^+$ ]; found 620.2656.



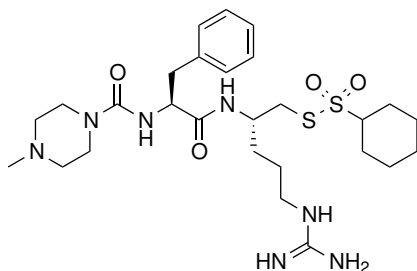
Inhibitor Arg-CH $_2$ SSO $_2$ -(p-NO $_2$ )Ph (**13**): Prepared according to general procedure 10 on a 0.217 mmol (155 mg) scale. Purified by reverse phase automated column chromatography (Linear gradient 5 $\rightarrow$ 40% MeCN, 0.1% TFA, 40 min) using a Biotage<sup>®</sup> SNAP Ultra C18, 60 g column yielding the title compound as a TFA salt after lyophilisation which was >98% pure by HPLC. Yield: 110 mg, 0.132 mmol, 61%.

$^1\text{H}$  NMR (400 MHz, Methanol- $d_4$ )  $\delta$  8.47 (d,  $J$  = 8.5 Hz, 2H), 8.19 (d,  $J$  = 8.5 Hz, 2H), 8.10 (d,  $J$  = 8.7 Hz, 1H), 7.31-7.17 (m, 5H), 4.33 (dd,  $J$  = 9.2, 6.4 Hz, 1H), 4.21-4.02 (m, 2H), 4.00-3.90 (m, 1H), 3.45 (br s, 2H), 3.22-2.98 (m, 7H), 2.95-2.90 (m, 1H), 2.98-2.80 (m, 2H), 2.88 (s, 3H), 1.69-1.40 (m, 4H).

$^{13}\text{C}$  NMR (101 MHz, Methanol- $d_4$ )  $\delta$  175.08, 158.60, 158.54, 152.09, 150.82, 138.65, 130.30, 129.62, 129.56, 127.86, 125.97, 58.21, 54.08, 43.62, 42.36, 41.89, 41.42, 38.93, 31.52, 26.22.

LC-MS (Linear gradient 5 $\rightarrow$ 95% MeCN, 0.1% TFA, 40 min): Rt (min): 13.24 (ESI-MS (m/z): 635.33 (M+H $^+$ )).

HRMS: calcd. for  $\text{C}_{27}\text{H}_{39}\text{N}_8\text{O}_6\text{S}_2$ , 635.2428 [M+H $^+$ ]; found 635.2407.





Inhibitor Arg-CH<sub>2</sub>SSO<sub>2</sub>-cHex (**17**): Prepared according to general procedure 10 on a 0.220 mmol (148 mg) scale. Purification directly by preparative HPLC (Linear gradient 5→40% MeCN, 0.1% TFA, 80 min) returned the title compound as a TFA salt after lyophilisation which was >99% pure by HPLC Yield: 93 mg, 0.118 mmol, 54%.

<sup>1</sup>H NMR (400 MHz, Methanol-d<sub>4</sub>) δ 8.11 (d, *J* = 8.8 Hz, 1H), 7.34-7.19 (m, 5H), 4.40 (dd, *J* = 9.6, 5.9 Hz, 1H), 4.18-4.02 (m, 3H), 3.51-3.38 (m, 2H), 3.32 (tt, *J* = 11.8, 3.4 Hz, 1H), 3.25-3.04 (m, 7H), 2.96 (dd, *J* = 13.8, 9.6 Hz, 1H), 2.87 (s, 3H), 2.29-2.20 (m, 2H), 1.95-1.88 (m, 2H), 1.75-1.59 (m, 4H), 1.59-1.46 (m, 3H), 1.46-1.31 (m, 2H), 1.29-1.19 (m, 1H).

<sup>13</sup>C NMR (101 MHz, Methanol-d<sub>4</sub>) δ 175.17, 158.63, 158.59, 138.84, 130.32, 129.55, 127.81, 71.96, 58.29, 54.10, 50.12, 42.37, 42.00, 41.29, 39.02, 31.87, 27.57, 27.43, 26.32, 26.22, 26.09, 26.05.

LC-MS (Linear gradient 5→95% MeCN, 0.1% TFA, 40 min): Rt (min): 13.65 (ESI-MS (m/z): 596.33 (M+H<sup>+</sup>)).

HRMS: calcd. for C<sub>27</sub>H<sub>46</sub>N<sub>7</sub>O<sub>4</sub>S<sub>2</sub>, 596.3047 [M+H<sup>+</sup>]; found 596.3028.

## 10.0.4 Stability testing

### pH dependent aqueous stability

All stability tests were carried out in 0.1 M sodium phosphate buffer of the corresponding pH with Ac-Phe-OH used as an internal standard. 70 μL of Ac-Phe-OH solution (1 mg/mL, 0.07 mg) in phosphate buffer of the relevant pH was added to 25 μL of inhibitor solution (9.8 mM stock solution in DMSO) and diluted with 880 μL of pH buffer solution giving a final volume of 1 mL with 5% DMSO. The sample was shaken for 10 min. before being centrifuged at 13,200 rpm for 5 min. and transferred to an HPLC vial for sampling. The sample was then analysed over 12 hrs. by HPLC using 12×60 min. gradients from (5→95% MeCN, 0.1% TFA, 60 min.) yielding 12 measurements 1 h apart, a blank sample containing 5% DMSO in phosphate buffer was run after the analysis for subtraction from the baseline. The inhibitor peak was integrated against the internal standard, data normalised to show percentage degradation with time and plotted on a scatter graph for visualisation.

### Stability towards benzyl amine and benzyl mercaptan

approximately 0.3 mg of the thiosulfonate inhibitor for testing was accurately weighed out and dissolved in 1 mL MeCN. To this was added either 10 equivalents of benzyl mercaptan

(for thiol stability testing) or 10 equivalents of benzyl amine (for amine stability testing) in an HPLC vial. The vial was then analysed by HPLC using 13×60 min. gradients from (5→95% MeCN, 0.1% TFA, 60 min.) yielding 13 measurements 1 h apart, a blank sample containing only MeCN was run after the analysis for subtraction from the baseline. The first measurement was defined as time 0. The inhibitor peak area was then taken for each time point and normalised against time 0 to show percentage degradation with time. This was then plotted on a scatter graph for visualisation.

## 10.0.5 Biology

### Papain assay

All tested inhibitors were prepared as described and were >98% pure as determined by analytical HPLC. Papain from papaya latex was purchased as a lyophilised powder from Sigma Aldrich<sup>®</sup>. The substrate N $\alpha$ -benzoyl-L-arginine 4-nitroanilide hydrochloride was purchased from Sigma Aldrich<sup>®</sup>. As assay buffer sodium phosphate (100 mM, pH 6.5) containing EDTA (1.5 mM) was used. Compounds were dissolved in DMSO (for molecular biology grade) which was purchased from Sigma Aldrich<sup>®</sup>. A CLARIOstar microplate reader with Corning<sup>®</sup> 96 well UV- transparent plates was used.

### Method

Each inhibitor was screened by three separate experiments, each conducted in duplicate. The Papain stock solution was prepared in assay buffer (40  $\mu$ M), the solution shaken for 10 min. and then centrifuged at 13,200 rpm for 3 min. The substrate stock solution (Bz-L-Arg-pNA) was prepared in DMSO (50 mM mM). Stock solutions of inhibitors were prepared in DMSO (9.8 mM) and a 1:1 dilution made with buffer solution (4.95 mM), from which relevant serial dilutions with 1:1 DMSO:Buffer solution were made to achieve the desired concentrations of inhibitor solutions for the assay. To each well was added inhibitor solution (4.0  $\mu$ L), buffer solution (172.0  $\mu$ L) and papain solution (20.0  $\mu$ L) followed by thorough mixing (15 times by pipette upon papain addition). For the positive control a DMSO/buffer solution (1:1) was used instead of the inhibitor solution and the negative control (blank) was taken by replacing papain solution with assay buffer. After 1 h incubation with shaking (shaker plate set to 100 rpm) a sample (98.0  $\mu$ L) was taken from each well and added to wells containing substrate solution (2.0  $\mu$ L), Mixed thoroughly by pipette (15 times), the plate covered with a lid and centrifuged at 1000 rpm for 1 min (to remove any air bubbles) and the subsequent liberation of *p*-nitroaniline was measured over a 1h. time frame with 1 measurement per well per minute. Final concentrations in the wells were: enzyme: 4  $\mu$ M; substrate: 1.0 mM; inhibitor: Doubling dilution range starting

at 10  $\mu\text{M}$  giving: 10  $\mu\text{M}$ , 5  $\mu\text{M}$ , 2.5  $\mu\text{M}$ , 1.25  $\mu\text{M}$ , 0.625  $\mu\text{M}$ , 0.3125  $\mu\text{M}$ , 156.25 nM, 78.125 nM, 39.0625 nM.

Processing: As the liberation of *p*-nitroaniline was linear, the response was measured by applying a line of best fit to the blank corrected data and taking the gradient as a measure of the response. The gradients were normalized against the gradient of the positive control (100% response in the absence of inhibitor). The normalized response was used to calculate the  $\text{IC}_{50}$  value by plotting  $\log(\text{inhibitor})$  *vs.* normalized response with the GraphPad Prism software suite.

## **S. mansoni bio-assay**

### **Preparation of schistosomula**

*S. mansoni* life cycle is maintained at the Institute of Tropical Medicine, University Hospital Tübingen/Germany. *S. mansoni* infected vector snails were exposed to day-light to induce shedding of cercariae. Schistosomula were obtained by mechanical transformation of cercariae by vortexing following published procedures.<sup>101</sup> Schistosomula were kept in 48-well plates in schistosomula culture medium (SCM; phenol-red free medium 199 [M199; catalog number 11043-023; Gibco], 5.5 mM d-glucose, 200 U/ml penicillin, 200  $\mu\text{g}/\text{ml}$  streptomycin, 1% heat-inactivated FCS [iFCS]) at 400 schistosomula/1 ml SCM/well for 24 h to allow maturation before being further processed.

### **Schistosomula in vitro drug susceptibility assay**

All compounds were tested in vitro against mature schistosomula. Thiosulfonate compounds were dissolved in DMSO at a stock concentration of 10 mM. Mefloquine hydrochloride (MQ, Sigma-Aldrich) and auranofin (AU, Sigma-Aldrich) were dissolved in DMSO at 36 mM and 14.7 mM, respectively. All compounds were stored at -20 celsius until further use. 96 well, sterile, flat bottom plates were pre-dosed with compounds at respective concentrations. Drug dilutions were done in SCM. Schistosomula were distributed at a density of 100 worms/well. The number of worms was counted for each individual well. Final volume/well was 225  $\mu\text{L}$ . In vitro cultures were kept at 37 °C and 5%  $\text{CO}_2$ . Each experiment included MQ and AU to control for schistosomula inhibition. Worm viability during in vitro culture was controlled using 'medium only' and DMSO (1%), respectively. Viability of worms per drug concentration was analysed by an inverted microscope (Nikon Eclipse Ti) using 40 $\times$  magnification. On the basis of motility and morphology, the parasites were classified as viable (movement and normal appearance) or dead (no movement within 10 s and/or severe morphological changes of any kind compared to the morphology of untreated parasites, e.g., granularity, blebbing). Data analysis: Viability of schistosomula is reported as the proportion of viable schistosomula to the total number of schistosomula

per respective well. Viability in % and log concentration of the drug were used to estimate the  $IC_{50}$  by GraphPad Prim 6 applying the built-in 4 parametric regression analysis to model curve fit.

# Bibliography

- [1] James C Powers, Juliana L Asgian, Özlem Doğan Ekici, and Karen Ellis James. Irreversible Inhibitors of Serine, Cysteine, and Threonine Proteases. *Chemical Reviews*, 102(12):4639–4750, dec 2002.
- [2] Neil D Rawlings and Alan J Barrett. Merops: the peptidase database. *Nucleic acids research*, 28(1):323–325, 2000.
- [3] Vito Turk, Boris Turk, and Dušan Turk. Lysosomal cysteine proteases: facts and opportunities. *The EMBO journal*, 20(17):4629–4633, 2001.
- [4] Neil D Rawlings and Alan J Barrett. Merops: the peptidase database. *Nucleic acids research*, 27(1):325–331, 1999.
- [5] Hans-Hartwig Otto and Tanja Schirmeister. Cysteine proteases and their inhibitors. *Chemical reviews*, 97(1):133–172, 1997.
- [6] Małgorzata Grudkowska and Barbara Zagdanska. Multifunctional role of plant cysteine proteinases. *Acta Biochimica Polonica - English Edition*, pages 609–624, 2004.
- [7] Stephen CJ Cole, A Keith Charnley, and Richard M Cooper. Purification and partial characterisation of a novel trypsin-like cysteine protease from metarhizium anisopliae. *FEMS microbiology letters*, 113(2):189–195, 1993.
- [8] Andrew Hotson and Mary Beth Mudgett. Cysteine proteases in phytopathogenic bacteria: identification of plant targets and activation of innate immunity. *Current opinion in plant biology*, 7(4):384–390, 2004.
- [9] MA Curtis, J Aduse-Opoku, and M Rangarajan. Cysteine proteases of porphyromonas gingivalis. *Critical Reviews in Oral Biology & Medicine*, 12(3):192–216, 2001.
- [10] MJ North, JC Mottram, and GH Coombs. Cysteine proteinases of parasitic protozoa. *Parasitology Today*, 6(8):270–275, 1990.

- [11] Mohammed Sajid and James H McKerrow. Cysteine proteases of parasitic organisms. *Molecular and biochemical parasitology*, 120(1):1–21, 2002.
- [12] Hiroyuki Sorimachi, Shoichi Ishiura, and Koichi Suzuki. Structure and physiological function of calpains. *Biochemical Journal*, 328(3):721–732, 1997.
- [13] Mary E McGrath. The lysosomal cysteine proteases. *Annual review of biophysics and biomolecular structure*, 28(1):181–204, 1999.
- [14] I.G. Kamphuis, K.H. Kalk, M.B.A. Swarte, and J. Drenth. Structure of papain refined at 1.65 resolution. *Journal of molecular biology*, 179(2):233–256, 1984.
- [15] Boris Turk, Dusan Turk, and Guy S Salvesen. Regulating cysteine protease activity: essential role of protease inhibitors as guardians and regulators. *Current pharmaceutical design*, 8(18):1623–1637, 2002.
- [16] Dušan Turk, Gregor Gunčar, Marjetka Podobnik, and Boris Turk. Revised Definition of Substrate Binding Sites of Papain-Like Cysteine Proteases. *Biological Chemistry*, 379(2):137–147, 1998.
- [17] R Coulombe, P Grochulski, J Sivaraman, R Menard, JS Mort, and Miroslaw Cygler. Structure of human procathepsin l reveals the molecular basis of inhibition by the prosegment. *The EMBO journal*, 15(20):5492–5503, 1996.
- [18] Kailash C Pandey, Puran S Sijwali, Ajay Singh, Byoung-Kuk Na, and Philip J Rosenthal. Independent intramolecular mediators of folding, activity, and inhibition for the plasmodium falciparum cysteine protease falcipain-2. *Journal of Biological Chemistry*, 279(5):3484–3491, 2004.
- [19] Kailash C Pandey, David T Barkan, Andrej Sali, and Philip J Rosenthal. Regulatory elements within the prodomain of falcipain-2, a cysteine protease of the malaria parasite plasmodium falciparum. *PloS one*, 4(5):e5694, 2009.
- [20] Kai Tao, Nancy A Stearns, JM Dong, QI Wu, and G Gary Sahagian. The proregion of cathepsin l is required for proper folding, stability, and exit. *Archives of Biochemistry and Biophysics*, 311(1):19–27, 1994.
- [21] John W Cuozzo, Kai Tao, Qi-long Wu, Wen Young, and G Gary Sahagian. Lysine-based structure in the proregion of procathepsin l is the recognition site for mannose phosphorylation. *Journal of Biological Chemistry*, 270(26):15611–15619, 1995.
- [22] Sonia Verma, Rajnikant Dixit, and Kailash C Pandey. Cysteine proteases: modes of activation and future prospects as pharmacological targets. *Frontiers in pharmacology*, 7:107, 2016.

- [23] Amir R Khan and Micael NG James. Molecular mechanisms for the conversion of zymogens to active proteolytic enzymes. *Protein Science*, 7(4):815–836, 1998.
- [24] John P Dalton, Paul J Brindley, Sheila Donnelly, and Mark W Robinson. The enigmatic asparaginyl endopeptidase of helminth parasites. *Trends in parasitology*, 25(2):59–61, 2009.
- [25] Bernd A Schröder, Christian Wrocklage, Andrej Hasilik, and Paul Saftig. The proteome of lysosomes. *Proteomics*, 10(22):4053–4076, 2010.
- [26] John P Dalton and Jan Dvorak. Activating the cathepsin b1 of a parasite: a major route with alternative pathways? *Structure*, 22(12):1696–1698, 2014.
- [27] T Schechter. On the size of the active site in proteases. i. papain. *Biochem. Biophys. Res. Commun.*, 27:157–162, 1967.
- [28] Brian E. Cuthers, Cynthia Barrett, James T. Palmer, and Robert M. Rydzewski. pH dependence of inhibitors targeting the occluding loop of cathepsin B. *Bioorganic Chemistry*, 30(4):264–275, 2002.
- [29] Heidrun Kirschke, Alan J Barrett, and Neil D Rawlings. Proteinases 1: lysosomal cysteine proteinases. *Protein profile*, 2(14):1581–1643, 1995.
- [30] Heidrun Kirschke, Peter Wikstrom, and Elliott Shaw. Active center differences between cathepsins l and b: the s1 binding region. *FEBS letters*, 228(1):128–130, 1988.
- [31] J Barrett and H Kirschke. Cathepsin B, Cathepsin H, and cathepsin L. *Methods in enzymology*, 80 Pt C:535–561, 1981.
- [32] Sarah A Gillmor, Charles S Craik, and Robert J Fletterick. Structural determinants of specificity in the cysteine protease cruzain. *Protein Science*, 6(8):1603–1611, 1997.
- [33] Mary E McGrath, Ann E Eakin, Juan C Engel, James H McKerrow, Charles S Craik, and Robert J Fletterick. The crystal structure of cruzain: a therapeutic target for chagas’ disease. *Journal of molecular biology*, 247(2):251–259, 1995.
- [34] Matthias Schuster, Volker Kasche, and Hans-Dieter Jakubke. Contributions to the s<sup>2</sup>-subsite specificity of papain. *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology*, 1121(1-2):207–212, 1992.
- [35] Robert Ménard, Euridice Carmona, Céline Plouffe, Dieter Brömme, Yasuo Konishi, Jean Lefebvre, and Andrew C Storer. The specificity of the s1’subsite of cysteine proteases. *FEBS letters*, 328(1-2):107–110, 1993.

- [36] Paul J Berti and Andrew C Storer. Alignment/phylogeny of the papain superfamily of cysteine proteases. *Journal of molecular biology*, 246(2):273–283, 1995.
- [37] Boris Turk, Vito Turk, and Dušan Turk. Structural and functional aspects of papain-like cysteine proteinases and their protein inhibitors. *Biological chemistry*, 378(3-4):141–150, 1997.
- [38] Suzanne Afonso, Linda Romagnano, and Bruce Babiarz. The expression and function of cystatin c and cathepsin b and cathepsin l during mouse embryo implantation and placentation. *Development*, 124(17):3415–3425, 1997.
- [39] Yvonne MC Henskens, Enno CI Veerman, and AV Nieuw Amerongen. Cystatins in health and disease. *Biological Chemistry-Hoppe Seyler*, 377(2):71–86, 1996.
- [40] Anders O Grubb. Cystatin c-properties and use as diagnostic marker. *Advances in clinical chemistry*, 35:63–99, 2001.
- [41] R E Esser, R A Angelo, M D Murphey, L M Watts, L P Thornburg, J T Palmer, J W Talhouk, and R E Smith. Cysteine proteinase inhibitors decrease articular cartilage and bone destruction in chronic inflammatory arthritis. *Arthritis & Rheumatism*, 37(2):236–247, 1994.
- [42] P A Hill, D J Buttle, S J Jones, A Boyde, M Murata, J J Reynolds, and M C Meikle. Inhibition of bone resorption by selective inactivators of cysteine proteinases. *J. Cell Biochem.*, 56(1):118–130, 1994.
- [43] J Schedel, C A Seemayer, T Pap, M Neidhart, S Kuchen, B A Michel, R E Gay, U Muller-Ladner, S Gay, and W Zacharias. Targeting cathepsin L (CL) by specific ribozymes decreases CL protein synthesis and cartilage destruction in rheumatoid arthritis. *Gene Therapy*, 11(13):1040–1047, 2004.
- [44] Fabien Lecaille, Jadwiga Kaleta, and Dieter Brömme. Human and parasitic papain-like cysteine proteases: Their role in physiology and pathology and recent developments in inhibitor design. *Chemical Reviews*, 102(12):4459–4488, 2002.
- [45] Gong-Jun Tan, Zheng-Ke Peng, Jin-Ping Lu, and Fa-Qing Tang. Cathepsins mediate tumor metastasis. *World journal of biological chemistry*, 4(4):91–101, 2013.
- [46] Shiqing Yan, M Sameni, and BF Sloane. Cathepsin b and human tumor progression. *Biological chemistry*, 379(2):113–123, 1998.
- [47] WJ Kruszewski, R Rzepko, J Wojtacki, J Skokowski, A Kopacz, K Jaśkiewicz, and K Drucis. Overexpression of cathepsin b correlates with angiogenesis in colon adenocarcinoma. *Neoplasma*, 51(1):38–43, 2004.



- [48] SM Wang, L Li, Wei Zhang, DR Li, and BJ Tang. Relationship between cathepsin b and invasion and metastasis of ovarian carcinoma cells. *Zhonghua fu chan ke za zhi*, 45(8):598–602, 2010.
- [49] Raghu Kalluri and Robert A Weinberg. The basics of epithelial-mesenchymal transition. *The Journal of clinical investigation*, 119(6):1420–1428, 2009.
- [50] Sogu  Coulibaly, Herwig Schwihla, Magnus Abrahamson, Adriana Albini, Christa Cerni, Jason L Clark, Ken M Ng, Nobuhiko Katunuma, Otto Schlappack, Josef Gl ssl, et al. Modulation of invasive properties of murine squamous carcinoma cells by heterologous expression of cathepsin b and cystatin c. *International journal of cancer*, 83(4):526–531, 1999.
- [51] Laurent Dumartin, Hannah J Whiteman, Mark E Weeks, Deepak Hariharan, Branko Dmitrovic, Christine A Iacobuzio-Donahue, Teresa A Brentnall, Mary P Bronner, Roger M Feakins, John F Timms, et al. Agr2 is a novel surface antigen that promotes the dissemination of pancreatic cancer cells through regulation of cathepsins b and d. *Cancer research*, 71(22):7091–7102, 2011.
- [52] Olga Vasiljeva, Anna Papazoglou, Achim Kr ger, Harald Brodoefel, Matvey Korovin, Jan Deussing, Nicole Augustin, Boye S Nielsen, Kasper Almholt, Matthew Bogyo, et al. Tumor cell-derived and macrophage-derived cathepsin b promotes progression and lung metastasis of mammary cancer. *Cancer research*, 66(10):5242–5250, 2006.
- [53] Kai Tao, Jin Li, Jason Warner, Kathy MacLeod, Fred R Miller, and G Gary Sahagian. Multiple lysosomal trafficking phenotypes in metastatic mouse mammary tumor cell lines. *International journal of oncology*, 19(6):1333–1339, 2001.
- [54] Bernadette C Victor, Arulselvi Anbalagan, Mona M Mohamed, Bonnie F Sloane, and Dora Cavallo-Medved. Inhibition of cathepsin b activity attenuates extracellular matrix degradation and inflammatory breast cancer invasion. *Breast Cancer Research*, 13(6):R115, 2011.
- [55] Nimali P Withana, Galia Blum, Mansoureh Sameni, Clare Slaney, Arulselvi Anbalagan, Mary B Olive, Bradley N Bidwell, Laura Edgington, Ling Wang, Kamiar Moin, et al. Cathepsin b inhibition limits bone metastasis in breast cancer. *Cancer research*, 72(5):1199–1209, 2012.
- [56] Claudia D Andl, Kelsey M McCowan, Gillian L Allison, and Anil K Rustgi. Cathepsin b is the driving force of esophageal cell invasion in a fibroblast-dependent manner. *Neoplasia (New York, NY)*, 12(6):485, 2010.

- [57] Vasilena Gocheva, Wei Zeng, Danxia Ke, David Klimstra, Thomas Reinheckel, Christoph Peters, Douglas Hanahan, and Johanna A Joyce. Distinct roles for cysteine cathepsin genes in multistage tumorigenesis. *Genes & development*, 20(5):543–556, 2006.
- [58] Jacqueline M Lankelma, Daniëlle M Voorend, Temo Barwari, Joris Koetsveld, Anne H Van der Spek, Alexander PNA De Porto, Geeske Van Rooijen, and Cornelis JF Van Noorden. Cathepsin l, target in cancer treatment? *Life sciences*, 86(7-8):225–233, 2010.
- [59] Steven L Teitelbaum. Osteoclasts; culprits in inflammatory osteolysis. *Arthritis research & therapy*, 8(1):201, 2005.
- [60] Steven L Teitelbaum. Bone resorption by osteoclasts. *Science*, 289(5484):1504–1508, 2000.
- [61] JM Delaisse, Yves Eeckhout, and Gilbert Vaes. Inhibition of bone resorption in culture by inhibitors of thiol proteinases. *Biochemical Journal*, 192(1):365–368, 1980.
- [62] JM Delaisse, A Boyde, E Maconnachie, NN Ali, CHJ Sear, Yves Eeckhout, Gilbert Vaes, and SJ Jones. The effects of inhibitors of cysteine-proteinases and collagenase on the resorptive activity of isolated osteoclasts. *Bone*, 8(5):305–313, 1987.
- [63] JM Delaisse, Philippe Ledent, and Gilbert Vaes. Collagenolytic cysteine proteinases of bone tissue. cathepsin b,(pro) cathepsin l and a cathepsin l-like 70 kda proteinase. *Biochemical Journal*, 279(1):167–174, 1991.
- [64] Hisao Kakegawa, Takeshi Nikawa, Kahori Tagami, Hiroshi Kamioka, Koji Sumitani, Terushige Kawata, Marinka Drobnič-Kosorok, Brigita Lenarčič, Vito Turk, and Nobuhiko Katunuma. Participation of cathepsin l on bone resorption. *FEBS letters*, 321(2-3):247–250, 1993.
- [65] Tetsuya Inaoka, Graeme Bilbe, Osamu Ishibashi, Ken-ichi Tezuka, Masayoshi Kumegawa, and Toshio Kokubo. Molecular cloning of human cdna for cathepsin k: novel cysteine proteinase predominantly expressed in bone. *Biochemical and biophysical research communications*, 206(1):89–96, 1995.
- [66] A Littlewood-Evans, T Kokubo, O Ishibashi, T Inaoka, B Wlodarski, JA Gallagher, and G Bilbe. Localization of cathepsin k in human osteoclasts by in situ hybridization and immunohistochemistry. *Bone*, 20(2):81–86, 1997.
- [67] Patrick Garnero, Olivier Borel, Inger Byrjalsen, Mercedes Ferreras, Fred H Drake, Michael S McQueney, Niels T Foged, Pierre D Delmas, and Jean-Marie Delaissé.

- The collagenolytic activity of cathepsin k is unique among mammalian proteinases. *Journal of Biological Chemistry*, 273(48):32347–32352, 1998.
- [68] Wa'el Kafienah, Dieter Brömme, David J Buttle, Lisa J Croucher, and Anthony P Hollander. Human cathepsin k cleaves native type i and ii collagens at the n-terminal end of the triple helix. *Biochemical Journal*, 331(3):727–732, 1998.
- [69] Zhenqiang Li, Yoshiyuki Yasuda, Weijie Li, Matthew Bogyo, Norman Katz, Ronald E Gordon, Gregg B Fields, and Dieter Brömme. Regulation of collagenase activities of human cathepsins by glycosaminoglycans. *Journal of Biological Chemistry*, 279(7):5470–5479, 2004.
- [70] Bartholomew J Votta, Mark A Levy, Alison Badger, Jeremy Bradbeer, Robert A Dodds, Ian E James, Scott Thompson, Mary J Bossard, Thomas Carr, Janice R Connor, et al. Peptide aldehyde inhibitors of cathepsin k inhibit bone resorption both in vitro and in vivo. *Journal of Bone and Mineral Research*, 12(9):1396–1406, 1997.
- [71] Francesca Lazner, Maxine Gowen, Durda Pavasovic, and Ismail Kola. Osteopetrosis and osteoporosis: two sides of the same coin. *Human molecular genetics*, 8(10):1839–1846, 1999.
- [72] Riku Kiviranta, Jukka Morko, Hannele Uusitalo, Hannu T Aro, Eero Vuorio, and Juho Rantakokko. Accelerated turnover of metaphyseal trabecular bone in mice overexpressing cathepsin k. *Journal of Bone and Mineral Research*, 16(8):1444–1452, 2001.
- [73] Jukka Morko, Riku Kiviranta, Sara Hurme, Juho Rantakokko, and Eero Vuorio. Differential turnover of cortical and trabecular bone in transgenic mice overexpressing cathepsin k. *Bone*, 36(5):854–865, 2005.
- [74] Juho Rantakokko, Hannele Uusitalo, Timo Jämsä, Juha Tuukkanen, Hannu T Aro, and Eero Vuorio. Expression profiles of mrnas for osteoblast and osteoclast proteins as indicators of bone loss in mouse immobilization osteopenia model. *Journal of Bone and Mineral Research*, 14(11):1934–1942, 1999.
- [75] Shinji Kakudo, Hiroshi Mano, Miho Shiokawa, Yoshihisa Mori, Masayoshi Kumegawa, and Yoshiyuki Hakeda. Concanavalin a directly stimulates bone-resorbing activity of osteoclasts and their gene expression of cathepsin k/oc-2. *Biochemical and biophysical research communications*, 234(3):600–604, 1997.
- [76] Gerold Holzer, Helge Noske, Thomas Lang, Lukas Holzer, and Ulrike Willinger. Soluble cathepsin k: a novel marker for the prediction of nontraumatic fractures? *Journal of Laboratory and Clinical Medicine*, 146(1):13–17, 2005.

- [77] Dong Wang, W Li, M Pechar, P Kopečková, D Brömme, and J Kopeček. Cathepsin k inhibitor–polymer conjugates: potential drugs for the treatment of osteoporosis and rheumatoid arthritis. *International journal of pharmaceutics*, 277(1-2):73–79, 2004.
- [78] David N Deaton and Sanjay Kumar. Cathepsin k inhibitors: their potential as anti-osteoporosis agents. *Progress in medicinal chemistry*, 42:245–375, 2004.
- [79] Henry G Bone, Michael R McClung, Christian Roux, Robert R Recker, John A Eisman, Nadia Verbruggen, Carolyn M Hustad, Carolyn DaSilva, Arthur C Santora, and B Avery Ince. Odanacatib, a cathepsin-k inhibitor for osteoporosis: a two-year study in postmenopausal women with low bone density. *Journal of Bone and Mineral Research*, 25(5):937–947, 2010.
- [80] Avery Ince, Michael McClung, Henry Bone, Nadia Verbruggen, Andrea Rybak-Feiglin, Arthur Santora, et al. A randomized, double-blind, placebo-controlled study of a cathepsin-k inhibitor in the treatment of postmenopausal women with low bmd: 18 month results. *Bone*, 42:S74, 2008.
- [81] S Kumar, L Dare, JA Vasko-Moser, IE James, SM Blake, DJ Rickard, S-M Hwang, T Tomaszek, DS Yamashita, RW Marquis, et al. A highly potent inhibitor of cathepsin k (relacatib) reduces biomarkers of bone resorption both in vitro and in an acute model of elevated bone turnover in vivo in monkeys. *Bone*, 40(1):122–131, 2007.
- [82] S Adami, J Supronik, T Hala, JP Brown, P Garnerio, S Haemmerle, CE Ortmann, F Bouisset, and U Trechsel. Effect of one year treatment with the cathepsin-k inhibitor, balicatib, on bone mineral density (bmd) in postmenopausal women with osteopenia/osteoporosis. *Journal of Bone and Mineral Research*, 21:S24, 2006.
- [83] Olga Vasiljeva, Thomas Reinheckel, Christoph Peters, Dusan Turk, Vito Turk, and Boris Turk. Emerging roles of cysteine cathepsins in disease and their potential as drug targets. *Current pharmaceutical design*, 13(4):387–403, 2007.
- [84] Philip J Rosenthal. Cysteine proteases of malaria parasites. *International journal for parasitology*, 34(13-14):1489–1499, 2004.
- [85] Alexandra Grote, Conor R. Caffrey, Karina M. Rebello, David Smith, John P. Dalton, and Sara Lustigman. Cysteine proteases during larval migration and development of helminths in their final host, 2018.
- [86] Bruno Gryseels, Katja Polman, Jan Clerinx, and Luc Kestens. Human schistosomiasis. *The Lancet*, 368(9541):1106–1118, 2006.

- [87] AA Sabah, Cathy Fletcher, GMJD Webbe, and MJ Doenhoff. Schistosoma mansoni: chemotherapy of infections of different ages. *Experimental parasitology*, 61(3):294–303, 1986.
- [88] Livia Pica-Mattoccia and Donato Cioli. Sex-and stage-related sensitivity of schistosoma mansoni to in vivo and in vitro praziquantel treatment. *International journal for parasitology*, 34(4):527–533, 2004.
- [89] Sanaa Botros, Livia Pica-Mattoccia, Samia William, Naglaa El-Lakkani, and Donato Cioli. Effect of praziquantel on the immature stages of schistosoma haematobium. *International journal for parasitology*, 35(13):1453–1457, 2005.
- [90] Donato Cioli. Praziquantel: is there real resistance and are there alternatives? *Current Opinion in Infectious Diseases*, 13(6), 2000.
- [91] Sanaa S Botros and James L Bennett. Praziquantel resistance. *Expert Opinion on Drug Discovery*, 2(sup1):S35–S40, oct 2007.
- [92] Samia William, T A Day, Sanaa Botros, L F Tao, James L Bennett, Adel Farghally, Magdi Ismail, and Aiesha Metwally. Resistance to praziquantel: direct evidence from Schistosoma mansoni isolated from Egyptian villagers. *The American Journal of Tropical Medicine and Hygiene*, 60(6):932–935, jun 1999.
- [93] Paul J Brindley, Bernd H Kalinna, John P Dalton, Sharon R Day, Joanna Y.M Wong, Mark L Smythe, and Donald P McManus. Proteolytic degradation of host hemoglobin by schistosomes. *Molecular and Biochemical Parasitology*, 89(1):1–9, oct 1997.
- [94] Jan Dvořák, Pavla Fajtová, Lenka Ulrychová, Adrian Leontovyč, Liliana Rojo-Arreola, Brian M. Suzuki, Martin Horn, Michael Mareš, Charles S. Craik, Conor R. Caffrey, and Anthony J. O’Donoghue. Excretion/secretion products from schistosoma mansoni adults, eggs and schistosomula have unique peptidase specificity profiles. *Biochimie*, 122:99 – 109, 2016. A potpourri of proteases and inhibitors: from molecular toolboxes to signaling scissors.
- [95] Jan Dvořák, Susan T. Mashiyama, Simon Braschi, Mohammed Sajid, Giselle M. Knudsen, Elizabeth Hansell, Kee-Chong Lim, Ivy Hsieh, Mahmoud Bahgat, Bryony Mackenzie, Katalin F. Medzihradzsky, Patricia C. Babbitt, Conor R. Caffrey, and James H. McKerrow. Differential use of protease families for invasion by schistosome cercariae. *Biochimie*, 90(2):345 – 358, 2008. Cellular proteolysis.
- [96] Ciaran P. Brady Paul J. Brindley Burton J. Bogitsh, John P. Dalton. Gut-associated immunolocalization of the schistosoma mansoni cysteine protease, smc11 and smc12. *Journal of Parasitology*, 87(2):237 – 241 – 5, 2001.

- [97] Anja Michel, Hossam Ghoneim, Maristella Resto, Mo-Quen Klinkert, and Werner Kunz. Sequence, characterization and localization of a cysteine proteinase cathepsin l in schistosoma mansoni. *Molecular and Biochemical Parasitology*, 73(1):7 – 18, 1995.
- [98] Mo-Quen Klinkert, Richard Felleisen, Gisela Link, Andreas Ruppel, and Ewald Beck. Primary structures of sm31/32 diagnostic proteins of schistosoma mansoni and their identification as proteases. *Molecular and Biochemical Parasitology*, 33(2):113 – 122, 1989.
- [99] Conor R. Caffrey, Jason P. Salter, Kimberley D. Lucas, Dustin Khiem, Ivy Hsieh, Kee Chong Lim, Andreas Ruppel, James H. McKerrow, and Mohammed Sajid. SmCB2, a novel tegumental cathepsin B from adult Schistosoma mansoni. *Molecular and Biochemical Parasitology*, 121(1):49–61, 2002.
- [100] Conor R. Caffrey and Andreas Ruppel. Cathepsin b-like activity predominates over cathepsin l-like activity in adult schistosoma mansoni and s. japonicum. *Parasitology Research*, 83(6):632–635, Jun 1997.
- [101] Maha-hamadien Abdulla, Kee-chong Lim, Mohammed Sajid, James H McKerrow, and Conor R Caffrey. Schistosomiasis Mansoni: Novel Chemotherapy Using a Cysteine Protease Inhibitor. *PLoS Medicine*, 4(1):130–138, jan 2007.
- [102] World Health Organization. *World malaria report 2018*. World Health Organization, 2018.
- [103] William E. Collins and Geoffrey M. Jeffery. Plasmodium malariae: Parasite and disease. *Clinical Microbiology Reviews*, 20(4):579–592, 2007.
- [104] Dhanpat K Kochar, Vishal Saxena, Narvachan Singh, Sanjay K Kochar, S Vijay Kumar, and Ashis Das. Plasmodium vivax malaria. *Emerging infectious diseases*, 11(1):132–134, jan 2005.
- [105] William E. Collins and Geoffrey M. Jeffery. Plasmodium ovale: Parasite and disease. *Clinical Microbiology Reviews*, 18(3):570–581, 2005.
- [106] Robert W Snow, Carlos A Guerra, Abdisalan M Noor, Hla Y Myint, and Simon I Hay. The global distribution of clinical episodes of plasmodium falciparum malaria. *Nature*, 434(7030):214–217, 2005.
- [107] D Schellenberg, C Menendez, E Kahigwa, F Font, C Galindo, C Acosta, J A Schellenberg, J J Aponte, J Kimario, H Urassa, H Mshinda, M Tanner, and P Alonso. African children with malaria in an area of intense plasmodium falciparum transmis-

- sion: features on admission to the hospital and risk factors for death. *The American Journal of Tropical Medicine and Hygiene*, 61(3):431–438, 1999.
- [108] Brandy L. Salmon, Anna Oksman, and Daniel E. Goldberg. Malaria parasite exit from the host erythrocyte: A two-step process requiring extraerythrocytic proteolysis. *Proceedings of the National Academy of Sciences*, 98(1):271–276, 2001.
- [109] P A Winstanley. Chemotherapy for falciparum malaria: The armoury, the problems and the prospects. *Parasitology Today*, 16(4):146–153, 2000.
- [110] T K Mutabingwa. Artemisinin-based combination therapies (acts): Best hope for malaria treatment but inaccessible to the needy! *Acta Tropica*, 95(3):305–315, 2005.
- [111] P J Rosenthal. Proteases of malaria parasites: new targets for chemotherapy. *Emerging infectious diseases*, 4(1):49–57, 1998.
- [112] Doron C. Greenbaum, Zachary Mackey, Elizabeth Hansell, Patricia Doyle, Jiri Gut, Conor R. Caffrey, Julia Lehrman, Philip J. Rosenthal, James H. McKerrow, and Kelly Chibale. Synthesis and structure-activity relationships of parasitocidal thiosemicarbazone cysteine protease inhibitors against *Plasmodium falciparum*, *Trypanosoma brucei*, and *Trypanosoma cruzi*. *Journal of Medicinal Chemistry*, 47(12):3212–3219, 2004.
- [113] Philip J Rosenthal. Falcipains and other cysteine proteases of malaria parasites bt - cysteine proteases of pathogenic organisms. pages 30–48. Springer US, Boston, MA, 2011.
- [114] Lawrence H Bannister. Looking for the exit: How do malaria parasites escape from red blood cells? *Proceedings of the National Academy of Sciences*, 98(2):383 LP – 384, jan 2001.
- [115] Puran S Sijwali, Bhaskar R Shenai, Jiri Gut, Ajay Singh, and Philip J Rosenthal. Expression and characterization of the plasmodium falciparum haemoglobinase falcipain-3. *Biochemical Journal*, 360(2):481 LP – 489, dec 2001.
- [116] Palakodeti V N Dasaradhi, Reshma Korde, Jennifer K Thompson, Charu Tanwar, Tapas C Nag, Virander S Chauhan, Alan F Cowman, Asif Mohmmed, and Pawan Malhotra. Food vacuole targeting and trafficking of falcipain-2, an important cysteine protease of human malaria parasite plasmodium falciparum. *Molecular and Biochemical Parasitology*, 156(1):12–23, 2007.
- [117] Philip J Rosenthal, Puran S Sijwali, Ajay Singh, and Bhaskar R Shenai. Cysteine proteases of malaria parasites: targets for chemotherapy. *Current pharmaceutical design*, 8(18):1659–1672, 2002.

- [118] Philip J Rosenthal, JH McKerrow, M Aikawa, H Nagasawa, and JH Leech. A malarial cysteine proteinase is necessary for hemoglobin degradation by plasmodium falciparum. *The Journal of clinical investigation*, 82(5):1560–1566, 1988.
- [119] Philip J Rosenthal, Jed E Olson, Garson K Lee, James T Palmer, Jeffrey L Klaus, and David Rasnick. Antimalarial effects of vinyl sulfone cysteine proteinase inhibitors. *Antimicrobial agents and chemotherapy*, 40(7):1600–1603, 1996.
- [120] Anis Rassi Jr, Anis Rassi, and José Antonio Marin-Neto. Chagas disease. *The Lancet*, 375(9723):1388–1402, 2010.
- [121] José Rodrigues Coura and Pedro Albajar Viñas. Chagas disease: a new worldwide challenge. *Nature*, 465(7301\_suppl):S6, 2010.
- [122] Kevin M Bonney and David M Engman. Chagas heart disease pathogenesis: one mechanism or many? *Current molecular medicine*, 8(6):510–8, sep 2008.
- [123] Rodolfo Viotti, Carlos Vigliano, Bruno Lococo, Maria Gabriela Alvarez, Marcos Petti, Graciela Bertocchi, and Alejandro Armenti. Side effects of benznidazole as treatment in chronic chagas disease: fears and realities. *Expert review of anti-infective therapy*, 7(2):157–163, 2009.
- [124] Deborah A Nicoll-Griffith. Use of cysteine-reactive small molecules in drug discovery for trypanosomal disease. *Expert opinion on drug discovery*, 7(4):353–366, 2012.
- [125] Ana M Tomás and John M Kelly. Stage-regulated expression of cruzipain, the major cysteine protease of trypanosoma cruzi is independent of the level of rna. *Molecular and biochemical parasitology*, 76(1-2):91–103, 1996.
- [126] Ana M Tomas, Michael A Miles, and John M Kelly. Overexpression of cruzipain, the major cysteine proteinase of trypanosoma cruzi, is associated with enhanced metacyclogenesis. *European journal of biochemistry*, 244(2):596–603, 1997.
- [127] Ana Paula CA Lima, Paulo C Almeida, Ivarne LS Tersariol, Veronica Schmitz, Alvin H Schmaier, Luiz Juliano, Isaura Y Hirata, Werner Müller-Esterl, Jair R Chagas, and Julio Scharfstein. Heparan sulfate modulates kinin release by trypanosoma cruzi through the activity of cruzipain. *Journal of Biological Chemistry*, 277(8):5875–5881, 2002.
- [128] Juan C Engel, Patricia S Doyle, Ivy Hsieh, and James H McKerrow. Cysteine protease inhibitors cure an experimental trypanosoma cruzi infection. *Journal of Experimental Medicine*, 188(4):725–734, 1998.



- [129] Berta M Franke de Cazzulo, Javier Martínez, Michael J North, Graham H Coombs, and Juan-Jose Cazzulo. Effects of proteinase inhibitors on the growth and differentiation of trypanosoma cruzi. *FEMS microbiology letters*, 124(1):81–86, 1994.
- [130] Peng-Yu Yang, Min Wang, Cynthia Y. He, and Shao Q. Yao. Proteomic profiling and potential cellular target identification of K11777, a clinical cysteine protease inhibitor, in Trypanosoma brucei. *Chem. Commun.*, 48(6):835–837, 2012.
- [131] M Sajid, SA Robertson, LS Brinen, and JH McKerrow. Cruzain—the path from target validation to the clinic, cysteine proteases of pathogenic organisms, ed., robinson mw, dalton jp, 2011.
- [132] Katrien Brak, Iain D. Kerr, Kimberly T. Barrett, Nobuhiro Fuchi, Moumita Deb-nath, Kenny Ang, Juan C. Engel, James H. McKerrow, Patricia S. Doyle, Linda S. Brinen, and Jonathan A. Ellman. Nonpeptidic Tetrafluorophenoxymethyl Ketone Cruzain Inhibitors as Promising New Leads for Chagas Disease Chemotherapy. *Journal of Medicinal Chemistry*, 53(4):1763–1773, feb 2010.
- [133] Christian Beaulieu, Elise Isabel, Angélique Fortier, Frédéric Masse, Christophe Mel-lon, Nathalie Methot, Momar Ndao, Deborah Nicoll-Griffith, Doris Lee, Hyeram Park, and W. Cameron Black. Identification of potent and reversible cruzipain in-hibitors for the treatment of Chagas disease. *Bioorganic & Medicinal Chemistry Letters*, 20(24):7444–7449, dec 2010.
- [134] Uwe Machon, Christian Büchold, Martin Stempka, Tanja Schirmeister, Christoph Gelhaus, Matthias Leippe, Jiri Gut, Philip J. Rosenthal, Caroline Kisker, Matthias Leyh, and Carsten Schmuck. On-Bead Screening of a Combinatorial Fumaric Acid Derived Peptide Library Yields Antiplasmodial Cysteine Protease Inhibitors with Unusual Peptide Sequences. *Journal of Medicinal Chemistry*, 52(18):5662–5672, sep 2009.
- [135] Zhaozhao Li, Girish S Patil, Zbigniew E Golubski, Hitoshi Hori, Kamin Tehrani, J E Foreman, David D Eveleth, Raymond T Bartus, and James C Powers. Peptide .alpha.-keto ester, .alpha.-keto amide, and .alpha.-keto acid inhibitors of calpains and other cysteine proteases. *Journal of Medicinal Chemistry*, 36(22):3472–3480, oct 1993.
- [136] James T. Palmer, David Rasnick, Jeffrey L. Klaus, and Dieter Bromme. Vinyl Sulfones as Mechanism-Based Cysteine Protease Inhibitors. *Journal of Medicinal Chemistry*, 38(17):3193–3196, aug 1995.
- [137] Katrien Brak, Patricia S. Doyle, James H. McKerrow, and Jonathan A. Ellman. Identification of a new class of nonpeptidic inhibitors of cruzain. *Journal of the*

*American Chemical Society*, 130(20):6404–6410, 2008.

- [138] Radim Vicik, Verena Hoerr, Melanie Glaser, Martina Schultheis, Elizabeth Hansell, James H. McKerrow, Ulrike Holzgrabe, Conor R. Caffrey, Alicia Ponte-Sucre, Heidrun Moll, August Stich, and Tanja Schirmeister. Aziridine-2,3-dicarboxylate inhibitors targeting the major cysteine protease of *Trypanosoma brucei* as lead trypanocidal agents. *Bioorganic & Medicinal Chemistry Letters*, 16(10):2753–2757, may 2006.
- [139] Dennis S Yamashita, Ward W Smith, Baoguang Zhao, Cheryl A Janson, Thaddeus A Tomaszek, Mary J Bossard, Mark A Levy, Hye-Ja Oh, Thomas J. Carr, Scott K Thompson, Carl F Ijames, Steven A Carr, Michael McQueney, Karla J. D’Alessio, Bernard Y Amegadzie, Charles R Hanning, Sherin Abdel-Meguid, Renee L. DesJarlais, John G Gleason, and Daniel F Veber. Structure and Design of Potent and Selective Cathepsin K Inhibitors. *Journal of the American Chemical Society*, 119(46):11351–11352, nov 1997.
- [140] Bryan T. Mott, Rafaela S. Ferreira, Anton Simeonov, Ajit Jadhav, Kenny Kean-Hooi Ang, William Leister, Min Shen, Julia T. Silveira, Patricia S. Doyle, Michelle R. Arkin, James H. McKerrow, James Inglese, Christopher P. Austin, Craig J. Thomas, Brian K. Shoichet, and David J. Maloney. Identification and Optimization of Inhibitors of Trypanosomal Cysteine Proteases: Cruzain, Rhodesain, and TbCatB. *Journal of Medicinal Chemistry*, 53(1):52–60, jan 2010.
- [141] Je-Tae Woo, Sanae Sigeizumi, Kohji Yamaguchi, Kikuo Sugimoto, Takeo Kobori, Tomoko Tsuji, and Kiyosi Kondo. Peptidyl aldehyde derivatives as potent and selective inhibitors of cathepsin L. *Bioorganic & Medicinal Chemistry Letters*, 5(14):1501–1504, jul 1995.
- [142] Xiaohui Du, Chun Guo, Elizabeth Hansell, Patricia S. Doyle, Conor R. Caffrey, Tod P. Holler, James H. McKerrow, and Fred E. Cohen. Synthesis and Structure–Activity Relationship Study of Potent Trypanocidal Thio Semicarbazone Inhibitors of the Trypanosomal Cysteine Protease Cruzain. *Journal of Medicinal Chemistry*, 45(13):2695–2707, jun 2002.
- [143] Péter Ábrányi-Balogh, László Petri, Tímea Imre, Péter Szijj, Andrea Scarpino, Martina Hrast, Ana Mitrović, Urša Pečar Fonovič, Krisztina Németh, Hélène Barreteau, et al. A road map for prioritizing warheads for cysteine targeting covalent inhibitors. *European journal of medicinal chemistry*, 160:94–107, 2018.
- [144] Jiewen Du, Xin Yan, Zhihong Liu, Lu Cui, Peng Ding, Xiaoqing Tan, Xiuming Li, Huihao Zhou, Qiong Gu, and Jun Xu. cBinderDB: a covalent binding agent

- database. *Bioinformatics*, 33(8):1258–1260, 12 2016.
- [145] Malte Gersch, Johannes Kreuzer, and Stephan A. Sieber. Electrophilic natural products and their biological targets. *Nat. Prod. Rep.*, 29:659–682, 2012.
- [146] Edson dos Santos, Fernando Gonçalves, Paulo Prado, Daniele Sasaki, Dênis de Lima, and Maria Macedo. Synthesis Method for Thiosulfonate and Report of Its Insecticidal Activity in *Anagasta kuehniella* (Lepidoptera: Pyralidae). *International Journal of Molecular Sciences*, 13(12):15241–15251, nov 2012.
- [147] Chunpo Ge, Hao Wang, Baoxin Zhang, Juan Yao, Xinming Li, Weimin Feng, Panpan Zhou, Yawen Wang, and Jianguo Fang. A thiol–thiosulfonate reaction providing a novel strategy for turn-on thiol sensing. *Chem. Commun.*, 51(80):14913–14916, 2015.
- [148] Yasushi K. Nakamura, Tomoaki Matsuo, Kayoko Shimoi, Yoshiyuki Nakamura, and Isao Tomita. S -Methyl Methanethiosulfonate, Bio-antimutagen in Homogenates of Cruciferae and Liliaceae Vegetables. *Bioscience, Biotechnology, and Biochemistry*, 60(9):1439–1443, jan 1996.
- [149] Bandaru S. Reddy. Chemopreventive effect of S-methylmethane thiosulfonate and sulindac administered together during the promotion/progression stages of colon carcinogenesis. *Carcinogenesis*, 20(8):1645–1648, aug 1999.
- [150] Muneerah Smith, Roger Hunter, Nashia Stellenboom, Daniel A. Kusza, M. Iqbal Parker, Ahmed N.H. Hammouda, Graham Jackson, and Catherine H. Kaschula. The cytotoxicity of garlic-related disulphides and thiosulfonates in WHCO1 oesophageal cancer cells is dependent on S-thiolation and not production of ROS. *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1860(7):1439–1449, jul 2016.
- [151] Ariel A Petruk, Silvina Bartesaghi, Madia Trujillo, Darío A Estrin, Daniel Murgida, Balaraman Kalyanaraman, Marcelo A Marti, and Rafael Radi. Molecular basis of intramolecular electron transfer in proteins during radical-mediated oxidations: Computer simulation studies in model tyrosine–cysteine peptides in solution. *Archives of biochemistry and biophysics*, 525(1):82–91, 2012.
- [152] Samie R Jaffrey, Hediye Erdjument-Bromage, Christopher D Ferris, Paul Tempst, and Solomon H Snyder. Protein S-nitrosylation: a physiological signal for neuronal nitric oxide. *Nature Cell Biology*, 3(2):193–197, feb 2001.
- [153] Adela Rodríguez-Romero, Alejandra Hernández-Santoyo, Luis del Pozo Yauner, Adrián Kornhauser, and D Alejandro Fernández-Velasco. Structure and inactivation of triosephosphate isomerase from *entamoeba histolytica*. *Journal of Molecular Biology*, 322(4):669 – 675, 2002.

- [154] Norman S Millian and Timothy A Garrow. Human betaine–homocysteine methyltransferase is a zinc metalloenzyme. *Archives of biochemistry and biophysics*, 356(1):93–98, 1998.
- [155] W T Lowther, N Brot, H Weissbach, J F Honek, and B W Matthews. Thiol-disulfide exchange is involved in the catalytic mechanism of peptide methionine sulfoxide reductase. *Proceedings of the National Academy of Sciences*, 97(12):6463–6468, jun 2000.
- [156] Kazutsugu Matsumoto, Benjamin G. Davis, and J. Bryan Jones. Chemically Modified “Polar Patch” Mutants of Subtilisin in Peptide Synthesis with Remarkably Broad Substrate Acceptance: Designing Combinatorial Biocatalysts. *Chemistry European Journal*, 8(18):4129–4137, sep 2002.
- [157] Elena Gabriele, Chiara Ricci, Fiorella Meneghetti, Nicola Ferri, Akira Asai, and Anna Sparatore. Methanethiosulfonate derivatives as ligands of the STAT3-SH2 domain. *Journal of Enzyme Inhibition and Medicinal Chemistry*, 32(1):337–344, 2017.
- [158] Nataraj S Pagadala, Khajamohiddin Syed, and Jack Tuszynski. Software for molecular docking: a review. *Biophysical reviews*, 9(2):91–102, 2017.
- [159] Martin Stahl and Matthias Rarey. Detailed analysis of scoring functions for virtual screening. *Journal of medicinal chemistry*, 44(7):1035–1042, 2001.
- [160] Hezekiel Mathambo Kumalo, Soumendranath Bhakat, and Mahmoud E S Soliman. Theory and applications of covalent docking in drug discovery: Merits and pitfalls. *Molecules*, 20(2):1984–2000, 2015.
- [161] Tiejun Cheng, Qingliang Li, Zhigang Zhou, Yanli Wang, and Stephen H Bryant. Structure-based virtual screening for drug discovery: a problem-centric review. *The AAPS journal*, 14(1):133–141, 2012.
- [162] Elizabeth Yuriev, Jessica Holien, and Paul A Ramsland. Improvements, trends, and new ideas in molecular docking: 2012–2013 in review. *Journal of Molecular Recognition*, 28(10):581–604, 2015.
- [163] Oleg Trott and Arthur J. Olson. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, 31(2):455–461, 2010.
- [164] John J Irwin and Brian K Shoichet. Zinc—a free database of commercially available compounds for virtual screening. *J Chem Inf Model*, 45(1):177–182, 2005.

- [165] Colin R. Groom, Ian J. Bruno, Matthew P. Lightfoot, and Suzanna C. Ward. The Cambridge Structural Database. *Acta Crystallographica Section B*, 72(2):171–179, Apr 2016.
- [166] Helen M. Berman, John Westbrook, Zukang Feng, Gary Gilliland, T. N. Bhat, Helge Weissig, Ilya N. Shindyalov, and Philip E. Bourne. The Protein Data Bank. *Nucleic Acids Research*, 28(1):235–242, 01 2000.
- [167] Molecular Networks GmbH, www.mn-am.com. *CORINA Classic, Corina Classic, Generation Of Three-Dimensional Molecular Models*, 4.2.0 edition, September 2018.
- [168] D.S. BIOvIA. Discovery studio modeling environment. san diego, dassault systemes, release, 4., 2015.
- [169] Daan MF Van Aalten, R Bywater, John BC Findlay, Manfred Hendlich, Rob WW Hooft, and Gert Vriend. Prodrq, a program for generating molecular topologies and unique molecular descriptors from coordinates of small molecules. *Journal of computer-aided molecular design*, 10(3):255–262, 1996.
- [170] Ghemical, <http://bioinformatics.org/ghemical/ghemical/index.html>. Last checked 11th August 2019.
- [171] Marcus D. Hanwell, Donald E. Curtis, David C. Lonie, Tim Vandermeersch, Eva Zurek, and Geoffrey R. Hutchison. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *Journal of Cheminformatics*, 4(1):17, 2012.
- [172] Noel M O’Boyle, Michael Banck, Craig A James, Chris Morley, Tim Vandermeersch, and Geoffrey R Hutchison. Open Babel: An open chemical toolbox. *Journal of Cheminformatics*, 3(1):33, 2011.
- [173] Open babel: The open source chemistry toolbox, August 2019.
- [174] Open Babel Wiki, [http://openbabel.org/wiki/Main\\_Page](http://openbabel.org/wiki/Main_Page), Last checked 18th May 2019.
- [175] Thomas A Halgren. Merck molecular force field. i. basis, form, scope, parameterization, and performance of mmff94. *Journal of computational chemistry*, 17(5-6):490–519, 1996.
- [176] Garrett M Morris, Ruth Huey, William Lindstrom, Michel F Sanner, Richard K Belew, David S Goodsell, and Arthur J Olson. Autodock4 and autodocktools4: Automated docking with selective receptor flexibility. *J Comput Chem*, 30(16):2785–2791, Dec 2009.

- [177] Schrödinger, LLC. The PyMOL molecular graphics system, version 1.8. November 2015.
- [178] Arwin J Brouwer, Anika Jonker, Paul Werkhoven, Ethan Kuo, Nan Li, Nerea Galastegui, Johan Kemmink, Bogdan I Florea, Michael Groll, Herman S Overkleeft, and Rob M J Liskamp. Peptido Sulfonyl Fluorides as New Powerful Proteasome Inhibitors. *Journal of Medicinal Chemistry*, 55(24):10995–11003, dec 2012.
- [179] Christian Dubiella, Haissi Cui, Malte Gersch, Arwin J Brouwer, Stephan a Sieber, Achim Krüger, Rob M J Liskamp, and Michael Groll. Selective Inhibition of the Immunoproteasome by Ligand-Induced Crosslinking of the Active Site. *Angewandte Chemie (International ed. in English)*, 53(44):11969–11973, sep 2014.
- [180] Arwin J Brouwer, Natalia Herrero Alvarez, Adriano Ciaffoni, Helmus van de Langemheen, and Rob M J Liskamp. Proteasome inhibition by new dual warhead containing peptido vinyl sulfonyl fluorides. *Bioorganic & Medicinal Chemistry*, 24(16):3429–3435, August 2016.
- [181] Raik Artschwager, David J. Ward, Susan Gannon, Arwin J. Brouwer, Helmus van de Langemheen, Hubert Kowalski, and Rob M. J. Liskamp. Potent and Highly Selective Inhibitors of the Proteasome Trypsin-like Site by Incorporation of Basic Side Chain Containing Amino Acid Derived Sulfonyl Fluorides. *Journal of Medicinal Chemistry*, 61(12):5395–5411, jun 2018.
- [182] Arwin J. Brouwer, Tarik Ceylan, Tima Van Der Linden, and Rob M.J. Liskamp. Synthesis of  $\beta$ -aminoethanesulfonyl fluorides or 2-substituted taurine sulfonyl fluorides as potential protease inhibitors. *Tetrahedron Letters*, 50(26):3391–3393, jul 2009.
- [183] Arwin J Brouwer, Tarik Ceylan, Anika M Jonker, Tima van der Linden, and Rob M J Liskamp. Synthesis and biological evaluation of novel irreversible serine protease inhibitors using amino acid based sulfonyl fluorides as an electrophilic trap. *Bioorganic & Medicinal Chemistry*, 19(7):2397–2406, apr 2011.
- [184] Salvador Guardiola, Roger Prades, Laura Mendieta, Arwin J. Brouwer, Jelle Streefkerk, Laura Nevola, Teresa Tarragó, Rob M.J. Liskamp, and Ernest Giralt. Targeted Covalent Inhibition of Prolyl Oligopeptidase (POP): Discovery of Sulfonylfluoride Peptidomimetics. *Cell Chemical Biology*, 25(8):1031–1037, aug 2018.
- [185] Hideaki Tsuge, Tomohiro Nishimura, Yukio Tada, Tetsuji Asao, Dusan Turk, Vito Turk, and Nobuhiko Katunuma. Inhibition mechanism of cathepsin l-specific inhibitors based on the crystal structure of papain–clik148 complex. *Biochemical and Biophysical Research Communications*, 266(2):411 – 416, 1999.

- [186] Iain D. Kerr, Ji H. Lee, Christopher J. Farady, Rachael Marion, Mathias Rickert, Mohammed Sajid, Kailash C. Pandey, Conor R. Caffrey, Jennifer Legac, Elizabeth Hansell, James H. McKerrow, Charles S. Craik, Philip J. Rosenthal, and Linda S. Brinen. Vinyl sulfones as antiparasitic agents and a structural basis for drug design. *Journal of Biological Chemistry*, 284(38):25697–25703, 2009.
- [187] Adéla Jílková, Pavlína Řezáčová, Martin Lepšík, Martin Horn, Jana Váchová, Jindřich Fanfrlík, Jiří Brynda, James H. McKerrow, Conor R. Caffrey, and Michael Mareš. Structural Basis for Inhibition of Cathepsin B Drug Target from the Human Blood Fluke, *Schistosoma mansoni*. *Journal of Biological Chemistry*, 286(41):35770–35781, oct 2011.
- [188] H. B. Bürgi, J. D. Dunitz, and E. Shefter. Chemical reaction paths. IV. Aspects of  $O \cdots C = O$  interactions in crystals. *Acta Crystallographica Section B*, 30(6):1517–1527, Jun 1974.
- [189] Eric P Lodge and Clayton H Heathcock. Acyclic stereoselection. 39. the origin of diastereofacial selectivity in additions to chiral aldehydes and ketones: trajectory analysis. *Journal of the American Chemical Society*, 109(9):2819–2820, apr 1987.
- [190] Mauro Mileni, Satwik Kamtekar, David C Wood, Timothy E Benson, Benjamin F Cravatt, and Raymond C Stevens. Crystal structure of fatty acid amide hydrolase bound to the carbamate inhibitor urb597: discovery of a deacylating water molecule and insight into enzyme inactivation. *Journal of molecular biology*, 400(4):743–754, 2010.
- [191] David Rasnick. Small synthetic inhibitors of cysteine proteases. *Perspectives in Drug Discovery and Design*, 6(1):47–63, dec 1996.
- [192] Ryu Sato, Takehiko Goto, Yuji Takikawa, and Saburo Takizawa. Convenient synthesis of aromatic thiosulfonates from aromatic sulfinates with elemental sulfur in amines. *Synthesis*, 1980(08):615–615, 1980.
- [193] Jérôme Lacour, David Monchaud, Jiri Mareda, France Favarger, and Gérald Bernardinelli. Synthesis, enantiomeric conformations, and stereodynamics of aromatic ortho-substituted disulfones. *Helvetica Chimica Acta*, 86(1):65–81, 2003.
- [194] Scott Sweeney. Final year project work, October 2015. Personal correspondence.
- [195] Jean-Pierre Mahieu, Martine Gosselet, Bernard Sebillé, and Yves Beuzard. Synthesis of new thiosulfonates and disulfides from sulfonyl chlorides and thiols. *Synthetic Communications*, 16(13):1709–1722, 1986.

- [196] Hien Thi Pham, Ngoc-Lan Thi Nguyen, Fritz Duus, and Thi Xuan Thi Luu. Ultrasound-accelerated synthesis of asymmetrical thiosulfonate s-esters by base-promoted reaction of sulfonyl chlorides with thiols. *Phosphorus, Sulfur, and Silicon and the Related Elements*, 190(11):1934–1941, 2015.
- [197] Stuart Warren Jonathan Clayden, Nick Greeves. *Organic Chemistry*. Oxford University Press, 2 edition, 2012.
- [198] Kiyoko Fujiki, Naoki Tanifuji, Yohei Sasaki, and Taku Yokoyama. New and Facile Synthesis of Thiosulfonates from Sulfinato/Disulfide/I<sub>2</sub> System. *Synthesis*, 2002(03):343–348, 2002.
- [199] Erica R Kiemle, Matthew Wathier, Paul Bichler, and Jennifer A Love. Total Synthesis of K777 SI. *Organic letters*, 7(11):956–963, 2016.
- [200] Gerjan de Bruin, Eva J. van Rooden, David Ward, Charlotte Wesseling, Adrianus M. C. H. van den Nieuwendijk, Constant A. A. van Boeckel, Christoph Driessen, Alexei F. Kisselev, Bogdan I. Florea, Mario van der Stelt, and Herman S. Overkleeft. Asymmetric Synthesis of Lysine Analogues with Reduced Basicity, and their Incorporation into Proteasome Inhibitors. *European Journal of Organic Chemistry*, 2017(39):5921–5934, oct 2017.
- [201] Harold B Brooks, Sandaruwan Geeganage, Steven D Kahl, Chahrzad Montrose, Sitta Sittampalam, Michelle C Smith, and Jeffrey R Weidner. Basics of enzymatic assays for hts. In *Assay Guidance Manual [Internet]*. Eli Lilly & Company and the National Center for Advancing Translational Sciences, 2012.
- [202] Michael G Acker and Douglas S Auld. Considerations for the design and reporting of enzyme assays in high-throughput screening applications. *Perspectives in Science*, 1(1-6):56–73, 2014.
- [203] Philip W Iversen, Brian J Eastwood, G Sitta Sittampalam, and Karen L Cox. A comparison of assay performance measures in screening assays: signal window, z'factor, and assay variability ratio. *Journal of biomolecular screening*, 11(3):247–252, 2006.
- [204] Jessy Aziz, Samir Messaoudi, Mouad Alami, and Abdallah Hamze. Sulfinato derivatives: dual and versatile partners in organic synthesis. *Org. Biomol. Chem.*, 12:9743–9759, 2014.
- [205] Holly Woolven, Carlos González-Rodríguez, Isabel Marco, Amber L Thompson, and Michael C Willis. Dabco-bis (sulfur dioxide), dabso, as a convenient source of sulfur dioxide for organic synthesis: utility in sulfonamide and sulfamide preparation. *Organic letters*, 13(18):4876–4878, 2011.



- [206] Atsuko Nishiguchi, Kazuhiro Maeda, and Shokyo Miki. Sulfonyl Chloride Formation from Thiol Derivatives by N -Chlorosuccinimide Mediated Oxidation. *Synthesis*, 2006(24):4131–4134, nov 2006.
- [207] Malcolm D King, Rodney E Sue, Rodney H White, and David J Young. The thermal desulfination of allylic sulfonyl halides. *Tetrahedron Letters*, 38(25):4493 – 4496, 1997.
- [208] Stephanie D Meyer and Stuart L Schreiber. Acceleration of the dess-martin oxidation by water. *The Journal of Organic Chemistry*, 59(24):7549–7552, 1994.
- [209] Abijah M Nyong and Jon D Rainier. The diastereoselective synthesis of quaternary substituted thioindolines from sulfur ylide intermediates. *The Journal of organic chemistry*, 70(2):746–748, 2005.
- [210] Wanchai Pluempanupat, Parinthorn Temyarasilp, Michael Widhalm, and Warinthorn Chavasiri. Camphorquinone: a new and efficient oxidant for the preparation of 2-thio-substituted benzothiazoles from alcohols by oxidation-reduction condensation. *Journal of Sulfur Chemistry*, 35(4):418–430, 2014.
- [211] Takehiko Nishio. Direct conversion of alcohols into thiols. *Journal of the Chemical Society, Perkin Transactions 1*, (10):1113–1117, 1993.
- [212] B. P. Bandgar, V. S. Sadavarte, and L. S. Uppalla. Remarkably Fast Direct Synthesis of Thiols from Alcohols under Mild Conditions. *Chemistry Letters*, (11):1304–1305, 2000.
- [213] Marc H. Dresden and AndréM. Deelder. Schistosoma mansoni: Thiol proteinase properties of adult worm “hemoglobinase”. *Experimental Parasitology*, 48(2):190–197, oct 1979.
- [214] Margaret M. Wasilewski, K.C. Lim, Jennifer Phillips, and James H. McKerrow. Cysteine protease inhibitors block schistosome hemoglobin degradation in vitro and decrease worm burden and egg production in vivo. *Molecular and Biochemical Parasitology*, 81(2):179–189, oct 1996.
- [215] Dj Musil, D Zucic, D Turk, RA Engh, I Mayr, R Huber, T Popovic, V Turk, T Towatari, and N Katunuma. The refined 2.15 ax-ray crystal structure of human liver cathepsin b: the structural basis for its specificity. *The EMBO journal*, 10(9):2321–2330, 1991.
- [216] Youngchool Choe, Francesco Leonetti, Doron C Greenbaum, Fabien Lecaille, Matthew Bogyo, Dieter Brömme, Jonathan A Ellman, and Charles S Craik. Substrate profiling of cysteine proteases using a combinatorial peptide library identifies

- functionally unique specificities. *Journal of Biological Chemistry*, 281(18):12824–12832, 2006.
- [217] Patrik Kadesch, Thomas Quack, Stefanie Gerbig, Christoph G. Grevelding, and Bernhard Spengler. Lipid Topography in *Schistosoma mansoni* Cryosections, Revealed by Microembedding and High-Resolution Atmospheric-Pressure Matrix-Assisted Laser Desorption/Ionization (MALDI) Mass Spectrometry Imaging. *Analytical Chemistry*, 91:4520–4528, 2019.
- [218] Kim Retra, Saskia deWalick, Marion Schmitz, Maria Yazdanbakhsh, Aloysius GM Tielens, Jos FHM Brouwers, and Jaap J van Hellemond. The tegumental surface membranes of *Schistosoma mansoni* are enriched in parasite-specific phospholipid species. *International journal for parasitology*, 45(9-10):629–636, 2015.
- [219] Franz Meyer, Haruko Meyer, and Ernest Bueding. Lipid metabolism in the parasitic and free-living flatworms, *Schistosoma mansoni* and *Dugesia dorotocephala*. *Biochimica et Biophysica Acta (BBA)-Lipids and Lipid Metabolism*, 210(2):257–266, 1970.
- [220] Paul P Geurink, Wouter A van der Linden, Anne C Mirabella, Nerea Gallastegui, Gerjan de Bruin, Annet E M Blom, Mathias J Voges, Elliot D Mock, Bogdan I Florea, Gijs A van der Marel, Christoph Driessen, Mario van der Stelt, Michael Groll, Herman S Overkleeft, and Alexei F Kisselev. Incorporation of non-natural amino acids improves cell permeability and potency of specific inhibitors of proteasome trypsin-like sites. *Journal of medicinal chemistry*, 56(3):1262–1275, feb 2013.
- [221] Hugo E Gottlieb, Vadim Kotlyar, and Abraham Nudelman. NMR Chemical Shifts of Common Laboratory Solvents as Trace Impurities. *The Journal of Organic Chemistry*, 62(21):7512–7515, oct 1997.
- [222] Joanne L Fraser, Richard FW Jackson, and Barry Porter. Preparation of enantiomerically pure homophenylalanine derivatives by palladium-catalysed coupling reactions. *Synlett*, 1994(05):379–380, 1994.
- [223] Tsuyoshi Shinozuka, Kousei Shimada, Satoshi Matsui, Takahiro Yamane, Mayumi Ama, Takeshi Fukuda, Motohiko Taki, Yuko Takeda, Eri Otsuka, Michiko Yamato, et al. Potent and selective cathepsin k inhibitors. *Bioorganic & medicinal chemistry*, 14(20):6789–6806, 2006.
- [224] Yancey D Ward, David S Thomson, Leah L Frye, Charles L Cywin, Tina Morwick, Michel J Emmanuel, Renée Zindell, Daniel McNeil, Younes Bekkali, Marc Girardot, et al. Design and synthesis of dipeptide nitriles as reversible and potent cathepsin s inhibitors. *Journal of medicinal chemistry*, 45(25):5471–5482, 2002.

- [225] Charles L Cywin, Raymond A Firestone, Daniel W McNeil, Christine A Grygon, Kathryn M Crane, Della M White, Peter R Kinkade, Jerry L Hopkins, Walter Davidson, Mark E Labadia, et al. The design of potent hydrazones and disulfides as cathepsin s inhibitors. *Bioorganic & medicinal chemistry*, 11(5):733–740, 2003.

# Appendix

DrawToDock Script

```

#!/bin/bash
#System requirements and notes: designed and tested on MacOS.
Require install of Open Babel and install of Python 2.7. for
script to operate.
#input parameters here:

pH=6.5 # Input the pH for protonation of ligands.
steps=3000 # Input the number of energy minimisation steps
used for generation of ligand 3D coordinates (for conversion
to .pdb)

echo Running ligand conversion ... # all echo commands give a
readout to the user in terminal to update on the process
echo
echo parameters:
echo protonation state \(\pH\) = $pH
echo energy minimisation steps = $steps
echo Energy minimisation method: MMF94\, steepest descending
#these parameters can be altered in line 19, this is assumed
to be the standard required.

for f in ligand_*.cdx; do #this for loop converts all .cdx
files starting with the name ligand_ to corresponding .pdb
files
    b=`basename $f .cdx`
    echo
    echo Processing $b to .pdb format ...
    echo
    obabel $f -O ${b}.pdb --gen3d -p $pH --minimize --steps
$steps --sd # This line contains the presets used by Open
Babel for file conversion
    echo
    echo $b conversion complete.
    echo
done

echo All ligands converted to .pdb

for f in ligand_*.pdb; do #this for loop converts all .pdb
files starting with the name ligand_ to corresponding .pdbqt
files
    b=`basename $f .pdb`
    echo
    echo Processing $b to .pdbqt format ...
    echo
    python ./prepare_ligand4.py -l ${f} # this line calls on
the python script to prepatate .pdbqt files from the
corresponding .pdb files. python version 2.7 is required for
this.
    echo

```

```

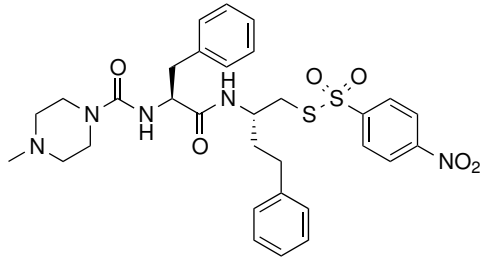
    echo $b conversion to .pdbqt complete.
    echo
done
echo All ligands converted to .pdbqt
for f in ligand_*.pdbqt; do #this for loop rund the docking
simulation for each of the previously prepared ligand .pdbqt
files
    b=`basename $f .pdbqt`
    echo Docking ligand $b ...
    mkdir -p $b
    ./vina --config config.txt --ligand $f --out ${b}/${b}
_out.pdbqt --log ${b}/${b}_log.txt #this is the vina
executable to run docking. vina.exe must be present in the
directory for this to function along with a configuration file
named config.txt
done

# The following code is to 'clean' the apperance of the
results and files generated
mkdir chemdraw\ files # makes a new directory named Chemdraw
files
mv *.cdx chemdraw\ files # moves all Chemdraw files to new
directory
mkdir pdb\ files # makes a new directory named pdb files
mv ligand_*.pdb pdb\ files # moves all pdb files to new
directory
mkdir pdbqt\ files # makes a new directory named pdbqt files
mv ligand_*.pdbqt pdbqt\ files # moves all .pdbqt files to new
directory
mkdir results # makes a new directory named results
mv ligand_* results # moves all docking result files to new
directory

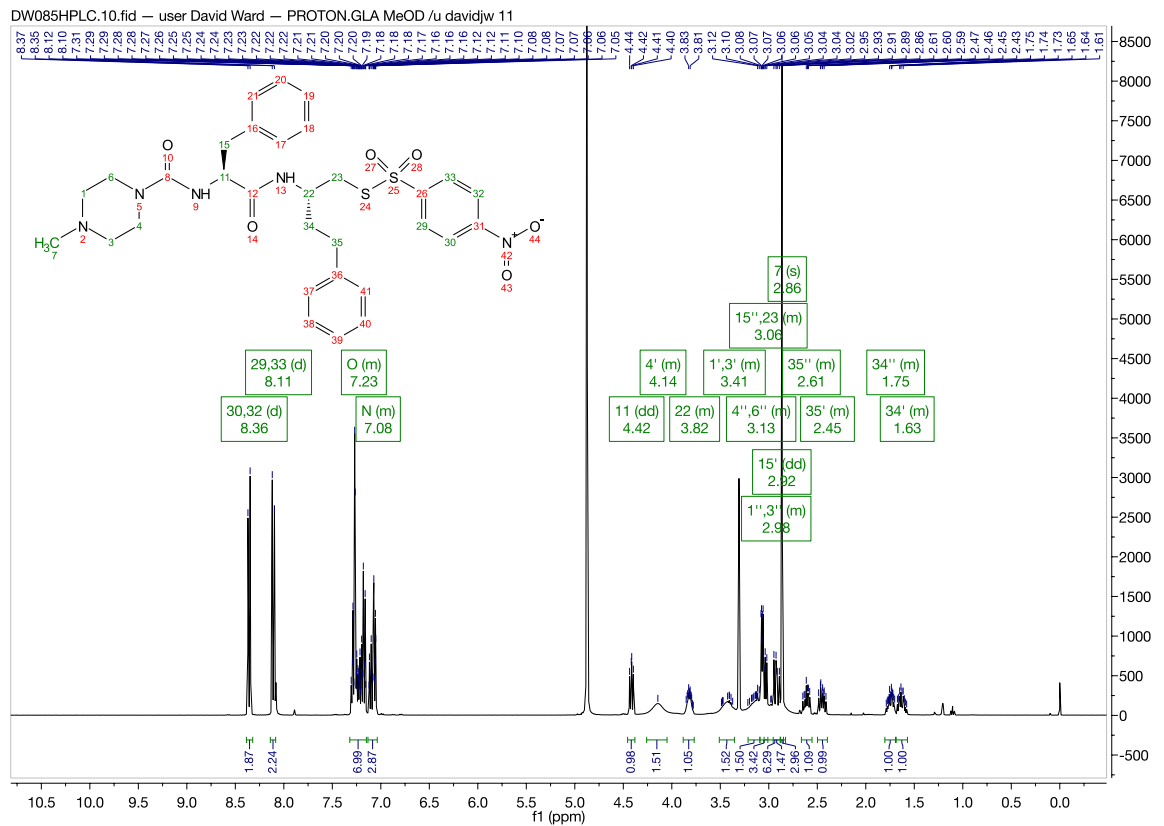
echo All ligands docked.
echo Job Complete.
cat DrawtoDock1.0.sh >> scriptlog.txt # saves a copy of the
DrawtoDock1.0.sh script used to a log file for future
reference
mv scriptlog.txt results # Moves the DrawtoDock script log
file into the results folder.

```

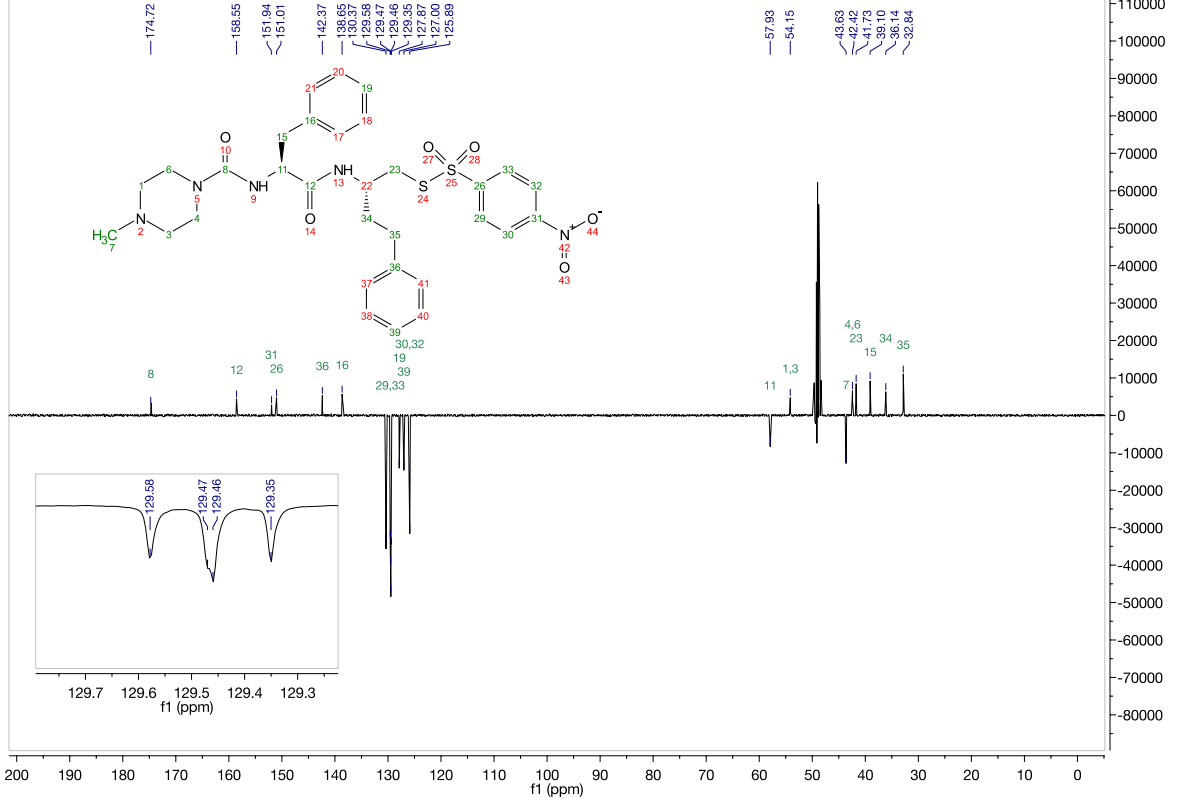
# NMR of Final Inhibitors



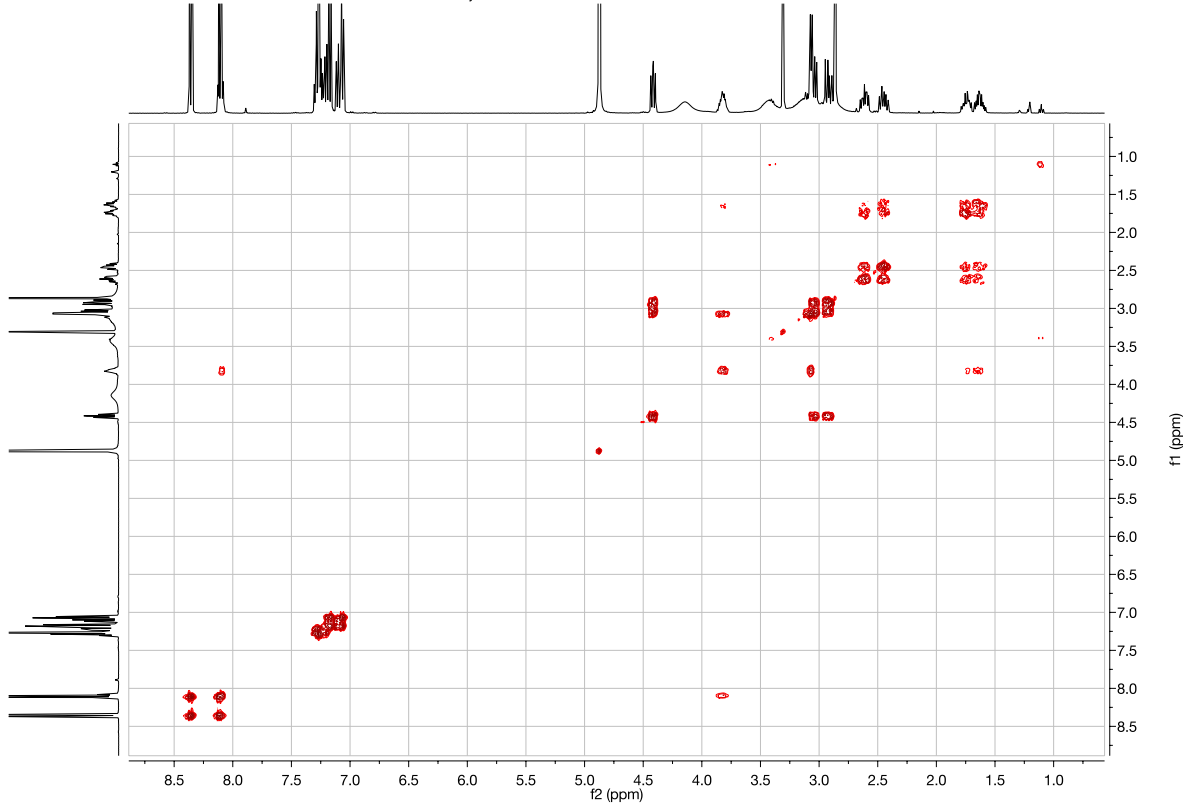
Inhibitor 7



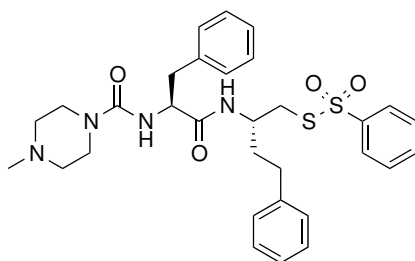
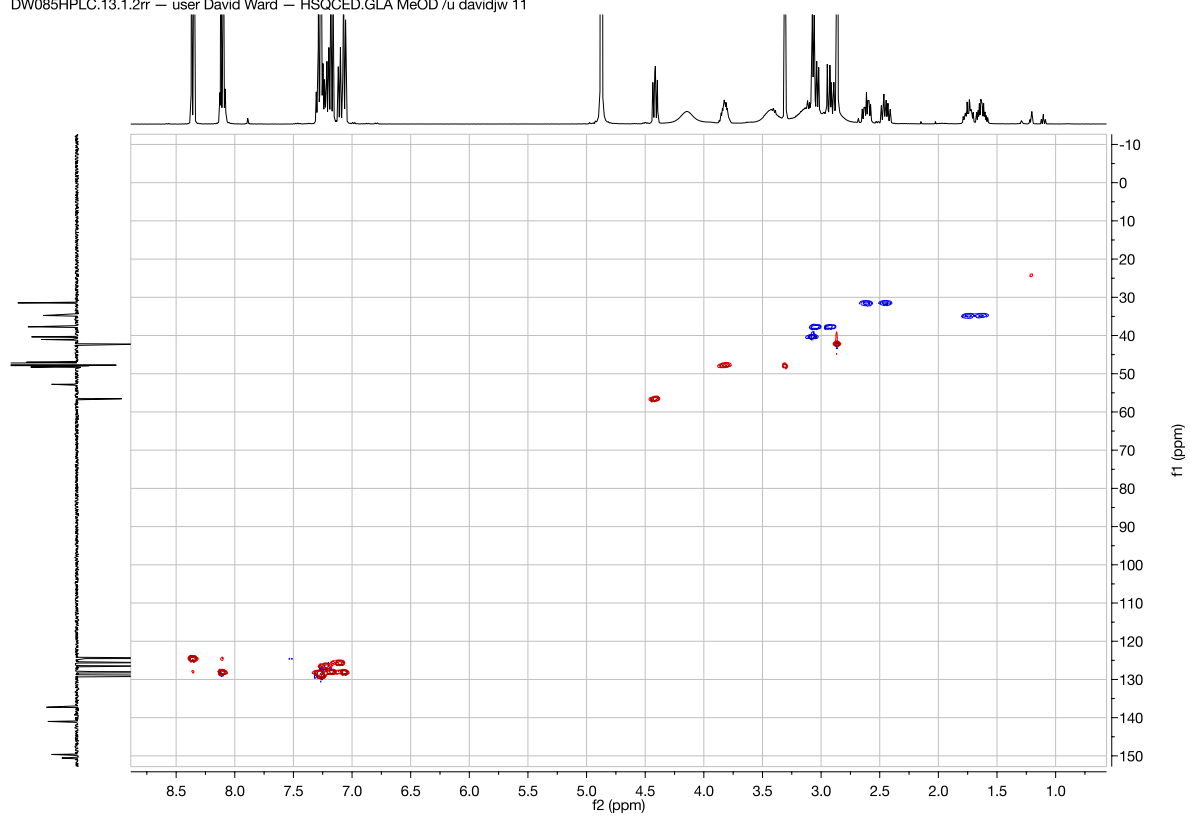
DW085HPLC.12.fid — user David Ward — C13DEPTQ.GLA MeOD /u davidjw 11



DW085HPLC.11.1.2rr — user David Ward — COSY.GLA MeOD /u davidjw 11



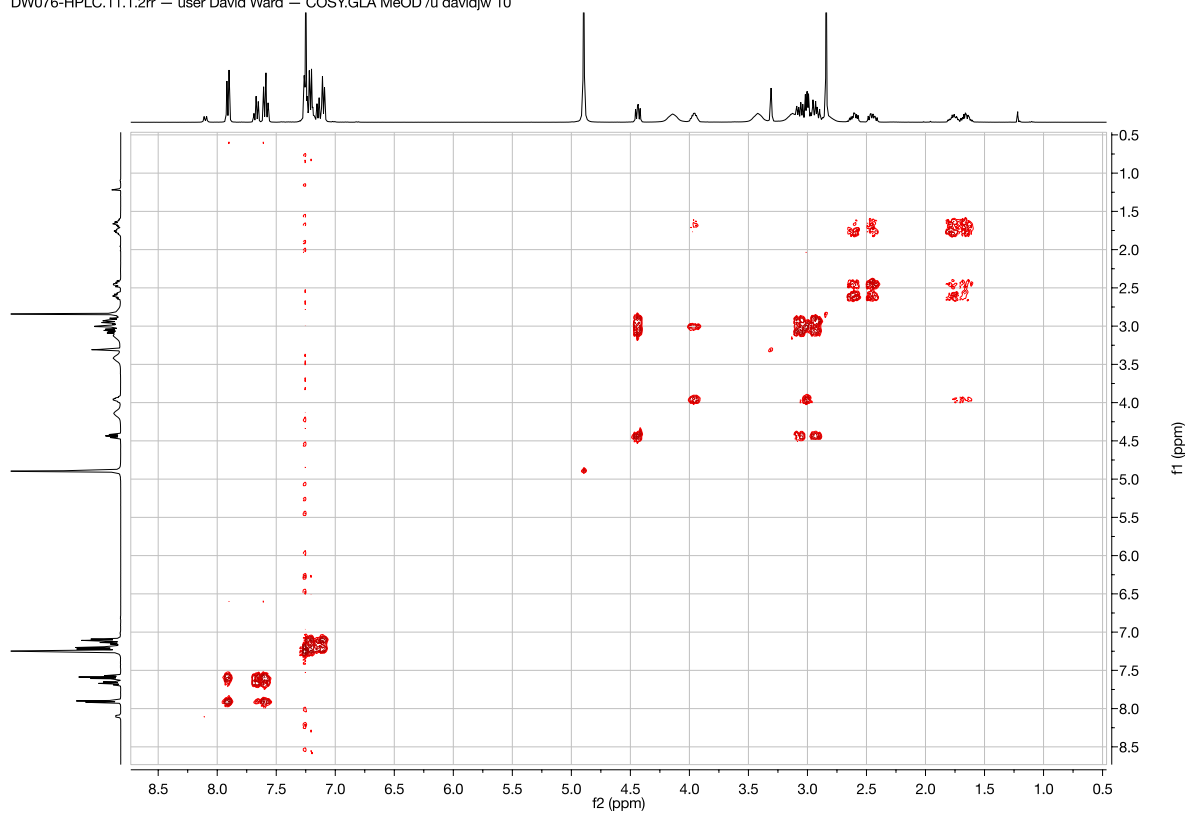




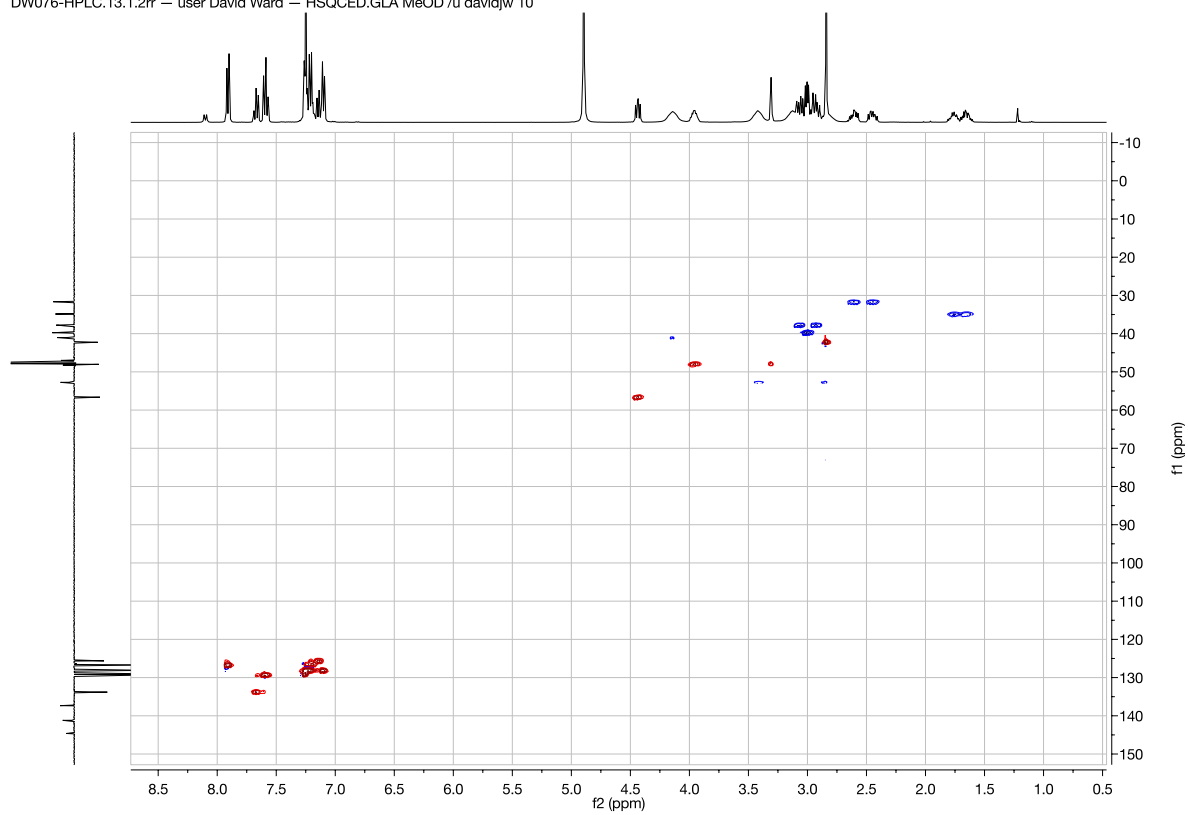
**Inhibitor 8**

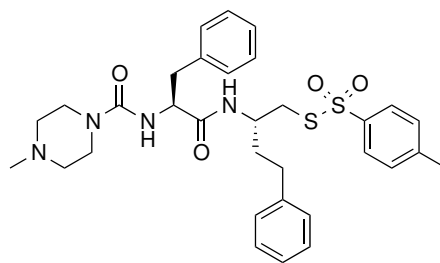


DW076-HPLC.11.1.2rr — user David Ward — COSY.GLA MeOD /u davidjw 10

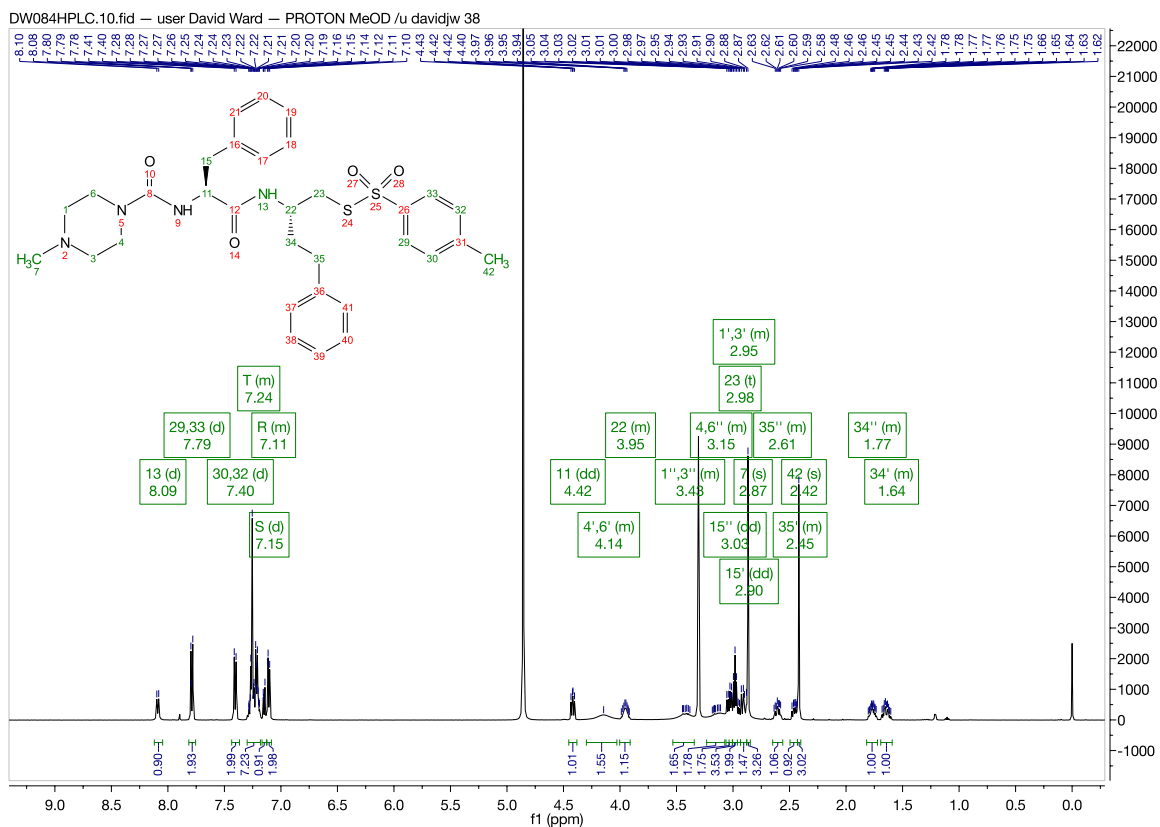


DW076-HPLC.13.1.2rr — user David Ward — HSQCED.GLA MeOD /u davidjw 10

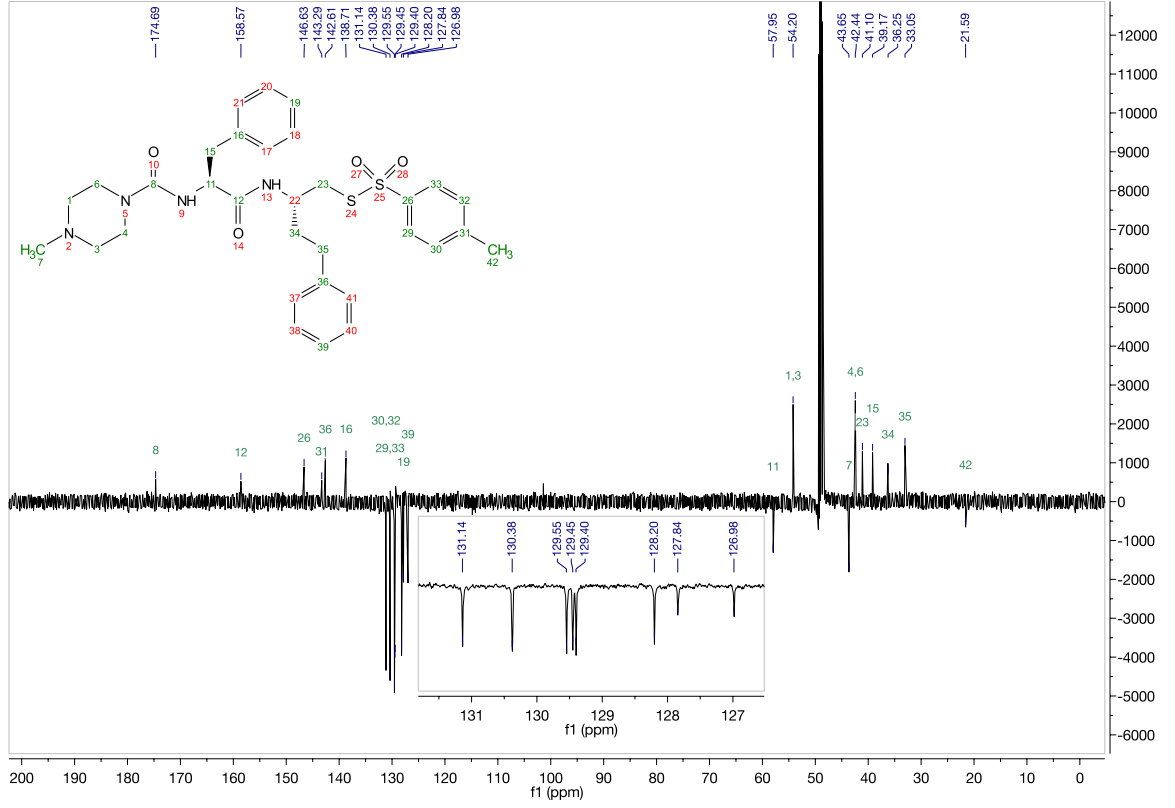




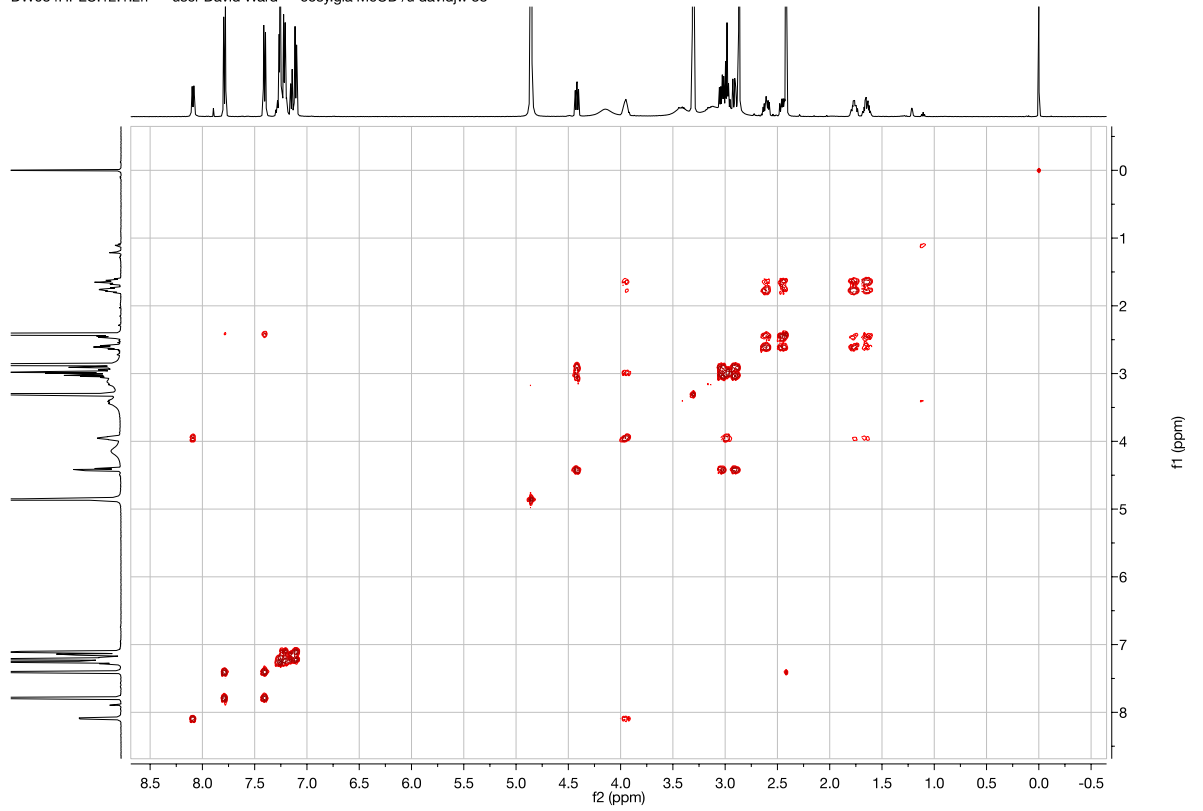
Inhibitor 9

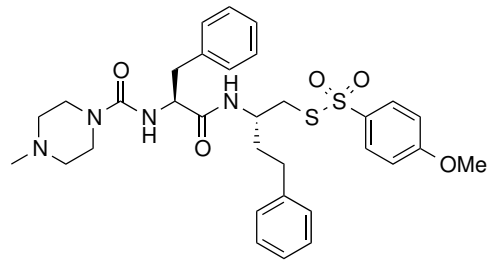


DW084HPLC.13.fid — user David Ward

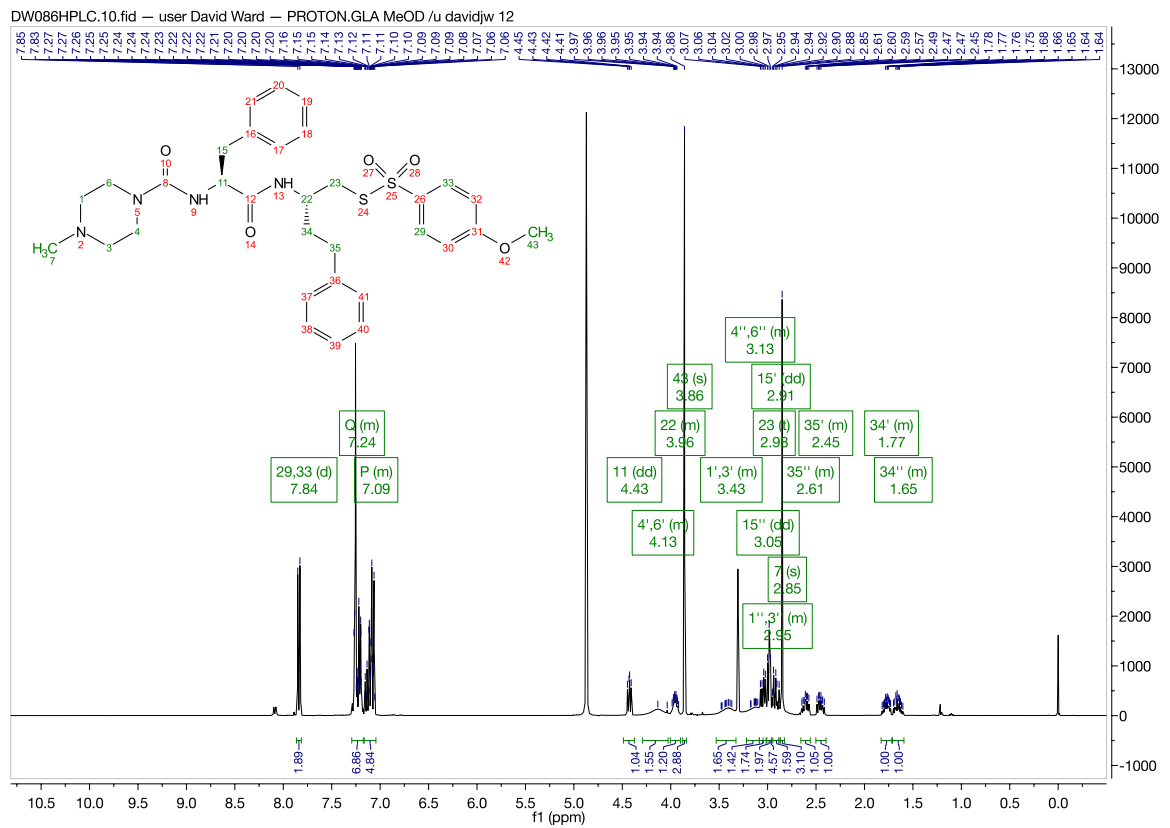


DW084HPLC.12.1.2rr — user David Ward — cosy.gla MeOD /u davidjw 38

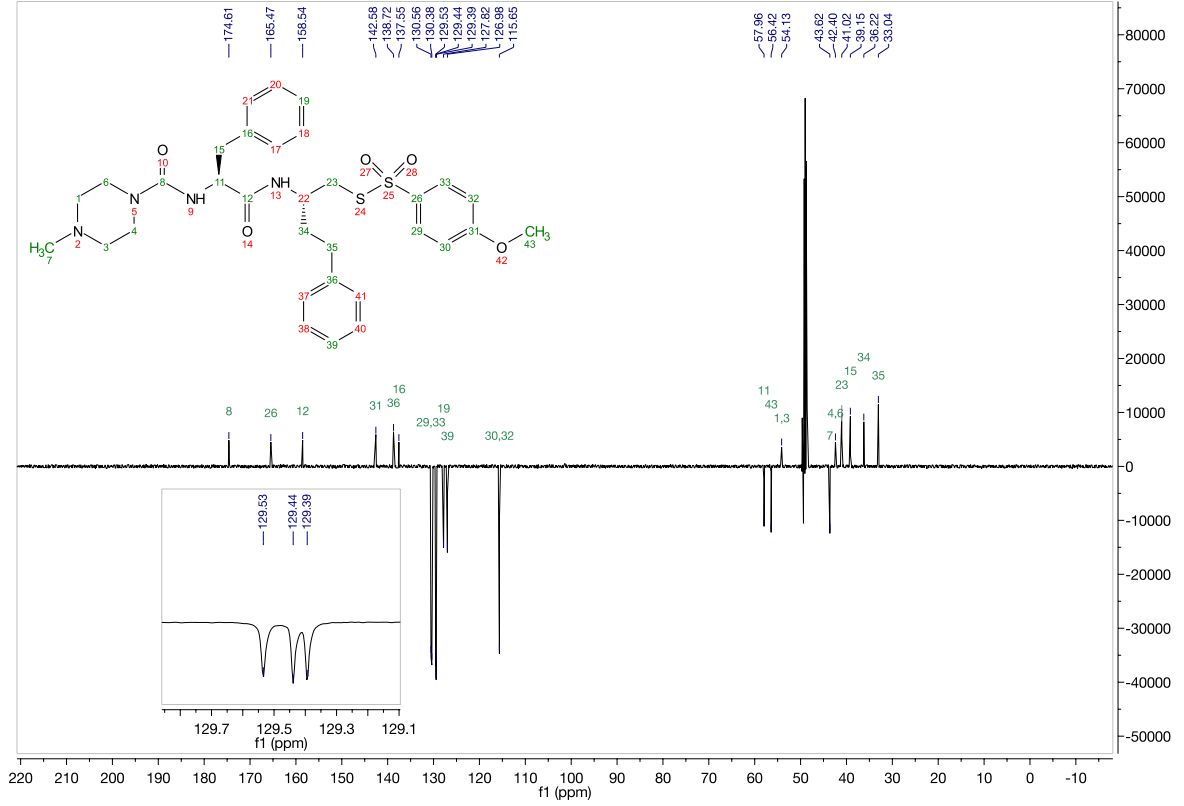




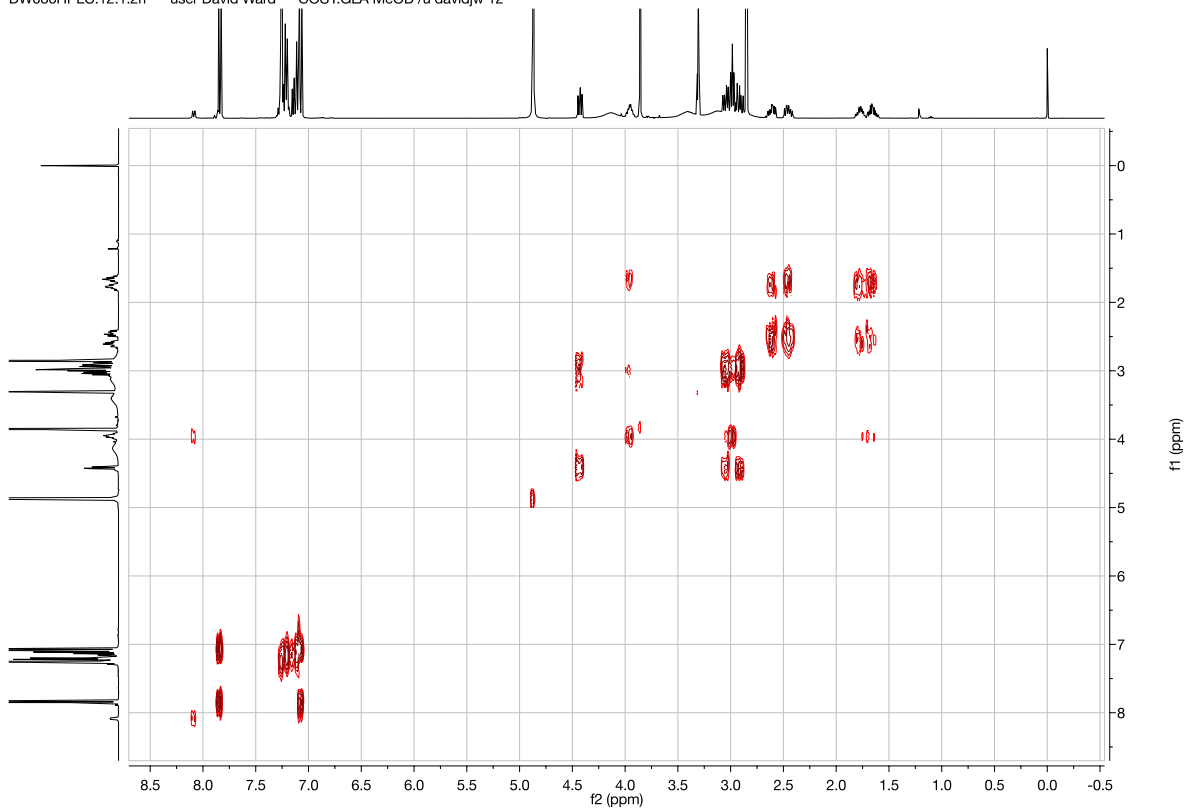
Inhibitor 10

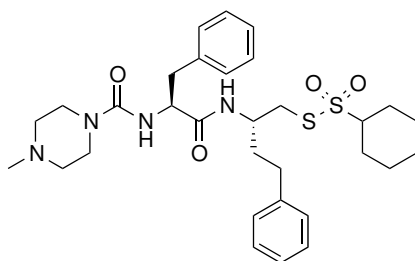
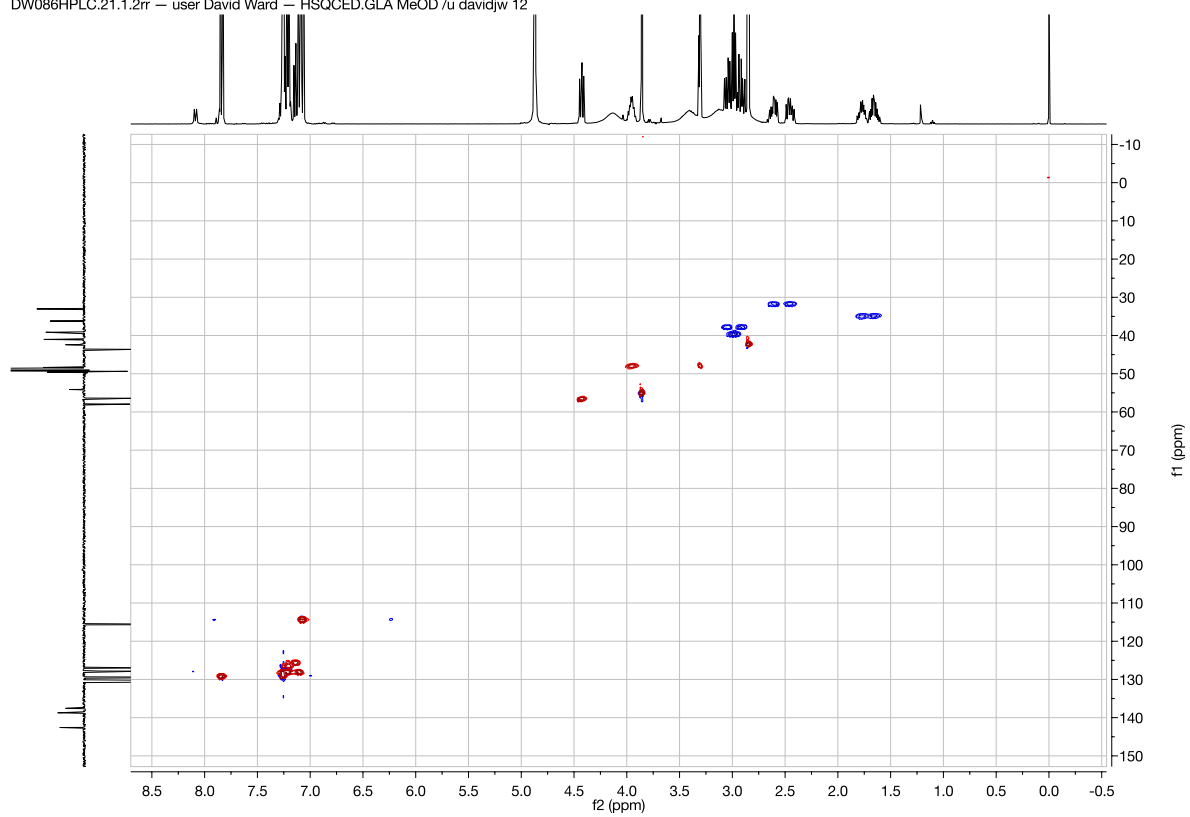


DW086HPLC.20.fid — user David Ward — C13DEPTQ.GLA MeOD /u davidjw 12



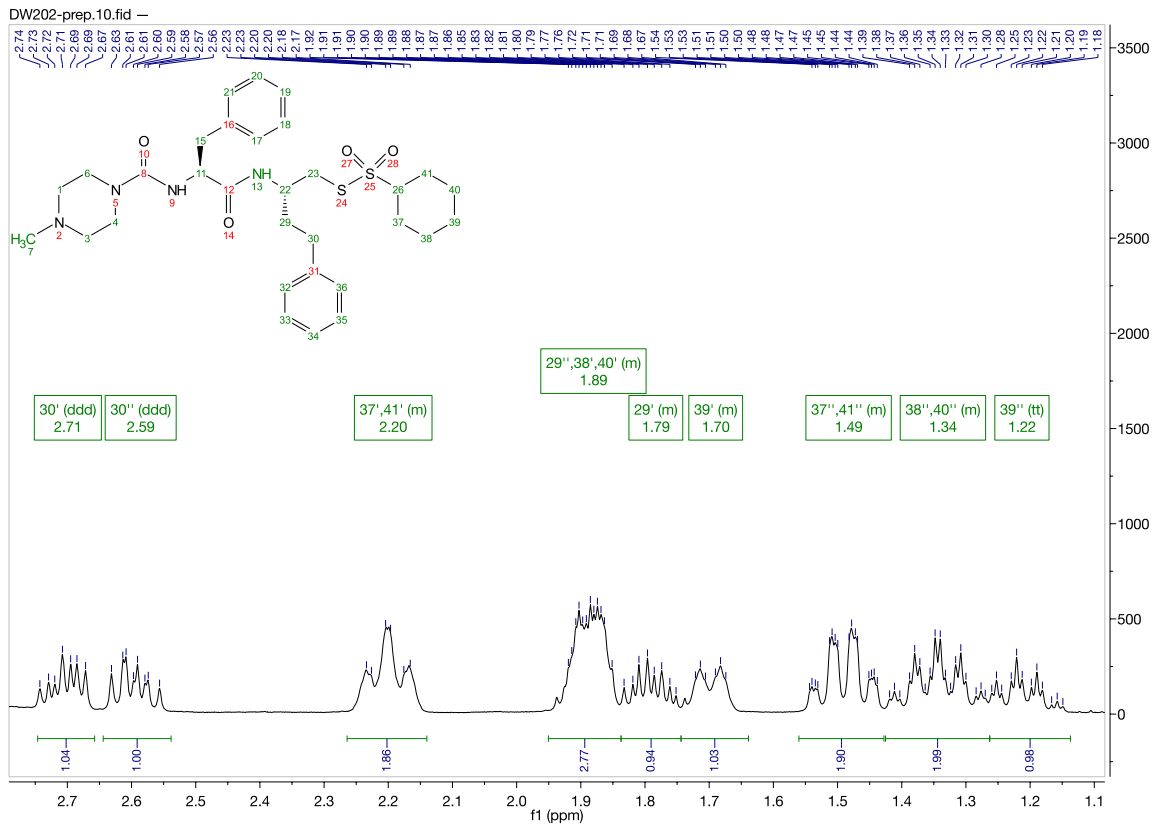
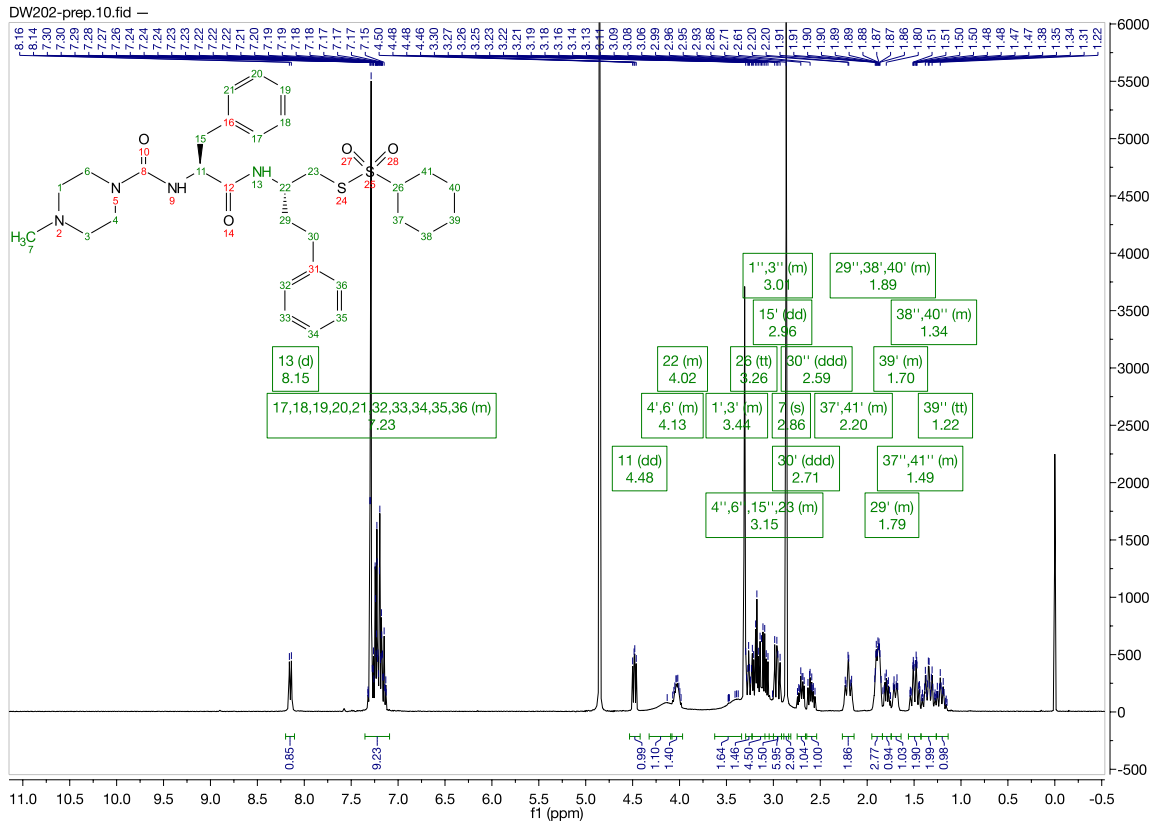
DW086HPLC.12.1.2rr — user David Ward — COSY.GLA MeOD /u davidjw 12



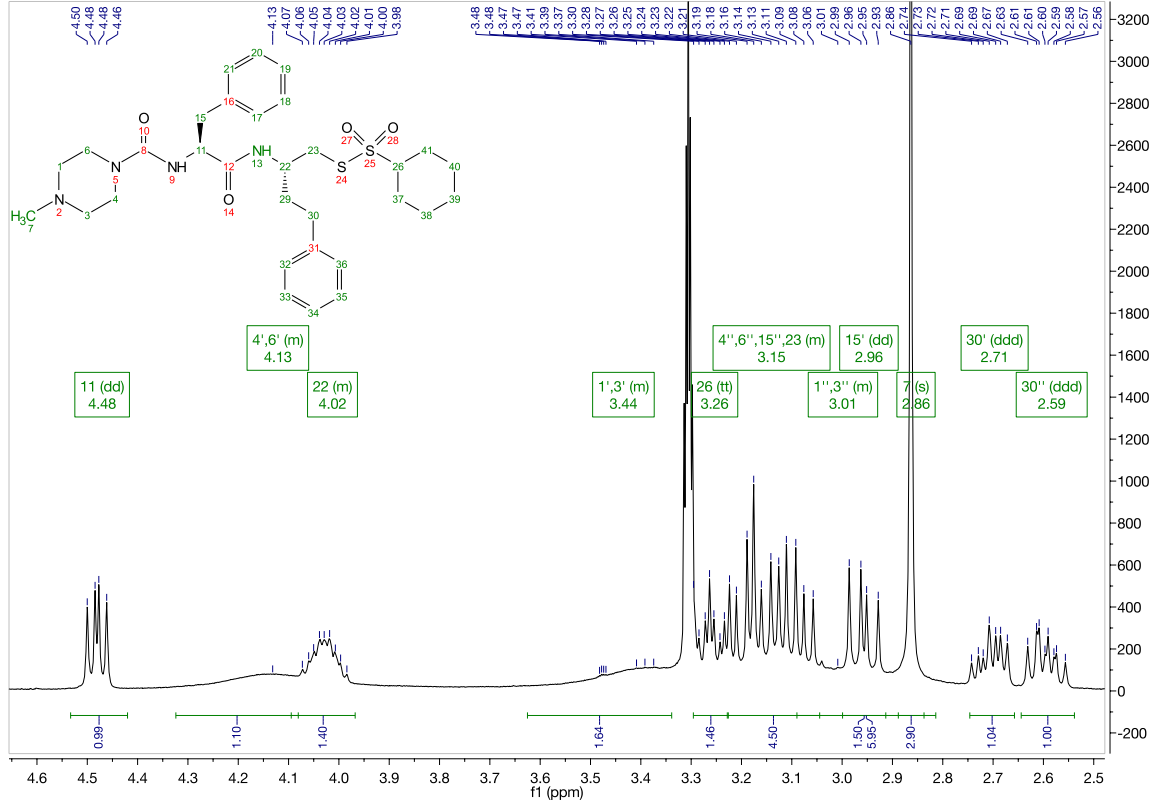


**Inhibitor 11**

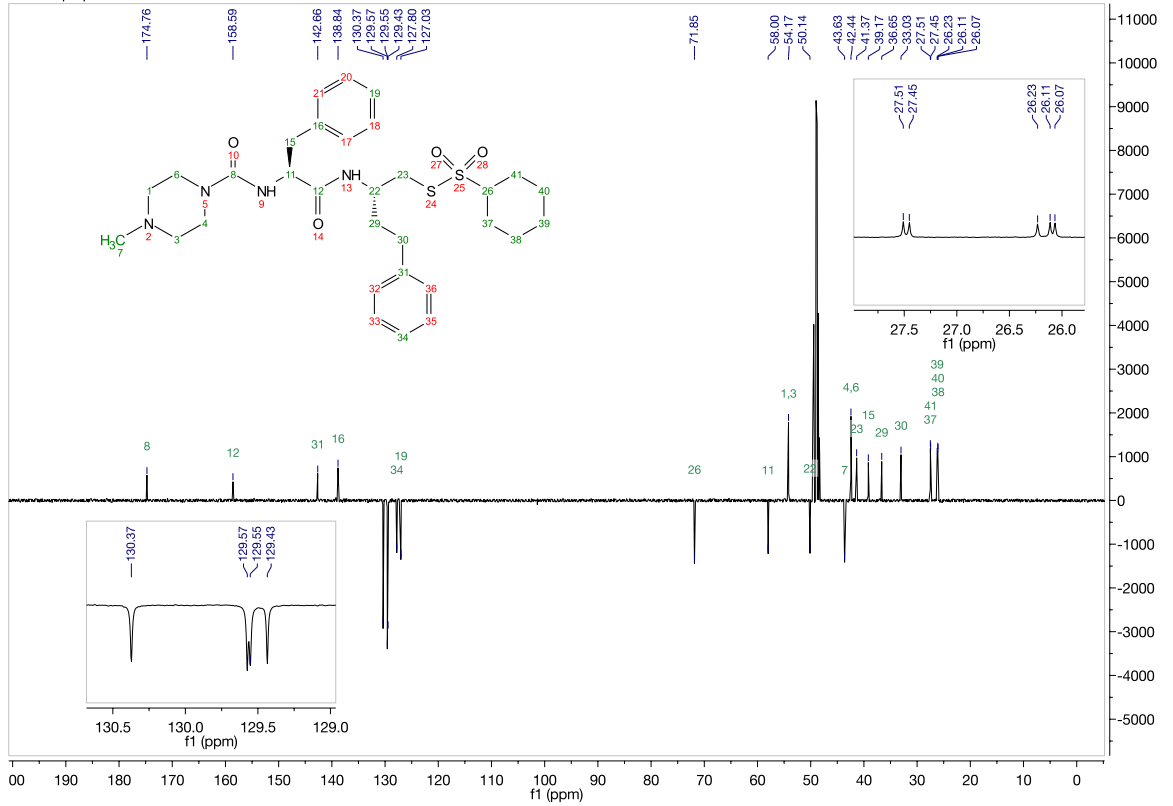


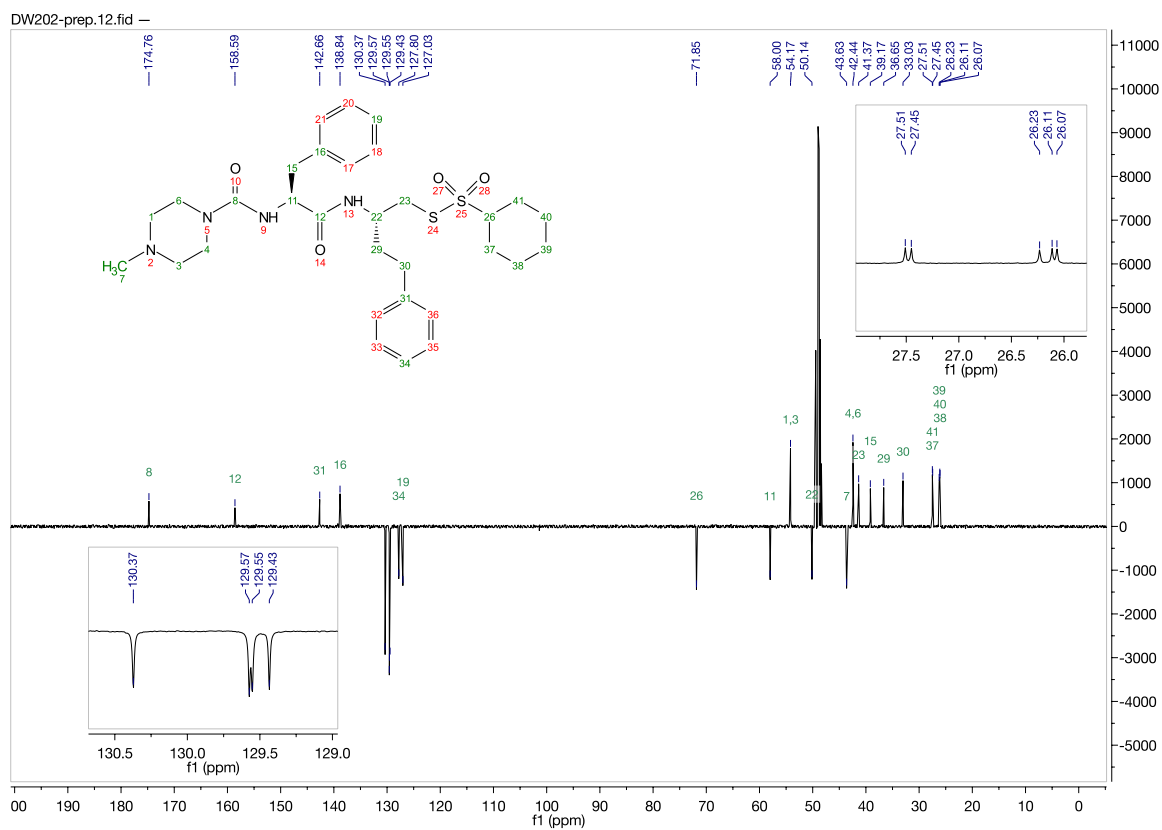
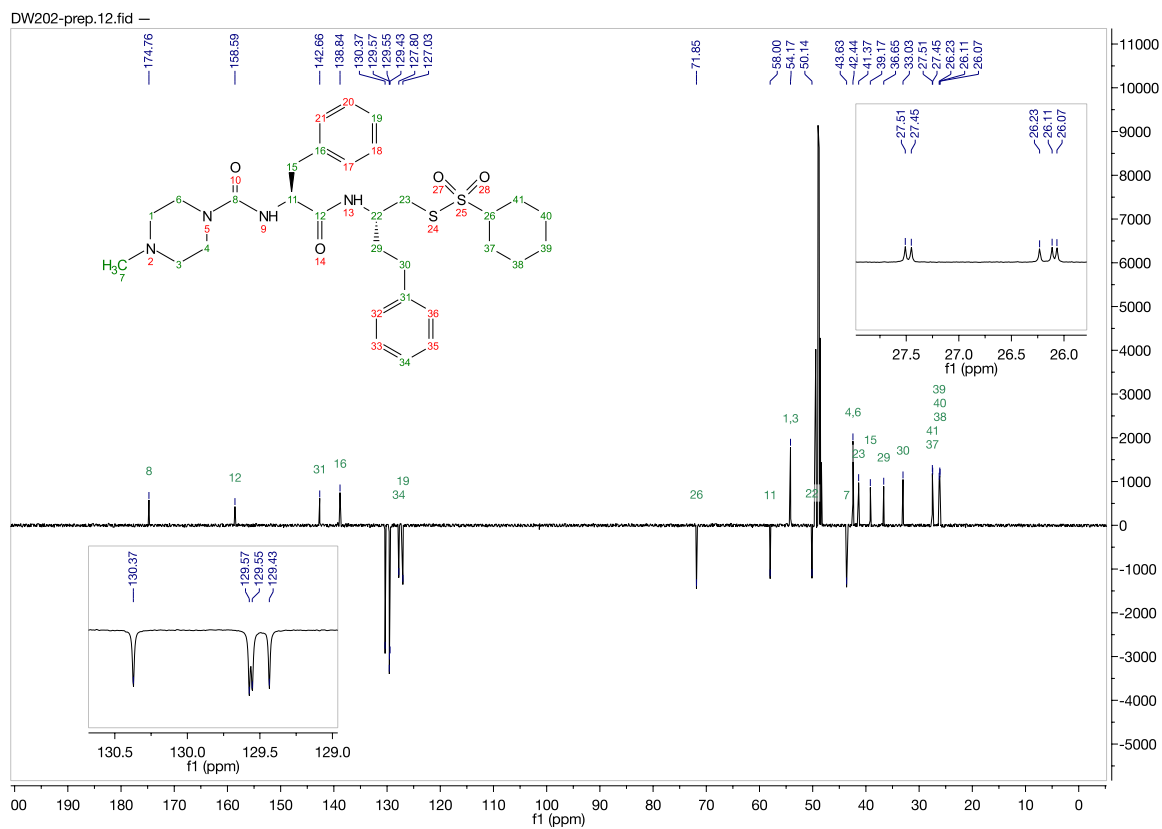


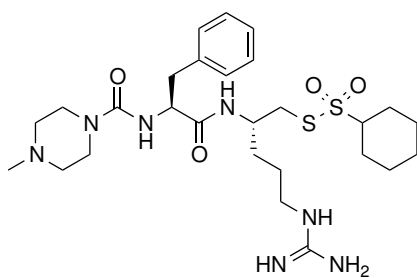
DW202-prep.10.fid —



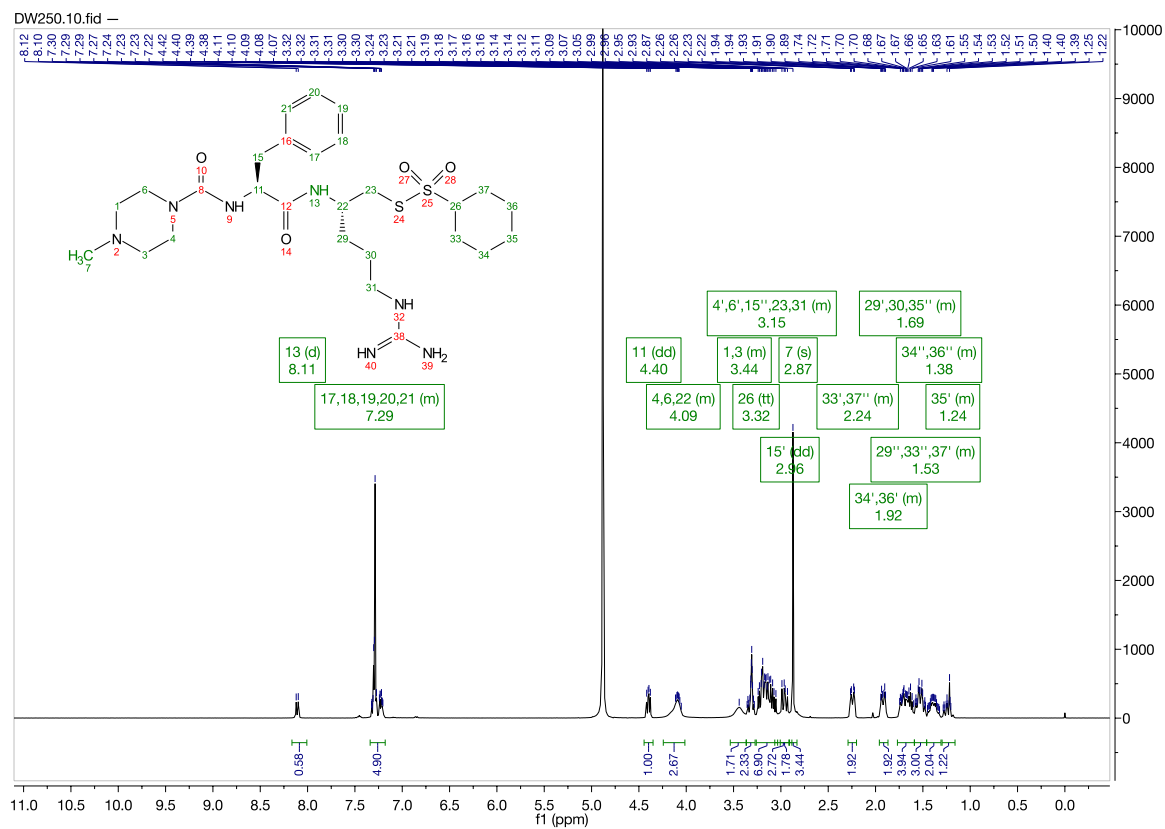
DW202-prep.12.fid —



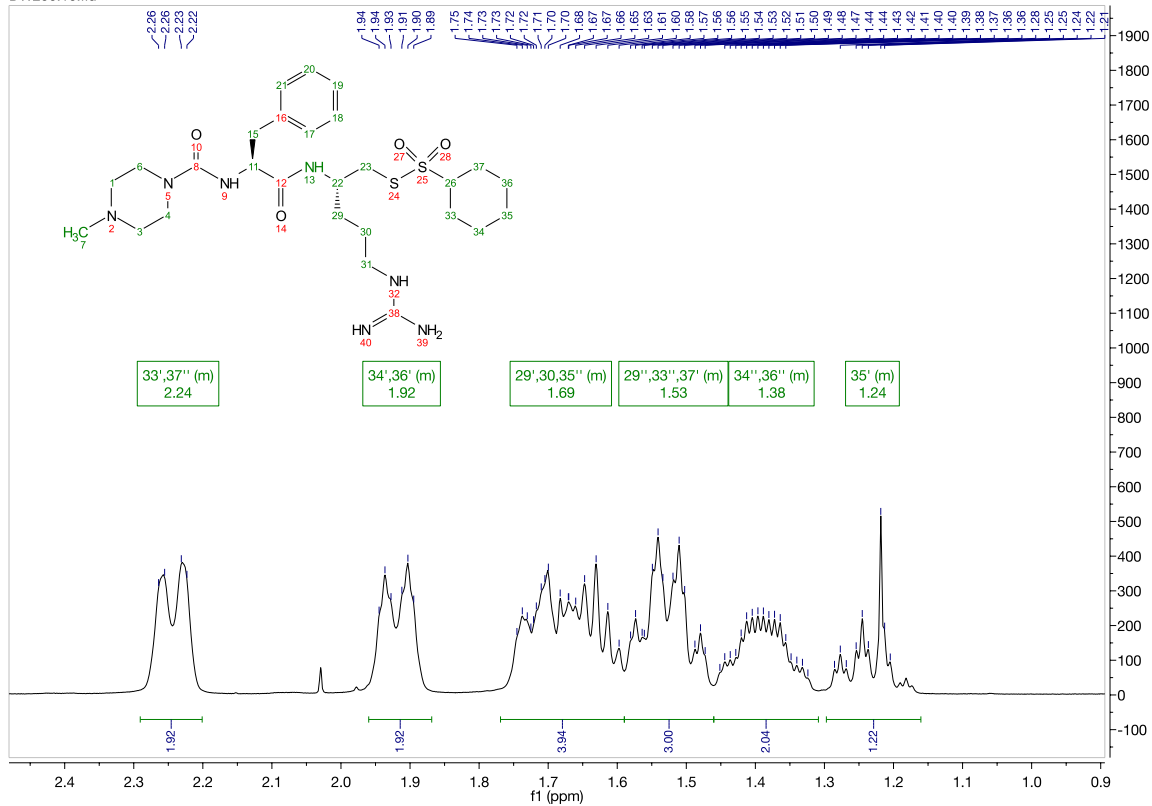




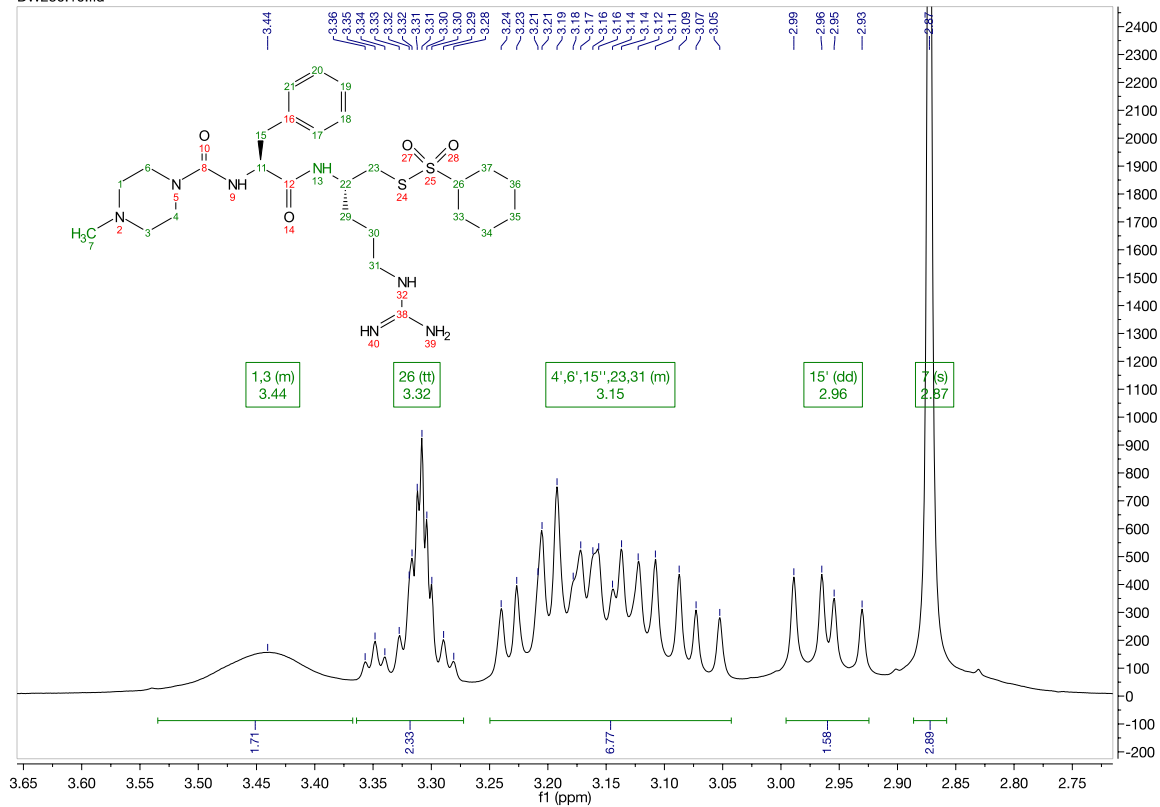
Inhibitor 17

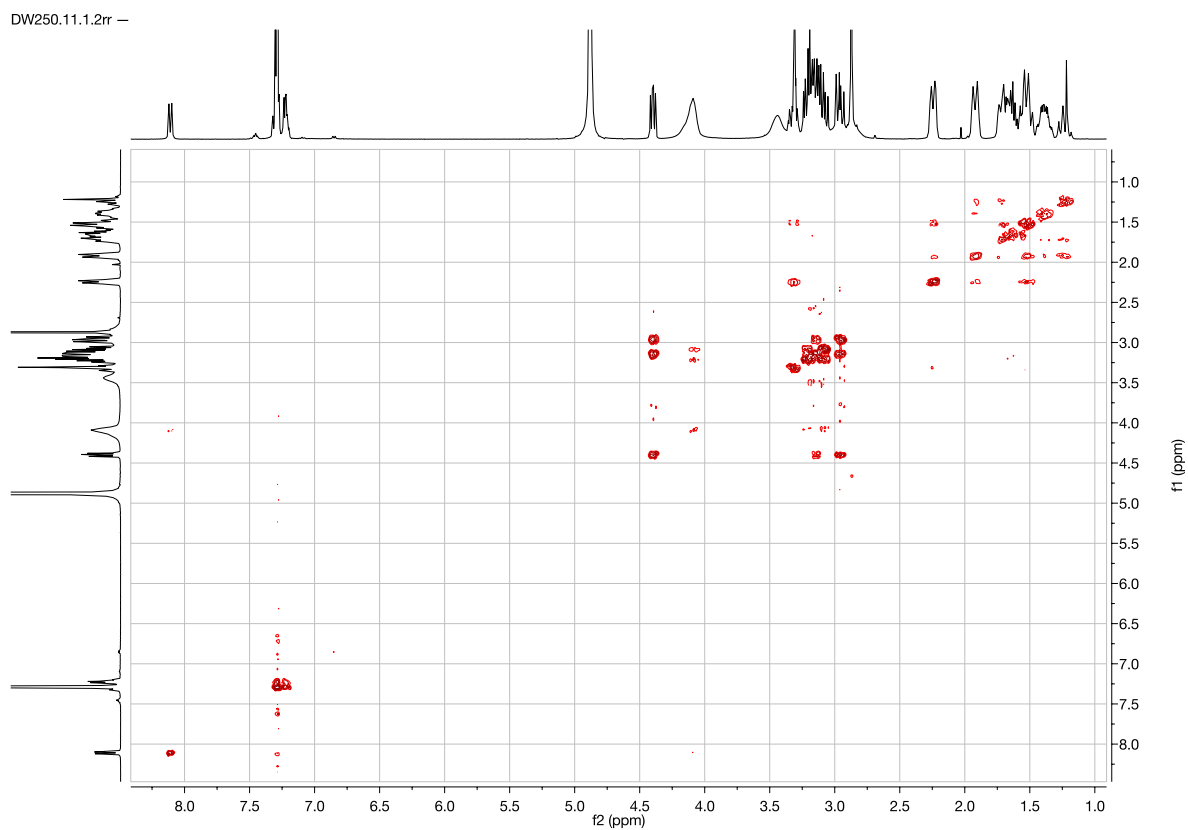
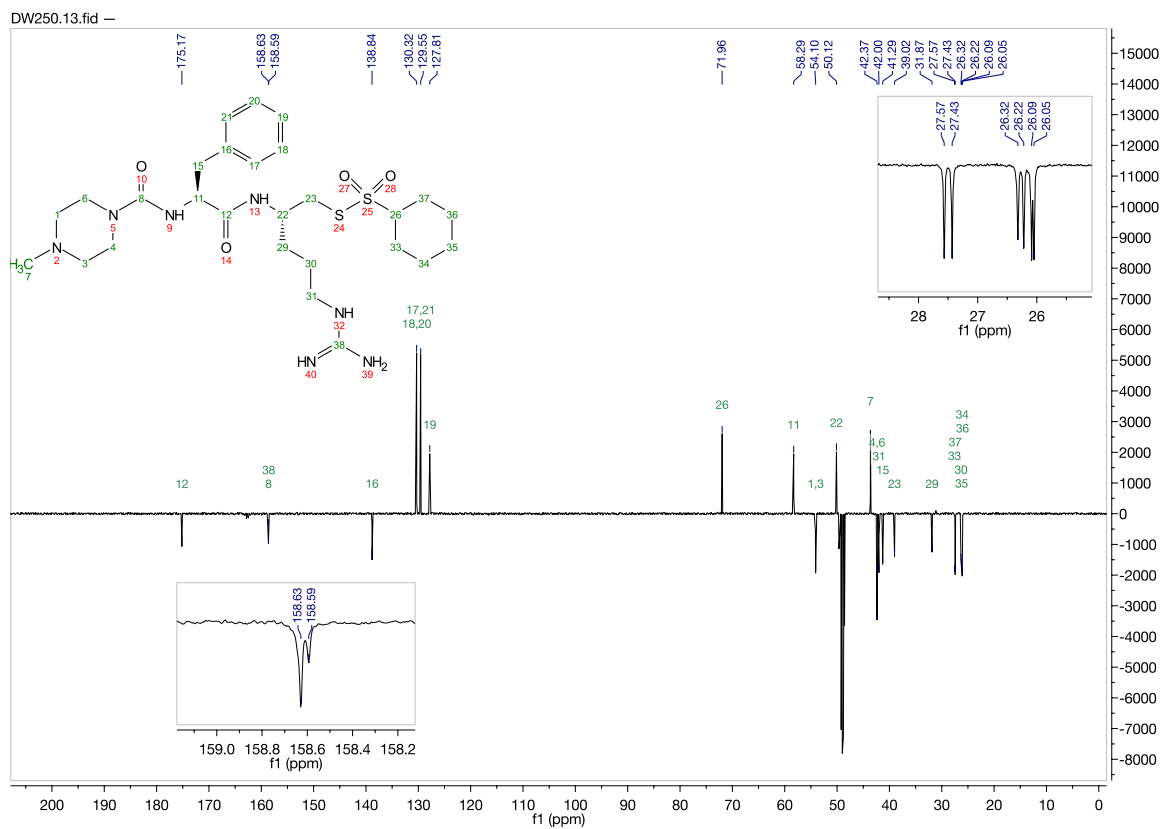


DW250.10.fid -

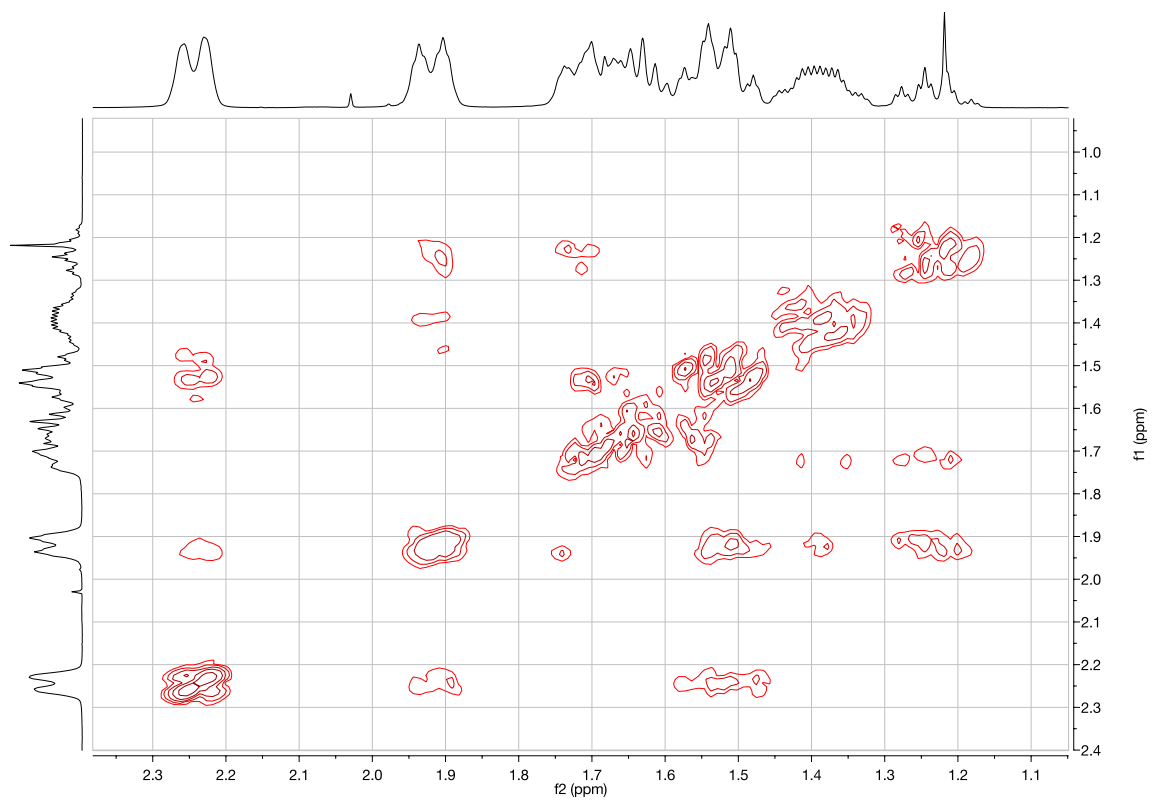


DW250.10.fid -

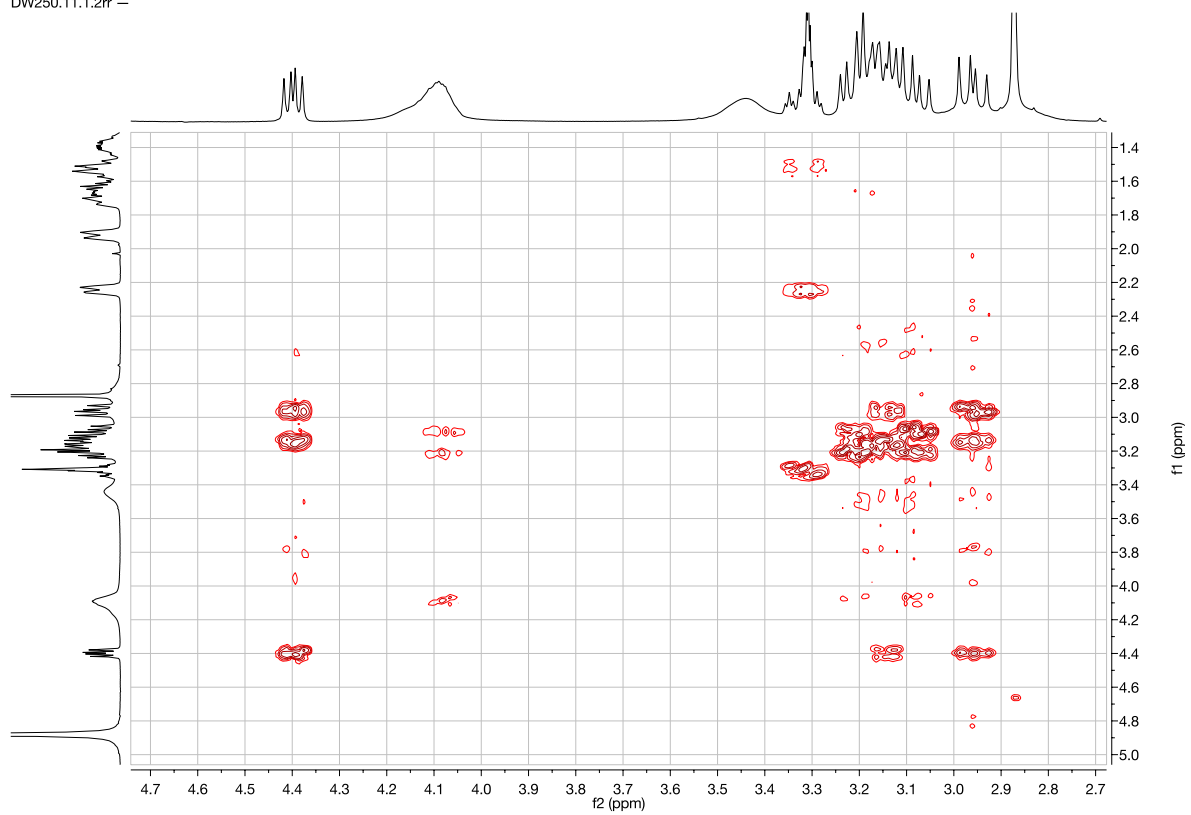




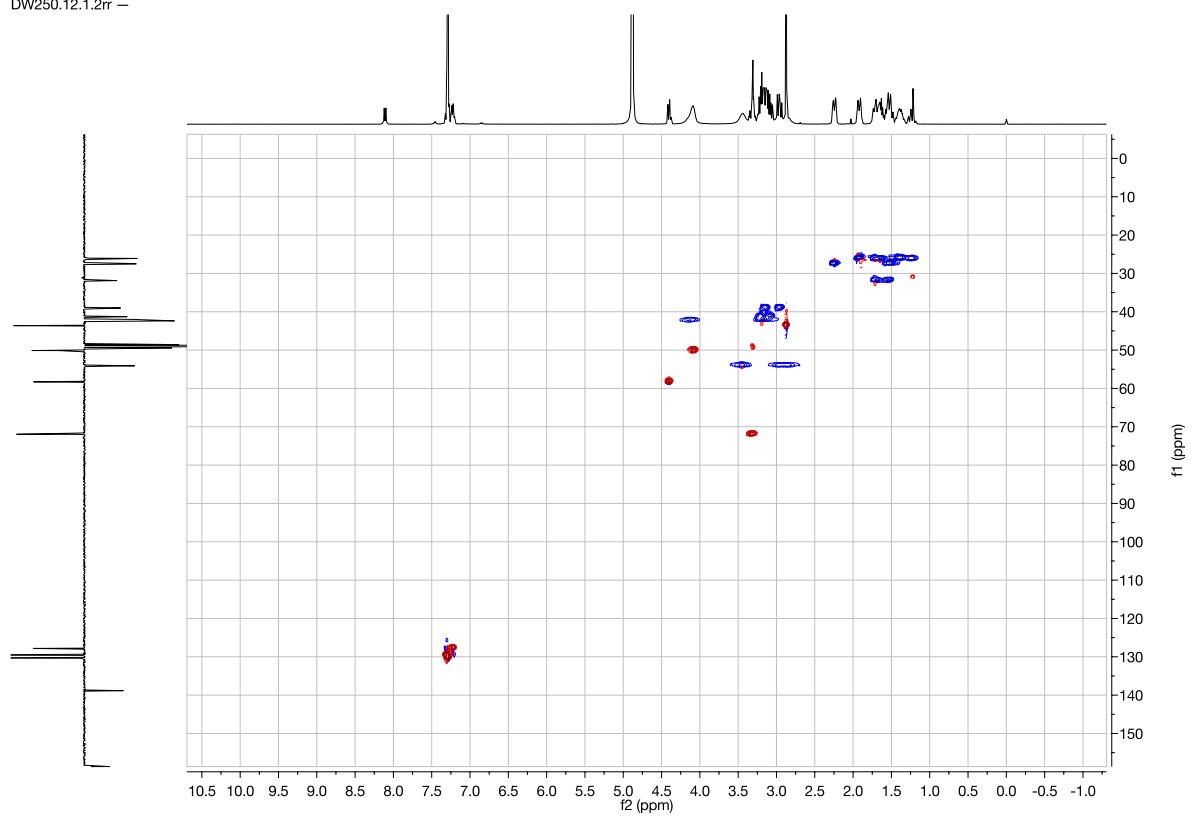
DW250.11.1.2rr -



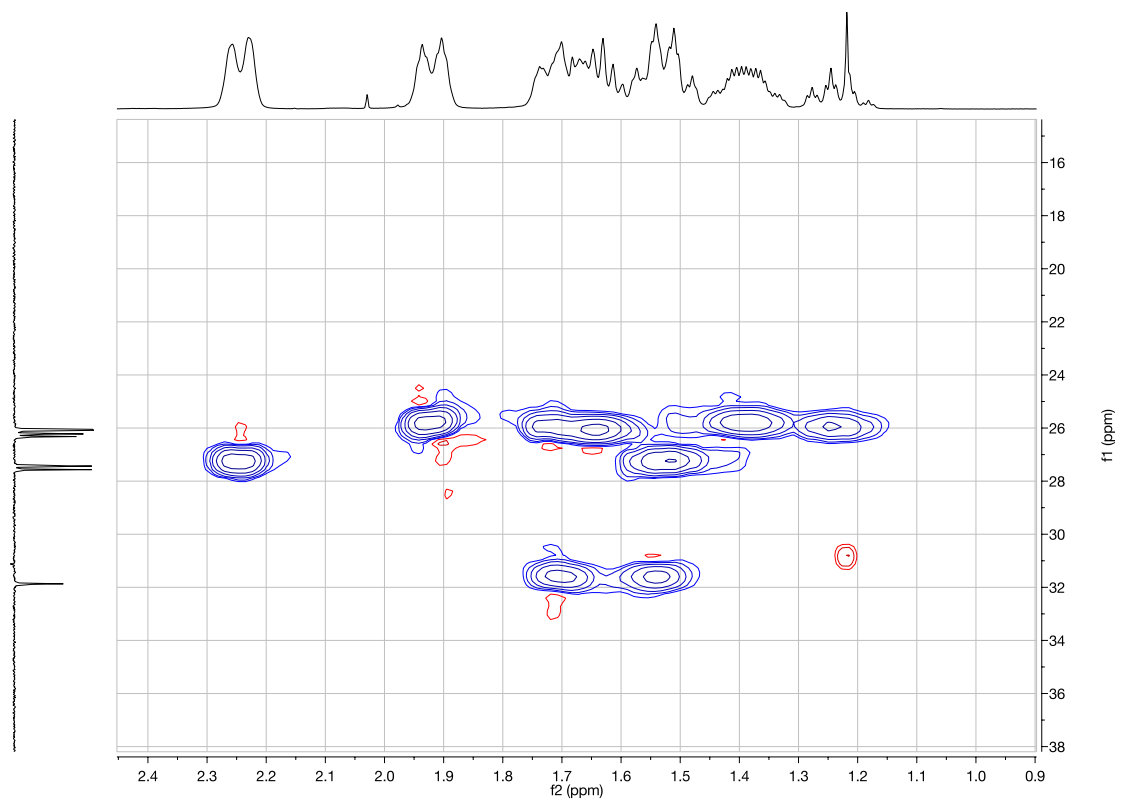
DW250.11.1.2rr -



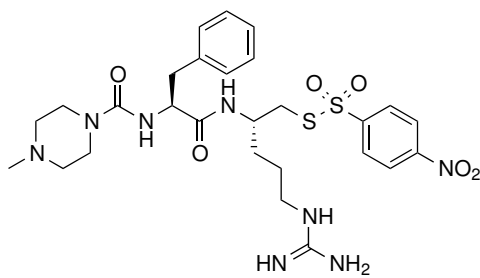
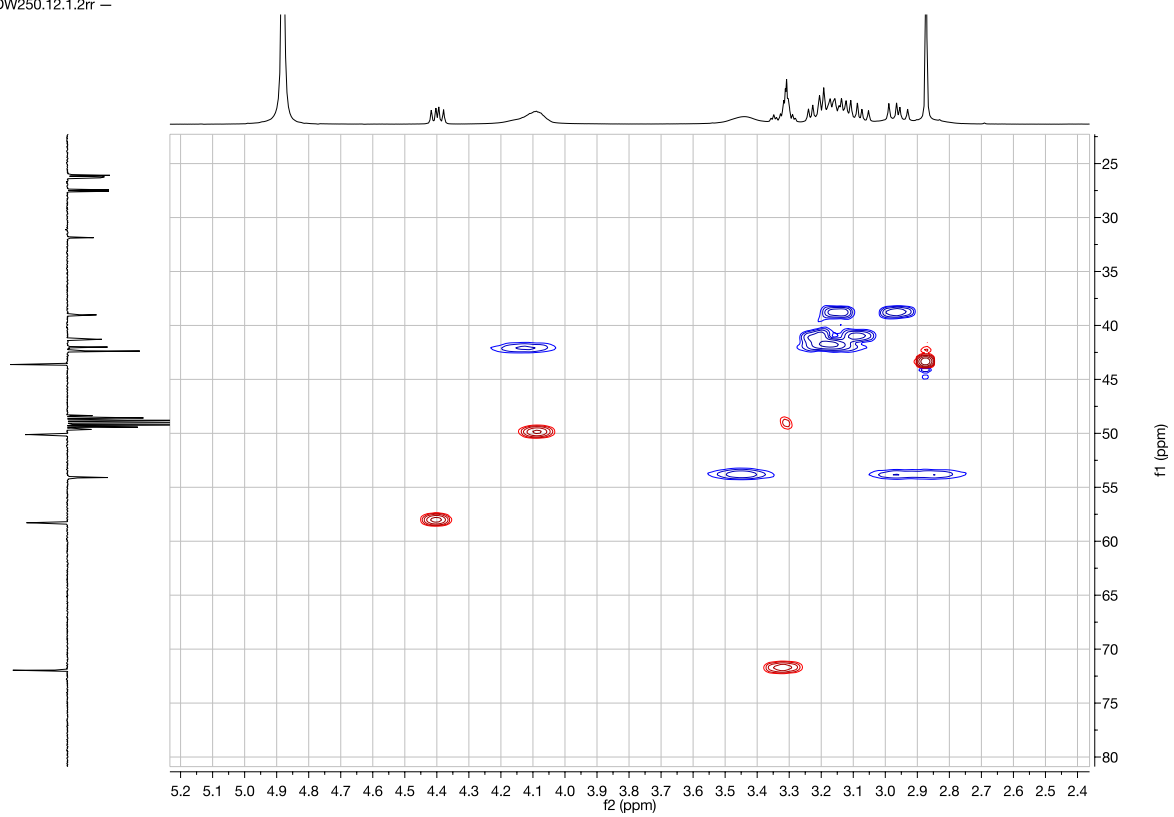
DW250.12.1.2rr -



DW250.12.1.2rr -

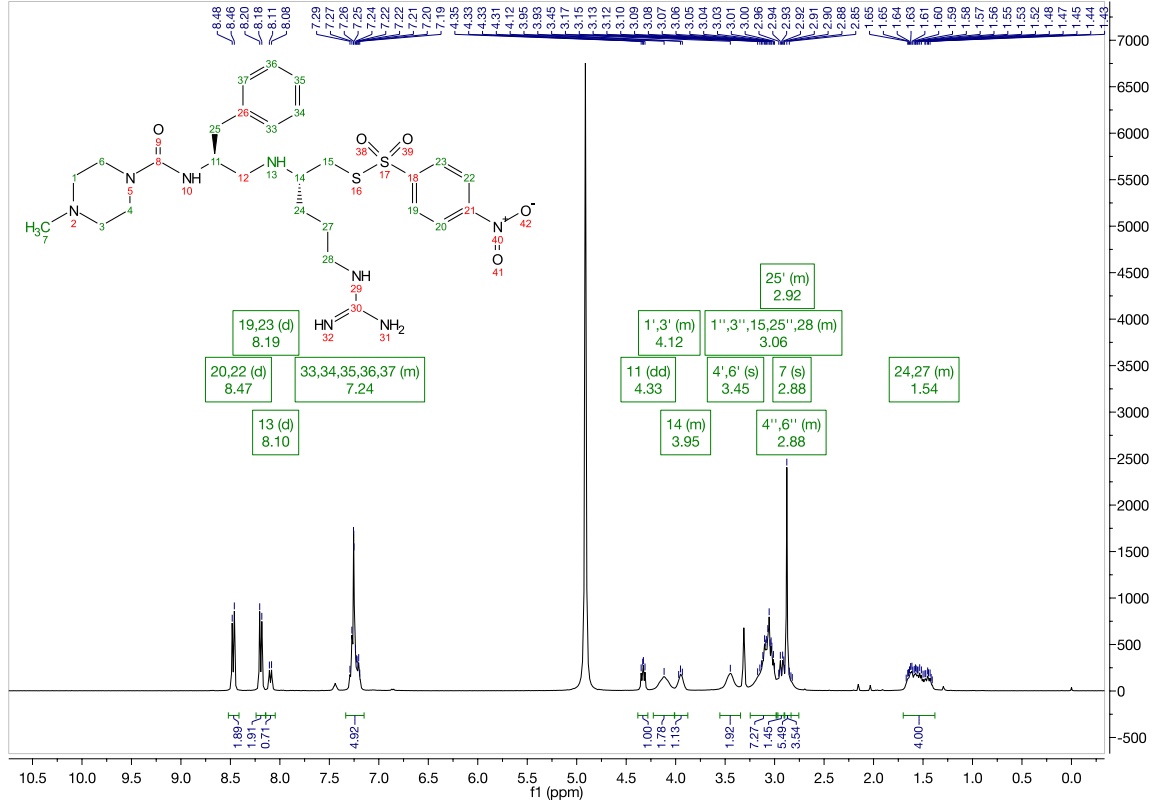




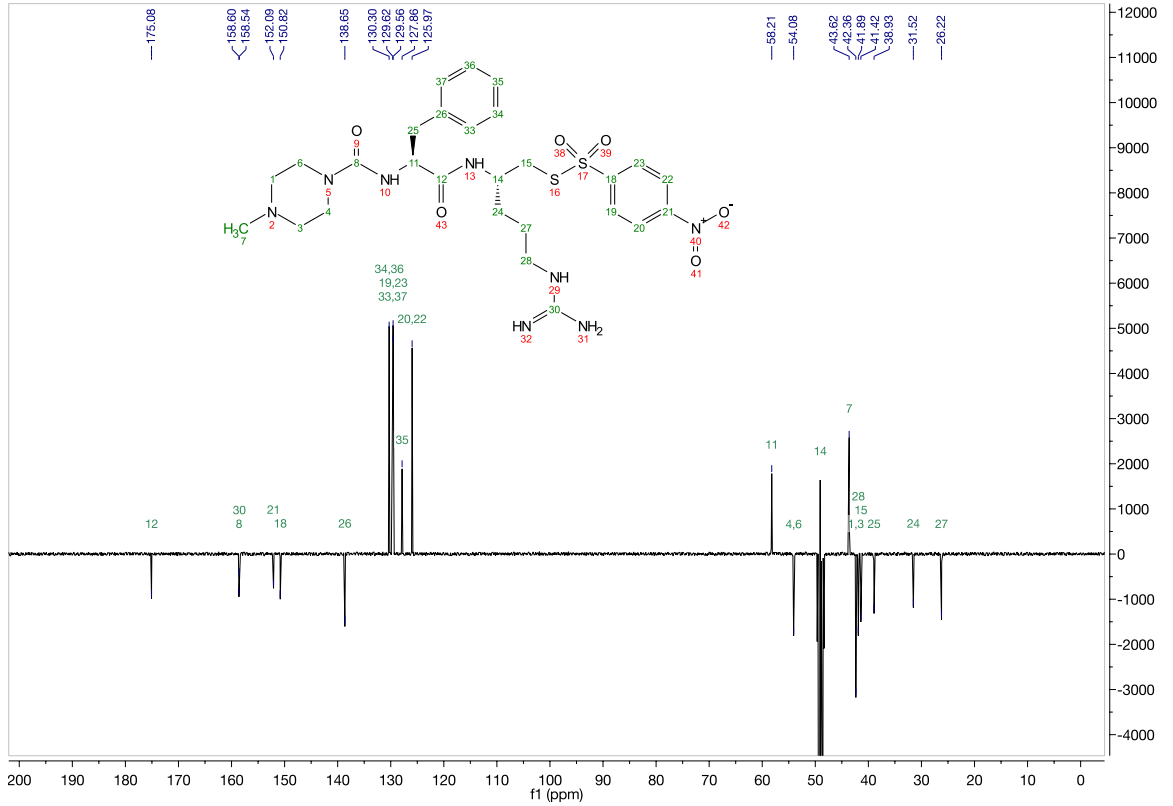


**Inhibitor 13**

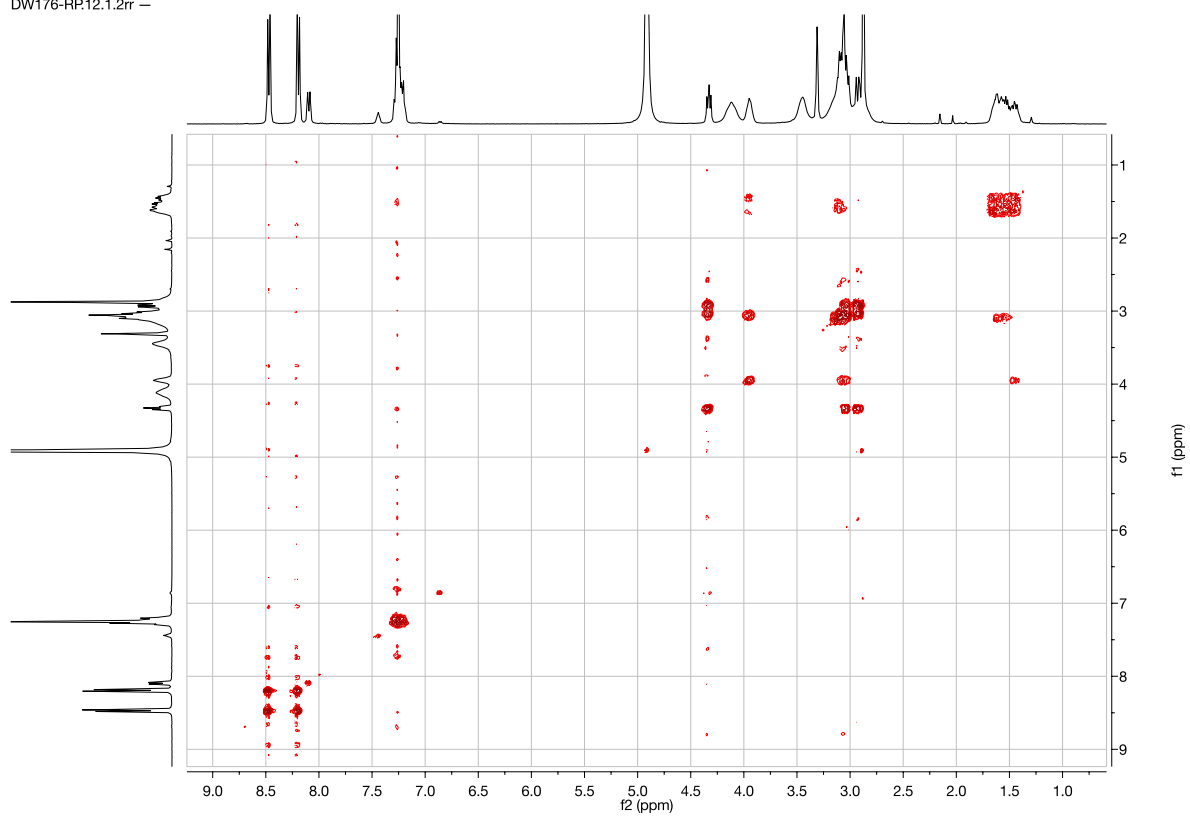
DW176-PP.10.fid —



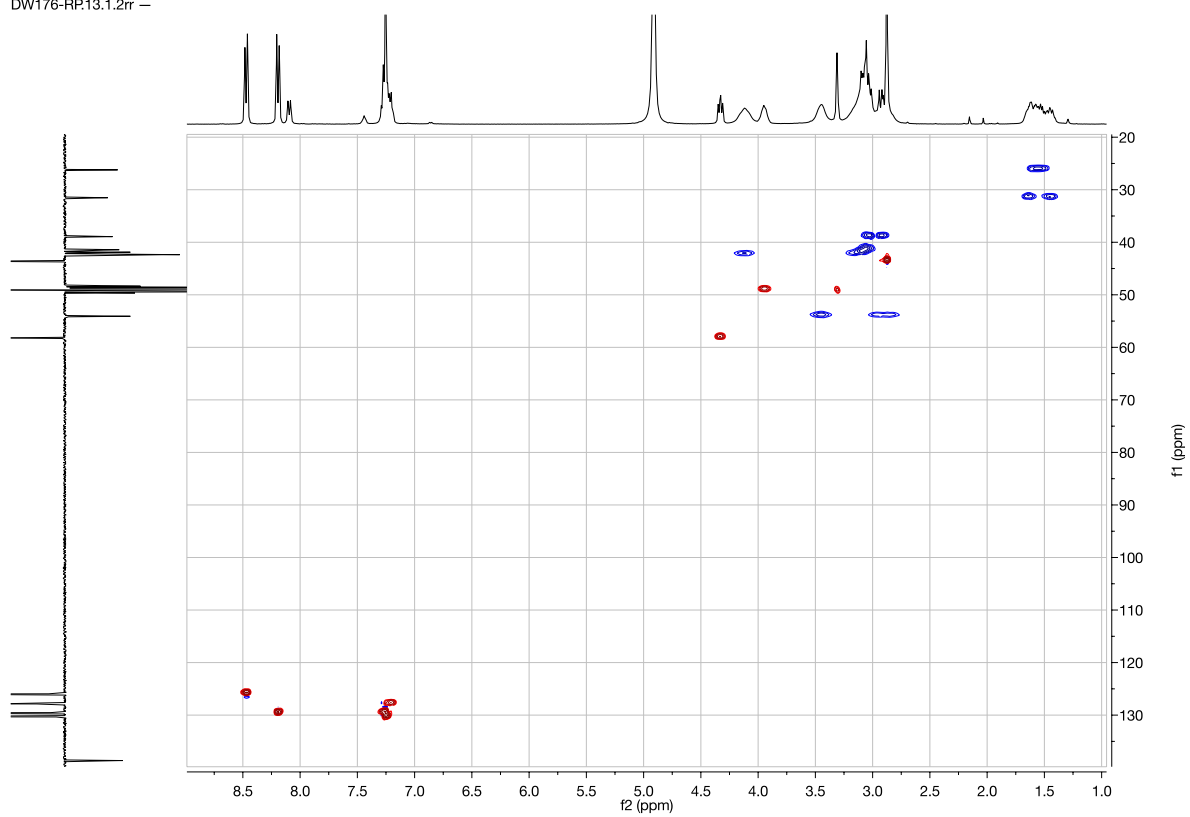
DW176-PP.11.fid —

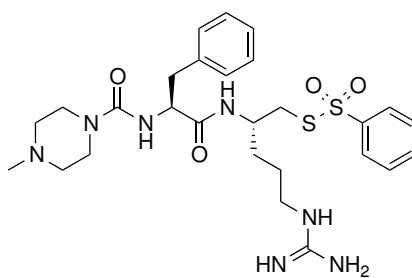


DW176-RP:12.1.2rr -

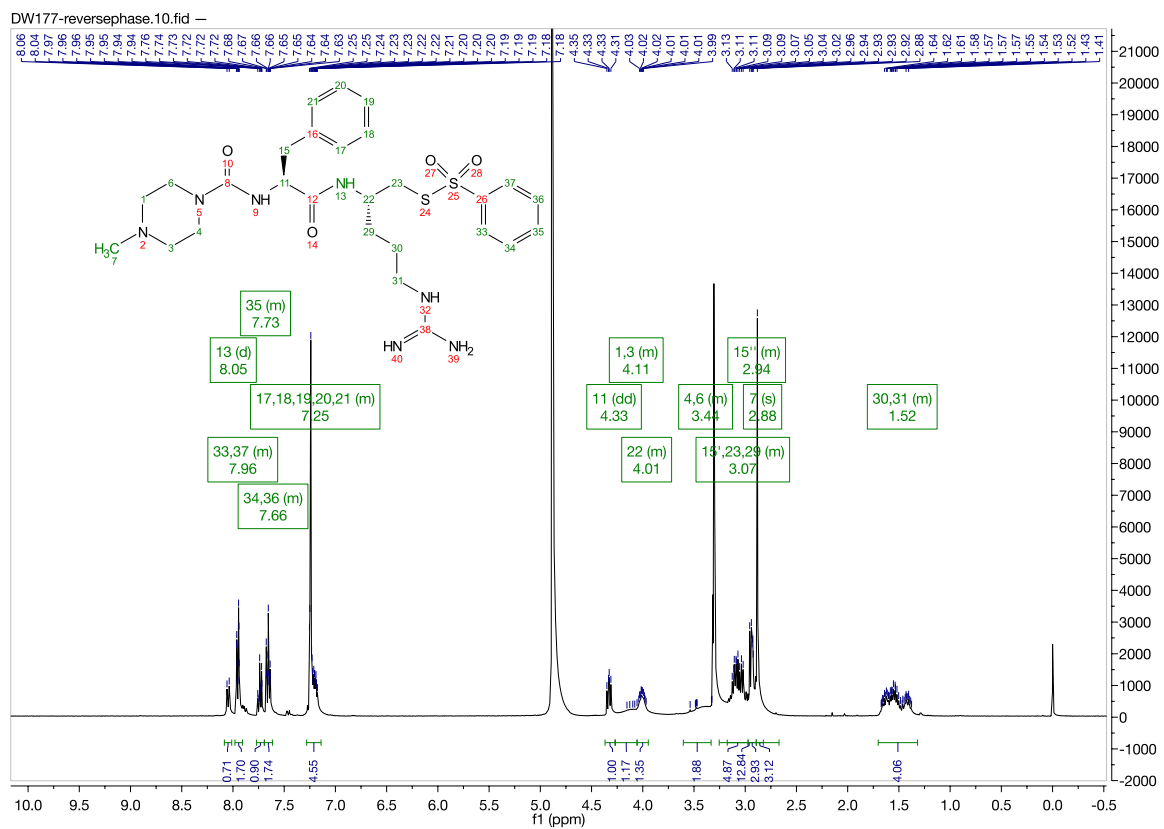


DW176-RP:13.1.2rr -

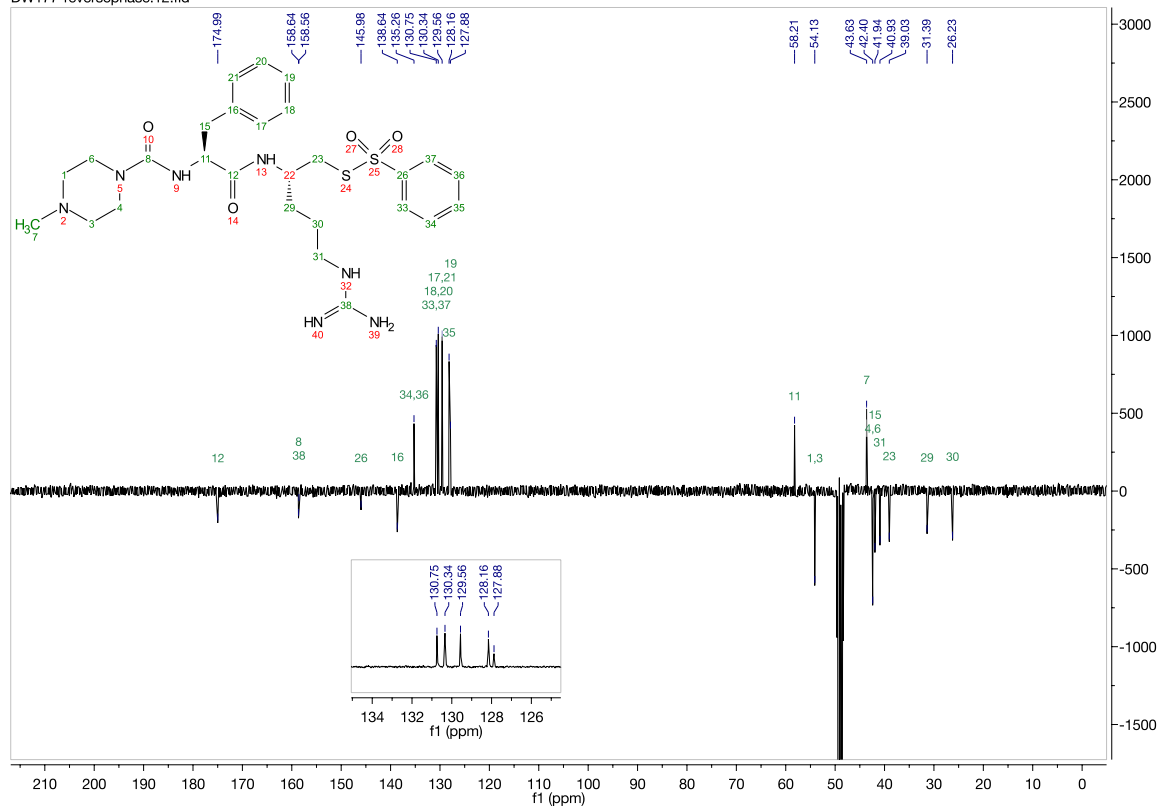




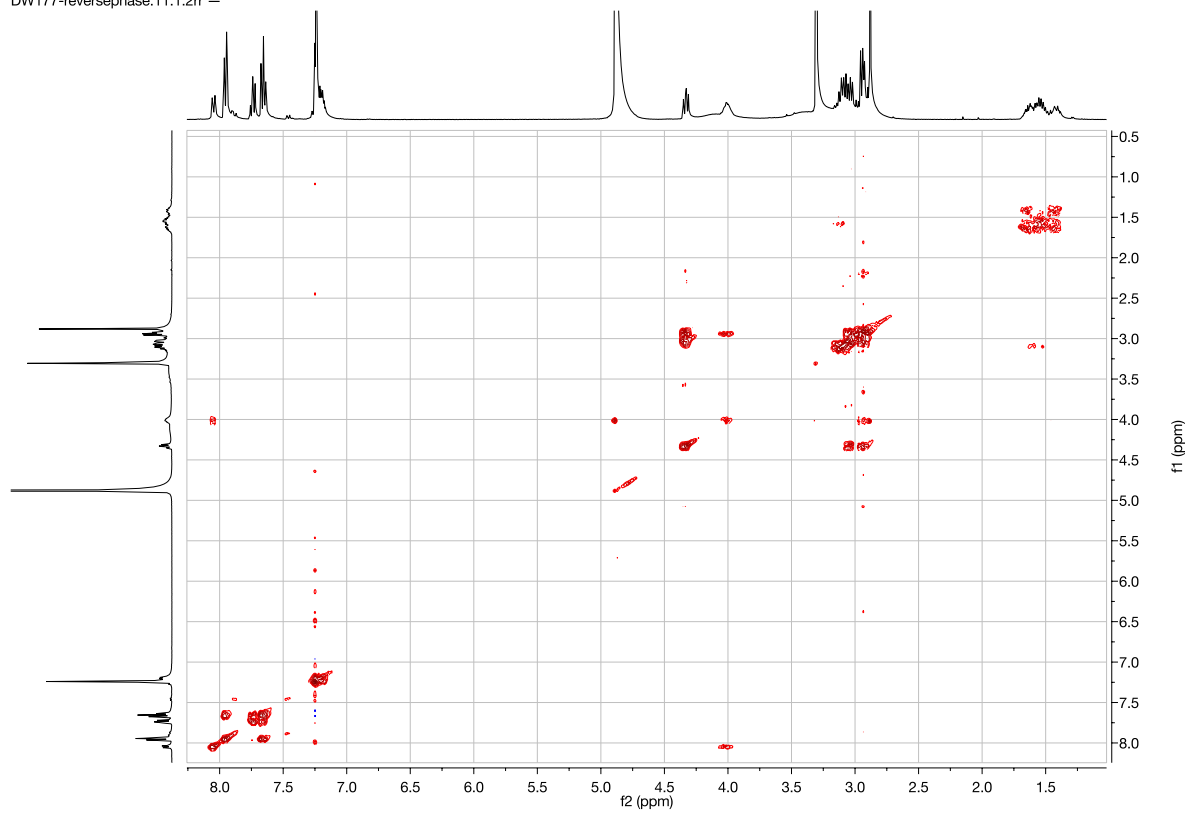
Inhibitor 14

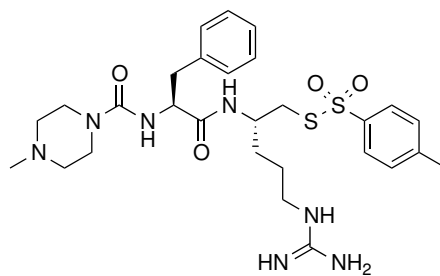


DW177-reversephase.12.fid -

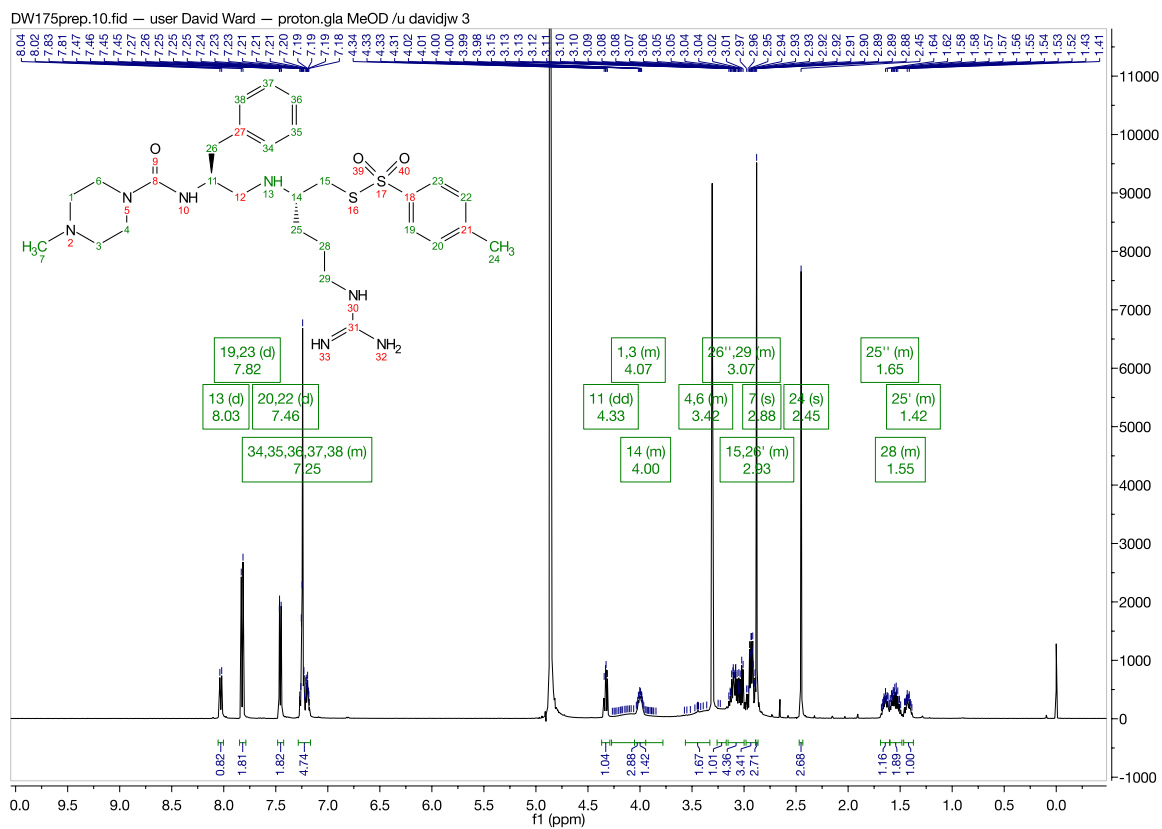


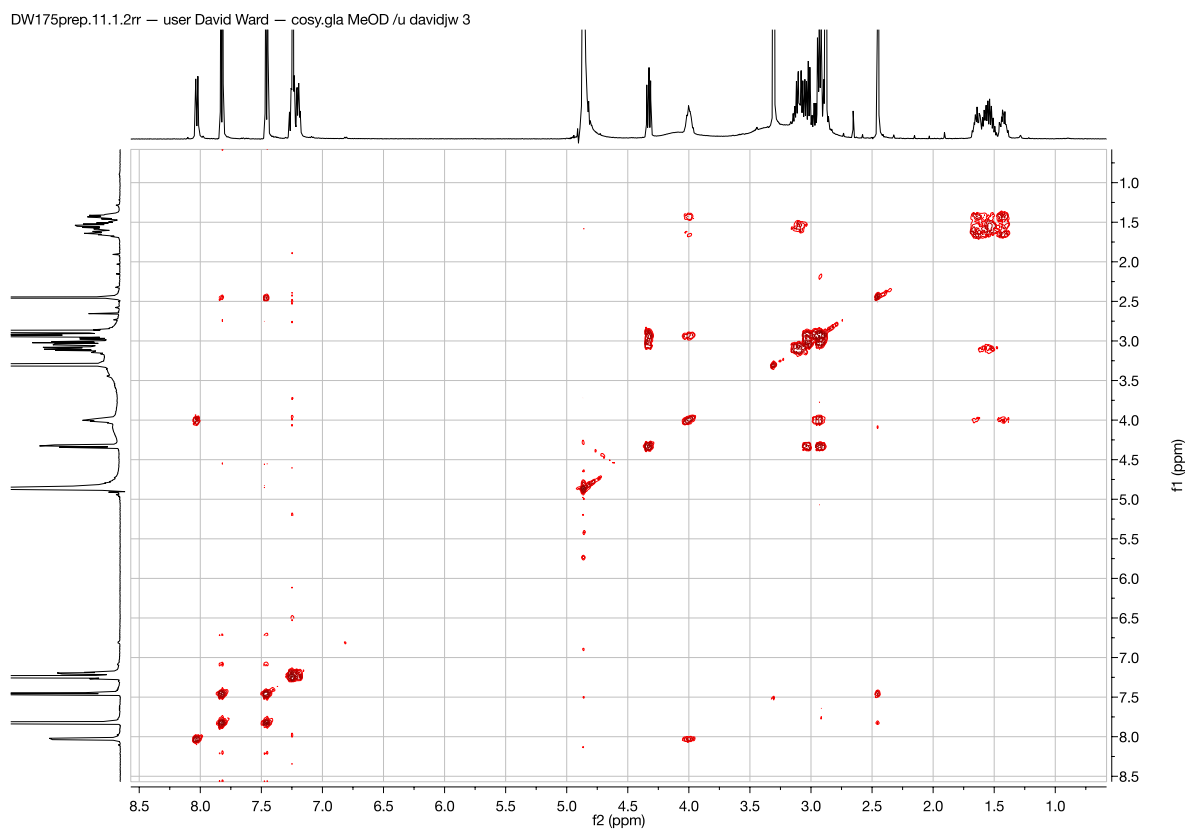
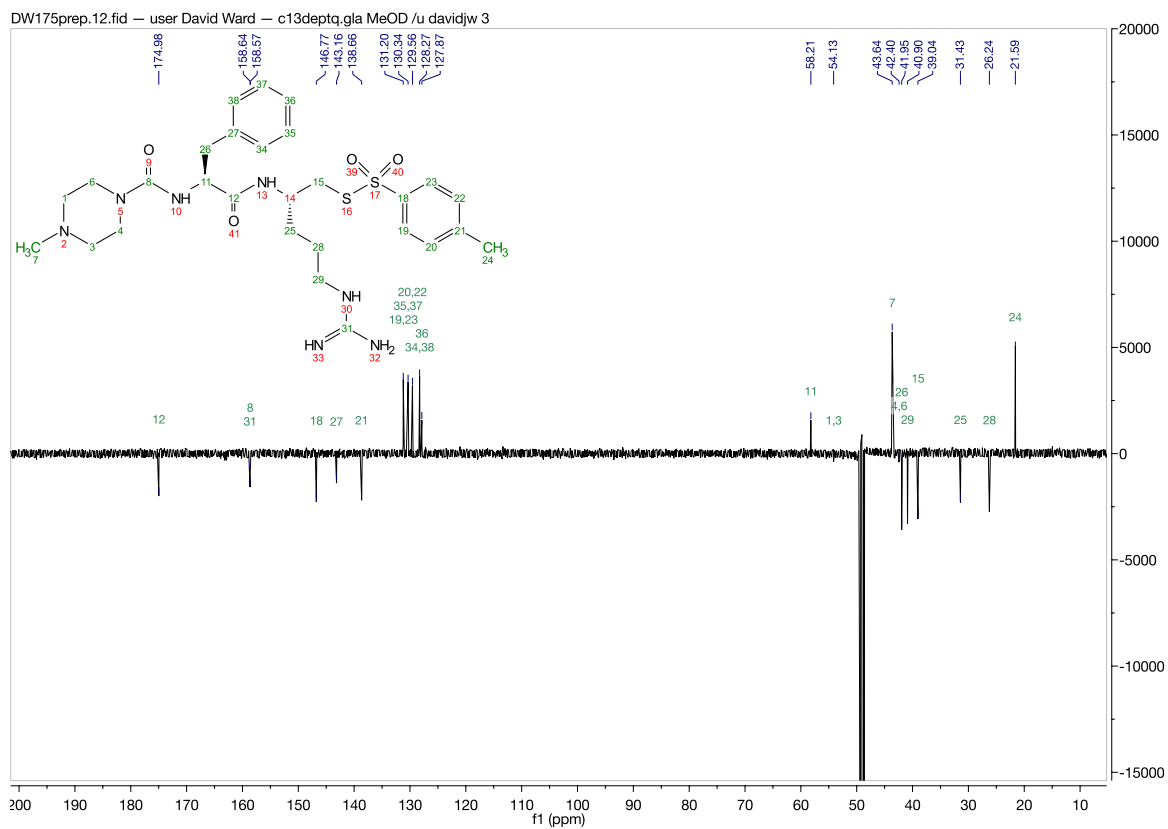
DW177-reversephase.11.1.2r -

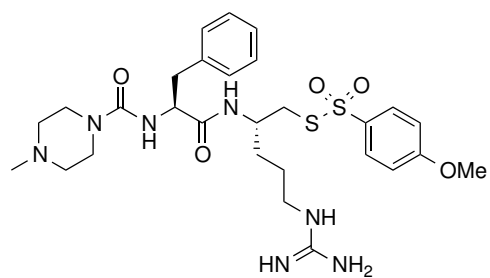




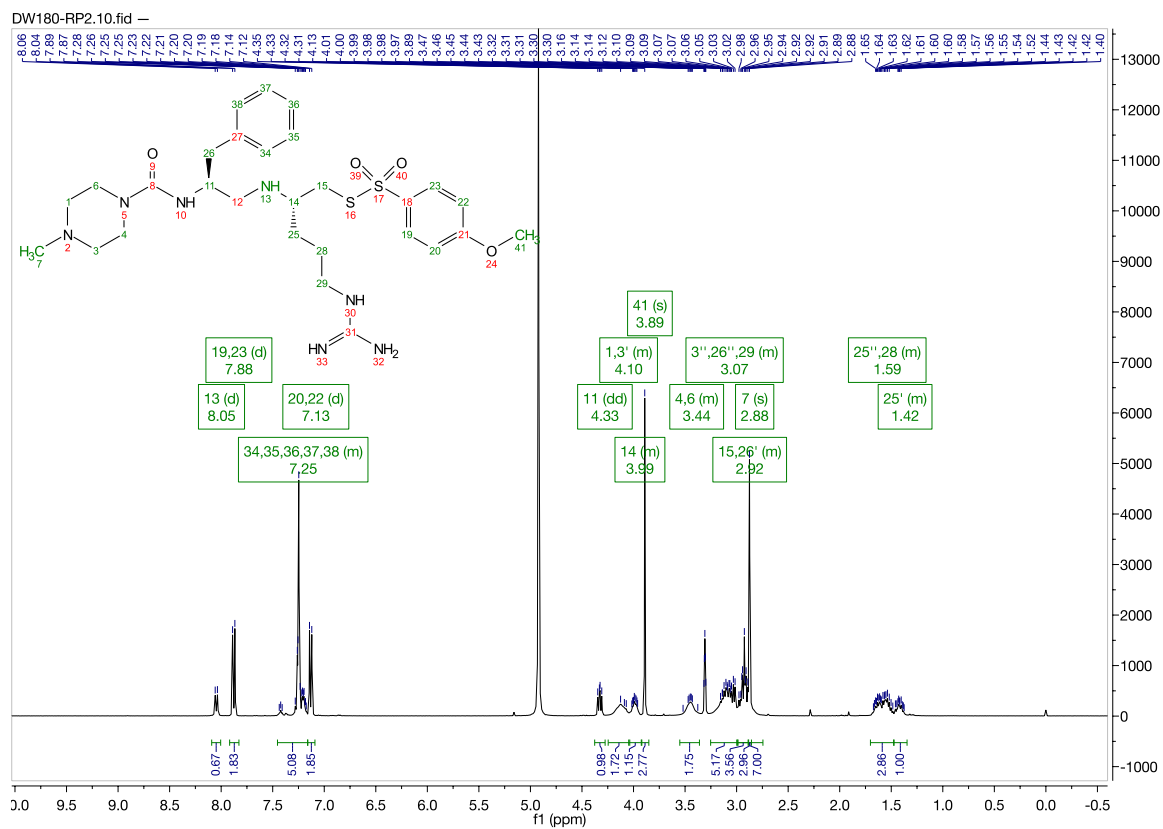
Inhibitor 15







Inhibitor 16





DW180-RP2.12.fid

