

Chemical probes of surface layer biogenesis in *Clostridium difficile*

By
Thi Hoai Tam Dang

A thesis submitted to Imperial College London in candidature for the degree of
Doctor of Philosophy of Imperial College.

Department of Chemistry
Imperial College London
Exhibition Road
London
SW7 2AZ
Februar 2011

Abstract

The bacterium *Clostridium difficile* is responsible for recent epidemics of gastroenteritis and currently causes over twice as many deaths per year as the other major hospital ‘superbug’, MRSA. Since the bacterium is resistant to conventional antibiotics there is an urgent need to develop novel therapies. *C. difficile* secretes a family of proteins that are held in place on the cell wall by non-covalent forces, producing a proteinaceous coat (the surface layer or S-layer) that surrounds the entire cell. These S-layer proteins (SLPs) are immunogenic in humans and play a role in binding to host cells. Synthesis of the *C. difficile* S-layer involves site-specific proteolytic cleavage of the SlpA precursor by an as yet unidentified *C. difficile* protease. Identification of the protease that processes SlpA has proven challenging, due in part to a lack of established genetic tools in *C. difficile*.

Here the development of novel chemical probes that can disrupt S-layer formation by inhibiting the protease are described, and found to be powerful *chemical proteomic* tools for both protease identification and exploring the process of S-layer formation. Screening and inhibition experiments were first performed to identify novel synthetic irreversible protease inhibitors combining an electrophilic warhead with a specific sequence element matching the SlpA cleavage. These compound series were shown to possess structure-dependent activity, and inhibited cultures were also more sensitive to lysozyme-induced cell lysis, suggesting that correct processing and assembly of the S-layer is important for cell envelope integrity. Optimised inhibitors were further developed into ‘activity-based probes’ (ABPs) carrying an affinity tag, and were successfully used to isolate and identify *de novo* the key protease involved in cleavage of SlpA, Cwp84, using a combination of different labelling and proteomics approaches. These probes also permitted identification of Cwp84 activity across a wide range of clinical strains. In later work, the scope of the warhead element was explored in more detail, and the potential antibacterial effects of the inhibitors were investigated in both wild type and genetically engineered strains. Finally, the contributions of this work in terms of both chemical technology and *C. difficile* biology are critically assessed.

Acknowledgements

I would like to thank everyone who has made this thesis possible as well as having made my time at Imperial College and in London an unforgettable experience.

First and foremost, I would like to thank my supervisors: Dr. Ed Tate not just for offering me a PhD position in his lab, but also for his endless patience to guide, support and motivate me throughout the three years. Thanks for your countless and always valued advice as well as your help and listening when I was in trouble; Prof. Neil Fairweather for providing me a microbiology working corner in Flowers and teaching me about *C. difficile*.

For assistance in technical aspects, I must thank Dr. Robert Fagan for teaching me all required microbiological techniques to handle bacteria and turning me from an organic chemist into a “*C. difficile* killer” and half-microbiologist, Dr. Lucia de la Riva Perez for the wonderful collaboration and constantly encouraging me as a good friend, and Dr. Paul Hitchin for practical training in proteomics. I am indebted for the help of Dr. William Heal with helpful discussion, valuable advice in the chemistry lab and sometimes beyond the research-related matters. You seem always to have a right answer for any of my concern.

Many thanks to the vital proof-reading team: Will, Neki, Megan, Alex, Lucy, Chris, Helen and Lucia, your time and effort are much appreciated. A general thank you as well goes to all the members of the Tate, Fairweather and Leatherbarrow research groups for making my time in London in general and at Imperial specifically, very productive and enjoyable. I will miss a lot how much fun we had during the conference stays, the together cooking and the special meals not just Vietnamese food but also European cheese and wine.

For financial support, I am very grateful to the Chemistry Department at Imperial College which, courtesy of EPSRC, provided the funding for my PhD and the RSC, GdCH, SGM, Marie Curie and Biochemical Society for funding the travel and stay of my numerous conferences and providing me the chance to present my work national and international-wide.

Finally, and most importantly, my biggest thank you goes to my family in Germany, my husband in Vietnam and my friends for all their warmly love and unconditional support.

Publications and Presentations Arising from this Thesis

Publications

W. P. Heal, T. H. T. Dang and E. W. Tate " Activity-based probes: discovering new biology and new drug targets ", *Chem. Soc. Reviews*. **2011**, 40, 246-257.

T. H. T. Dang, L. de la Riva, R. P. Fagan, E. M. Storck, W. P. Heal, C. Janoir, N. F. Fairweather and E. W. Tate "Chemical probes of surface layer formation in *Clostridium difficile*", *ACS Chem. Biol.* **2010**, 5 (3), 279–285.

Oral Presentations

T. H. T. Dang et al. Chemical probes of surface layer biogenesis in *Clostridium difficile*. *Biological and Medicinal Chemistry Symposium for Postgraduates, Cambridge, UK (2009)* †

T. H. T. Dang et al. Activity-based probes reveal a putative S-layer processing complex in *C. difficile*. *Activity-based probes meeting, Leiden, Holland (2009)*

Poster Presentations

T. H. T. Dang et al. Activity-based probes reveal an S-layer processing complex in *Clostridium difficile*. *Chemical biology ISACS, San Francisco, USA (2010)*

T. H. T. Dang et al. Activity-based probes reveal an S-layer processing complex in *Clostridium difficile*. *SET for Britain, London, UK (2010)*

T. H. T. Dang et al. Activity-based probes reveal a putative S-layer processing complex in *C. difficile*. *Pfizer poster competition, London, UK (2009)*

T. H. T. Dang et al. Novel chemical probes of Surface-Layer formation in *Clostridium difficile*. *Clostrpath 2009, Rome, Italy (2009)*

T. H. T. Dang et al. Novel chemical probes of Surface-Layer formation in *Clostridium difficile*. *42nd IUPAC congress, Glasgow, UK (2009)*

T. H. T. Dang et al. Novel chemical probes of Surface-Layer formation in *Clostridium difficile*.
RSC South East Regional Organic Meeting, UCL, UK (2009)

T. H. T. Dang et al. Novel chemical probes of Surface-Layer formation in *Clostridium difficile*
Biological and Medicinal Chemistry Symposium for Postgraduates, Cambridge, UK (2008)

T. H. T. Dang et al. Novel chemical probes of Surface-Layer formation in *Clostridium difficile*.
BioCamp Novartis, UK (2008)

T. H. T. Dang et al. Secretion of cell surface proteins in *Clostridium difficile*. 2nd EuCheMS
Congress in Torino, Italy (2008)

†Presentation awarded a prize due to content and delivery

Contents

Abstract	2
Acknowledgements	3
Publications and Presentations Arising from this Thesis	4
Publications.....	4
Oral Presentations	4
Poster Presentations.....	4
Contents.....	6
Abbreviations	12
Chapter 1 Introduction.....	17
1.1 Chemical proteomics.....	18
1.1.1 Strengths and limitations of chemical proteomics	20
1.2 Activity-based protein (or proteome/proteomic) profiling (ABPP)	21
1.3 Activity-Based Probes (ABPs)	22
1.3.1 Designing ABPs.....	22
1.3.2 Click chemistry	26
1.3.3 Affinity enrichment of labelled proteins.....	29
1.4 Application of ABPs to different protease families.....	31
1.4.1 Proteases.....	31
1.4.2 Serine and Threonine proteases.....	33
1.4.3 Cysteine proteases.....	34
1.4.4 Metalloproteases and aspartic proteases	37
1.5 Mass spectrometry analysis	38

1.5.1	MS instruments.....	39
1.5.2	Bioinformatics/database search approaches	43
1.5.3	MS-based proteomic strategies for protein identification and PTM characterisation	44
1.6	<i>Clostridium difficile</i>	48
1.6.1	<i>C. difficile</i> virulence factors.....	51
1.6.2	The S-layer of <i>Clostridium difficile</i>	51
1.6.3	Proteases in <i>C. difficile</i>	56
1.7	Project aims and outline	59
1.7.1	Structure of this thesis.....	60
Chapter 2 Discovery of E-64 and leupeptin as the first inhibitors of S-layer processing in <i>C. difficile</i>		61
2.1	Screening of commercial inhibitors	62
2.1.1	Inhibition assay	62
2.1.2	Protease inhibitors.....	63
2.1.3	First inhibition results	71
2.2	E-64.....	74
2.2.1	Isolation, characterisation and structure determination	74
2.2.2	Inhibitory activity and specificity	74
2.3	Conclusions.....	76
Chapter 3 Synthetic inhibitors of S-layer processing in <i>C. difficile</i>		78
3.1	Potential protease targets in SlpA maturation	79
3.2	Inhibitor design	80
3.3	Synthesis of E-64 analogues.....	85
3.4	Activity of E-64 analogues.....	89

3.5	Determination of the location of the target protease(s)	93
3.6	Activity vs. concentration.....	95
3.7	Conclusion	96
Chapter 4 Inhibitor analogues and bioorthogonal ligation		98
4.1	ABP design and synthesis	99
4.2	Activity of chemical tag appended analogues	101
4.3	Activity vs. Concentration	103
4.4	Chemical labelling	104
4.4.1	Click ligation in labelled <i>C. difficile</i>	104
4.4.2	Click assay development.....	105
4.4.3	Ligation results.....	109
4.5	Conclusion	112
Chapter 5 Directly labelled activity-based probes		116
5.1	Biotinylated ABPs	117
5.1.1	Activity of biotinylated analogues	118
5.1.2	Labelling effect of biotinylated ABPs	120
5.1.3	Stability of inhibitors in culture	124
5.1.4	Minimal inhibition time required for detectable labelling or inhibitory effects 126	
5.1.5	Strain dependence	127
5.2	Fluorescent ABPs.....	129
5.2.1	Design of fluorescent ABPs	129
5.2.2	Synthesis of fluorescent ABPs.....	132
5.2.3	Activity and labelling effect of fluorescence appended ABPs	135
5.2.4	Strain dependence	137

5.3	Conclusion	139
Chapter 6 Target enrichment, identification and proteomic study.....		140
6.1	Enrichment of labelled proteins in <i>C. difficile</i>	141
6.1.1	Pull down of labelled proteins with NeutrAvidin agarose resin	141
6.1.2	Pull down of labelled proteins with Streptavidin Dynabeads	144
6.2	Identification of protease(s) by MALDI-TOF-MS and PMF.....	145
6.3	Identification of proteases by LC-MS/MS	149
6.4	In-depth proteomic study using LC-MS/MS with the wild type 630.....	152
6.5	Cwp13 may be an off-target of epoxysuccinyl ABPs.....	156
6.6	Effect of Cwp84 on other cell wall proteins.....	158
6.7	Proteomic studies with a <i>C. difficile</i> 630 Δ cwp84 mutant.....	158
6.8	Conclusion	162
Chapter 7 Exploring the scope and potential antibacterial activity of irreversible cysteine protease inhibitors in <i>C. difficile</i>		165
7.1	Warheads against cysteine protease inhibitors.....	166
7.2	Inhibitor design	169
7.3	Inhibitors with the warhead appended to the N-terminus	171
7.3.1	Synthesis	171
7.3.2	Activity and labelling effect	173
7.4	Inhibitors with the warhead appended to the C-terminus.....	179
7.4.1	Synthesis of Michael acceptor inhibitors.....	179
7.4.2	Synthesis of AOMK- and CMK- inhibitors	181
7.4.3	Activity and labelling effect of inhibitors with the warhead appended to C-terminus.....	183
7.5	Growth inhibition of <i>C. difficile</i>	184

7.5.1	Growth inhibition by Z-Ala-CMK and EtEP-R-Pra in solution culture	184
7.5.2	Growth inhibition of 630 and 630 Δ cwp84 by antibiotics	186
7.6	Conclusion	187
Chapter 8 Conclusion and Perspectives.....		191
8.1	Technology development.....	192
8.1.1	Design.....	192
8.1.2	Synthesis of inhibitors/ABPs	193
8.1.3	Activity assay.....	196
8.1.4	Labelling effect.....	199
8.1.5	Proteomics	200
8.2	Impacts of this study on current understanding of <i>C. difficile</i> biology.....	201
8.2.1	Cwp84 is the target protease responsible for SlpA internal cleavage.....	201
8.2.2	Cwp13 and other cell wall proteins	203
8.2.3	Epoxysuccinyl inhibitors/ABPs with Arg as AA ¹ is a special case	204
8.2.4	Could Cwp84 be a potential therapeutic target for CDI treatment?.....	205
8.2.5	Other potential targets in <i>C. difficile</i>	206
Chapter 9 Materials and methods.....		207
9.1	Microbiology	208
9.1.1	Buffers.....	208
9.1.2	Bacteria strains and culture conditions	208
9.1.3	SDS-PAGE	210
9.1.4	Western immunoblotting	212
9.1.5	NeutrAvidin TM -HRP immunoblotting	212
9.1.6	Cell fractionation.....	214
9.1.7	CuAAC ligation.....	214

9.1.8	Enrichment of captured proteins.....	215
9.1.9	Expression and purification of recombinant protein.....	216
9.1.10	Antibiotic inhibition using Etest.....	220
9.2	Proteomics	221
9.2.1	Proteomics using Imperial College facilities	221
9.2.2	Proteomics using the facilities of the Proteomics Centre of the University of Leicester.....	222
9.3	Chemistry	223
9.3.1	Reagents and Equipment.....	223
	Solid Phase Resins:.....	224
	Amino Acids:	224
	Reagents:	224
9.3.2	General procedures	224
9.3.3	Synthesis and characterisation	226
Chapter 10	References	252
Chapter 11	Appendix	278

Abbreviations

1D	One-dimensional
2D	Two-dimensional
A	Adenine (DNA base)
ABP	Activity-based probe
AfBP	Affinity-based probes
ABPP	Activity-based protein profiling
ADH	Alcohol dehydrogenase from yeast- Uniprot P00330
Amp	Ampicillin
AOAA	Acyloxyacetamide
AOMK	Acyloxymethyl ketone
<i>BamH I</i>	Restriction enzyme derived from <i>Bacillus amyloliquefaciens</i>
BCS	Bathocuproine disulfonate
BHI	Brain-heart infusion
Bis-Tris	Bis(2-hydroxyethyl)amino-tris(hydroxymethyl)methane
Boc	t-Butyloxycarbonyl
BSA	Bovine Serum Albumin
C	Cytosine (DNA base)
CAA	Chloroacetamide
CC	Click chemistry
CDI	<i>Clostridium difficile</i> infection
CID	Collision-induced dissociation
CL	Chloramphenicol
CM	Clindamycin
CMK	Chloromethyl ketone
C-PH	Cysteine N-terminal PH domain
C-terminus	Carboxy terminus of a protein
CuAAC	Copper (I) catalysed alkyne-azide cycloaddition
DCM	Dichloromethane
DDBJ	DNA Data Bank of Japan
DIC	<i>N,N'</i> -Diisopropylcarbodiimide
DIPEA	<i>N,N'</i> -Diisopropylethylamine

DMF	Dimethylformamide
DMSO	Dimethyl sulfoxide
DNA	Deoxyribose Nucleic Acid
DTT	Dithiothreitol
ECD	Electron-capture dissociation
<i>EcoR I</i>	Restriction enzyme derived from <i>Escherichia coli</i>
EDT	1,2-ethanedithiol
EDTA	Ethylenediaminetetraacetic acid
ELISA	Enzyme-linked immunosorbent assay
EM	Electron multiplier
EM	Erythromycin
EMBL	European Molecular Biology Laboratory
ESI	Electrospray ionisation
ETD	Electron-transfer dissociation
EtOH	Ethanol
FAA	Fluoroacetamide
FMK	Fluoromethyl ketone
Fmoc	9-Fluorenylmethyl carbonyl
FP	Fluorophosphonate
FRET	Förster Resonance Emission Transfer
FT-ICR	Fourier transform ion cyclotron resonance
GST	Glutathione-S-transferase
HAA	Haloacetamide
HATU	<i>O</i> -(7-Azabenzotriazol-1-yl)- <i>N,N,N',N'</i> -tetramethyluronium hexafluorophosphate
HBTU	<i>O</i> -(Benzotriazol-1-yl)- <i>N,N,N',N'</i> -tetramethyluronium hexafluorophosphate
HCTU	<i>O</i> -(6-Chlorobenzotriazol-1-yl)- <i>N,N,N',N'</i> -tetramethyluronium hexafluorophosphate
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His-tag	Polyhistidine affinity tag
HMK	Halomethyl ketone
HMW	High molecular weight

HOBt	1-hydroxybenzotriazol
HPLC	High-pressure liquid chromatography
<i>hν</i>	photon
ICE	interleukin-1β converting enzyme
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IT	Ion trap
LB	Luria-Bertani
LC-MS	Liquid chromatography-mass spectrometry
LDS	Lithium dodecyl sulfate
LIT	Linear ion trap
LMW	Low molecular weight
MA	Michael acceptor
MALDI	Matrix-Assisted Laser Desorption/Ionisation
MeOH	Methanol
MH ⁺	Mass plus a proton
MIC	Minimum inhibitory concentration
Mmt	Monomethoxytrityl
MP	Metalloprotease
MRM	Multiple reaction monitoring
Mts	2,4,6,-trimethylbenzenesulfonate
MS	Mass spectrometry
MW	Molecular Weight
MZ	Metronidazole
MX	Moxifloxacin
NCBI	National Center for Biotechnology Information
NMM	N-methyl morpholine
NMR	Nuclear Magnetic Resonance
N-terminus	Amino terminus of a protein
OD	Optical Density
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
PEG	Poly(ethylene glycol)
PMF	Peptide mass fingerprint
Pra	L-Propagyl-glycine

PRIDE	Proteomics Identifications Database
PSD	post-source decay
PTM	Post translational modification
Q	Quadrupole
rcf	Relative Centrifugal Force
RMSD	Root Mean Square Deviation
RNA	Ribose Nucleic Acid
RP-HPLC	Reverse-Phase High-Performance Liquid Chromatography
rpm	Revolutions per minute
RT	Room Temperature
SAR	Structure activity relationship
SAXS	Small-angle X-ray scattering
SDS-PAGE	Sodium Dodecyl Sulphate – Polyacrylamide Gel Electrophoresis
SLP	Surface layer protein
SPPS	Solid Phase Peptide Synthesis
T	Thymine (DNA base)
TAMRA	5(6)-Carboxytetramethylrhodamine
TAT	Twin arginine transport
TBME	tert-butyl methyl ether
TBS-T	Tris-Buffered Saline – TWEEN
TBTA	Tris-(triazolyl)benzylamine
tBu	tert-butyl
TC	Tetracycline
TCEP	Tris-(2-cyanoethyl)phosphine
TEV	Tobacco Etch Virus
TFA	Trifluoroacetic acid
TIS	Triisopropyl silane
ToF	Time-of-Flight
Tris	Tris(hydroxymethyl)aminomethane
Trt	Trityl
UV	Ultraviolet
WH	Warhead

Amino acid abbreviation:

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

Chapter 1 Introduction

Identification, characterisation and assignment of function to members of a proteome associated with specific biological processes in the cell is a central theme in post-genomic biology. Although conventional genomic [1-2] and proteomic [3-6] methods have provided many useful insights into quantification of transcripts and proteins, they are only indirect indicators of biological function, and suffer from a limited capacity to measure the catalytic activity of proteins involved in post-translational mechanisms [7]. Consequently, new and more efficient methods that can provide a deeper understanding of protein activity in a particular complex biological system would enable access to system-level functional information about biochemical networks that are important, for example, in the complex search for viable drug targets. In this study, the application of one of such method, termed activity-based protein (or proteome) profiling (ABPP) [8-10] is discussed, in particular its implementation in the discovery and functional characterisation of proteases involved in the formation of the surface layer of *Clostridium difficile* [11]. This study provides compelling evidence that, when coupled with chemical proteomic methods, ABPP is a powerful tool for studying S-layer biogenesis in this challenging pathogen that lacks robust technologies for genetic engineering, and may reveal potential therapeutic targets involved in colonisation of the host.

1.1 Chemical proteomics

Advances in genomic sequencing have brought biological research into a new era, in which scientists are facing new challenges in cataloguing and assigning functions to gene products. Although conventional genomic data analysis and proteomic study grant access to useful information on protein expression level and hypothesis regarding protein function, this method cannot be used for evaluation of protein function on a global scale. This is especially the case for proteins that are regulated by posttranslational processes or complex arrays such as substrate co-localisation regulation, allosteric control and co-regulation of endogenous inhibitors. To move away from inherent limitations of these methods, the field of chemical proteomics has introduced several strategies to permit the identification, quantification and manipulation of the full repertoire of enzyme activity in systems.

Chemical proteomics [12-13] is a broad area that employs numerous different experimental techniques. The three main approaches are: the global proteomics approach, the compound-centric approach, and activity- or affinity-based protein profiling (ABPP). A comparison of these approaches is illustrated in **Figure 1. 1**. These methods have distinct advantages and disadvantages, and the choice of approach depends on the specific application[14].

The global proteomics approach is applied to determine changes in protein abundance and activity in response to drug treatment [15-18]. Cellular responses are evaluated in a global and unprejudiced manner after the biological system (which may be cells or a whole organism) is treated with a drug. However, this strategy is restricted to the most abundant proteins (affected either by inhibited signalling pathways or by a stress response and/or housekeeping) because observed changes are limited by the depth of the analysis. Several studies [19-21] have made use of the advantages of this strategy to study proteins of interest.

The compound-centric approach is typically applied to characterise the mechanism of action of a known bioactive compound with its interacting proteins [13-14]. Usually this compound is immobilised onto an inert and biocompatible solid support prior to incubation with the biological extract to identify interacting proteins. Unfortunately, this method is still limited; it fails to provide data on protein function and is difficult to apply in living systems.

Moreover there are significant synthetic challenges, for example the difficulty of introducing the functional groups required for the immobilisation process. Publications using this approach have focused on kinase inhibitors [22-24], signalling molecules [25], cyclic nucleotides [26] or specific peptide sequences [27-28].

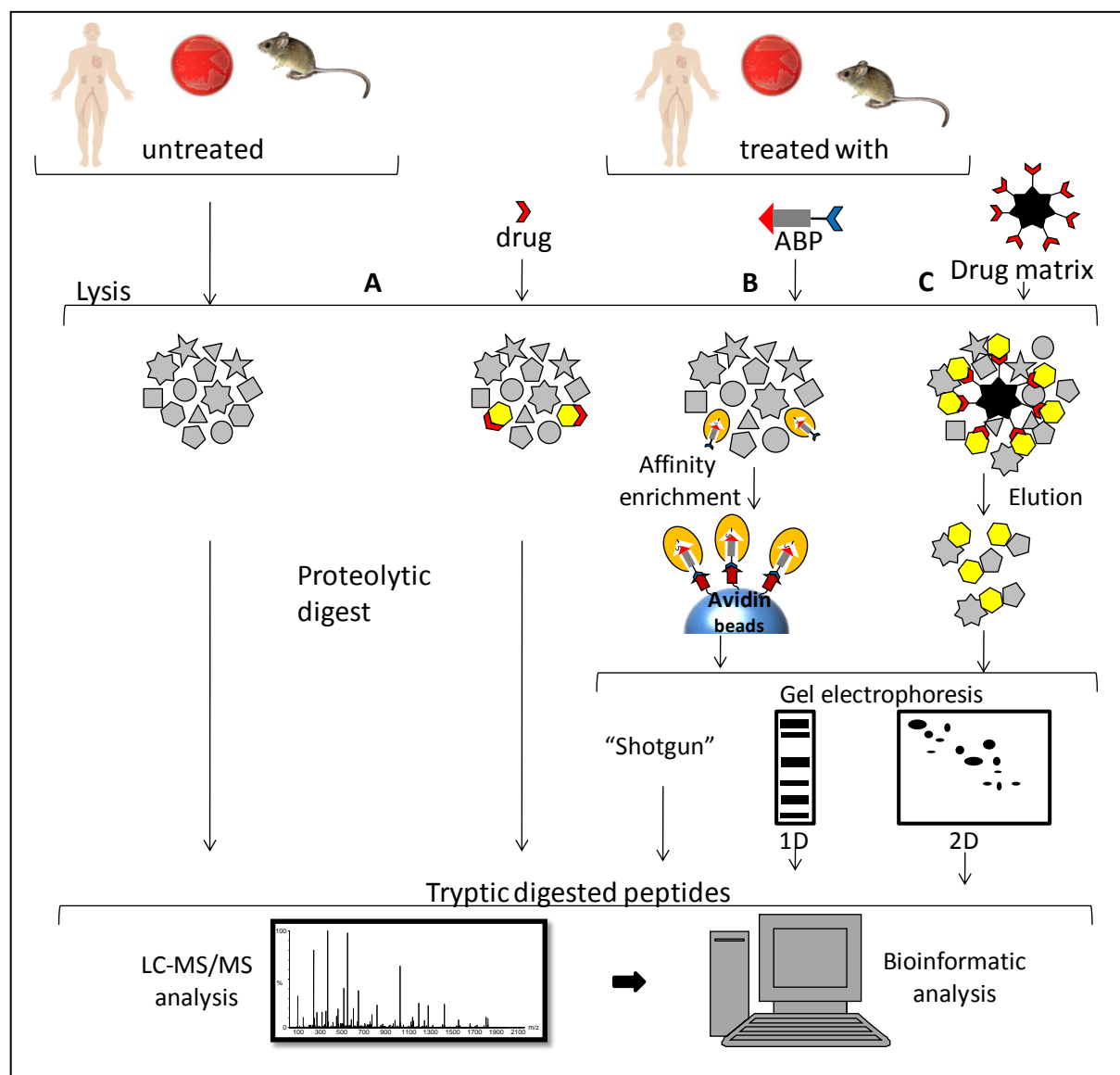


Figure 1. 1: Comparison of three different chemical proteomics approaches: A) global proteomics approaches for determining changes in protein abundance and activity in response to drug treatment, B) ABPP involves specific targeting and subsequent purification of drug target protein classes with ABPs, and C) in the compound-centric approaches, bioactive compounds (drugs) are modified and immobilised on a solid support (drug matrix). Proteins that interact with this matrix are then analysed.

ABPP analyses the enzymatic activity of a particular protein family under certain conditions (for example in a disease). As this approach is the main topic of this thesis, a detailed review follow in section 1.2.

1.1.1 Strengths and limitations of chemical proteomics

Chemical proteomics enables the study of proteins in a highly dynamic system affected by small changes (e.g. the variation of bacterial culture conditions, cell treatments with drugs or inhibitors, stress, changing metabolic and regulatory processes, *etc.*) [29]. The different states of these changes can be monitored and investigated quantitatively in a unique and highly sensitive manner [30]. In addition, the identification of proteins by chemical proteomics usually comes with functional annotation. Such approaches also provide further information about uncharacterised proteins discovered by other means (e.g.: putative proteins from genomic analysis).

Furthermore, chemical proteomics approaches are flexible and allow for the application of probes to entire proteomes or sub-proteomes of any cell type (e.g.: mammalian cells, plants, bacteria and viruses) without the restriction of having to engineer individual recombinant proteins [13]. Therefore, these proteins are studied in their natural, un-engineered state and environment, including post translation modification and endogenous abundance levels, providing a genuine broad and in depth understanding of the system under investigation. This can include, at the molecular level, aspects of an enzyme's mechanism or a protein's interactions.

Despite numerous advantages, this technique still has some limitations where improvements are needed. One prominent limitation is the relatively high amount of cellular material required and therefore the high consumption of probes, which can be expensive to synthesise, follows. However, this demand is reduced with the most recent developments depending on the working system and the abundance of the active enzymes. A second limitation is the dependence on the type of lysis operation, as only proteins that are present in the lysates can be captured, identified and investigated; proteins such as membrane proteins prove challenging. Nevertheless, the latest solubilisation technologies aid significantly in circumventing this problem and have allowed the successful identification of type I membrane proteins (e.g.: receptor tyrosine kinases [22, 31-32]).

A third limitation is the high level of background that derives from abundant proteins interacting *via* nonspecific, hydrophobic or charged surface interactions with the proteins of interest. This problem can be reduced to a certain level by the careful choice of affinity

parameters in the pull down step and adding negative controls. Another challenge of this technique, which comes quite early in the study, is the synthetic hurdle in accessing the probes, which are crucial to the whole process.

1.2 Activity-based protein (or proteome/proteomic) profiling (ABPP)

ABPP [8-10], one of the emerging chemical proteomic strategies that utilises the ability of forming covalent bond of a chemical probe directly with the active site of proteins, is applied to profile the functional state of proteins in a complex proteome. The principle of ABPP is to design those specific chemical probes in such a way that they modify the target enzyme(s) covalently at the active site with the objective of detecting the activity and identity of the target enzymes directly in a living system [33]. Importantly, activity-based probes (ABPs) permit the detection of enzyme activity independently from the enzyme abundance and therefore provide access to low-abundance proteins which are usually difficult to characterise by other means. While ABPs have been designed to target the catalytic residues on enzymes, another form of chemical probe – called affinity-based probes (AfBPs) – have been designed to target non-catalytic residues on proteins relying on highly selective tight binding to the target proteins [34] (**Figure 1. 2**). Several groups have developed AfBPs on the basis of well-documented photo-affinity reagents [35], these probes were designed from tight-binding reversible inhibitors of target proteins coupled to a photoreactive molecule. Carefully designed AfBPs can tune the labelling of target enzymes in an activity-dependent fashion and therefore can be potent alternative tools to ABPs.

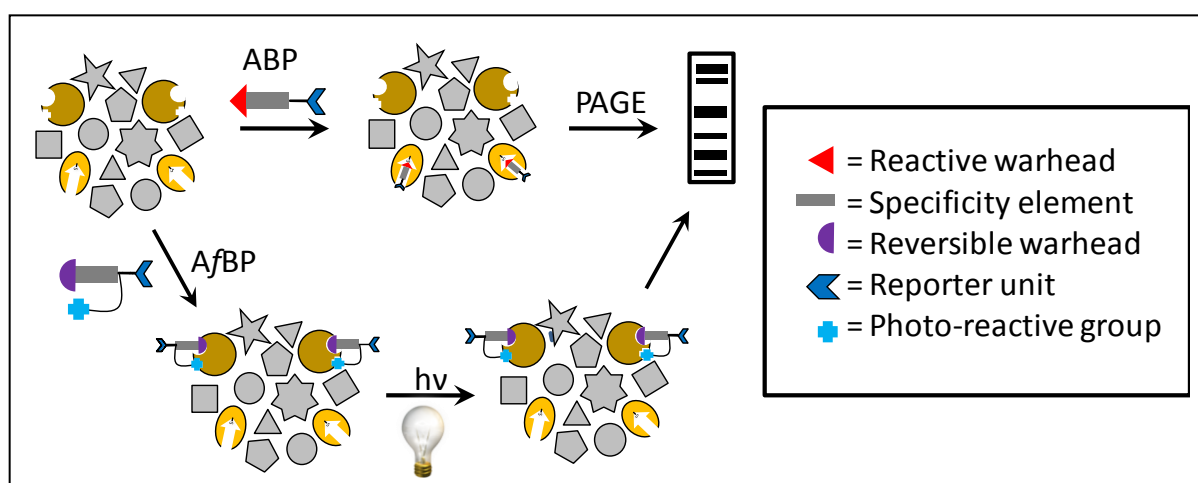


Figure 1. 2: ABPs versus AfBPs [36]. ABPs label enzymes *via* their catalytic mechanism, whilst AfBPs make use of protein affinity, stabilised by non-specific cross-linking.

The requirement of a catalytically active enzyme is the decisive property for the distinction of ABPs and AfBPs, as AfBPs can bind to a specific site which is not necessary the active site on the target enzyme, whilst ABPs only target the active site of the enzyme. In many cases, AfBPs have been applied to profile targets usually inaccessible to ABPs [37]. Because AfBPs lie outside the scope of this study, no further discussion is made of this type of probe.

1.3 Activity-Based Probes (ABPs)

ABPs are small molecules that typically possess structural components enabling the appropriate exploitation of the catalytic site chemistry of the target enzyme, including alkylation or acylation of a specific nucleophile group at the catalytic site. Because ABPs form a covalent bond with the enzyme at the active site, the catalytic activity is inactivated in the same manner as a “suicide substrate” or irreversible inhibitor. The detection of the bound proteins can be facilitated using a suitable tag or label. To date, ABPs have been applied for myriad enzyme classes including kinases [38-41], phosphatases [42-43], hydrolases [44], lipases [45], glycosidases [46-47], oxidoreductases [48-49] and proteases including cysteine proteases [50-53], serine proteases [54] and threonine proteases [55-56].

1.3.1 Designing ABPs

Designing an effective ABP requires certain fundamental knowledge of the catalytic site chemistry of the target enzymes. Therefore, the early stage developments of ABPs were mostly linked directly with the discovery and development of reversible and irreversible inhibitors. Conversion of inhibitors, which are usually derived from natural products [57] or a substrate mimetic [33], into ABPs are usually achieved by careful molecular engineering. Typically, ABPs consist of three parts: (1) a reactive group (the “warhead” or WH) for covalently binding and labelling the active site of the targeted enzymes; (2) a specificity element, which acts as recognition site and improves the selective labelling of a particular class or sub-class of enzymes, and (3) a reporter unit, which can be a tag or a label, for manipulation, measurement and identification of the labelled enzymes (**Figure 1. 3**). In addition to these components, ABPs may also contain a linker separating the specificity element from the reporter unit to minimise steric hindrance of the tag/label to the active part of the probe. Each part of an ABP is discussed in detail in the following sections.

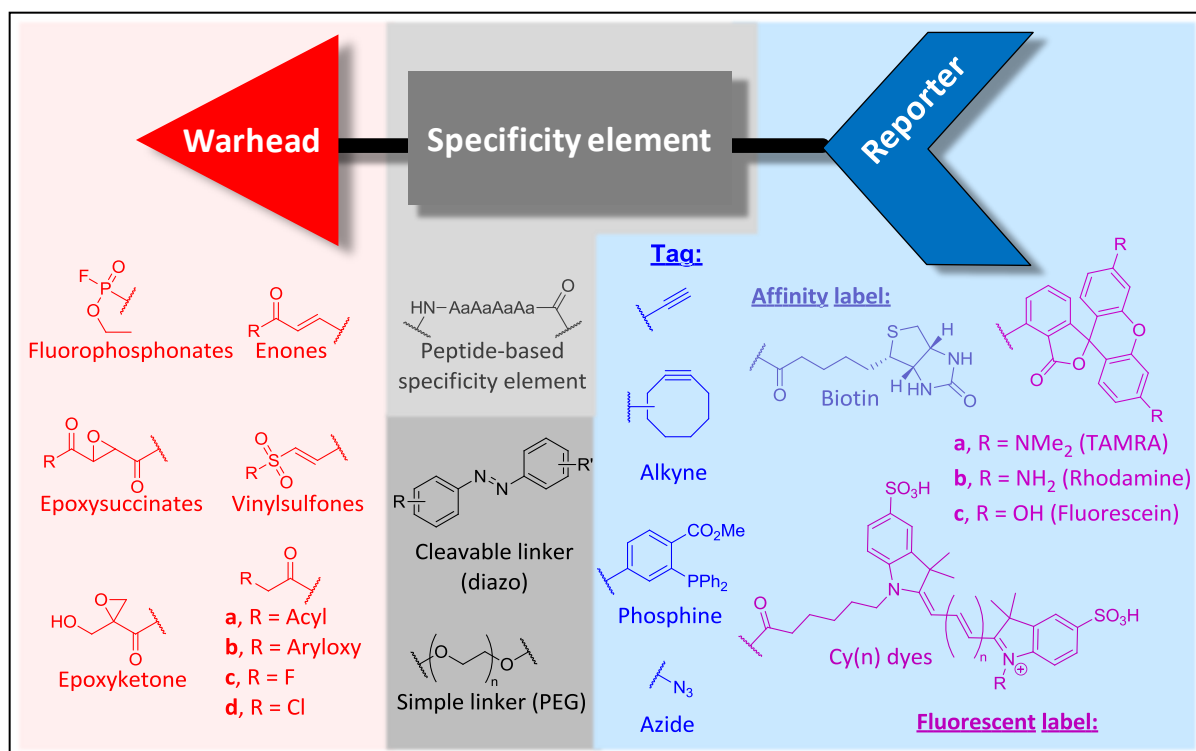


Figure 1. 3: ABP anatomy with three key features and their representative examples.

Design of ABPs follows two basic strategies: directed and non-directed [29, 33]. Development of the former is intended to target specific enzymes within a mechanistically related family, e.g. to target only serine proteases, not lipases, or a particular enzyme (or few enzymes) within a class. Usually, directed ABPs are based upon known inhibitors, which are either mechanism-based inhibitors containing reactive groups conferring a chemical preference for a particular enzyme class, or broad-spectrum inhibitors containing high affinity specificity element directing the probes to enzymes that share common catalytic site chemistry. Consequently, the target enzymes have to be characterised to a certain degree to provide necessary information about mechanism and structure. In contrast, non-directed ABPs are deliberately designed to be broad-spectrum, e.g. to target all serine hydrolases (including serine proteases, lipases, etc.), so as not to limit the discovery of potential new enzymes. This method usually applies one or more non-specific site-directed electrophiles and an array of specificity elements to screen a wide range of mechanistically distinct enzymes in a complex proteomes. Both strategies are widely used and their application will be described in later sections in targeting class-selective ABPs.

1.3.1.1 Warheads

Selectivity of a reactive group (warhead) is perhaps the greatest challenge in designing this part of the probe, because it has to be active and provide the required covalent modification to the specific residue on the target enzyme (e.g.: a nucleophilic amino acid residue in the active site) and simultaneously, it has to be inert towards other potential reactive species (e.g.: thiols, amines, alcohols) within the complex proteome. Because the results of ABPP experiments depend on the ABPs used, the tailoring of the warhead in respect of integrating the enzyme chemistry can define the specificity to certain catalytic class of enzyme of interest. Therefore, distinction is possible between enzymes e.g.: proteases, phosphatases and glycosidases. Typically, warheads are electrophilic traps such as sulfonates [48, 58], epoxides [51, 59-60] or Michael acceptors [61-62], that react irreversibly forming new covalent bonds with the nucleophilic residues at the active site of the enzyme. Affinity and inherited chemical reactivity are the decisive factors to define the reactivity of the warhead towards the enzyme active site. Adjusting these factors properly allows the selectivity between enzymes to be manipulated. For example, although the inactivation of serine and cysteine proteases relies on a catalytic amino acid nucleophile in the active site, they possess a different nucleophilic residue and distinct catalytic mechanisms. Therefore design of a warhead that reacts with one class and not the other is possible.

1.3.1.2 Specificity elements

Highly reactive warheads are often promiscuous and therefore they address a relative broad set of targets. In order to increase the selectivity of ABPs and narrow down the target panel, a specificity element is often incorporated into the probe to direct the probe to a desired well-defined enzyme or enzyme family. Normally, these specificity elements are a short peptide (ca. two to four amino acids), or peptide-like structures, that can be chemically linked with the warhead, or are a subunit within the structure of the parent inhibitor [63-64]. In some cases, very large peptides or proteins have been used to gain a higher degree of target specificity [33]. A prominent example is the modification of recombinant ubiquitin with an electrophile warhead to probe proteases that process ubiquitin *in vitro* [46, 65-66]. This kind of probe is very useful when target specificity requires protein-binding domains.

1.3.1.3 Reporter units

The purpose of a reporter unit on an ABP is to provide a “handle” for manipulating, visualising, identifying and quantifying the probe-modified enzymes. This reporter unit can be either directly appended to the probe via a linker or introduced at a later stage using bioorthogonal ligation [67-69] or click chemistry [36]. For the sake of distinguishing both cases, in this thesis a “tag” is a “clickable” group (e.g. an alkyne or azide) used for indirect (“two step”) labelling, whilst a “label” refers to a label (e.g. fluorophore, affinity label) directly attached to the ABP.

A “label”

Labels for ABPs fall into three categories: affinity, fluorescent and radioactive. Biotin, an affinity label, is the most commonly used, because it is compatible with SDS-PAGE (Sodium Dodecyl Sulphate – Polyacrylamide Gel Electrophoresis) and can be detected simply by Western blotting [63]. This label is also preferred over the others (fluorescent and radioactive labels), because it is cost effective and environmentally friendly, and facilitates the subsequent purification of the labelled enzymes on beads. However, detection by Western blot might be not the best approach for quantitative analysis due to some inherent limitations of the Western blot method.

Fluorescent [45, 51, 70-71] and radioactive [60, 72] labels facilitate detection by direct scanning of gels after SDS-PAGE analysis with a fluorescent scanner or a phosphor imager [63]. Despite the higher cost compared to biotin, they benefit from several advantages such as time saving due fewer handling steps, greater sensitivity and greater dynamic range. Moreover, fluorescent labels can have other advantages, if different, non-overlapping excitation/emission fluorescent dyes are incorporated, allowing multiplexing of samples labelled with different coloured ABPs but visualised on the same gel for comparative evaluation [71, 73-74].

A “tag”

The physical properties of the appended label such as polarity, size, charge, structure and chemical reactivity may have a large impact on the behaviour of the probe as a whole, including inhibitory effect and cellular uptake [63]. Especially large labels may suffer from detrimental effect on enzyme recognition/labelling and PAGE analysis [37, 75]. To avoid

these potential limitations these labels may be replaced by a “clickable” tag allowing subsequent introduction of a capture reagent after the ABPs have been administered, using bioorthogonal ligation [67-68] or click chemistry [36]. The tags are small, predominantly alkyne or azide functional groups, and are typically linked directly into the specificity element, and therefore expected to have minimal effect on the probes [76]. In addition, tags are largely inert towards endogenous biological functionalities, allowing specific and effective bioorthogonal ligation with a capture reagent that contains a corresponding cyclisation partner, e.g. alkyne or azide, and a reporting group such as biotin, a fluorophore or a radioactive group, or a combination of such groups.

Linker

The linker connects the warhead/specificity element to the labels or tagged capture reagent. Its primary purpose is to create enough space between the connected components to prevent steric hindrance that might impede the accessibility of the warhead/specificity element to the active site, or the accessibility of the affinity label for purification purposes. Linkers can be either non-cleavable or cleavable.

Non-cleavable spacers can be simple long-chain alkyl or poly-ethylene glycol (PEG). The alkyl linkers can be tailored to confer enhanced hydrophobicity and ease the permeability of probes into live cells or tissues. In contrast, PEG linkers provide better solubility to hydrophobic ABPs in aqueous environments. The choice of a suitable linker depends on the requirement of the ABPP application.

Cleavable linkers facilitate the selective release of the labelled enzymes after purification on beads. The cleavage event can be either chemical (e.g.: cleavage of disulfide or diazo linker [77-80]) or enzymatic, utilising the relationship of proteases (e.g.: TEV protease [81-82]) and their corresponding peptide substrates. Application of cleavable linkers circumvents harsh elution conditions and minimises non-specific pull down of proteins, enhancing signal and reducing noise.

1.3.2 Click chemistry

Defined by K. Barry Sharpless and his co-workers in 2001 [83], “Click” chemistry (CC) has boomed in many applications [84-86]. In the last few years, CC has been adopted

extensively in ABPP ranging from clicking a tag with the capture reagent [87-89] to introducing linker unit [90-91] or rapid modular synthesis of ABPs and AfBPs [89, 92-93]. Originally, the term CC referred to a set of reactions that are modular, easy to perform, selective, stereospecific, giving very high yields and producing no (or only non-toxic) byproducts. Among them, azide-alkyne cycloadditions occupy a special place as their chemistries facilitate various biological applications [76, 94] (**Figure 1. 4**).

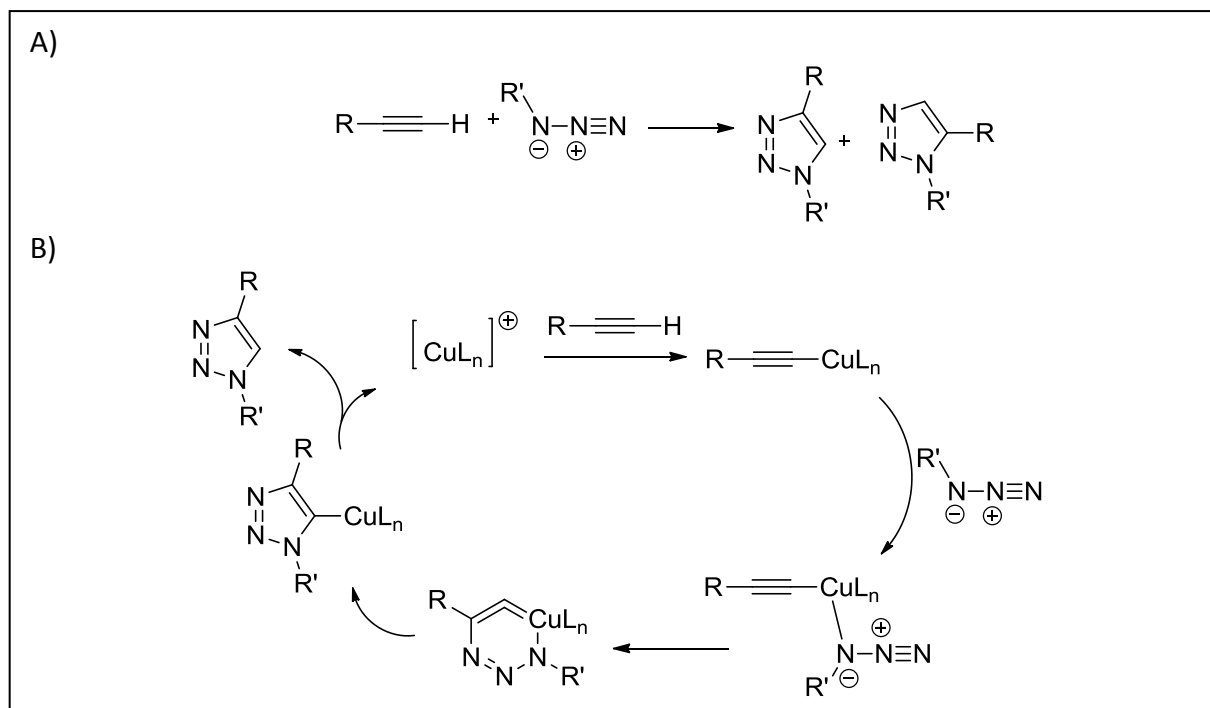


Figure 1. 4: Click reaction: A) General uncatalysed Huisgen reaction. B) Proposed mechanism of the copper-catalysed Huisgen cycloaddition of alkynes and azides [95]

Pioneering work by the groups of Cravatt, Overkleeft, Bogoy and Yao have utilised CC to synthesise probes by linking the reporter unit with the reactive moiety (**Figure 1. 5A**). For example, Cravatt *et al.* generated probes for selective labelling of carboxylesterase-1 (CE-1) from mouse heart proteomes by simply linking rhodamine or biotin and rhodamine together with the natural product FR182877 using CC [90]. In a similar approach, Yao and colleagues utilised CC to build up a whole combinatorial library of hydroxamate-containing small molecule probes against matrix metalloproteases [93]. Fuwa *et al.* generated a library of photo activatable probes by clicking biotinylated photo reactive molecules to alkyne-incorporated benzodiazepine for the study of the enzyme γ -secretase [92]. Another elegant strategy is the synthesis of inhibitor scaffolds on a microarray using solid phase peptide

synthesis (SPPS) and then introducing the desired reporter unit into the probes using CC [34].

Another application of CC in ABPP is two-step labelling (**Figure 1. 5B**), which was first reported independently by Cravatt and Overkleeft [96-97]. Using an azide-derivatised phenyl sulphonate probe, Cravatt and co-workers targeted GSTO1-1 enoyl CoA hydrolase (ECH-1) and aldehyde dehydrogenase (ALDH-1) in treated cell lysates from COS7 cells [96]. After this initial success, his group applied this principle of bioorthogonal ligation to other systems, e.g.: the profiling of fatty acid amide hydrolases by selectively labelling in animal brain, liver and kidney [98]; probing a glycolytic enzyme in cancer cells with a probe containing natural product scaffold MJE3 [99]; or inhibition and activation of cytochrome P450 enzymes [100]. The Overkleeft laboratory, on the other hand, employed functionalised vinyl sulfone probes to investigate proteasomes in cell lysates as well as in live cells [97]. Most of this work utilised copper (I) catalysed alkyne-azide cycloaddition (CuAAC), which can be suitable for detecting labelled protein samples, but not usually on live cells because of the cytotoxicity of copper. To override these problems, the Bertozzi lab worked on a copper-free variation, a strain-promoted alkyne-azide cycloaddition (SPAAC) reaction [76]. In this work, the researchers showed that the incorporation of azido-sugar onto the cell surface as glycoprotein through metabolic labelling could be visualised in living cells. Other groups have followed the approach and used similar click reagents to study cell-surface glycans via non-natural sugars [101], or other mammalian cell- surface proteins using lipolic acid ligase [102].

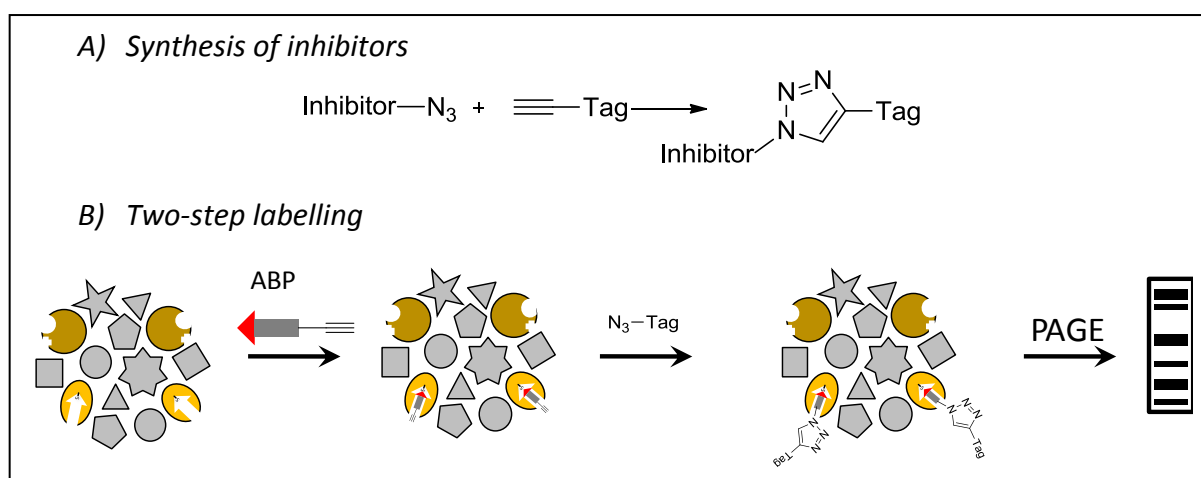


Figure 1. 5: Application of CC in ABPP. A) Synthesis of inhibitors. B) Two-step labelling.

1.3.3 Affinity enrichment of labelled proteins

Affinity pull down facilitates the enrichment of labelled proteins and separates them from non-labelled proteins, significantly reducing the complexity of the investigated proteome before mass spectrometric analysis (see **Figure 1. 1**) [103-105]. Here in this work, affinity enrichment is used for identifying the most catalytically active proteins. Although pull down assays are widely applied in chemical biology, their use still has some limitations, as discussed below.

1.3.3.1 Solubility of target proteins

Depending on the purpose of the study, preservation of the nature of labelled proteins, such as folding and assembly, may be desirable. In that case, it is necessary to avoid intensive agitation and degradation and therefore experiments should be implemented under non-denaturing conditions. Soluble proteins such as cytosolic proteins do not tend to be problematic; however, insoluble proteins such as intact membrane proteins and insoluble cell components require more aggressive conditions such as sonication and strong detergents to solubilise the proteins [106-107]. It is possible that during the sample preparation some proteins are not solubilised and therefore are not pulled down.

1.3.3.2 Abundance of the target protein

One crucial requirement for any successful chemical proteomics experiments is sufficient expression of the target protein and the presence of potential off-target proteins in the cell; proteins with very low abundance in the cell may be missed. Another factor contributing to success is the strong binding of the ABPs to the target proteins. Besides the specific covalent bond between ABPs and target proteins, there is also non-specific binding (probably *via* non covalent, hydrophobic or charged surface interactions) of ABPs with abundant proteins and non-specific binding of abundant proteins to the affinity beads [13-14]. Consequently, the amount of enriched proteins does not give clear insight into the type of the binding (specific or non-specific) or the function of the protein (target or off-target).

1.3.3.3 Background binding of proteins

A number of background proteins can co-purify with the target proteins and cause a high background signal. These “contaminant” proteins could be those that bind non-specifically to the beads or to the target protein through second- or third-shell protein-protein interactions (“piggybacking”) [13]. Using increasingly stringent washing methods could

reduce the level of background proteins; however, it could also remove lower-abundance target proteins. Therefore, it is important to choose a suitable wash method for the pull down experiment.

Usually, appropriate negative controls are introduced to enable the identification of background proteins by comparing proteins captured with non-treated samples and treated samples (either with active or inactive ABPs). This strategy could be problematic in the case of more complex proteomes due to the potentially high level of background, requiring other strategies, e.g.: the competition binding approach [32, 108], a stable isotope labelling strategy [12, 109] or serial affinity chromatography [110-111].

1.3.3.4 Biotin binding proteins

Biotin, known as Vitamin H, is an essential enzyme cofactor. It is acquired from various sources, for example, it is biosynthesised by human intestinal microflora. Many bacteria transport biotin into the cell via an active uptake mechanism, as it is an essential micronutrient for normal cellular function, growth and development [112-113]. Therefore biotin appended ABPs might be delivered more easily into the cell [114-115] and may target further potential proteins inside the cell. However, the key feature of biotin, responsible for its wide use in molecular biology, is its very high affinity for the protein avidin [116]. Attaching biotin to a protein of interest directly or via inhibitors/ABPs enables identification and/or characterisation of these proteins using protein tools such as avidin, or related analogues streptavidin and neutravidin [117], [118].

Avidin is a tetrameric glycoprotein isolated from egg white and other tissues (birds, reptiles and amphibia) [117]. It exhibits very strong binding to biotin in one of the strongest non-covalent bonds known. The carbohydrate moieties of avidin, a source of binding to lectin and sugars, account for about 10% of its total mass.

Streptavidin isolated from *Streptomyces avidinii* possesses a high affinity for biotin, similar to that of avidin [118]. Unlike avidin, streptavidin has no glycosylation moiety and therefore it exhibits less non-specific binding. In addition, streptavidin is much less soluble in water than avidin [119].

NeutrAvidin is a deglycosylated form of avidin [120], commercialised by Thermo Scientific and Invitrogen. Because it lacks the glycosylation site, NeutrAvidin shows very low non-specific binding. In addition, NeutrAvidin lacks the Arg-Tyr-Asp sequence, the motif that binds to cell surface proteins, possessed by streptavidin, and therefore it exhibits even less non-specific binding [120]. NeutrAvidin agarose resin is prepared by covalently coupling NeutrAvidin protein onto beaded agarose (polysaccharide based), which is stable at pH 2-11 and leach-resistant.

A comparison of these three biotin binding proteins is given in **Table 1. 1**. Affinity for biotin is similar in all cases; however, NeutrAvidin has the highest specificity and the lowest non-specific binding.

Comparison of biotin-binding proteins.			
	Biotin-Binding Protein		
	Avidin	Streptavidin	NeutrAvidin
Molecular Weight (of tetramer)	67K	53K	60K
Biotin-binding Sites	4	4	4
Isoelectric Point (pI)	10	6.8-7.5	6.3
Affinity for Biotin (K_d)	10^{-15} M	10^{-15} M	10^{-15} M
Nonspecific Binding	High	Low	Lowest

Table 1. 1: A comparison of biotin-binding proteins [121].

1.4 Application of ABPs to different protease families

1.4.1 Proteases

Proteases are proteolytic enzymes that catalyse the selective hydrolysis of peptide bonds (**Figure 1. 6**). Proteases form one of the largest and most important groups of enzymes. They are involved in various important physiological processes such as protein turnover, digestion, cell differentiation and growth, cell signalling and migration, fertilization, blood coagulation and wound healing, immunological defence, and apoptosis [122]. Due to their ubiquity and roles in many biological processes, inhibitors of such proteases allow the regulation of undesired protease activity and can be used for therapeutic applications or as tools for biological studies [123].

Proteases have been classified into evolutionary families and clans by Rawlings and Barrett (1993, 1994) [124-125] based on similarities in tertiary structure, the order of catalytic residues in the primary sequence or common sequence motifs around the catalytic residues. This classification led to the development of the MEROPS database of proteases (<http://merops.sanger.ac.uk>) [126]. This database contains the listing of all known protease sequences of different families and clans.

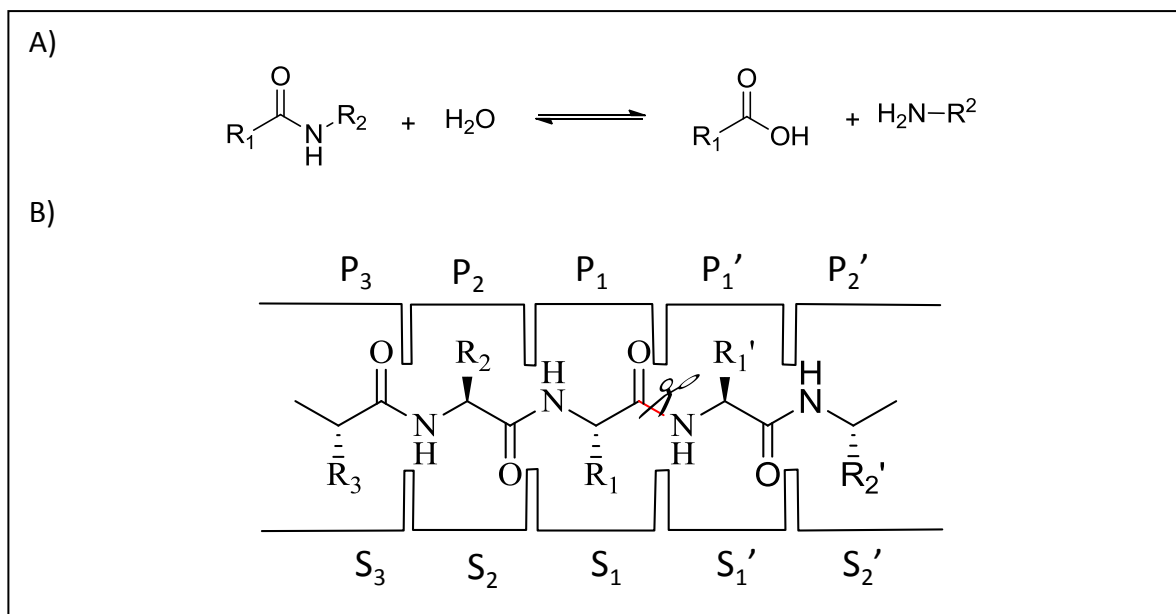


Figure 1. 6: A) Equation of peptide hydrolysis. B) Standard nomenclature for substrate residues (P_n) and their corresponding binding sites (S_n) [127].

Generally, proteases are divided in four major classes: aspartic, serine/threonine, cysteine, and metallo proteases [124-125]; they possess similar mechanisms of action, which involve an initial attack of a nucleophile onto a substrate carbonyl function. This nucleophile is a water molecule in the case of aspartic and metalloproteases, or a serine, cysteine or threonine in the remain classes [122].

Most proteases are substrate sequence-specific, so that their chemical and physical properties define the possible binding of amino acid side chains of the substrates. **Figure 1. 6** shows the standard nomenclature introduced by Schlechter and Berger [127] to designate substrate or inhibitor residues (e.g. P_3 , P_2 , P_1 , P_1') that bind to the corresponding protease subsites (S_3 , S_2 , S_1 , S_1').

Although proteases have long been known to be important in human pathological process and approximately 10% of drugs currently on the market target proteases, their activity and

regulation in disease are still not fully understood. Furthermore, the function of a vast number of encoded proteases predicted by genomic analysis still remains unknown. Application of ABPP could aid further understanding of proteases activities as well as implication and regulation in disease pathologies, or even identification of novel proteases as potential therapeutic targets. The application of ABPP for each protease class is discussed in the following sections.

1.4.2 Serine and Threonine proteases

Comprising approximately 1% of the predicted protein products of mammalian genomes, the serine hydrolases known as serine proteases are an extremely large family of enzymes [128-129]. Sharing a common catalytic mechanism involving a nucleophilic serine residue, this enzyme family catalyses the hydrolysis of ester and amide bonds and play a significant role in a huge number of physiological and pathological processes including diabetes [130], cancer [131-132], inflammation [133-134] and angiogenesis [135].

Fluorophosphonates (FP) - broad spectrum irreversible inhibitors that target over 100 distinct serine hydrolases including proteases, peptidases, lipases, esterases and amidases - have been used as a powerful tool to characterise this class of enzyme in complex proteomes [10, 132]. These inhibitors benefit from several advantages such as selectivity towards serine hydrolases [136] and no cross-reactivity with other classes including cysteine, aspartic and metalloproteases [137]. Utilising these advantages, Cravatt and colleagues applied a panel of FP-biotin or fluorescein probes to characterise catalytically active serine hydrolases in human and mouse proteomes [10, 74, 132, 138]. Labelling was observed only with active hydrolases and not with inactive forms (serine nucleophile mutant, zymogen or inhibitor-bound hydrolase) [10, 132]. Despite these successes FP-based ABPs suffer from a drawback: broad reactivity prevents the application of these probes to lower-abundance enzymes. Recognising this limitation, a new generation of probes was developed named diphenylphosphonates possessing similar reactivity and mechanism towards serine hydrolases as FP-based probes but with narrower selectivity to trypsin-like serine proteases such as plasmin, β -tryptase and trypsin itself [133, 139-140]. Notably, adding a P2 proline residue into the probe design led to a significant increase of the

reactivity and selectivity of these probes towards tryptase compared to trypsin, confirming the important role of the probe design [133].

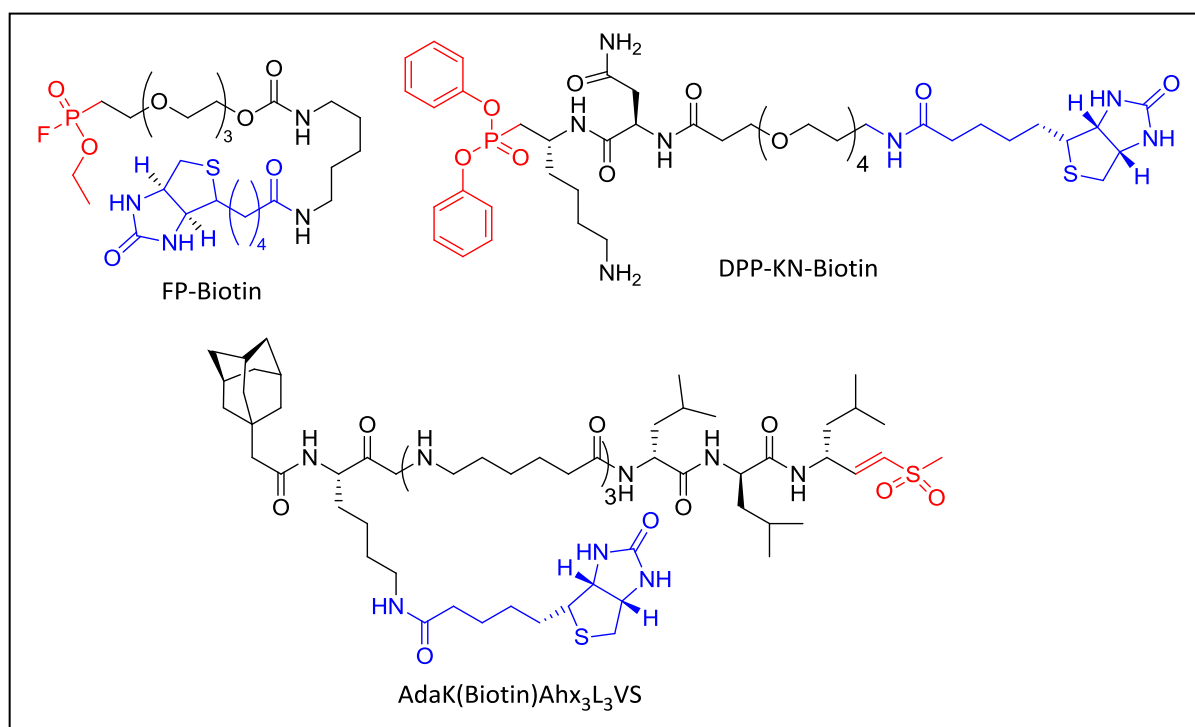


Figure 1. 7: Structure of ABPs for serine proteases. FP-ABPs possess a relative broad-spectrum reactivity [138], whilst DPP-ABPs displays more selectivity to trypsin-like proteases [133, 139-140]. AdaK(Biotin)Ahx₃L₃VS probe was applied to label proteasomes in live cells [75].

Another target of serine hydrolase family is the proteasome, a primary cytosolic protein degradation machinery present in almost all living cells [141]. Three types of active sites were found in all eukaryotic proteomes: β 1 as caspase-like, β 2 as trypsin-like and β 5 as chymotrypsin-like [75, 142]. In addition, a threonine protease-like activity at the N-terminus was found from mutational and structural studies of the proteasome [75]. Several groups have been working on inhibition of proteasome both in live cells and *in vitro* using vinyl sulfone-based probes [56, 97, 142]. For example, Ploegh and co-workers targeted β -subunits and the Overkleeft lab has developed site-specific probes towards β 1 and β 5 units [142], while Kessler and colleagues concentrated on substrate specificity using peptide-based vinyl sulfone probes [75].

1.4.3 Cysteine proteases

Cysteine proteases represent the second largest class of proteases involved in various human pathologies such as arthritis [143], apoptosis [144], asthma [145], osteoporosis [146]

and cancer [145, 147]. Many electrophilic inhibitors that specifically target the nucleophilic cysteine thiol in the catalytic triad have been developed for cysteine proteases including diazo- and fluoromethylketones (FMKs) [148], vinylsulfones [149], epoxides [51, 59, 150-151] and acyloxymethyl ketones (AOMKs) [152-154].

ABPs based on the epoxysuccinyl inhibitor E-64 [155], a natural product with many valuable properties such as a peptide-like structure and potent inhibitory activity towards several papain-like cysteine proteases, were studied extensively by several groups [51, 53, 72, 156-157]. Bogyo and co-workers have developed bifunctional probes DCG-03 and -04 for profiling the activities of enzymes from crude extracts of cell and tissues, identifying multiple cathepsins [157]. Extending this study, this group has generated several second-generation libraries of diversified, “double-headed” epoxysuccinyl probes for investigating the selectivity among cysteine proteases, or containing a cell-permeable unit for *in situ* monitoring of enzyme function [72]. Furthermore, development of isotope-code ABPs enables quantitative analysis of cysteine protease activity [158]. Significantly, these studies highlight the ability of appropriately applied ABP design to tune the selectivity for individual cathepsins. Recently, Bogyo and colleagues have optimised aza-peptidyl epoxide ABPs that specifically label legumain, a lysosomal cysteine protease, for whole body non invasive imaging [51].

Ploegh and co-workers have also recognised the potential value of E-64 and developed an azido-tagged E-64 based probe to target cathepsin B in primary macrophages [156]. Cathepsin B is an abundant papain-like cysteine protease that has been shown to contribute to microbial infections, apoptosis and cancer [147, 159]. In this study, the investigators applied two-step non-radioactive labelling using the Staudinger ligation to evaluate cathepsin activity in living cells [156]. In addition, affinity enrichment of labelled proteins enabled subsequent mass-spectrometry analysis and fluorescent imaging allowed visualisation of the subcellular activity of cathepsin B during infection of macrophages by *Salmonella typhimurium* within sub-cellular compartments of individual cells.

AOMK probes display an exceptional class-wide reactivity with cysteine proteases and were studied by several groups. Bogyo and co-workers have reported AOMK probes that targeted two major sub-classes of cysteine proteases: the CD clans containing caspase-3, legumain

and arginine- and lysine-gingipains and the CS clans containing cathepsin B and L [53, 160]. The specificity of these probes towards certain enzymes was achieved by diversification of the single amino acid or an extended peptide attached to the AOMK warhead in the library of ABPs. These chemical tools were also applied to label multiple cysteine proteases of the above mentioned clans in crude cell extracts. Recently, this group has reported elegant experiments showing dynamic labelling of specific cysteine proteases using fluorescent cell-permeable quenched AOMK probes [161].

Vinyl sulfone based probes, if specifically designed, could also target cysteine proteases in addition to their use for serine protease profiling discussed above. For example, Bogyo and co-workers have developed a vinyl sulfone based probe, FY01, containing free α -amino dipeptides to target cathepsin C in rat liver extract [162]. Cathepsin C, expressed in most mammalian cells, is a lysosomal cysteine protease also known as dipeptidyl aminopeptidase I. This protease has a unique characteristic within the CA clan of the papain family: the ability to remove sequentially dipeptides from the amino-terminus of larger peptides. This property was integrated in the design of FY01. In the latest study, FY01 was found in another application for high throughput screening to identify novel enzyme inhibitors [52].

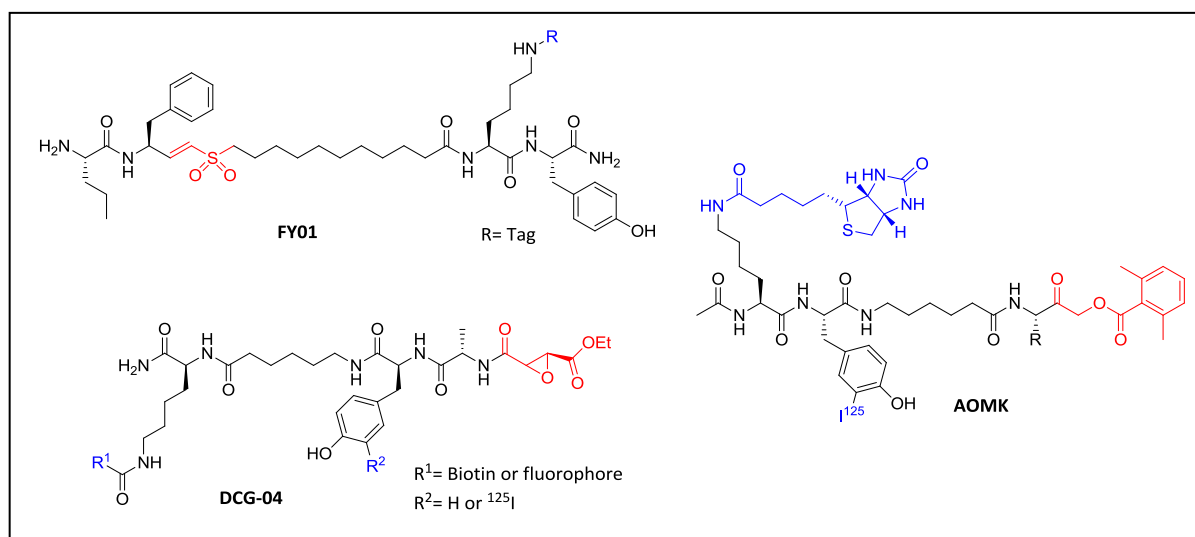


Figure 1. 8: Structure of ABPs for cysteine proteases. FY01 is a vinyl sulfone based probe targets capthepsin C [162]. AOMK probes with variation at P2 position (indicated as R in the structure) can label specifically CD and CS clans of cysteine proteases [53]. DCG-04 has led to identification of multiple cathepsins in crude cell extracts [157].

1.4.4 Metalloproteases and aspartic proteases

Metalloproteases (MPs) and aspartic proteases differ from the protease classes discussed in previous sections in that they do not covalently bond to their substrates but utilise an activated water molecule for catalysis [163-164]. Therefore, ABP design for these proteases is more challenging, as mechanism-based inactivation by electrophilic reagents cannot be used. Alternatively, ABPs can exploit tight-binding, broad spectrum reversible inhibitors by incorporation of a photoreactive group into the inhibitor structures so that these ABPs bind to the target enzyme in a reversible manner.

MPs, a relatively large and diverse family of enzyme that uses a metal-activated water molecule, often a zinc ion (Zn^{2+}), for catalysis [163] regulate a number of physiological and pathophysiological processes including tissue remodelling [165], cancer [166-167], arthritis and heart disease [167]. The precise biological roles of MPs are still unclear and broad spectrum inhibitors of this enzyme class that entered clinical trials suffered from dose-limiting toxicity. Therefore, profiling and characterisation of MPs is a pressing need in the chemical biology and medicinal chemistry fields. Working towards this purpose, Cravatt and colleagues have generated probes that selectively bind and modify the matrix MP active site [168]. The core features of these probes include a hydroxamic acid moiety to chelate the conserved zinc atom in MP active site, a peptide-like scaffold containing a benzophenone group for photoinduced chemical cross-linking of the active sites, and a reporter unit (biotin or fluorophore) for manipulating the labelled target.

Similarly, Yao and colleagues reported a library of complementary peptidyl hydroxamate AfBPs with a diazirine moiety instead of the benzophenone group [169], to target thermolysin from yeast extracts in an activity-dependent manner. Interestingly, the probes generated by both laboratories targeted only the active form of MPs and not the inactive form such as inhibitor bound or zymogen. As a result, these probes can be used in competitive ABPP experiments to evaluate selectivity of AfBPs in whole proteomes.

However, single fluorescent probes reported by Saghatelian *et al.* for identifying MPs in lysates suffered from the narrow spectrum of MPs identified and suboptimal detection of the low level of MP activities due to inherent limitations of the method [168]. The Yao group has improved on this by generating a library of probes by click synthesis for activity-based

finger printing of matrix MPs [170]. These probes bear a succinyl warhead and a diverse P1 functionality to achieve enhanced specificities. In addition, these probes were applied to high throughput screening of MP activities in a protein array [38].

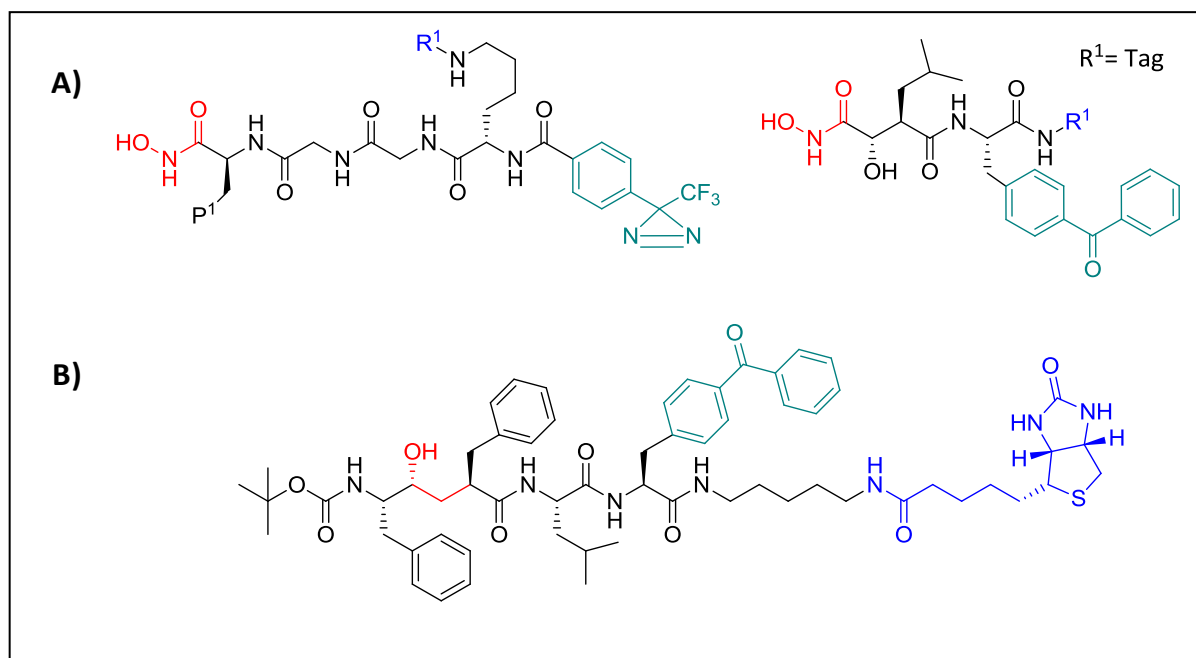


Figure 1. 9: Structure of ABPs for metalloproteases and aspartyl proteases. A) Representative photoreactive hydroxamate-based ABPs for metalloproteases [168-169]. B) A photoreactive hydroxyethylene-based ABP for aspartyl proteases [171].

ABPs used for profiling aspartic proteases were also started from the template of high-affinity, reversible inhibitors. Liu *et al.* at Merck have developed ABPs with similar key features as those for MPs, which contain a hydroxyethyl moiety to serve as a transition state mimic, a peptide-based scaffold possessing affinity to aspartic proteases, a benzophenone for photo-induced cross-linking and a biotin or fluorophore group for visualisation and identification [171-172]. These probes selectively labelled the solubilised γ -secretase from HeLa cell lysates. As with MP probes, the generated probes labelled only the active form of γ -secretase.

1.5 Mass spectrometry analysis

Mass spectrometry (MS) plays a crucial role in chemical proteomics due to its wide applicability in biology, for example in the characterisation of protein complexes [6], providing information about the localisation of proteins in cells and tissues [173], classifying bacterial species and subspecies [174], cataloguing the sub-cellular content of structures

and organelles [175], defining protein interactions and identifying sites of protein modification or post-translational modification (PTM) [176]. The use of MS for chemical proteomics is supported by rapid advances in MS instrumentation, fragmentation methods and computational analysis. In this section, the physical principles of each element of the mass spectrometer and its advances are introduced to provide an overview and an understanding of the methods used for the present study (described later).

1.5.1 MS instruments

MS measures accurately the mass-to-charge ratio (m/z) of a molecule. A general mass spectrometer is composed of three elements [173]: an ion source ensuring that sample molecules are ionised and brought into the gas phase, a mass analyser(s) separating the gas-phase ions on the basis of mass-to-charge ratio using electromagnetic fields in a vacuum, and a detector responsible for recording the presence of ions (**Figure 1. 10**).

Ion sources

There are two well established ionisation sources commonly used in proteomics: matrix-assisted laser desorption/ionisation (MALDI) [177] and electrospray ionisation (ESI) [178]. MALDI sources are typically based on a pulsed nitrogen UV laser and generate singly charged peptide ions, while ESI sources typically heat a needle to between 40 °C and 100 °C to assist nebulisation and evaporation of samples, thereby generating singly and multiply charged peptide ions (2^+ , 3^+ , 4^+). Both methods are sensitive, but MALDI is more tolerant than ESI to small amounts of contaminants, such as salts and detergents [5].

Mass analyser

The mass analyser is the core of MS technology. There are four commonly used types of analyser: quadrupole (Q) [179], ion trap (IT) [180], time-of-flight (ToF) [181] and Fourier transform ion cyclotron resonance (FT-ICR) [182-183]. Each of the analysers possesses distinct advantages and limitations due to their physical principles and performance standards, as well as their mode of operation. Depending on the purpose of the proteomics study, one can choose the appropriate analyser to support the chosen analytical strategies. Because no instrument can currently offer all capabilities simultaneously, the determination of which platform and strategy is preferred is restricted from the outset: the focus is either

on identification or quantification. In the case of prioritising identification, resolving power (to obtain improved separation of the various components) and mass accuracy are important. On the other hand, sensitivity, dynamic range and multiple reaction monitoring (MRM) capability are the emphasis in the case of an experiment focusing on quantification.

In ToF analysers, the velocity of the ion through a tube of specific length is related to the mass-to-charge ratio and therefore mass separation can be achieved by different time of flight of ions towards the detector [181]. ToF analysers can be operated with both MALDI and ESI ion sources and their performance has significantly improved in the last few years [5]. Several combinations of ToF with other analysers, such as Q-Q-ToF and ToF-ToF have been used to improve the sensitivity, resolution and mass accuracy [5, 184].

In IT analysers, effective trapping of ions in an oscillating electrical field leads to accumulation of these ions over time in a physical device [180]. Mass separation is acquired by tuning the oscillating field to eliminate only ions of a specific mass. IT possesses unrivalled sensitivity and fast data acquisition but is limited by low resolution and ion trapping capacity [185]. The dynamic range and overall sensitivity of IT instruments was improved by the development of the linear ion trap (LIT or LTQ), as this instrument offers a unique capability for analysis of PTMs [186].

In the quadrupole analyser, which consists of four rods forming two pairs running different current modes, ions are filtrated by high-pass mass between the first pair of rods and by low-pass mass between the other pair of rods. This filtration can be fine-tuned to overlap only in a specific mass-to-charge ratio interval. Ions with m/z values outside this interval will be excluded. Quadrupoles are usually applied in conjunction with other analysers (e.g.: Q-Q-ToF, Q-Q-Q, QQ-LIT) to expand either sensitivity or selectivity [5, 184].

In the FT-ICR analyser, the ions are trapped in orbits by a very strong magnetic field in a particle accelerator called a cyclotron, while being accelerated by an applied voltage [182-183]. The cyclotron frequency is related to m/z and Fourier transformation is required to extract individual ion frequencies. This technique allows measurement to accuracies in the low ppm to sub-ppm range. High resolution leads to better quality data and boosts the peak capacity, therefore this instrument can detect more signals compared to low resolving power instruments.

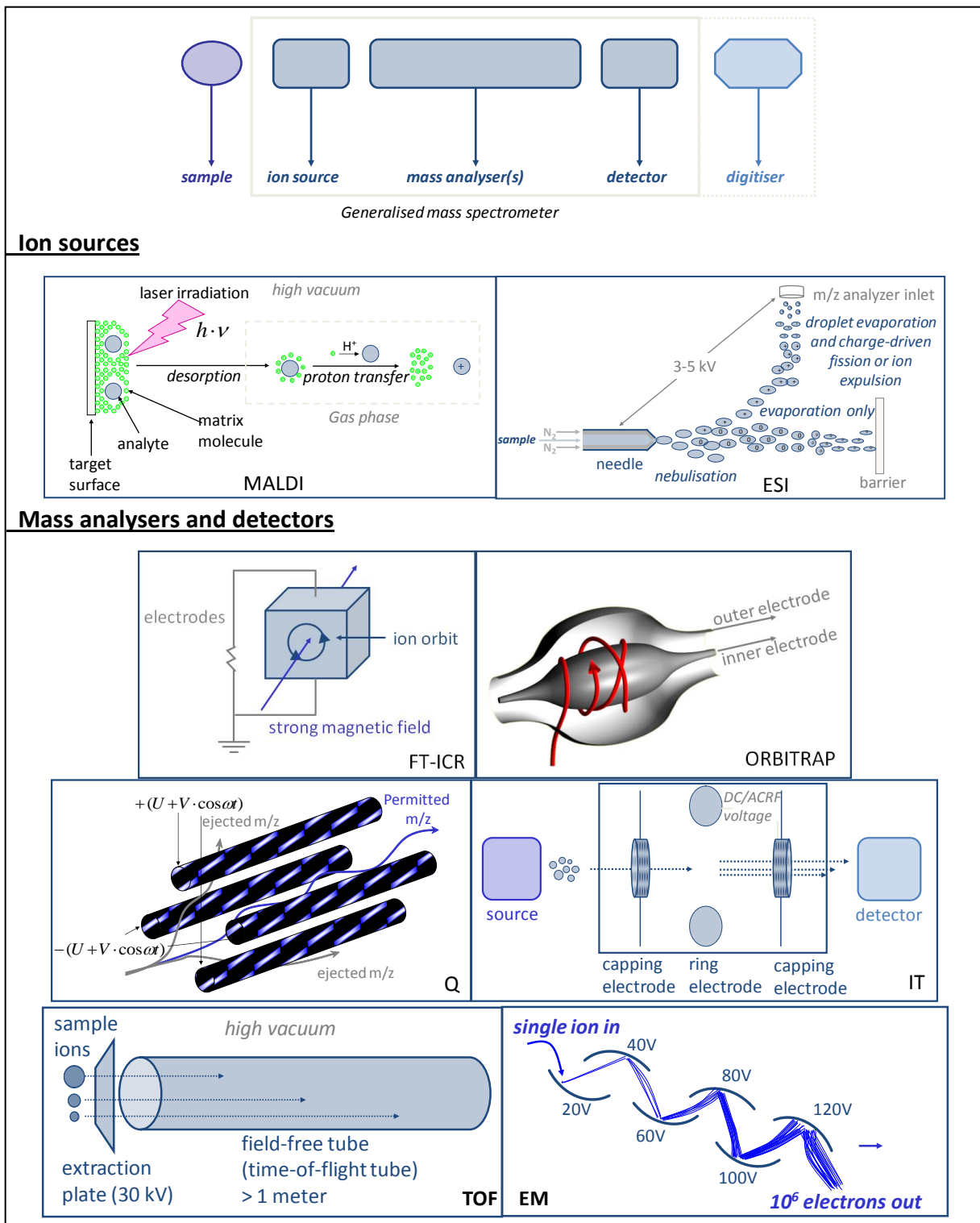


Figure 1. 10: Mass spectrometry instruments. The basic physical principles of each component of the mass spectrometer are shown. MALDI and ESI are ionisation sources and introduce sample into the analyser. Q analyser select by time-varying electric fields between four rods, which permit a stable trajectory only for ions of a particular desired m/z . In the IT, ions are trapped in an oscillating electrical field leading to accumulation of these ions over time. The FT-MS instrument also traps the ions, but does so with the help of strong magnetic fields. In Orbitrap, ions are captured and form an orbital harmonic oscillation along the axis of the field with a frequency characteristic of their m/z values. In TOF, the ions are accelerated to high kinetic energy and are separated along a flight tube as a result of their different velocities. At the detector (EM) ions are impinged on several Faraday cup dynodes, amplified and counted. Figure adapted from the lecture of Dr. Martens, EBI, Cambridge, UK.

In recent times, a new type of mass analyser, the Orbitrap [187], based on a new physical principle, was introduced to the field of proteomics [188]. It consists of an outer and inner coaxial electrode, which generate an electrostatic field. In the Orbitrap, ions are captured and form an orbital harmonic oscillation along the axis of the field with a frequency characteristic of their m/z values. This frequency is measured with very high precision and can then be calculated by Fourier transform. The performance of the Orbitrap is similar to the FT-ICR, but this instrument is cheaper (because it does not require an expensive superconducting magnet), much more robust and simpler to maintain [189-190]. Used in conjunction with a linear ion trap (LTQ), a new mass spectrometer named the LTQ-Orbitrap (introduced by Thermo Fisher Scientific [191]) demonstrated the combined advances of both techniques in robustness, sensitivity and MS/MS capability, possessing very high mass accuracy and high resolution capabilities, and has become a powerful tool in proteomics studies.

Detector

The most common type of detectors are electron multipliers (EM) [192], which are based on several Faraday cup dynodes with growing charges to generate an electron cascade based on a few incident ions.

MS mode of operation

Both MS mode and MS/MS mode are widely applied in proteomics. In MS mode, only one MS selection takes place and provides peptide mass spectra that allow subsequent identification of the parent proteins using peptide mass fingerprinting (PMF) [176, 193]. In contrast, MS/MS, also known as tandem-MS, employs two mass analysers in series (tandem) with some form of fragmentation taking place in between each MS selection [5]. The first mass analyser serves as an ion selector, by selectively passing through only ions of a given m/z . The second mass analyser positioned after fragmentation is triggered and exerted in its normal capacity as a mass analyser for the fragments. Tandem-MS allows determination of the amino acid sequence of a specific peptide, which can also be used for identification of parent proteins by PMF but with more confidence in the identification results. Various types of tandem-MS experiments are reviewed in several publications [5, 103, 175, 190, 194].

Fragmentation methods

Several fragmentation techniques are utilised in proteomics, such as collision-induced dissociation (CID) [195], post-source decay (PSD) [196], electron-capture dissociation (ECD) [197-198] and electron-transfer dissociation (ETD) [199]. CID is the most widely used method and relies on multiple collisions events of a gas-phase peptide with rare gas atoms to provide the peptide precursor with sufficient energy to fragment. CID typically causes backbone fragmentation [195]. On the other hand, PSD is based on a single unimolecular event activated by laser energy, in which a highly energetic (metastable) ion spontaneously fragments [196]. PSD also typically causes backbone fragmentation. ECD and ETD rely on a single impact of an electron with a peptide precursor either by capture of a thermal electron or transfer of electrons from radical anions to the protonated peptide/protein. This high-speed impact immediately imparts sufficient energy to fragment the precursor. Like CID and PSD, ETD and ECD also typically produce backbone fragmentation, however ECD is only workable in FT-ICR mass spectrometers, whereas ETD is used in traps.

1.5.2 Bioinformatics/database search approaches

Mass spectra contain the mass and charge as well as the intensity of the parent ion and its fragment ions. These data are submitted to a search engine such as Mascot [200], SEQUEST [201], X!Tandem [202] or OMSSA [203], which match the provided data to *in silico* fragmented peptide spectra or a peptide data base containing all experimentally-identified peptides and proteins [204]. Despite the huge number of all theoretically possible amino acid combinations, there is only a very minor subset of protein sequences that are formed in nature, and therefore a short peptide sequence is already highly protein-specific. However, it should be noted that the identified peptide can be also matched to several proteins, if these proteins are highly homologous. In that case, the number of unique identified peptides that are matched unambiguously to only a single protein is important. Subsequently, statistical validation of identified peptides and proteins provides scores that are used to assess the quality of matches.

Mascot and SEQUEST are the most widely used search algorithms, but their principles are fundamentally different [204]. Several studies had compared these algorithms and others to offer a starting point for researcher to choose an appropriate solution for their study [204-

206]. Mascot scoring algorithm is probability based and scores the best match of the observed and predicted spectra with the highest score and a significant match is typically a score of the order of 70 [200]. In contrast, SEQUEST algorithm scores the alignment between observed and predicted spectra using empirical and correlation measurements [201]. Traditionally, TOF-acquired spectra are interpreted with Mascot, whereas SEQUEST is used to sequence ion-trap spectra [206]. Both algorithms are fairly comparable and can complement each other [206].

There is a huge number of databases, both public or commercial, available to search engines. Some well known databases that are widely used in proteomic studies include EMBL Nucleotide DB (European Molecular Biology Laboratory) [207], DDBJ (DNA Data Bank of Japan) [208], GenBank [209], NCBI (National Center for Biotechnology Information) [210-211], Ensembl (Sanger Institute and EBI) [212], UniProt (Universal Protein Resource) [184], Swiss-Prot (Swiss Institute of Bioinformatics) [213], and PRIDE (Proteomics Identifications Database) [214]. Most of the databases are linked and easily navigated between each other. NCBI and EMBL have been used in the present study.

1.5.3 MS-based proteomic strategies for protein identification and PTM characterisation

Proteomic studies diverge in their objectives: they may focus on the identification of proteins in a complex proteome and possible the PTMs of those proteins or they may try to address the absolute quantification of proteins or quantitative change in protein abundance between samples. Strategies for protein identification and PTM characterisation have emerged in two different approaches illustrated in **Figure 1. 11**: a “top-down” approach (whole-protein analysis) and a “bottom-up” approach (analysis of enzymatically or chemically digested peptides).

1.5.3.1 Top-down proteomics

In this strategy, complex cellular lysates are separated or fractionated, then ionised into gas-phase intact ions or large protein fragments. Subsequently, these ions and fragments are measured by a high resolution mass analyser following their direct fragmentation without any pre-digestion. The strengths of this strategy are the potential acquisition of the whole protein sequence and the ability to trace and characterise possible PTMs of proteins [190,

215]. Furthermore, the possible variations and inconsistencies related with the digestion process can be avoided. However, this approach suffers from ambiguous and difficult data analysis resulting from the very complex spectra obtained from multiply charged proteins [216]. Further limitations such as expensive implementation costs (associated with the requirement for high resolving power instruments such as FT-ICR, LIT-FT-ICR or LIT-Orbitrap mass spectrometers), the size limit of analysed proteins (not larger than 50 kDa) and primitive bioinformatic tools disfavour this strategy compared with the bottom-up strategy. However, this approach is relatively young in its development and there is still scope for future improvements.

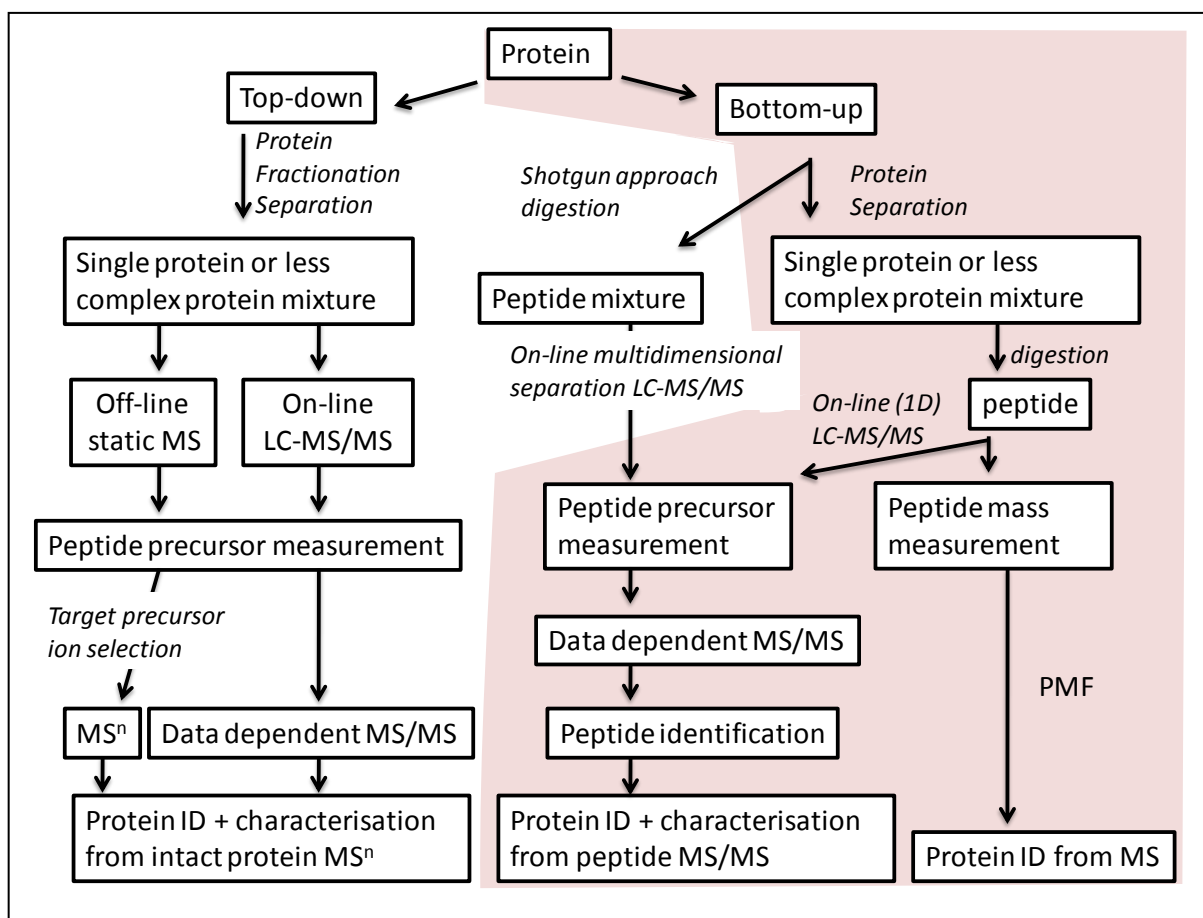


Figure 1. 11 : Workflows for MS-based protein identification and characterisation [190]. Bottom-up or top-down approaches can be used to analyse proteins. In the top-down approach, a protein mixture is fractionated and separated before analysis *via* either off-line static MS or on-line LC-MS/MS. In the bottom-up approach, a protein mixture may be separated before enzymatic (or chemical) digestion followed by either direct PMF-based acquisition or further peptide separation on-line coupled to tandem MS. Alternatively, the protein mixture can be directly digested and then separated by multidimensional chromatography on-line coupled to tandem MS analysis (the ‘shotgun’ approach). The highlighted workflow was applied in this study.

1.5.3.2 Bottom-up proteomics

Unlike the top-down approach, analytes for bottom-up proteomics are pre-digested before being introduced into the mass spectrometer. Two divergent workflows are commonly conducted: “break-then-sort” or “sort-then-break”.

The break-then-sort approach, also known as shotgun proteomics, deals with analytes that are digested without any pre-fractionation/separation [190, 215]. This approach has the advantage that it is conceptually and experimentally simple to implement, but it suffers from challenging data analysis (due to interfering peptide and protein sequence identities), as well as the massive complexity of the samples [216].

On the other hand, the sort-then-break approach offers samples with a much lower degree of complexity, because proteins are separated or fractionated prior to digestion followed by direct PMF [193] or further on-line separation using LC interfaced to a tandem-MS [215]. In addition, this approach may provide access to physical properties (e.g. approximate mass, isoelectric point) of the intact proteins in addition to their identities.

Peptide mass fingerprinting (PMF)

PMF [193] allows the identification of proteins by comparing the peptide masses generated during the MS analysis to the calculated peptide masses produced by *in silico* cleavage of proteins or gene sequences in the database. This method profits by saving the time and cost involved in the *de novo* peptide sequencing process.

1.5.3.3 Quantification method

Protein quantification aims to compare a stressed or perturbed state to an unperturbed reference sample. Several reviews describe extensively the strategies for quantification in MS-based proteomics [4-5, 173, 175, 194, 217-218]. There are several methods to determine the (relative) quantity of a protein using MS. Usually, these methods are categorised in three types: metabolic labelling (e.g.: ¹⁵N labelling [219], stable isotope labelling by amino acid in cell culture (SiLAC) [220]), chemical labelling (e.g.: isotope-code affinity tag (ICAT) [221], isobaric tags for relative and absolute quantification (iTRAQ) [222], standard peptide [223]) and label free quantification (e.g.: ion intensities [224], spectral count [204], divided indices [225-226]).

Stable isotope labelling approaches, whether metabolic or chemical labelling, share one common characteristic in that they are based on the mass differences of unlabelled and labelled peptides with stable isotopes, as these peptides exhibit the same properties during chromatography [218]. Metabolic labelling and chemical labelling are distinguished in their methodology: chemical labelling is introduced to the proteins or tryptic digested peptides through a chemical reaction, whilst metabolic labelling is achieved in the whole cell or organism during growth *via* the growth medium or feed.

Label-free quantification is based on the number of acquired spectra for each protein or on the ion intensities of identical peptides. These approaches are inexpensive, quick and effective for relative quantification of proteins [194, 218]. Spectral counting is one of the most applied, which operates under the assumption that if a protein is more abundant in one sample than another then the number of spectra corresponding to this protein will correlate linearly with its abundance during analysis [194, 218]. This is attainable if sample treatment is unbiased and samples equally represented. Although this quantitative approach is reliable for determining the relative abundance of proteins identified with many spectra (peptides), it is important to note that this approach is relatively unreliable for quantifying low-abundance proteins because such proteins will produce very few spectra leading to low spectral counts and consequently lower reproducibility [227]. For example, if only one spectrum was found for protein A in sample one and four spectra for protein A in sample two, it is not valid to conclude that the abundance of protein A is four-fold higher in sample two compared to sample one since the run-to-run variability in spectral count for a given protein is frequently of the same order of magnitude.

There are many ways to compare spectral abundance, such as unweighted spectral counting (a method of counting peptides in all instances when they are shared between proteins), quantitative value (a normalised version of the unweighted spectral count), number of unique spectra or unique peptides, number of assigned spectra (the total number of spectra that matched to a peptide in the protein), percentage of total spectra (the number of spectra matched to the protein as a percentage of the total number of spectra in the sample) and the relative abundance or fold change (the spectral count ratio of the sample/protein to the control sample/protein).

1.6 *Clostridium difficile*

Clostridium difficile is an anaerobic Gram-positive bacterium and produces spores that can survive for a long time in the environment. It was first isolated by Hall and O'Toole in 1935 and named *Bacillus difficilis* due to difficulties of isolation and culture [228]. However in the late 1970s this bacterium was identified as the cause of diarrhoea and colitis occurring mostly as a complication of antibiotic therapy [229]. During the past few years there has been renewed interest in *C. difficile*, due to the recognition that the disease is more common, more severe, and more resistant to standard treatment than previously thought [230-231]. Recognised now as the main cause of infectious diarrhoea, *C. difficile* is the most frequent nosocomial (hospital acquired) pathogen. It causes a wide range of gastrointestinal diseases ranging from mild diarrhoea to more life threatening pseudomembranous colitis (**Figure 1. 12**) [232].

C. difficile is resistant to many antibiotics and infection is associated with antimicrobial therapy. Many individuals are asymptotically colonised with *C. difficile*, but it is thought that the resident microflora restricts growth of *C. difficile* and limits bacterial growth and toxins expression. Antibiotic treatment can disrupt the normal bacterial gut flora, allowing *C. difficile* growth and toxin production. The acquisition of *C. difficile* infection (CDI) is through spread and transmission of spores, usually by the faecal-oral route through contact of the patients with the hospital environment or health care workers. *C. difficile* spores are highly resistant to heat and disinfectants and can persist in the hospital or nursing home environment for months or even years. Heat-resistance even helps spores to survive typical hospital laundering cycles and cross-contamination of bed linen during the wash cycle can occur. Once spores are ingested, they pass through the stomach unscathed because of their acid-resistance, and then germinate to form vegetative cells. The colonised *C. difficile* strain can be either non-toxigenic or toxigenic. Toxigenic *C. difficile* strains produce and secrete at least one of the two major toxins, toxin A (TcdA) and toxin B (TcdB) [233-237], which are the main pathogenicity factors of the bacterium. If the patients are infected by a non-toxigenic strain or their body can mount an antibody response to the toxin, they are protected from CDI.

The drugs of choice for treatment of CDI remain orally administered metronidazole or vancomycin, despite the dramatic increases in the severity and incidence during the past decade. About 20-25 % of patients will have a recurrence of symptoms when either therapy is discontinued [238], and the risk of a recurrent infection increases to about 40 % after the first recurrence and to more than 60 % after a later recurrence. In severely complicated cases, surgical removal of the colon is the only remaining life-saving measure. Recurrent infections can also be treated by regenerating the normal intestinal microbiota with a faecal transplant. Although this appears to be a highly effective treatment, it is logistically very complicated and rarely performed.

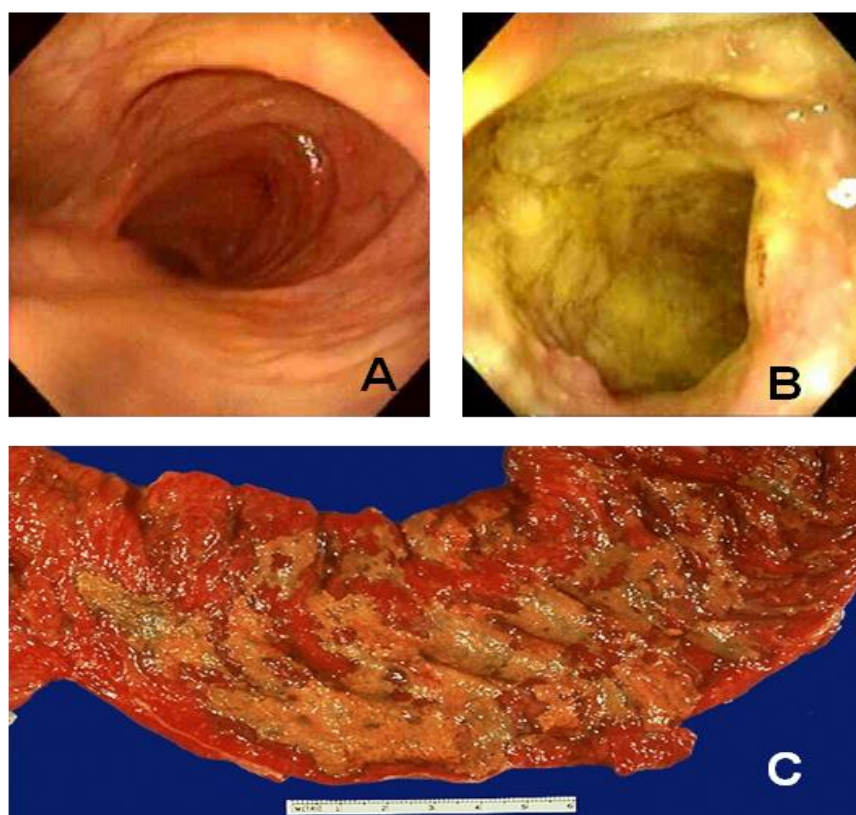


Figure 1. 12: A) Normal colon viewed with endoscope; B) Endoscopic view of pseudomembranous colitis, showing typical yellow pseudomembranes on the colonic mucosa; C) Cross section of a colon with pseudomembranous enterocolitis. Yellow-green plaques partially cover a hyperaemic mucosa, which itself is not eroded. Images are from The Internet Pathology Laboratory for Medical Education, Florida State University College of Medicine at <http://www.medlib.med.utah.edu/WebPath/ORGAN.html>

There is no doubt that *C. difficile* infections are increasing in frequency and severity. Prevalent endemic CDI and outbreaks have been reported in many hospitals in the UK and internationally. Reports on epidemic outbreaks of infection with hypervirulent *C. difficile* strains (ribotype 027) have shown that the rising rate of CDI has been largely assigned to the

presence of this strain. In US hospital and long-term care facilities, cases of CDI increase each year. The current estimate is ca. 500,000 cases a year, increased from <150000 in 2000. From these 500,000 CDI cases, 15,000 to 20,000 patients are estimated to die from CDI each year. The extra costs associated with infection are around £4000 per patient, by reason of treatment and tests required, and extra time spent in hospital [239], amounting to an estimated \$1.1 billion per year in US alone [240]. Increasing infection levels also found in Europe. For example, incidence of CDI in Saxony, Germany increased up to threefold between 2002 and 2006. In England and Wales, the number of deaths involving *C. difficile* rose by 28 % between 2006 and 2007. However, in the last few years, a marked decrease of CDI cases in England was observed due to recent national mandatory reform in hospitals including environmental decontamination, reducing the cross-infection opportunities by hand hygiene and barrier precautions.

The main group of people who are at high risk of CDI are elderly hospitalised patients. However *C. difficile* is not always a pathogen. It is one of dozens of *Clostridial* species that can be found in a low percentage in the healthy human colon and widespread in the environment [241-242]. Asymptomatic carriage is even more common in children [243], perhaps reflecting the different conditions in the immature intestine which might lead to less advantageous environmental conditions for bacterial pathogenicity. Molecular microbiologists have not paid great attention to *C. difficile* in the past, partly because of difficulties in handling anaerobes and more specifically the lack of genetic tools. However, this is changing with more laboratories intensively scrutinising *C. difficile*, aiming to understand in more detail the properties of this organism and its pathogenesis. For example, directed mutagenesis in *C. difficile* is more difficult than in other well studied species such as *E. coli*. There is no reliable technique for conditional gene deletion, however constitutive gene knockouts can be achieved using the CloStron system, utilising a bacterial retargeted mobile group II intron [244]. This allows creation of stable mutants in *Clostridium* species in a few steps including intron design, plasmid cloning, plasmid transfer to the *C. difficile* and integrant isolation. Notably, the complete genome sequence of *C. difficile* strain 630, a virulent and multidrug-resistant strain, was published in 2006 [245], enabling the use of standard bioinformatic and genetic analyses. Further sequencing projects are ongoing

with other virulent strains which will hopefully improve our understanding of this challenging organism.

1.6.1 *C. difficile* virulence factors

The main virulence factors of *C. difficile* are toxins A and B. These toxins catalyse the glucosylation and inactivation of Rho-GTPases, which are regulatory proteins with many functions in the eukaryotic cell. Both toxins are virulence factors, as permanent gene knock-out of either one or both toxins showed significantly reduced virulence activity, at least in the hamster model tested [246]. Biochemical, genetic and structural studies of these toxins have improved our understanding of their important role in pathogenesis. As this lies outside the scope of this thesis, details of the toxins are not discussed in this literature review but further reading can be found in the following publications: [246-249].

Other virulence factors have also been described, such as adhesins and hydrolytic enzymes [233, 250-251]. Among these, the surface layer proteins (SLPs) are additional potential virulence factors. They are proposed to aid the colonisation of the bacteria induce both inflammatory and antibody responses in the host. These proteins may adhere directly to the host cell or compete with the host flora.

1.6.2 The S-layer of *Clostridium difficile*

The S-layer of *C. difficile* has been studied by a number of groups. As in several other bacteria, the surface layer proteins are the predominant surface proteins in *C. difficile*. However, in contrast to most other species, where the S-layer is composed of one protein, the S-layer of *C. difficile* constitute of two distinct proteins: a low molecular weight (LMW) SLP and high molecular weight (HMW) SLP [252-254]. Interestingly, the sizes of these proteins vary between strains. The LMW SLP is 32–38 kDa and the HMW-SLP 42–48 kDa in size. Both the high and low molecular weight subunits are derived from a single gene, *slpA*, confirmed by sequencing of S-layer proteins and comparison with the *C. difficile* genome [255-256]. This gene is strongly transcribed during the entire growth phase [257] and it is estimated that up to 400 molecules of SlpA per cell per second are produced, translocated, and cleaved during the exponential growth phase. The translated gene product SlpA undergoes two rounds of post-translational cleavage, firstly to remove the signal sequence that targets the nascent chain for translocation across the cytoplasmic membrane, and then

internally (at position 355 in SlpA from strain 630) to release the two mature SLPs, LMW-SLP and HMW-SLP [258-260] (**Figure 1. 13**). It is presumed that removal of the signal peptide occurs prior to or during translocation across the plasma membrane, while the second cleavage takes place either within the plasma membrane or external to the membrane, i.e. within the cell wall. After cleavage, the mature proteins form a complex and reassemble to build the S-layer with a square symmetric outer layer and a hexagonal symmetric inner layer [252].

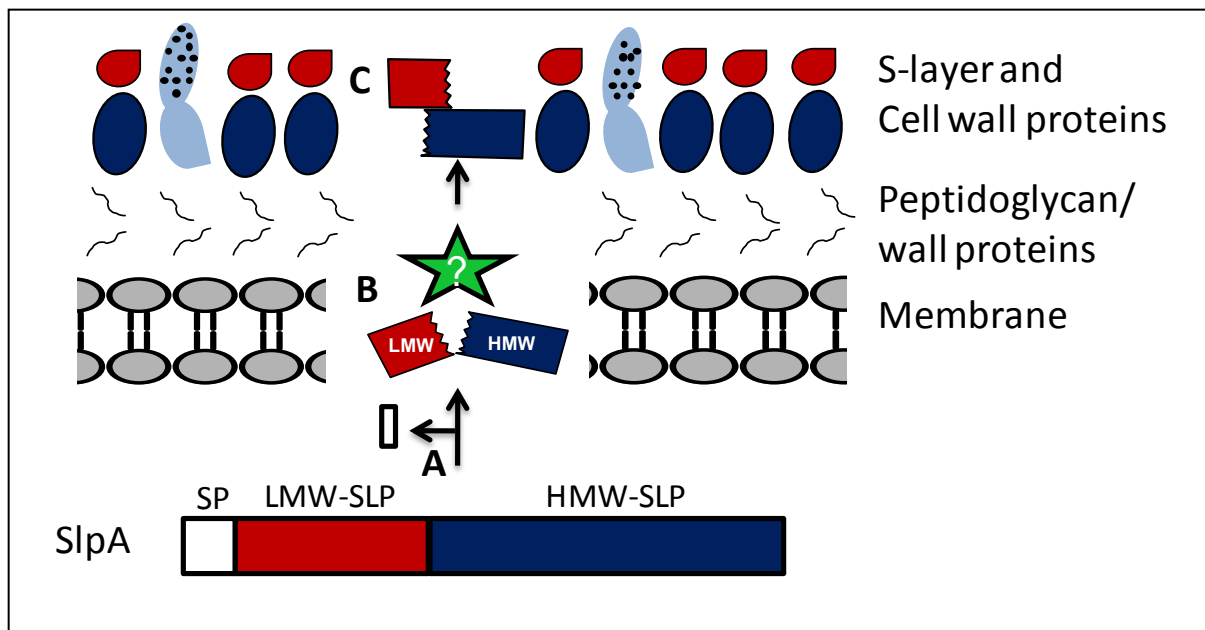







Figure 1. 13: Processing of the *C. difficile* S-layer. The SlpA precursor  is synthesised in the cytoplasm as a single polypeptide and, after targeting to the membrane (A), is translocated to the cell wall where proteolytic cleavage occurs (B) forming the two components of the S-layer, HMW SLP  and LMW SLP . Finally, the S-layer is assembled on the surface of the bacterial cell wall (C).  indicates an unidentified cysteine protease.  designates other cell wall proteins with cell wall binding domains (unspotted) and surface exposed domains (spotted).

Recently, the Fairweather laboratory reported the structure of the resultant LMW/HMW-SLP complex [261]. Using a combination of genetic and structural studies (X-ray crystallography of the truncated LMW-SLP and SAXS of full length LMW-SLP and LMW/HMW-SLP complex), several important points were identified. Firstly, the interaction between the two matured proteins, LMW- and HMW-SLPs, is non-covalent but very tight with an estimated affinity of 10 nM as the dimer. Gel electrophoresis under denaturing conditions (i.e. using SDS) breaks the interaction and revealed two distinct proteins on the gel after staining, whilst only one band (the complex) was found under native conditions.

Secondly, the structure of the LMW-SLP can be divided in three domains: domain 1, domain 2 and an interaction domain (**Figure 1. 14**). The interaction domain of the LMW SLP (C-terminal 62 residues; 260-321) was identified using ELISA (enzyme-linked immunosorbent assay) probing a series of deletion derivatives of the LMW SLP with the full length HMW SLP. A complementary series of deletions of the HMW-SLP identified the interaction domain of this protein. The LMW-SLP was further investigated by study of SAXS and X-ray structure. Domain 1 is small and contains both N-terminal (residues 1-88) and C-terminal (residues 239-249). Domain 2 is much larger and is the part that is exposed to the external environment. Thirdly, similar to LMW-SLP, the HMW-SLP also possesses an interaction domain, which is essential for the complex formation. However, this domain is located within the first 40 N-terminal residues. Therefore, the LMW- and HMW-SLPs bind to each other via their C- and N-terminal respectively (see **Figure 1. 14**).

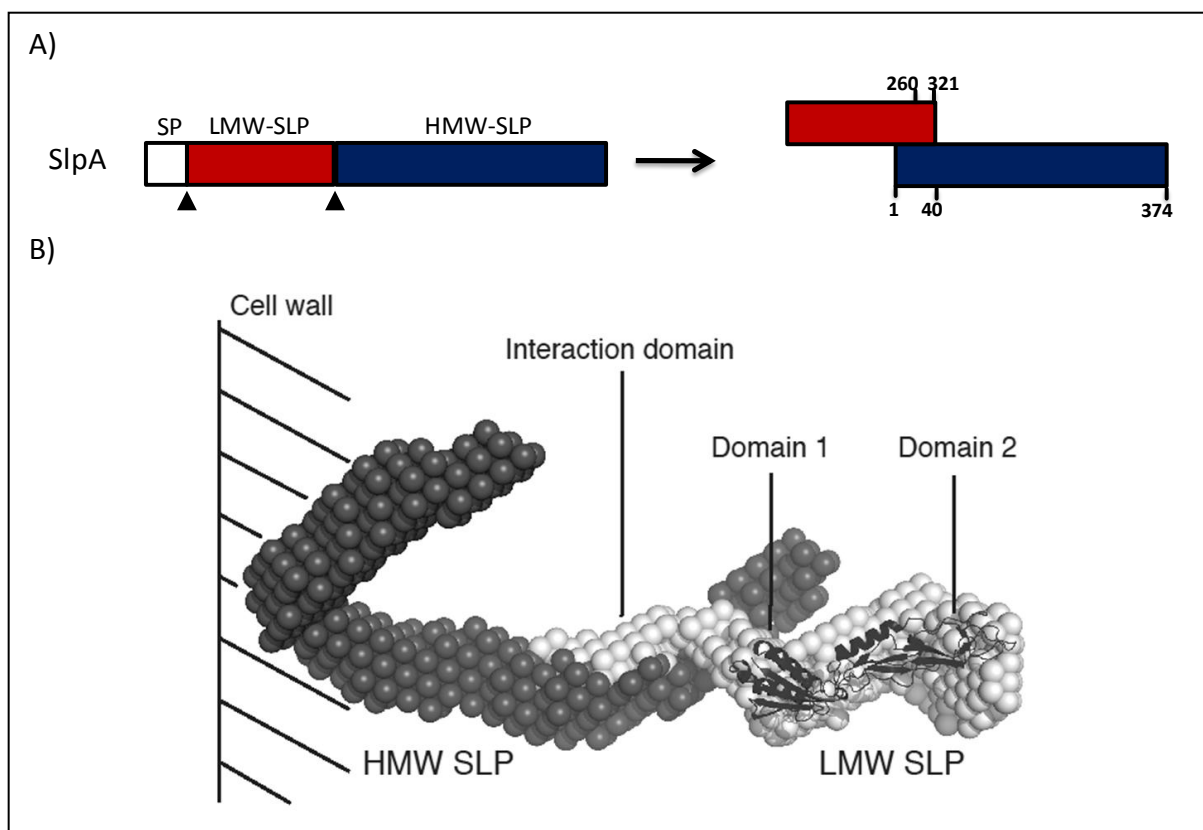


Figure 1. 14: A) Model of the interaction between the HMW SLP and LMW SLP. B) Proposed structure of the LMW/HMW-SLPs complex on the cell wall. A low resolution SAXS structure of the complex is shown overlaid with the crystal structure of the truncated LMW SLP. Three different domains of LMW-SLP are indicated. Figure is adapted from Fagan *et al.* [261].

Sequence analysis of *slpA* genes from several strains shows that the LMW–SLPs have a remarkable degree of variability, which could play a role in the antigenic variation of the bacteria as the LMW–SLP is immuno-dominant [252, 255, 262-263]. Despite the low level of sequence homology (largely in the sequence of domain 2) seen in the LMW–SLP across different strains, a highly conserved GKR motif is seen located near to the predicted cleavage site that generates the two mature SLPs [264-265]. This motif may be essential for the recognition by the protease which mediates this cleavage or it may perhaps be involved in formation of the two-dimensional lattice. The identity of the protease that performs this internal cleavage was unknown at the outset of the present study, and studies directed towards its identification are the topic of this thesis.

Strains	SlpA cleavage sequence																
	N-terminal						Cleavage position					C-terminal					
	P ₈	P ₇	P ₆	P ₅	P ₄	P ₃	P ₂	P ₁	P' ₁	P' ₂	P' ₃	P' ₄	P' ₅	P' ₆	P' ₇	...
R8366	G	K	R	V	T	T	K	S	A	A	K	A	S	I	A
R7404	G	K	R	L	E	T	K	S	A	D	I	I	A	D	A
R13702	G	K	R	L	E	T	K	S	A	D	I	I	A	D	A
Cd630	G	K	R	L	E	T	K	S	A	N	D	T	I	A	S
R12882	G	K	R	L	E	T	K	S	A	N	D	T	I	A	S
R13550	G	K	R	L	E	T	K	S	A	N	D	T	I	A	S
Cd 959	G	K	R	L	T	T	K	G	A	T	G	T	L	A	D
SE528	G	K	R	L	D	T	K	G	A	T	D	I	E	N	T
R13540	G	K	R	L	T	T	K	S	A	T	Q	A	S	K	T
R13549	L	N	T	A	S	T	Y	A	S	S	N	Y	P	E	L
R13699	G	K	R	L	N	T	L	S	A	S	S	A	K	T	I
R12884	L	Q	G	F	S	T	Y	R	A	T	N	Y	N	E	G
Cd 167	L	Q	G	F	S	T	Y	R	A	T	N	Y	N	E	G
Cd Y	R	L	Q	F	S	S	Y	G	K	F	S	T	D	E	A

Table 1. 2: Alignment of SlpA sequences at the LMW-HMW cleavage locus.

The HMW–SLPs are considerably more conserved in sequence, and all display significant homology to the cell wall-anchoring domain of autolysin CwlB of *Bacillus subtilis* [252, 264]. Using antisera to the HMW– and LMW–SLPs, it was shown that the anti-LMW–SLP antibodies do not always cross-react with the LMW–SLPs of other strains, whereas the HMW–SLPs are more antigenically conserved [252, 254-255, 266]. Antibodies to the LMW–

SLP are found in the sera of CDI patients whereas the HMW–SLP seems to be poorly immunogenic in *C. difficile* infections [262, 267-268].

Despite the increasing number of *C. difficile* strains isolated from clinical cases, only a few *slpA* genes have been sequenced and characterised [264-265, 269]. Alignment of *slpA* sequences, and in particular of the N-termini of a small number of HMW–SLPs as determined by Edman degradation and/or mass spectrometry, reveals sequence homology at the cleavage locus. Sequence analysis of approximately 20 strains (from different sources) reveals the putative cleavage site between the LMW and HMW subunits. Based on this homology, the SlpA cleavage sites of other *C. difficile* strains were predicted (**Table 1. 2**).

Methods to isolate the S-layer

Surface proteins can be isolated from the cell by various methods including treatment with LiCl, EDTA, urea or low pH glycine [270-274]. All of these treatments act on the cell wall but with different mode of actions. Therefore, they have different impact on the type and number of the released proteins. In a study undertaken by Anne Wright, a past member of Fairweather group, a comparison of these extraction methods was performed and revealed that LiCl and EDTA treatments are rather an inefficient approach to remove all surface layer proteins compared to urea and low pH glycine, judging by the released amount of LMW- and HMW-SLPs and other proteins [275]. Urea treatment released a significant number of other proteins besides the LMW- and HMW-SLPs, some of which may be due to cell lysis. Low pH glycine treatment resulted in equimolar amounts of LMW- and HMW-SLPs and fewer background proteins.

1.6.2.1 Other cell wall proteins

Whilst the SLPs are the most abundant in the surface layer, many other cell wall proteins have been postulated by analysis of the *C. difficile* 630 genome. These proteins belong to a family of genes related to *slpA* (about 29 genes) and share one common feature: all of these genes contain two or three Pfam 04122 motifs (known as putative cell wall binding motifs) and a signal sequence (**Figure 1. 15**) [255, 263, 276-277]. In addition to these common features, some of these proteins bear another domain which may confer a known or putative function. Among them, a protein of ~70 kDa was consistently co-purified in all strains [266, 272] and is now recognised to be Cwp2 [264]. Another cell wall protein, Cwp66,

has a two-domain structure: the C-terminal domain is exposed to the cell surface, and has been described as an adhesin, while the N-terminal domain has homology to the cell wall-anchoring domain and is thought to anchor Cwp66 to the cell wall of *C. difficile* [278]. Another putative adhesin is the cell wall protein CwpV, which is phase-variable [279]. About 5 to 7 % of *C. difficile* cells express CwpV. Finally, Cwp84 is identified as a cysteine protease [280], and is discussed in details in later section.

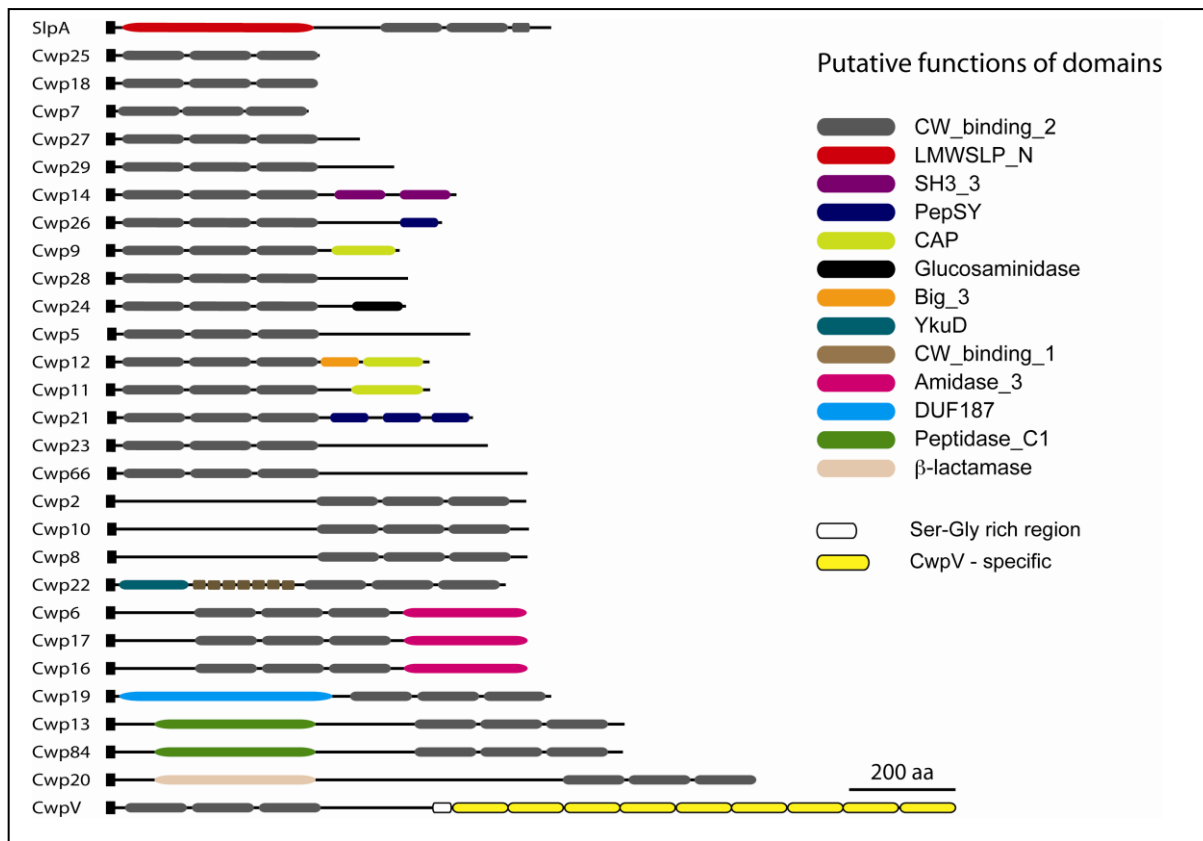


Figure 1. 15: Cell wall proteins (CWP) in *C. difficile* 630 [255]. Cysteine protease homolog domains in Cwp84 and Cwp13 are shown in green, and other conserved domains are coloured according to the key shown above (alignment performed and figure generated by Dr Jenny Emerson). SH3_3, Src homology (SH) regions; PepSY, a regulator of peptidase activity; CAP, pathogenesis-related 1 proteins, Big_3, brefeldin A-inhibited guanine nucleotide-exchange protein 3; YkuD, spores protein; DUF187, domain of unknown function 187.

1.6.3 Proteases in *C. difficile*

Despite the importance of *C. difficile* as a pathogen, apart from its two toxins and S-layer proteins, the organism is very poorly characterised. Currently, construction of mutants in *C. difficile* is very challenging, and alternative biochemical approaches towards characterisation of enzymes are employed. Only a few protease activities of *C. difficile* have been reported to date. Proteolytic activity has been observed in ten strains of *C. difficile*,

with degradation of gelatine, collagen and azocoll observed in all of those isolates [281]. Similar results were reported by Steffen and Hentges concerning production of gelatinase, collagenase, hyalurodinase and chondroitin sulfatase activities in *C. difficile* [282]. However the proteases responsible are not characterised *in vitro*, and their identities remain unknown.

The MEROPS database lists 139 known and putative peptidases in *C. difficile* with both known or unknown functions, and in addition 63 non-peptidase homologues. This database also classifies the peptidases in predicted protease classes where possible. **Table 1. 3** compiles the number of these peptidases in the corresponding classification.

Protease class	Number of peptidase found
Aspartic protease	5
Cysteine protease	17
Metalloprotease	112
Serine/Cysteine protease	17
Serine protease	41
Threonine protease	1
Unknown	7

Table 1. 3: Peptidases and non-peptidase homologues classified in the MEROPS database.

As noted previously, *C. difficile* synthesises two major toxins, toxin A and toxin B, both of which are responsible for the clinical manifestations of the disease. These toxins are large multi-domain proteins, whose actions depend on a complex uptake mechanism, including auto-proteolytic processing to release an N-terminal glycosyltransferase domain [283]. Both cysteine [283] and aspartate [284] protease activity have been proposed as responsible for the cleavage process of the toxin.

1.6.3.1 Putative cysteine proteases in the S-layer: Cwp84 vs. Cwp13

One interesting finding from studies of cell wall proteins of *C. difficile* was the presence of two proteins, Cwp13 and Cwp84, that have high homology to cysteine proteases (**Figure 1. 15**) [285]. Cwp13 remains uncharacterised, whereas recombinant (*E. coli* produced) Cwp84

was experimentally characterised *in vitro* as a cysteine protease with degrading activity on extracellular matrix proteins [280].

Comparison of both proteases reveals that they have relatively similar size of about 87 kDa. Both proteins possess a cysteine protease domain and three repeated cell wall binding domains Pfam 04122 [286] (**Figure 1. 15**). The genes have 63.2% nucleotide sequence identity. However, their locations in the genome are situated approximately 1.2Mb apart from each other. Cwp84 is found in the same genomic region as slpA and a number of other cell wall proteins including cwp2, cwp66 and cwp6, while cwp13 is not included in this gene cluster (**Figure 1. 16 A and B**). Sequence alignment of these proteases with well-known papain-like cysteine proteases such as human procathepsin L (1CS8) [287], rat cathepsin C (1JQP) [288], rat procathepsin B (1MIR) [289], procaricain (1PCI) (in the latex of *Carica papaya*) [290], cysteine protease Der p 1 (from a house dust mite allergen) (1XKG) [291], the *Plasmodium falciparum* cysteine protease Falcipain-2 (1YVB) [292], human procathepsin S (2COY) [293], procathepsin L1 from *Fasciola hepatica* (2O6X) [294], human procathepsin B (3PBH) [295] and human procathepsin K (7PCK) [296] revealed that the catalytic triads of both proteases are located in the same region of the polypeptide chain, even though the amino acid numbering is not identical for the two proteases (**Figure 1. 16B**). The Cwp84 catalytic triad is predicted to consist of C116-H263-S295, whereas the one for Cwp13 is predicted to be C109-H256-N288.

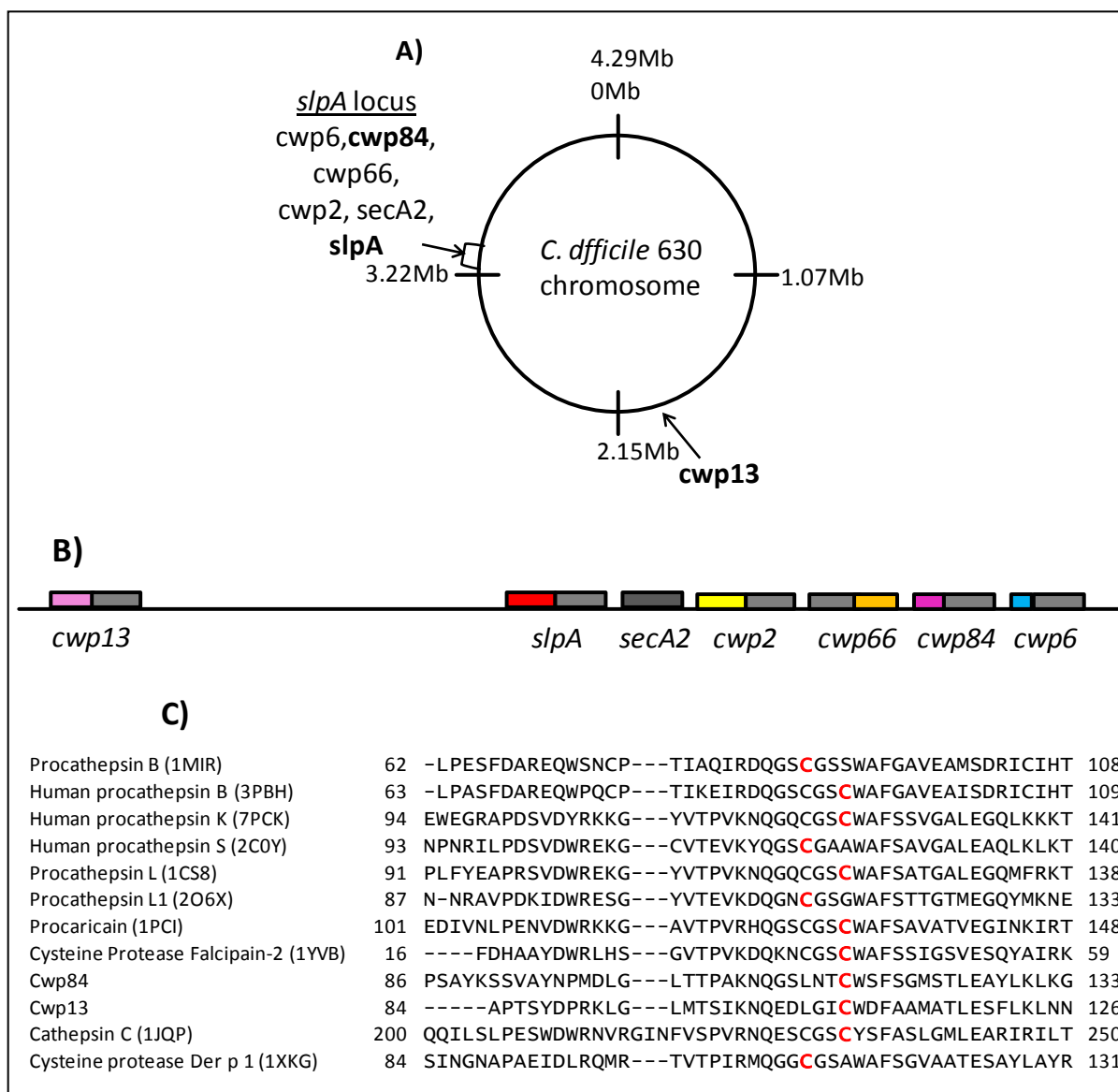


Figure 1. 16: A) Genomic region of Cwp13 and Cwp84. B) *SlpA* locus. C) Sequence alignment of Cwp84 and Cwp13 with well-known papain-like proteases. The cysteine residue in the catalytic triad is indicated in red. ClustalW was used to align the amino acid sequences which was obtained from RCSB protein data bank website.

1.7 Project aims and outline

Recognising the need for novel therapies for the drug-resistant *C. difficile*, this project aims to identify and characterise potential therapeutic targets in this bacterium using ABPP. As the synthesis and processing of the S-layer are key steps in the biogenesis of the cell wall, this thesis describes an in-depth chemical proteomic investigation of SlpA processing as a target for therapeutic intervention.

1.7.1 Structure of this thesis

Chapter 1: The first chapter introduces ABPP and relevant techniques that can contribute towards a chemical and microbiological toolbox for investigation the S-layer of *C. difficile*.

Chapter 2: This chapter describes the screening of commercial inhibitors leading to the discovery of the first inhibitors of S-layer processing. E-64 is discussed as a starting point for inhibitor design.

Chapter 3: Based on E-64, a series of inhibitors is designed, synthesised and tested *in vivo*.

Chapter 4: In this section active inhibitors are transformed into ABPs. Subsequently, the condition of bio-orthogonal ligation using these probes is developed and optimised.

Chapter 5: Direct label probes are developed and employed for target identification.

Chapter 6: The labelled target enzymes are enriched and identified using mass spectrometry.

Chapter 7: This chapter describes the exploration of further cysteine protease inhibitors in *C. difficile*.

Chapter 8: The main conclusions drawn from the work are presented and potential directions for future investigation are summarised in this section.

Chapter 9: This chapter records the experimental methods used throughout the thesis.

Chapter 10 and Chapter 11: These chapters contain the references and appendix figures.

The results generated by this study have two potential outcomes. Firstly, they may provide a new route to develop therapeutic targets against *C. difficile*. Secondly, the advances in chemical technology employed here may enable comparable insights into related species (*e.g. C. tetani*, the bacterium that causes tetanus) and related pathogenic processes (*e.g.* release and processing of bacterial toxins).

Chapter 2 Discovery of E-64 and leupeptin as the first inhibitors of S-layer processing in *C. difficile*

C. difficile secretes a family of proteins that are held in place on the cell wall by non-covalent forces, forming the S-layer, which surrounds the entire cell. These S-layer proteins (SLPs) are immunogenic in humans and play a role in binding to host cells [258]. Synthesis of the *C. difficile* S-layer involves site-specific proteolytic cleavage of the precursor SlpA into two components by one or more *C. difficile* proteases, that were unknown at the outset of the work reported here [255]. We hypothesised that processing of SlpA would be inhibited by addition of a protease inhibitor to a growing culture of *C. difficile*. The focus of this chapter is the series of experiments that led to the discovery of the first inhibitors of SlpA processing, and preliminary classification of the protease(s) involved in the cleavage of SlpA. As part of this work, an assay was developed to monitor the inhibition of SlpA cleavage, and this was used as the basis for inhibitor optimisation (Chapter 3) and ABP development (Chapter 4).

2.1 Screening of commercial inhibitors

2.1.1 Inhibition assay

As described in the previous chapter, the initial aim of the current project was to investigate S-layer biogenesis, and in particular the SlpA maturation events that lead to the LMW and HMW SLPs. In order to assign the protease(s) involved in the cleavage of SlpA to one of the protease families, a general method was developed to monitor directly the cleavage products. This measured the effect of the inhibitors on SlpA processing *in vivo* and provided the essential inhibition assay for this study. Ideally, an untreated culture will show only SlpA cleavage products, namely HMW-SLP and LMW-SLP, whereas a culture treated with inhibitor should reveal the presence of full-length SlpA.

Working with *C. difficile* presents certain challenges such as strictly anaerobic conditions for culturing the bacteria, the requirement for blood agar plates and inherent variability in the growth of cultures. These challenges were overcome as much as possible during the experiments undertaken. *C. difficile* was grown on a blood agar plate under anaerobic conditions for at least three days before use. However, this plate required renewal after ten days to ensure subsequent culture viability. Furthermore, all equipment such as tips, pipettes, microcentrifuge tubes and brain heart infusion (BHI) broth used for bacterial culture required incubation in an anaerobic cabinet prior use to allow any oxygen to be replaced with an anaerobic gas mixture (80% N₂, 10% CO₂, 10% H₂).

For the inhibition assay, a single colony of *C. difficile* was inoculated into BHI broth and grown anaerobically for 1 hour with vigorous shaking to obtain a homogeneous suspension. This culture was then aliquoted into microcentrifuge tubes containing specific inhibitors (usually at 100 µM, stock solution was in H₂O or DMSO). This ensured an equal starting point for all samples and limited culture variation at the early stage. The cultures were allowed to grow overnight (16–18 h) with agitation. Cells were harvested by centrifugation, fractionated when appropriate and analysed by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) and Western blotting against LMW-SLP antibody.

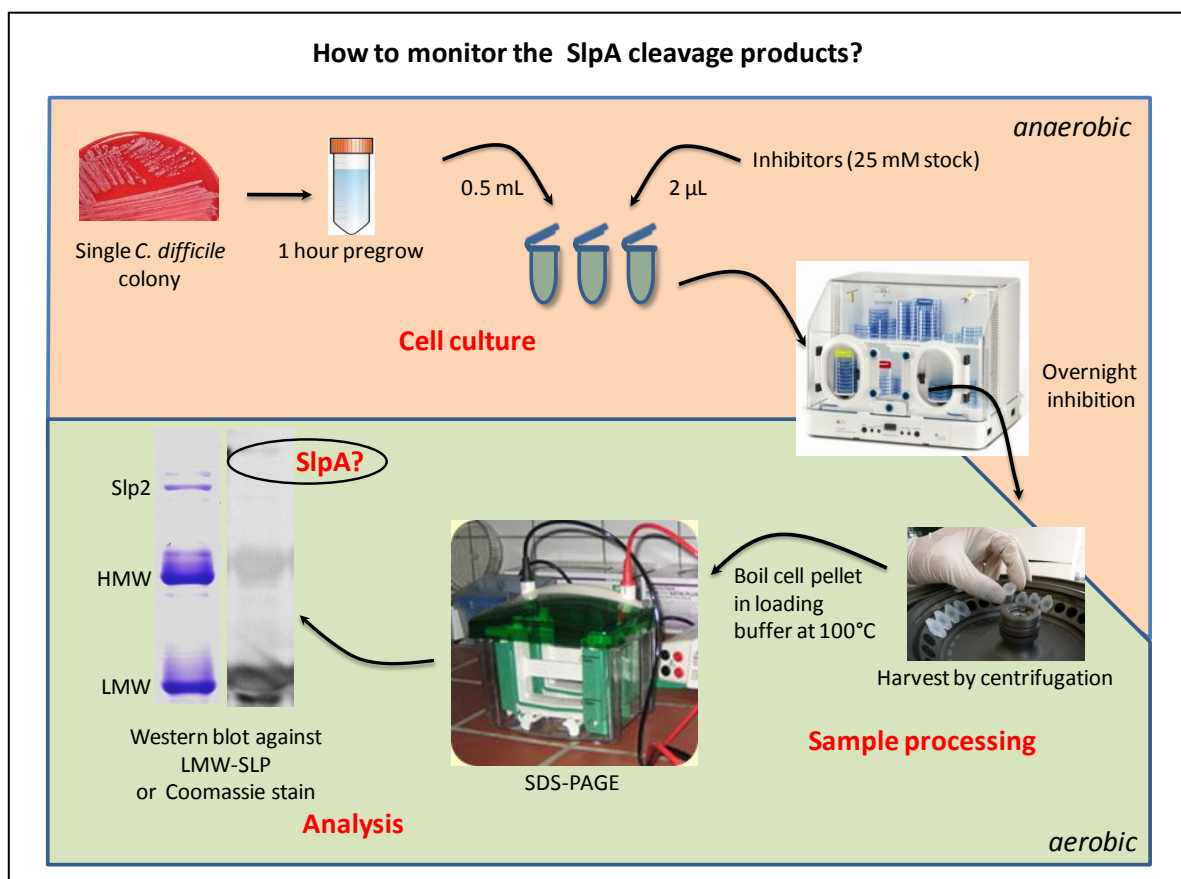


Figure 2. 1: The flow diagram of a general *in vivo* inhibition assay.

2.1.2 Protease inhibitors

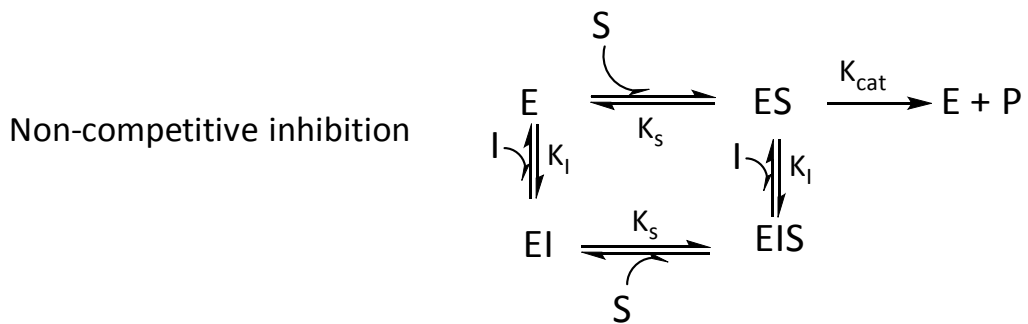
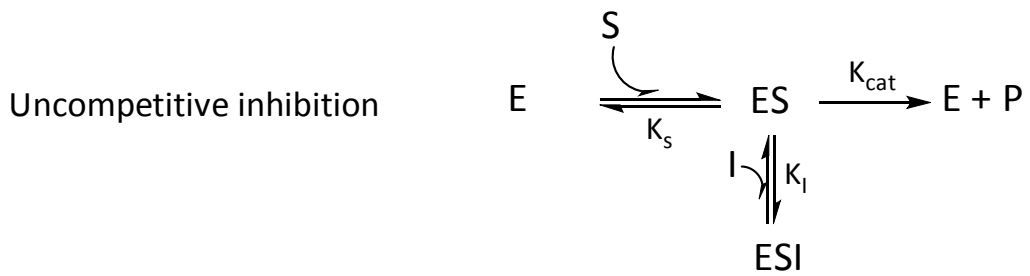
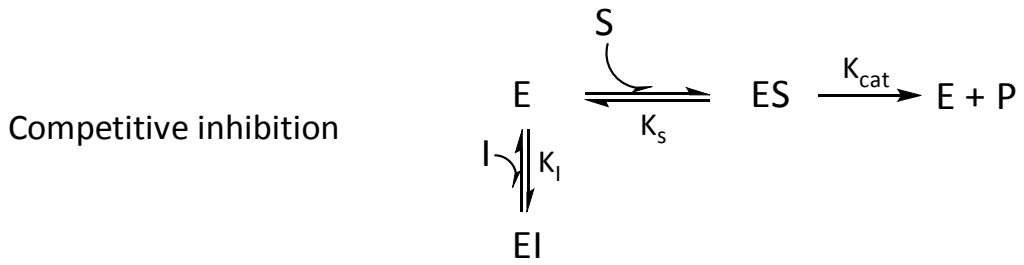
Protease inhibitors are molecules that bind to a protease to reduce its activity. They are ubiquitous and are found in multiple forms in all living organisms [297]. Protease inhibitors can be small non-proteinaceous molecules (typically found in microorganisms), peptides or proteins. Their gross function is to regulate undesirable protease activity, but protease inhibitors can be used as a powerful tool in basic as well as applied research, for example in the characterisation of binding sites, catalytic functional groups and transition-state geometries of proteases [298]; investigation of the physiological function of particular class of proteolytic enzymes within an organism [299]; and in drug discovery [300-301].

Because protease inhibitors can be used to target proteolytic enzymes crucial for disease propagation, development of such protease inhibitors into drugs has generated considerable research interest. Generally, protease inhibitors are developed from natural product screening or combinatorial library screening with subsequent lead optimisation [302] or by using substrate/transition state mimetic methods, including the truncation of a

peptidic substrate or substituting the cleavable amide bond by a non-cleavable isostere [303]. Substrate-based development can be taken further using receptor-based design, which requires the availability of three-dimensional structural information for the protease. With the aid of *in silico* modelling, protease inhibitors can be designed to fit into the active site of the receptor to gain higher selectivity [304].

Protease inhibitors are usually presented in two contrasting classes: reversible and irreversible. These two forms differ in their mode of inhibition (**Figure 2. 2**). Reversible protease inhibitors bind non-covalently to the protease to form an enzyme-inhibitor complex (EI), which is in equilibrium with the individual species; for example, these non-covalent interactions can include hydrogen bonds, van der Waals interactions and ionic bonds. These weak interactions combine to generate strong and specific binding between enzyme and inhibitor. There are three kinds of reversible inhibitors differentiated by their inhibition mode. *Competitive inhibitors* cannot bind to the enzyme (E) at the same time as the substrate (S) due to competition for access to the same binding site of the enzyme. The inhibition can be overcome by addition of excess substrate to disturb the equilibrium. *Uncompetitive inhibitors* are only able to bind to the enzyme-substrate complex (ES). Finally, *non-competitive inhibitors* are able to bind to the enzyme as well as the enzyme-substrate complex and does not affect binding of the substrate. Reversible inhibition can be overcome by dilution or dialysis, as it depends on the concentration of the inhibitor. In contrast, irreversible protease inhibitors also form an initial enzyme-inhibitor complex or enzyme-substrate-inhibitor complex, but subsequently react with the enzyme to form a covalent bond (EI*) and inactivate the enzyme permanently. Irreversible inhibitors often contain an electrophilic functional group, which reacts with a nucleophile in an amino acid side chain of the protease to form a covalent adduct. Furthermore, irreversible inhibitors act in a time-dependent manner, as the amount of active enzyme is reduced with the inhibitor-enzyme incubation time.

A) REVERSIBLE



B) IRREVERSIBLE

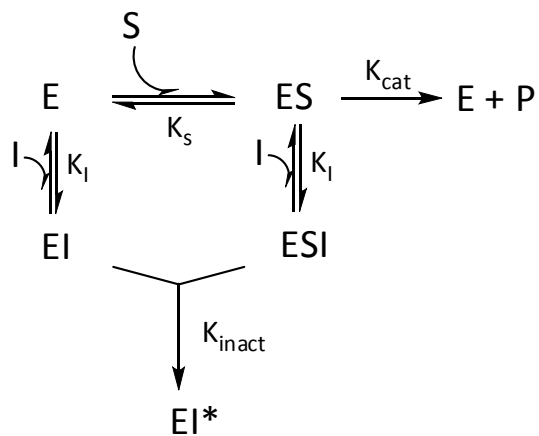


Figure 2. 2: Kinetic schema for reversible and irreversible inhibitors.

With regards to the protease(s) involved in SlpA cleavage in *C. difficile*, at the outset of this project little was known and the identification of these enzymes remained unclear. However, there are potential candidates such as Cwp84 and Cwp13, one of which is encoded in the same gene cluster as SlpA, that possess a cysteine protease domain (Section 1.6.3.1). Clues as to the correct classification of these proteases may be derived from their inhibitory activity in the presence of broad-spectrum protease inhibitors. For this purpose, a panel of commercially available protease inhibitors was selected. These inhibitors cover most of the protease families and clans. **Figure 2. 3** illustrates the structures of the chosen protease inhibitors and the details about each inhibitor are summarised in **Table 2. 1**.

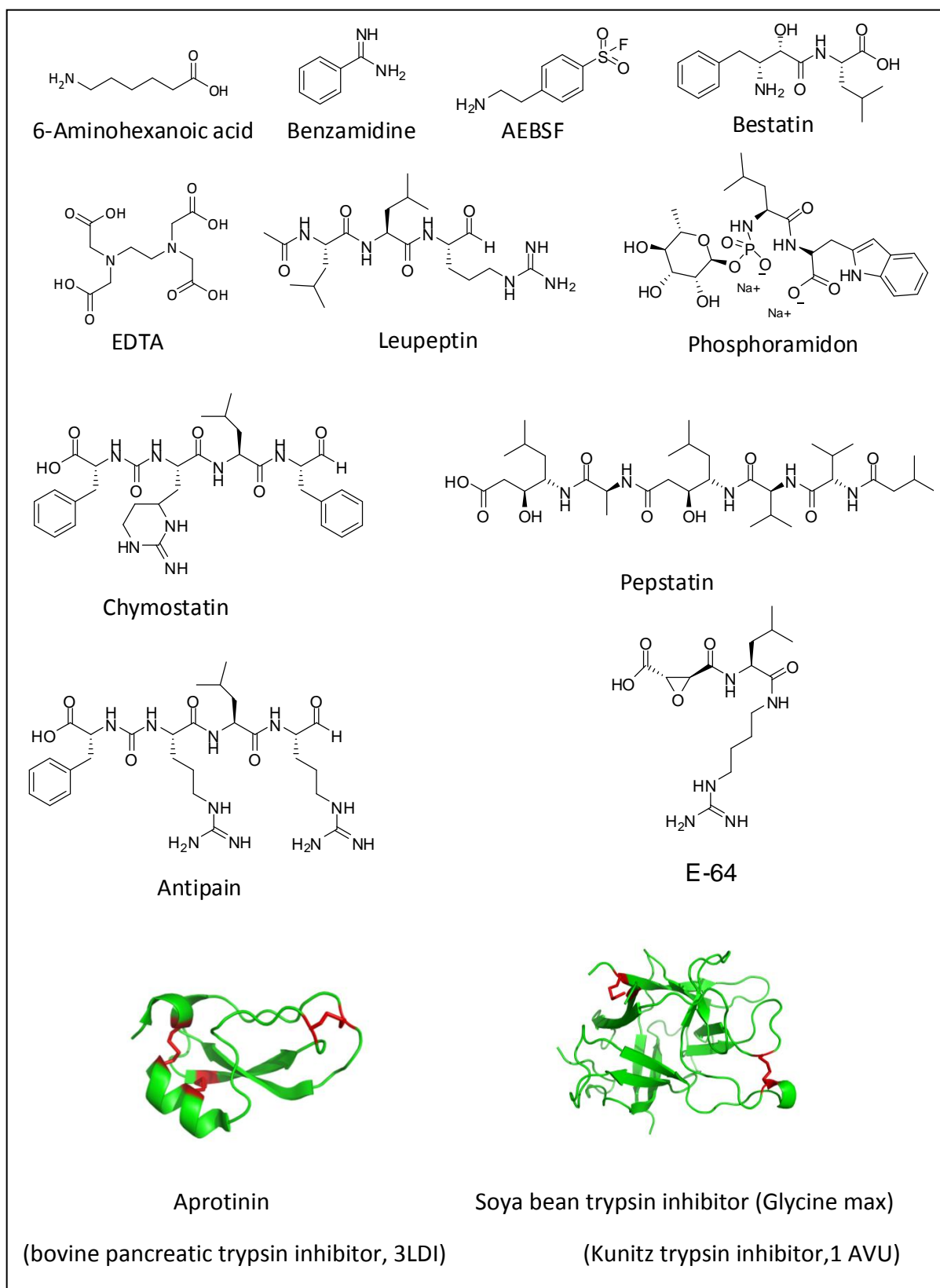


Figure 2. 3: Structure of chosen commercially available protease inhibitors (Sigma Aldrich library). Crystal structure of trypsin inhibitors are shown in green and disulfide bonds are highlighted in red.

Inhibitor	Type of target proteases	Inhibition mode	Origin	Application
ϵ -Aminocaproic acid (EACA)	Serine [17]	Reversible [305]	Synthetic	Marketed as Amicar [306] for treatment of bleeding disorders (fibrogenemia postoperative bleeding).
Benzamidine	Serine [307-309]	Reversible [307-309]	Synthetic	As effective as aprotinin in the prevention of glucagon degradation in human plasma (see below) [310]; additive in protein crystallography to prevent degradation of a protein of interest.
4-(2-Aminoethyl) benzenesulfonyl fluoride hydrochloride (AEBSF)	Serine [311-313]	Irreversible [311-313]	Synthetic	Preferred over other serine protease inhibitors (PMSF, DIFP) due to low toxicity and improved water solubility; used in preparation of tissue lysates in molecular biology [314-316].
Chymostatin	Serine [317]	Reversible [307-309]	<i>Streptomyces hygroscopicus</i> and <i>S. lavendulae</i>	Inhibits human leukocyte elastase [318], the cytosolic Ca^{2+} -dependent proteinase calpain [319], and other cytosolic multicatalytic proteinases [320] involved in antigen presentation [321] and in the degradation of ubiquitin-protein conjugates [322]

Aprotinin	Serine [323]	reversible[323]	bovine lung and bovine pancreas [324]	Marketed as Trasylol (Bayer) for treatment of various diseases including reducing bleeding during complex surgery. Withdrawn worldwide due to increased risk of complications or death (Bayer press release 2007).
Kunitz soya bean trypsin inhibitors	Serine [325-326]	Reversible [325-326]	soybean [327-328]	Inhibit plasma kallikreins and coagulation Factor Xa by a mechanism similar to trypsin inhibition.
Leupeptin	Serine and cysteine [329]	Reversible [329]	<i>Actinomycetes</i> [330]	Inhibits plasmin, trypsin [331], papain [332-333], thrombokinase and kallikrein, produces a curative effect for inflammation in oedema [333]
Bestatin	Serine and cysteine [334]	Reversible [334]	<i>Streptomyces olivoreticuli</i> [335]	Marketed as Ubenimex and used in the treatment of acute myelocytic leukemia [336]
Antipain	Serine and cysteine [308]	Reversible [308]	Actinomycetes [337]	Used in studies demonstrating the role of proteases in the process of cell transformation [338], and identifying proteases.[339] [340]
E-64	Cysteine [341]	Irreversible [341]	<i>Aspergillus japonicus</i> [155]	Used in <i>in vivo</i> studies due to specific inhibition, good cell permeability in cells and tissues,

				stability and low toxicity [342]
Pepstatin	Aspartyl [329]	Reversible [343]	<i>Actinomyces</i> [344]	Inhibits γ -secretase [308] and retroviral protease [345]; used to exploit proteases from several sources [346-347]
Phosphoramidon	Metalloendoproteinases [348]	Reversible [349]	Actinomycetes [348]	Inhibits many bacterial metalloendoproteinases, thermolysin and elastase [348], collagenase [308, 348, 350]. Does not inhibit trypsin, papain, chymotrypsin or pepsin [348]
Ethylenediaminetetraacetic acid (EDTA)	Metalloendoproteinases [351]	Reversible [352]	synthetic	Used as an anticoagulant in hematology [353], and in chelation therapy for treatment of atherosclerotic cardiovascular diseases [354]

Table 2. 1: Summary of selected commercially available protease inhibitors.

2.1.3 First inhibition results

Screening a panel of commercially available protease inhibitors at 100 μ M in the *in vivo* assay described above against the reference strain 630 revealed some three striking results (**Figure 2. 4**). First and foremost, a new band at 74 kDa apparent molecular weight appeared in the samples treated with E-64 and leupeptin (lanes 6 and 8); since this band cross-reacts with the anti-LMW-SLP antibody this was assumed to be an unprocessed SlpA band. This result strongly suggests that E-64 and leupeptin have a modest inhibitory effect on SlpA processing, and therefore, E-64 and leupeptin would represent the first ever reported inhibitors of S-layer processing in *C. difficile*. Secondly, EDTA disturbs bacterial growth at the given concentration completely, which might be due to removal of Ca^{2+} ions responsible for holding the S-layer/cell wall together [261], consequently it was not included in the gel. Thirdly, it is interesting to note that trypsin inhibitor caused extensive cell lysis compared to other inhibitors; this is visible on the Coomassie stained gel, which shows the release of proteins not normally found in the S-layer (**Figure 2. 4**, compare lanes 1 and 13). However, trypsin inhibitor does not inhibit SlpA processing under the conditions tested due to the absence of a strong band for full-length SlpA in the Western blot against LMW-SLP antibody.

It should be noted that there is a weak SlpA band in most lanes, visible by western blotting. This could be SlpA that has yet to be cleaved, or could reflect that fact that the bacteria may grow with a slight excess of unprocessed SlpA. However, it may be supposed that comparison of the amount of SlpA with the untreated sample (**Figure 2. 4**, lane 1) is a reliable (though rough) guide to inhibition level. In addition Western blot may be not suitable for precise quantitative analysis in the absence of a control band since the intensity of a particular band on the blot varies from one membrane to the next, and also depends on the length of the exposure time during imaging. This problem is reflected in the case of leupeptin and E-64, as the leupeptin coomassie band is actually at least as strong as E-64 but appeared weaker on Western blot. As a direct total protein stain the Coomassie gel could provide more comparable data, however unlike a Western blot it does not provide confirmation of protein identity.

Leupeptin is known as a reversible, competitive inhibitor of serine and cysteine proteases [329] (Section 2.1.2). Its activity applies to a wide range of proteases from trypsin [331],

thrombokinas and kallikrein to papain [332-333] and cathepsins B [355], H and L [356-357]. Leupeptin was determined to be *N*-acetyl(*N*-propionyl)-L-leucyl-L-leucyl-DL-argininal [358] and possesses an aldehyde group at the C-terminal position that is considered to be involved in the mechanism of inhibition [332].

E-64, on the other hand, is an irreversible cysteine protease inhibitor. It also inhibits papain and cathepsins B, H and L [341], but it does not have any inhibitory effect toward serine proteases like other cysteine protease inhibitors such as leupeptin and antipain [359]. Specifically, the functional thiol group of non-protease enzymes, such as L-lactate dehydrogenase or creatine kinase are not affected by E-64 [155]. Based on these data it seems likely that the protease (or proteases) involved in SlpA processing is a cysteine protease. Indeed, it later emerged that further in-depth investigation into structure activity relationships and binding modes provide sufficient evidence for classification of the target protease into a subfamily of cysteine proteases (see Chapter 6).

Having discovered the first inhibitors of S-layer processing in *C. difficile*, we next considered which should be selected as an initial lead compound for further investigation, taking into account the ease of access to the initial scaffold for inhibitor design, and potential for inhibitor optimisation to fine-tune reactivity for the target protein. Usually, small compounds that modulate target proteins reversibly are favoured as chemical tools or potential drugs because these compounds possess in most cases an attractive toxicity profile due to their elimination from the body/cell by dilution, excretion and metabolism. Irreversible inhibitors, on other hand, are seen as less useful for developing as drugs as they would be expected to modify a range of tissue constituents possessing amino or thiol groups besides the target proteins due to their alkylating or acylating chemistries. The undesired modification could lead to potentially serious side effects (e.g. DNA alkylation, carcinogenesis, immune response to modified proteins). Furthermore, these inhibitors inactivate the enzyme permanently and stay in the cell/body until the protein is degraded. Despite these potential problems there are many instances where irreversible inhibitors were successfully developed as drugs, sometimes inadvertently, and reached the market [360-362] (see Potashman and Duggan for an extended review on successful irreversible inhibitors as drugs [363]). Furthermore, irreversible inhibitors can be very useful for probing the catalytic nature of the enzyme concerned and can assist in identification by forming a

stable covalent bond with the functional group of the active site of the enzyme, labelling it irreversibly. In addition, irreversible inhibitors can sometimes be designed directly from the specific substrate of a particular enzyme. Combining these considerations, irreversible inhibitors are well-suited as starting points for the development of inhibitor analogues and activity based probes aiming at target identification. For this purpose, E-64 was chosen over leupeptin as the lead candidate due to its generally specific inhibitory effect and irreversible character.

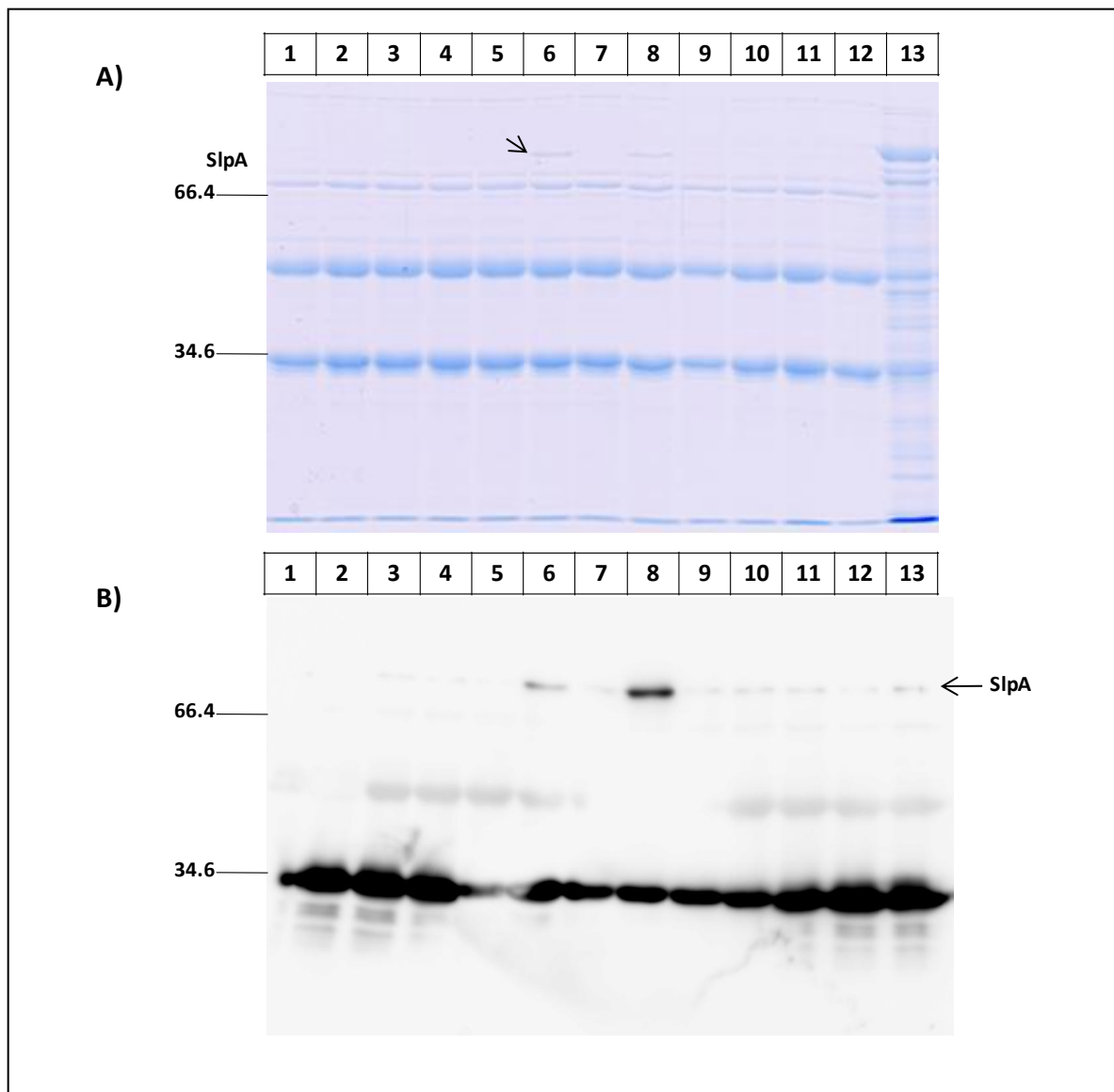


Figure 2. 4 : *In vivo* activity of inhibitors in *C. difficile* S-layer processing assay. A) Coomassie Blue stained gel. B) Western blot using anti-LMW SLP. Lanes of the gel show assays for the following inhibitors: Lane 1, no inhibitor; 2, 6-Aminohexanoic acid; 3, Benzamidine hydrochloride; 4, 4-(2-aminoethyl) benzenesulfonyl fluoride hydrochloride; 5, Chymostatin; 6, Leupeptin; 7, Bestatin; 8, E-64; 9, Pepstatin; 10, Phosphoramidon disodium salt; 11, Aprotinin; 12, Antipain; 13, Trypsin inhibitor from Glycine max (Soybean) type I-S.

2.2 E-64

2.2.1 Isolation, characterisation and structure determination

Inspired by the discovery of an aspartic protease inhibitor from *Penicillium cyclopium* [364] and the work of Umezawa *et al.* on a low molecular weight and highly reactive protease inhibitor from actinomycetes [337], Hanada and his co-workers isolated E-64 from a fungus, *Aspergillus japonicas* TPR-64 in 1977 [155]. E-64 was isolated as white needles with physicochemical properties as summarised in **Table 2. 2**. The structure of E-64 was determined by breaking the compound down in three parts using pronase E, each component analysed separately and confirmed by comparison with synthesised compounds. In addition, positive colour test and IR analysis provided further evidence for the guanidyl group, epoxide, 1,2-dicarboxyl motif and peptide bonds. The absolute structure was revealed to be *N*-[*N*-(*L*-3-*trans*-carboxyoxiran-2-carbonyl)-*L*-leucyl]-agmatine [365]. Hanada also found that E-64 reacts only with proteolytic enzymes possessing a functional thiol group such as papain, bromelains and ficin. In addition, his kinetic study suggested strongly that E-64 reacts irreversibly and almost stoichiometrically with the active thiol group [155].

Form	White needle
Mp	231–234 °C
Molecular formula	C ₁₅ N ₅ H ₂₇ O ₅ (MW: 357)
Optical rotation	[α] _D ²⁵ = +24.4 (<i>c</i> = 1, 0.1 N HCl)
Stability	Stable at pH range from 2 to 10
Solubility	Soluble: H ₂ O–MeOH, acetic acid, pyridine, DMSO. Insoluble: Acetone, EtOAc, CHCl ₃ , diethylether, benzene, petroleum ether.

Table 2. 2: Physicochemical properties of E-64.

2.2.2 Inhibitory activity and specificity

Continuing the characterisation of E-64, Hanada *et al.* investigated the structure activity relationship (SAR) of E-64 and its analogues with papain [366]. Their study revealed that the epoxydyl moiety is required for the activity against papain and the guanidyl moiety is not necessary for the inactivation of papain. Further modification of the leucyl moiety and esterification of the epoxydyl acid turned out to increase the papain inhibitory activity. This information was very useful; however knowledge of the enzyme-inhibitor interaction at the atomic level is more valuable for designing E-64 analogues with increased potency and

selectivity. For this purpose, Matsumoto *et al.* analysed the epoxysuccinyl inhibitory mechanism [367]. His study suggested a S_N2 -like reaction for the covalent bond formation between E-64 and the active thiol group. The nucleophile of the active thiol group (Cys-25) attacks from the backside of the E-64 epoxide on the C2 atom causing ring opening and changing the *S*-chirality of C2 atom to the (*R*)-form (**Figure 2. 5A**). This feature was found in most of cases of binding of E-64 to papain family proteases.

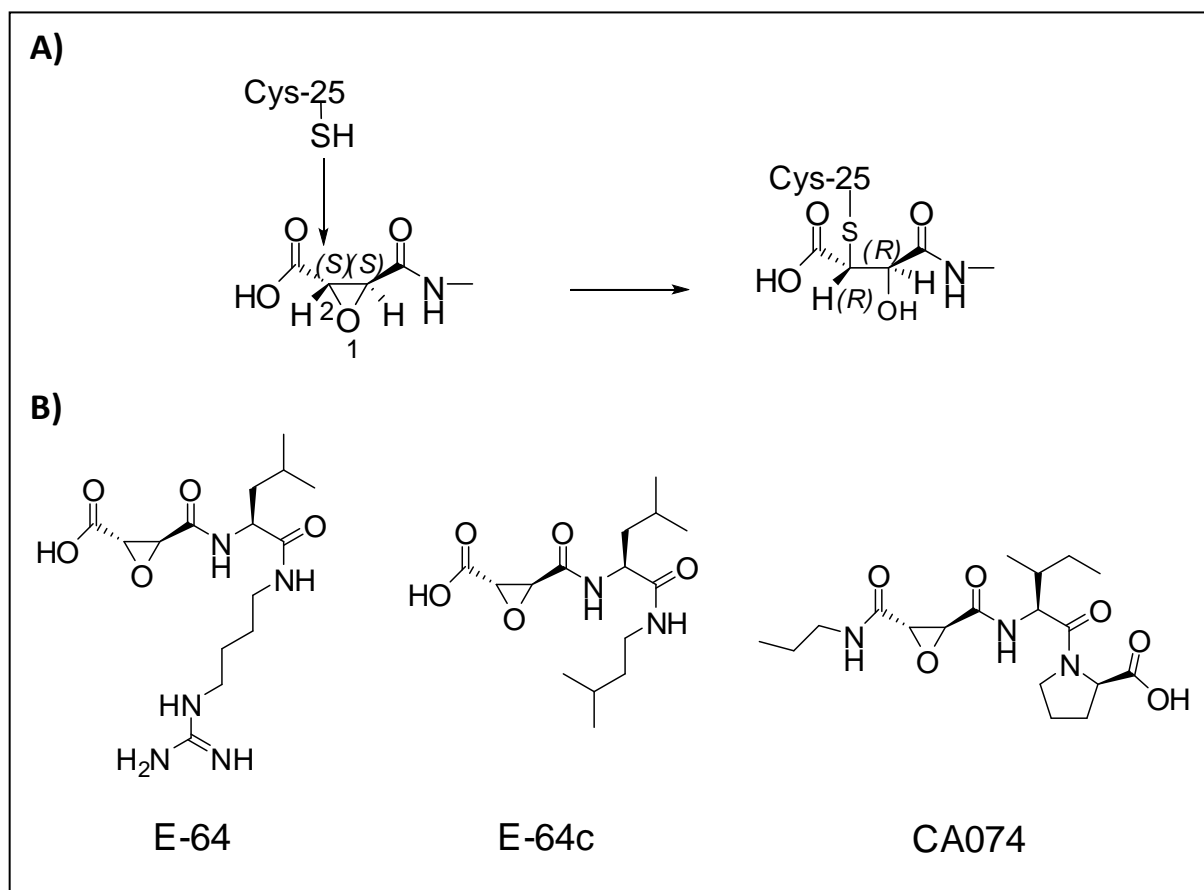


Figure 2. 5: A) Suggested S_N2 -like attack of the Cys-25 S_γ atom to an E-64 epoxy carbon atom. B) Chemical structure of E-64 and its analogues.

To date, several crystal structures of cysteine proteases inactivated by E-64 and its analogues *e.g.*: papain [368], cathepsins K [369] and B [370], mexicain [371], ervatamin [372] and actinidin [373], have been reported and it was confirmed that E-64 can bind to the active site of the enzymes in two different orientations. In most cases, E-64 binds into the *S*-subsite of the enzyme with a peptide backbone orientation opposite to that of the substrate. However, it can also bind into the *S* prime subsite as proposed in the cases of CA030 or CA074 with cathepsin B [370]. **Figure 2. 6** illustrates some examples of E-64 and its

analogues (chemical structures are displayed in **Figure 2. 5B**) bound to the active site of cysteine proteases derived from different sources, such as mammalian cells (human cathepsin K, 1ATK; rat mu-calpain, 1TLO), plant (papain, 1PE6; mexicain, 2BDZ; ervatamin-C, 2PRE; actinidin, 1AEC) and parasite (cathepsin B, 3HHI). It is interesting to note that the binding orientation of the inhibitors is very similar despite the dissimilarity of the binding pocket of each protease. Later work produced the so-called double-headed epoxysuccinyl inhibitors that can modulate the protease specificity, even at a high degree of similarity in the binding pocket [50, 303, 374-375]. Because the binding of E-64 is mostly at the S-subsite, the Sn-Pn main/side chain interaction plays an important role in determining the binding specificity of papain-family cysteine proteases. It was found that the P2 position of the substrate/inhibitor contributes significantly to the specificity towards peptide substrate/inhibitor. This amino acid is usually aromatic (Tyr, Phe, Trp) [371], as the S2 binding site is a deep hydrophobic pocket [376]. In contrast, the S1 and S1' sites of papain-like proteases provide a binding surface rather than a pocket, therefore they allow a relatively broad spectrum of substrate/inhibitor with no amino acid restrictions at the P1 and P1' sites [376-377]. One other interesting aspect is the stereochemistry of the epoxide moiety. A switch from an (*S,S*)- configuration, as present in E-64, to an (*R,R*)-form results in a dramatic change to the overall potency and specificity of inhibition [50, 359, 366, 374]. Usually the (*S,S*)-form shows significantly higher inhibitory activity than the corresponding (*R,R*)-diastereomer.

With the birth of chemical biology and its quickly evolving technology, E-64 became a favourite candidate for use as scaffold for ABP design. Its development for studying cysteine proteases was pioneered by the groups of Cravatt, Ploeg, Overcleft and Bogoy. There is still a lot of interesting work on-going based on this single molecule.

2.3 Conclusions

In this chapter the establishment of an assay that is effective for monitoring the inhibition of the SlpA cleavage process has been described. Using this assay, the first inhibitors of SlpA processing, E-64 and leupeptin, were discovered and reported. Furthermore, E-64 proved to be a more suitable candidate than leupeptin as an initial "lead", which was taken further into inhibitor design and development as described in the next chapter.

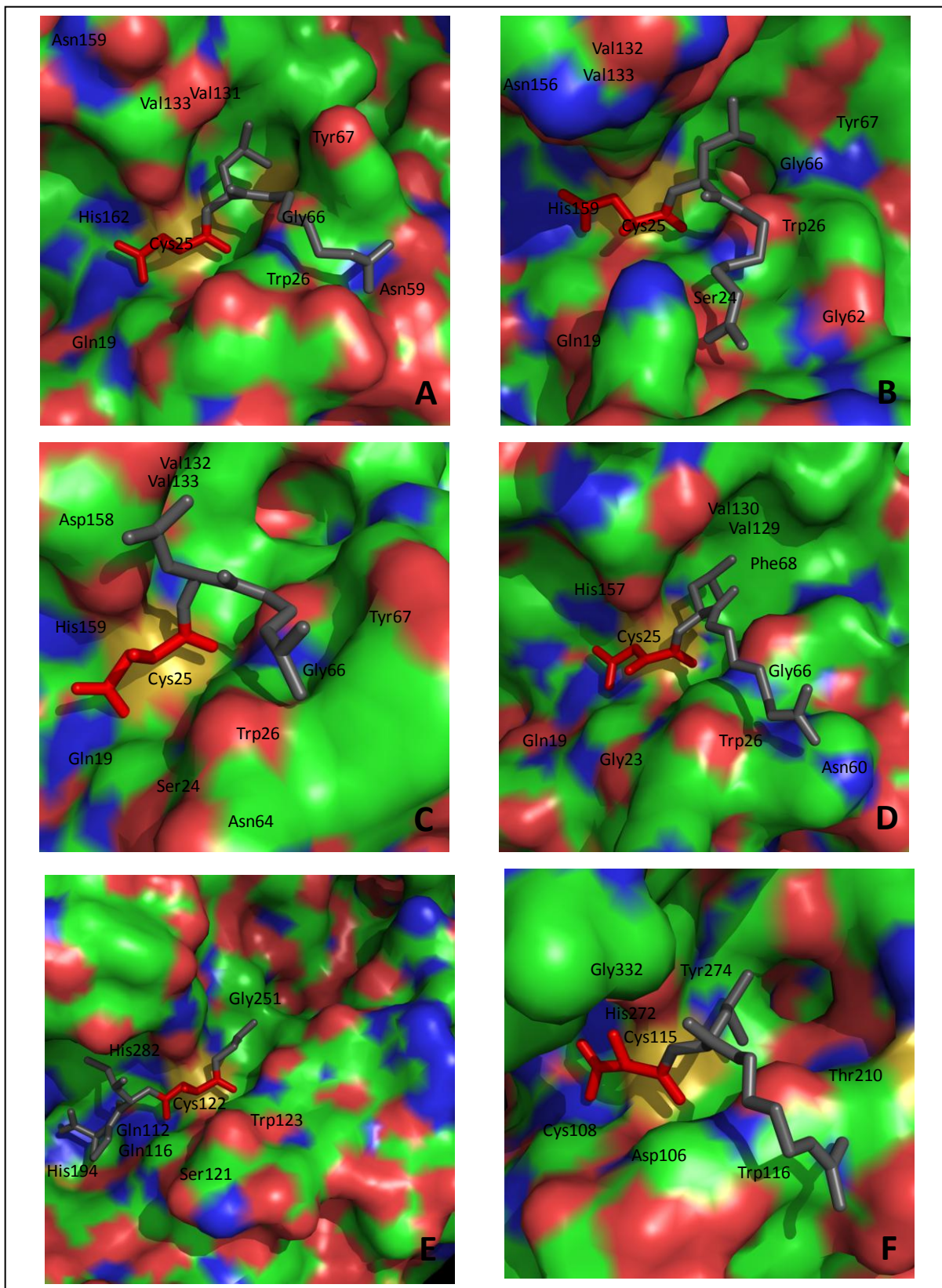


Figure 2. 6: Superposition of E-64 at the active site of different cysteine proteases. A, Cathepsin K – E-64 (mammalian cell, 1ATK); B, Mexicain – E-64 (plant, 2BDZ); C, papain – E-64c (plant, 1PE6); D, Ervatamin-C – E-64 (plant, 2PRE); E, Cathepsin B – CA074 (parasite, 3HHI); F, calpain – E-64 (mammalian cell, 1TLO). E-64 is presented as sticks (warhead is indicated in red and specificity element in gray), and the enzyme surface is represented as coloured by element C (green), N (blue), O (red) and S (gold).

Chapter 3 Synthetic inhibitors of S-layer processing in *C. difficile*

As discussed in Chapter 2, E-64 proved to be the most suitable starting point for inhibitor design. In Chapter 3, an in depth analysis of novel synthetic inhibitors is described. Starting from the E-64 scaffold and the cleavage site of the SlpA substrate, a series of inhibitors were designed. Analogues of E-64 were developed *via* versatile synthetic routes and tested for their activity and stability in order to identify more potent inhibitors of S-layer processing. These optimised inhibitors provided the basis for subsequent activity-based probe (ABP) development in this study (Chapter 4). Furthermore, the challenge of tracking down the target protease in a highly multifarious proteome mixture after cell lysis was simplified by implementing a technique for fractionating bacteria after inhibition into the S-layer and the remainder of the cell.

3.1 Potential protease targets in SlpA maturation

Cells possess a multitude of different enzymes and each of them catalyses a reaction of one substrate or a group of substrates. A range of enzymes takes part in a pathway and each enzyme activates one specific step in the pathway up to the final step reaching the desired product (**Figure 3. 1A**). Regulation results in a harmonious arrangement governed by the requirements of the cell. During development of enzyme inhibitors as drugs, the typical aim is maximal inhibition of a suitably selected target enzyme, causing either an accumulation of the substrates or a corresponding decrease in concentration of the products, which could result in a useful clinical response.

During biogenesis of SlpA in *C. difficile*, there are at least two proteases expected as the translated gene product SlpA undergoes two rounds of post-translational cleavage: the cleavage of the signal sequence (SP) that targets the nascent chain for translocation across the cytoplasmic membrane to the cell surface, and the internal cleavage that releases the mature LMW and HMW-SLPs. The protease involved in the first cleavage is presumably a signal peptidase [255]. Provided that signal peptide cleavage precedes and is essential for final maturation, inhibition of either of these proteases could cause the accumulation of full-length SlpA.

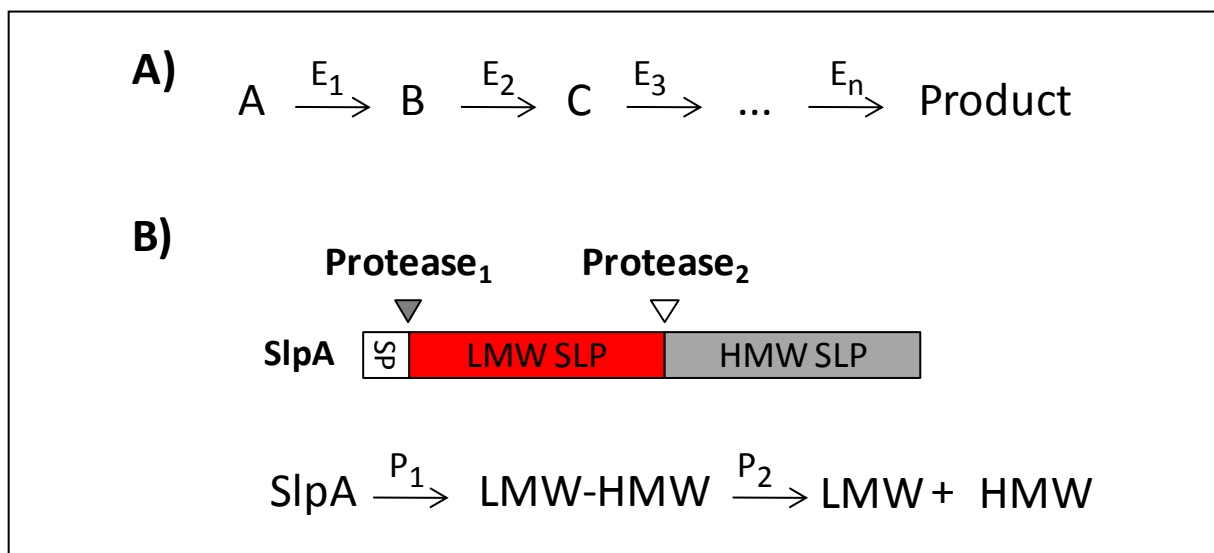


Figure 3. 1: A) Multi-enzyme pathway, e.g. in a cell. B) Cleavage of SlpA: removal of the signal peptide by protease₁ (a signal peptidase) and cleavage by protease₂ (the target enzyme for this study) to generate the LMW- and HMW-SLPs.

As all surface localised proteins in *C. difficile* contain a typical signal sequence in the precursor including SlpA [255, 277], there might be a general secretory pathway for surface proteins to export them across the plasma membrane. There are two pathways commonly responsible for general extracellular secretion: using N-terminal signal peptide as substrate for either a type I signal peptidase (SPase), a membrane-bound serine peptidase [379-380] after passing through a pore in the membrane with the support of the Sec machinery or a twin arginine transport (TAT) pathway [381]. Because the signal peptide of SlpA does not contain a twin arginine motif required for the TAT pathway, it is more likely that the signal peptidase is involved in the first cleavage.

At this stage of the project, it was not clear whether E-64 and leupeptin inhibited the signal peptidase (protease₁ in **Figure 3. 1B**) or the second protease (protease₂ in **Figure 3. 1B**). However, E-64 inhibits specifically only cysteine proteases, while signal peptidase are generally serine proteases [379]. It is tempting to speculate that E-64 and leupeptin target the second protease. As the aim of the project is to disrupt the specific cleavage of SlpA into LMW and HMW-SLPs, protease₂ was targeted for inhibitor design.

3.2 Inhibitor design

The aim of this part of the project was to develop a more potent inhibitor than E-64 that could be converted into an ABP for future study. Chapter 2 revealed two key factors that should be taken into consideration during inhibitor design. Firstly, the protease(s) involved in the cleavage process may be cysteine proteases, since the natural products E-64 and leupeptin (both known as cysteine protease inhibitors) can disrupt S-layer formation when the bacterium is grown in the presence of these inhibitors, whereas inhibitors of other protease classes cannot. Secondly, the cleaved locus of SlpA shows high homology between most strains sequenced to date, and it is thought that the protease involved in SlpA cleavage into LMW- and HMW-SLPs in these strains have the same mode of action. This implies the presence of a specific recognition element enabling binding into the active site of the proteases.

Examination of the structure of E-64 (**Figure 3. 2A**), reveals that this compound can be divided in two parts: an electrophilic epoxysuccinyl moiety (the warhead) and a short peptidic sequence (the specificity element). Similar to the canonical E-64 mode of action

discussed in Chapter 2, it was hypothesised that the warhead acts to alkylate the nucleophile of the active cysteine residue of the protease whereas the specificity element allows protease recognition. The influence of each of these two components on activity was thoroughly investigated.

A versatile synthetic route to E-64 mimics (**Figure 3. 2B**) was developed using solid phase peptide synthesis (SPPS), leading to a generic structure for these inhibitors (**Figure 3. 2C**) consisting of the two moieties previously described. The electrophilic warhead (indicated in red) in **Figure 3. 2C** is the epoxysuccinyl moiety and the specificity element (indicated in black) is the peptidic moiety thought to be mainly responsible for protease recognition. The peptidic chain terminates in a carboxamide group rather than a carboxylate group to avoid the associated negative charge in a cellular environment. This should act to minimise the possibility of metabolism and aid cell penetration and thus might enable the inhibitor to reach the site of the target protease. For future ABP development, a tag could be appended onto the peptidic unit, also on solid phase.

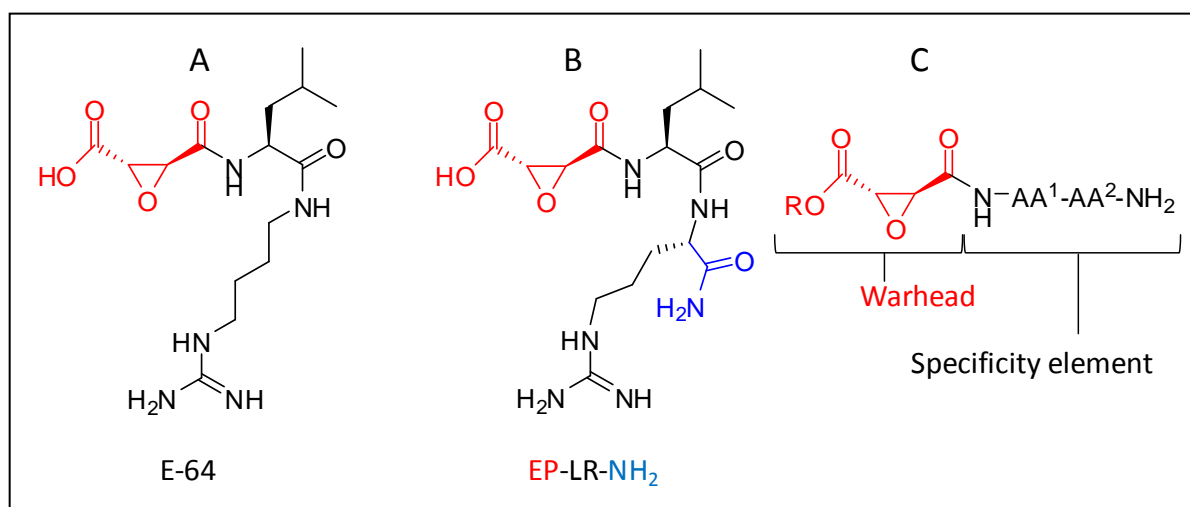


Figure 3. 2: Structures of inhibitors. A) E-64. B) EP-LR-NH₂, a mimic of E-64. C) Generic structure of inhibitors. Warheads are indicated in red and the specificity element is indicated in black.

For simplicity, a short and general nomenclature was applied to all inhibitors. The epoxide will be abbreviated as EP, whereas the EtEP presents the ethyl ester form. All EPs have an (*S,S*) configuration unless otherwise stated and all amino acids of the specific peptidic chain are represented by their single letter code. Hence under this nomenclature the E-64 mimic (**Figure 3. 2B**) will be named EP-LR-NH₂. E-64 and EP-LR-NH₂ differ from each other only in the carboxamide group indicated in blue (**Figure 3. 2 A and B**). To examine the space

requirement for the binding of the specificity element to the active site, glycine was inserted into the sequence to act as a spacer to extend the peptidic chain adding EP-LGR-NH₂ into the new design inhibitor list.

In order to assess the relative importance of these features of E-64 in facilitating cysteine protease inhibition, a series of modified E-64 mimics were synthesised. First, the warhead was modified, while the specificity element was kept as LR (**Figure 3. 4**). This included replacing the epoxysuccinyl moiety by an acetyl group to investigate the significance of the warhead motif (Ac-LR-NH₂); esterification of the epoxy carboxylate group with a view to improving cell penetration and potency (EtEP-LR-NH₂) [155]; removing the peptide appendage to demonstrate the necessity for a specificity element (EtEP-OH); and inverting the (*S,S*) stereochemistry present in E-64 to the (*R,R*) configuration since the stereochemistry of the epoxide has been shown in other studies to have a critical contribution to the inhibitory effect [155, 342, 366-367, 374]. It should be noted that this type of inhibitor has never previously been tested on *C. difficile*.

Second, the specificity element was tailored, while the warhead was kept as epoxysuccinyl. Since the discovery of E-64, research has primarily focused on the variation of the specificity element with the intention of improving the selectivity and potency [366, 371], and these studies suggest that the specificity element usually binds at and beyond the P2 position of the substrate as the epoxysuccinyl moiety occupies the P1 position. **Figure 3. 3** illustrates the binding mode of E-64 in actinidin [373] as an example of various other papain-like protease co-crystal structures and the suggested binding mode for the novel synthetic inhibitors designed in this study. It has been found that the S_n region varies in size from S2 to S5 across the cysteine proteases [367, 377, 382]. In the majority of cases, the inhibitor is bound in the S_n-subsite of the active centre and is positioned in the opposite sense to the substrate, i.e. the inhibitor binds in a C-to-N fashion along the S1-S2...S_n pocket whereas the substrate binds in an N-to-C fashion [366, 371]. Taking these findings into consideration, several inhibitors were synthesised in which the length of the peptidic chain mimicking the cleavage locus in SlpA was varied (EP-K-NH₂, EP-KT-NH₂, EP-KTE-NH₂, EP-KTEL-NH₂ (by reference to strain 630) and EP-KTT-NH₂, EP-KA-NH₂, EP-YT-NH₂ (by reference to strains 959 and 167) as shown in **Figure 3. 4**.

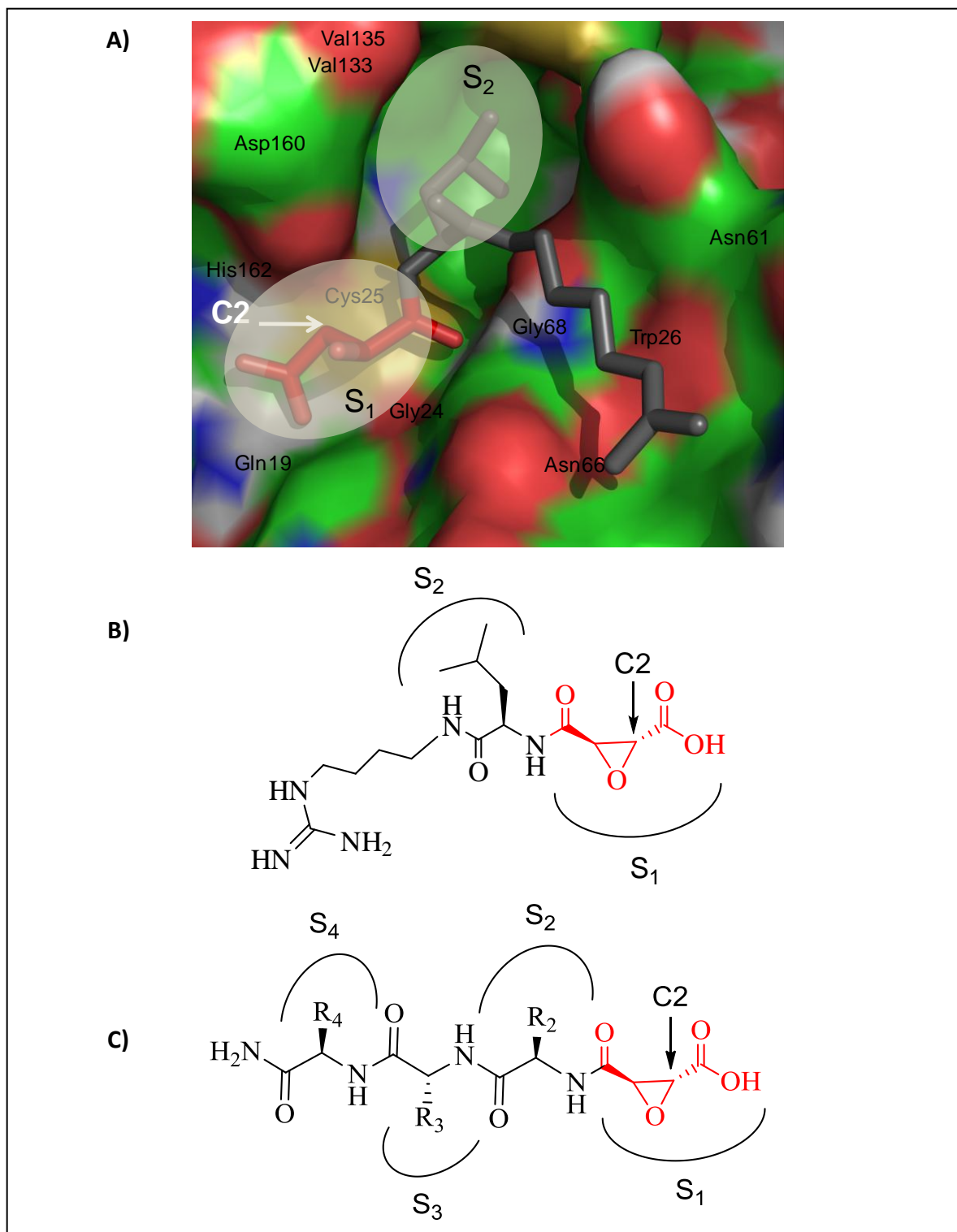


Figure 3. 3: Binding mode of E-64 and E-64 based inhibitors to cysteine proteases A) actinidin-E64 complex (1AEC). B) General binding mode of E-64. It should be noted that the backbone of E-64 binds to the S subsites in the opposite sense to the substrate. C) Proposed binding mode of E-64 based inhibitors.

It has usually been found that the P2 position of the substrate has a more significant effect than subsequent positions on the activity [50, 373], thus further inhibitors were produced in which the lysine in the AA¹ position which is corresponding to the P2 position of the

substrate was substituted by an alanine (a technique commonly applied to neutralise the effects of a particular residue side-chain); EP-A-NH₂, EP-AT-NH₂ and EP-ATT-NH₂); by a threonine ((present in the P3 position in the substrate, K-to-T); EP-T-NH₂); and by leucine or arginine ((inspired by the structure of E-64, K-to-L/R); (EP-R-NH₂, EP-L-NH₂, EP-LT-NH₂, EP-LTT-NH₂ and EP-LTTL-NH₂)).

To summarise, the inhibitor library designed consisted of 22 analogues covering most aspects of known E-64 structure-activity relationships (SARs) –These inhibitors were synthesised as described in the following section.

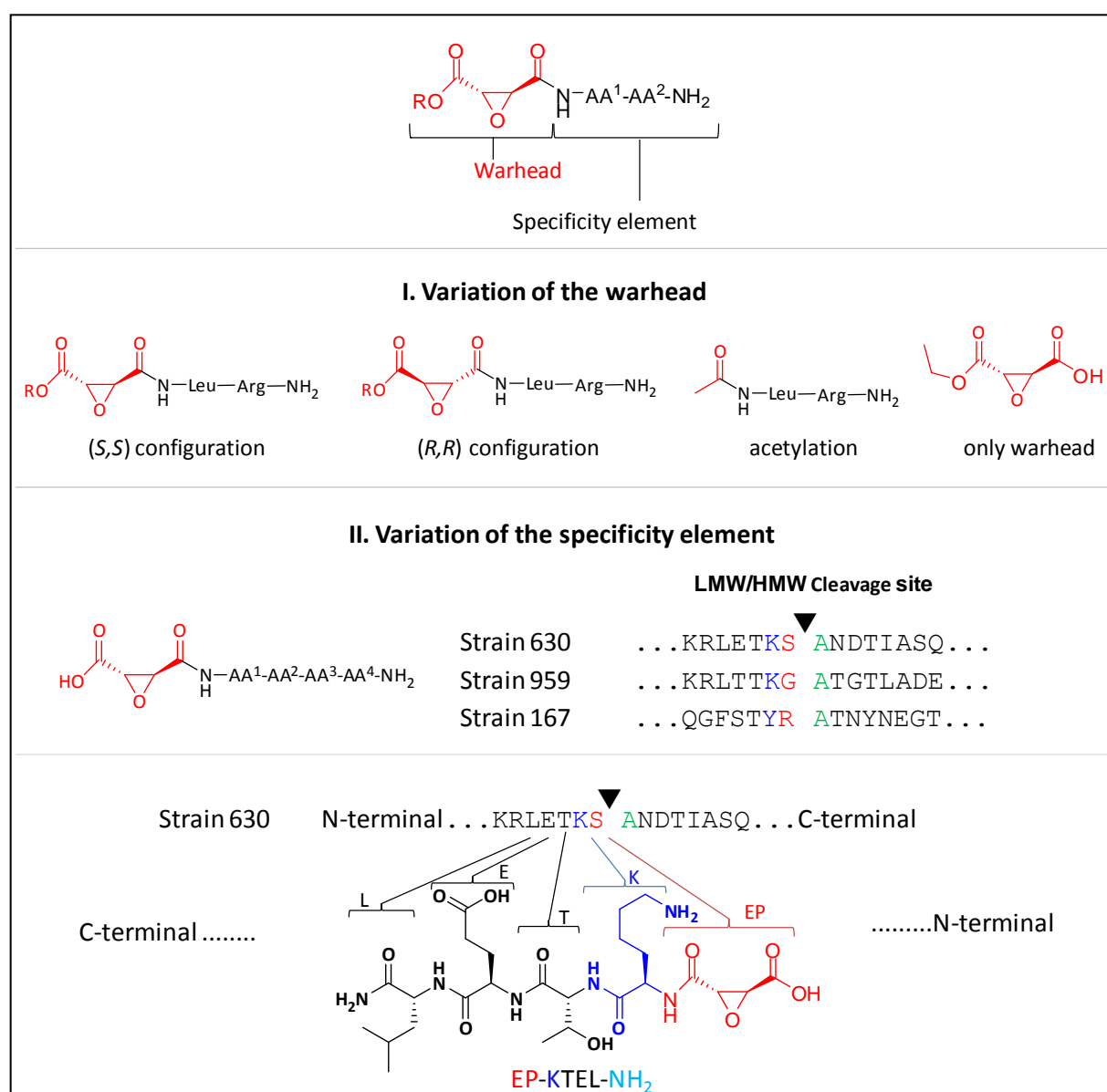
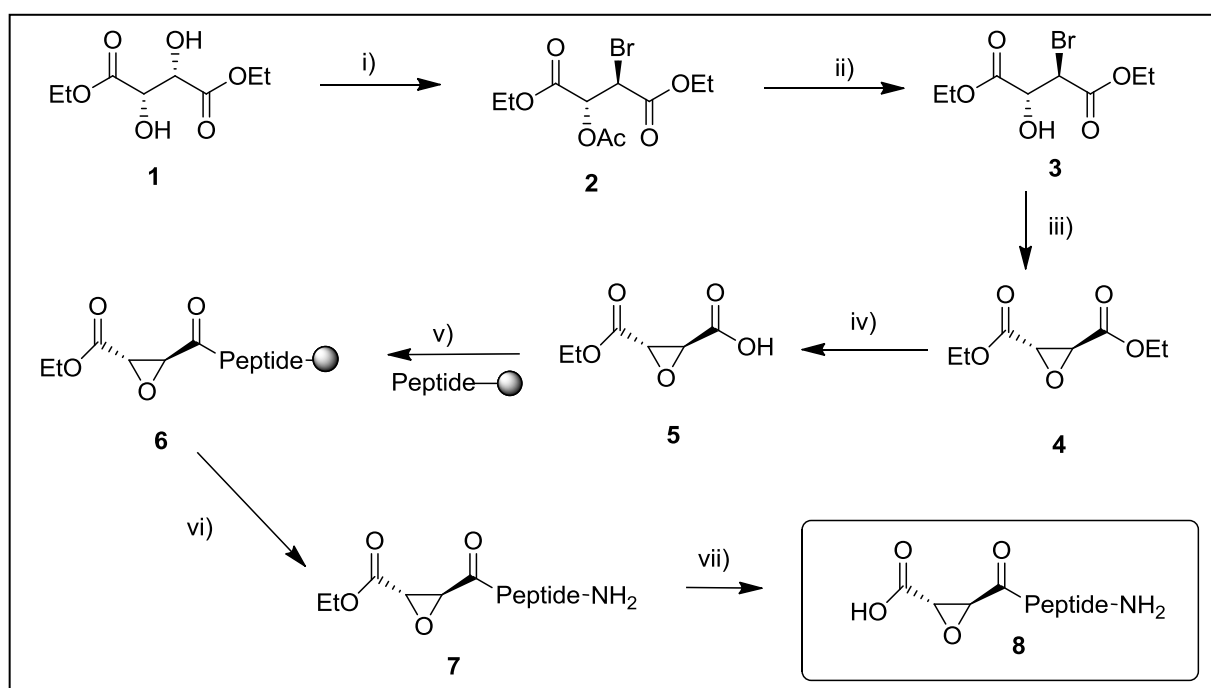


Figure 3. 4: Inhibitor design. Both components of the inhibitor, the warhead and the specificity element, were varied to investigate the influence of each components on the activity. Modification of warhead was concentrated on stereochemistry, while the one of specificity element was based on cleavage site locus of StpA.

3.3 Synthesis of E-64 analogues

The designed inhibitors discussed above were synthesised with the intention of turning them into ideal tools to evaluate the potency and selectivity towards the target protease. E-64 and its derivatives have historically been prepared using a short and flexible route in which the epoxysuccinyl moiety was synthesised using a modified method of Mori and Iwasawa [383]. This epoxide portion was subsequently appended to a peptide, where many modifications have also been reported [59, 151, 384-385].



Scheme 3.1: Synthesis of inhibitors. Reagents and conditions: i) HBr-HOAc, 0°C to RT, 16 h; ii) HCl-EtOH, Δ, 5 h; iii) DBU, Et₂O, 2 h; iv) KOH (1 eq), EtOH, 0 °C, 3 h; v) HCTU, DiPEA, 1 h; vi) TFA, TIS, H₂O (95:2.5:2.5), 2 h; vii) NaOH, 30 min. Yields of VII vary from 1 to 87% depending on the sequence.

The general route employed for the preparation of the E-64 analogues is illustrated in **Scheme 3.1**. The synthesis is a combination of solution and solid-phase chemistries. For simplicity, **Scheme 3.1** displays only the synthesis of (*S,S*)-configuration of epoxysuccinyl inhibitors, derived from (-) diethyl-D-tartrate, whilst (+) diethyl-L-tartrate yielded the analogue possessing an (*R,R*)-configuration.

Diethyl tartrate (**1**) was converted into the bromo ester (**2**) using HBr-HOAc, and this was subsequently deacetylated by HCl-EtOH treatment to yield the bromohydrin (**3**). A second Walden inversion was achieved by treatment with DBU in Et₂O at 4 °C to force ring closure and this afforded the epoxide (**4**) in 77 % yield. The di-ester was saponified with an

equimolar amount of KOH in EtOH, and after a dilute HCl work-up and EtOAc extraction, the desired epoxy mono-ester (**5**) was generated. Subsequent attempts at coupling the crude **5** onto the resin-bound peptide were unsuccessful despite what was previously reported to be a facile coupling step [59-60]. The coupling failure implied that even a small amount of acetic acid, formed during the work up by hydrolysing EtOAc through HCl, could be in competition with the epoxide mono-ester to prevent the coupling reaction from taking place. Hence the epoxide mono-ester was purified *via* HPLC and lyophilised to remove most residual water before coupling onto the resin-bound peptide. In addition, the mono-ester was seen to decompose over time even when stored at -20 °C, thus the warhead was stored as the epoxide di-ester and hydrolysed to the monoacid when required.

Standard automated Fmoc/^tBu solid-phase peptide chemistry was exploited in order to generate the peptide component of the E-64 mimics, where the amino acids were pre-activated *in situ* with HBTU and DiPEA. This provides a flexible system to enable the production of inhibitors with peptidic elements containing any combination of amino acids. However, this is not the only advantage to this strategy as once the desired peptide element has been made this also enables the straightforward incorporation of the epoxide warhead onto the peptide N-terminus and hence generate the desired inhibitors in a small number of steps in a typically high overall yield.

The key step in the synthesis was the coupling of **5** to the peptide portion of the inhibitors, and in particular, the activation of **5** prior to this step. Several strategies employed for the activation of such unusual compounds include the utilisation of DCC-HOBt [386-388] or DCC-p-nitrophenol [59], conversion to the NHS or pentafluorophenyl ester [389], and the use of phosphorous-based reagents such as diethylphosphoryl cyanide, BOP and PyBOP [365, 384]. During this study, it was determined that pre-activation using HCTU and DiPEA led to improved yields and required shorter coupling times (1h) in comparison with other coupling reagents *e.g.* PyBOP (5h).

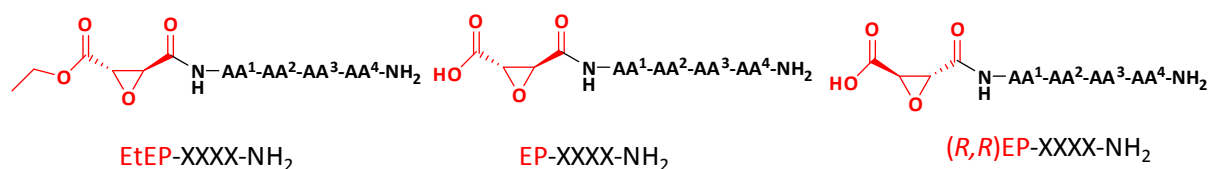
Once the epoxide peptide (**6**) was formed, the resin was washed thoroughly and dried before cleavage to obtain better yields. HPLC or LC-MS analysis revealed the crude material contained predominantly the ethyl epoxysuccinyl compound (**7**) however the hydrolysis product (**8**) was also isolated on occasion. When the epoxide mono-ester (**5**) used for

coupling to the peptide moiety contained a trace of acetic acid, an acetylated product of this pseudo peptide was also found. **7** was purified *via* HPLC and lyophilised to yield the ethyl epoxysuccinyl coupled peptidic inhibitors.

In E-64 the epoxysuccinyl moiety contains a free carboxylic acid, and thus the ester functionality in **7** was also hydrolysed with NaOH to give the acid (**8**) in quantitative yield. It was thought that it would be advantageous to eliminate the hydrolysis step to obtain the epoxy acids directly *via* coupling the epoxide di-acid to the free N-terminus of the resin-bound peptide instead of **5**. Unfortunately, the di-acid proved to be insoluble in DMF, contrary to the literature precedent [390]. This may be due to the difficulty of removing all the remaining water formed during epoxide di-acid precipitation, which was shown to remain despite heating and lyophilisation. The subsequent coupling was therefore not accessible.

Unfortunately, compounds possessing a single amino acid bearing a small or hydrophobic side chain such as alanine or leucine were also not accessible using the described method despite ESI-MS detection in the crude material since isolation by preparative HPLC proved problematic. Alternative approaches employed alaninamide or leucinamide hydrochloride with preactivated epoxide monoacid using either PyBop/DiPEA in DMF at room temperature or HOBt/DiPEA/EDC in chloroform/DMF at 0 °C [150]. However, these reactions were unsuccessful, and no product was observed in the crude ESI-MS spectrum. No further attempt was made to synthesise these compounds.

Stocks (10 mM or 25 mM) of the desired inhibitors accessible *via* the aforementioned synthetic route were prepared in water, or in a water/DMSO mixture for more hydrophobic compounds. It should be noted that the proportion of water/DMSO is depend on the hydrophobicity of the inhibitors and varied among inhibitors. However, the final level of DMSO in the culture was always less than 1% v/v. At this level DMSO does not have any influence on SlpA inhibition or cell growth (**Figure 3. 7**). Stock concentration calculations were based on the mass of the freeze-dried product, which may be inaccurate due to the accuracy of weighing relatively small amounts of material (± 0.1 mg) and potential additional mass of residual water or salts.



No.	Warhead	AA ¹	AA ²	AA ³	AA ⁴	Abbr. Name
TD 1	EP	Leu	Arg	-	-	EP-LR-NH ₂
TD 2	EtEP	-	-	-	-	EtEP-OH
TD 3	Ac	Leu	Arg	-	-	Ac-LR-NH ₂
TD 4	(R,R)EP	Leu	Arg	-	-	(R,R)EP-LR-NH ₂
TD 5	EtEP	Leu	Arg	-	-	EtEP-LR-NH ₂
TD 6	EP	Lys	-	-	-	EP-K-NH ₂
TD 7	EP	Lys	Thr	-	-	EP-KT-NH ₂
TD 8	EP	Lys	Thr	Glu	-	EP-KTE-NH ₂
TD 9	EP	Lys	Thr	Glu	Leu	EP-KTEL-NH ₂
TD 10	EP	Lys	Ala	-	-	EP-KA-NH ₂
TD 11	EP	Lys	Thr	Thr	-	EP-KTT-NH ₂
TD 12	EP	Tyr	Thr	-	-	EP-YT-NH ₂
TD 13	EP	Leu	Thr	-	-	EP-LT-NH ₂
TD 14	EP	Leu	Thr	Thr	-	EP-LTT-NH ₂
TD 15	EP	Leu	Thr	Thr	Leu	EP-LTTL-NH ₂
TD 16	EP	Leu	Gly	Arg	-	EP-LGR-NH ₂
TD 17	EP	Thr	-	-	-	EP-T-NH ₂
TD 18	EP	Ala	Thr	-	-	EP-AT-NH ₂
TD 19	EP	Ala	Thr	Thr	-	EP-ATT-NH ₂
TD 20	EP	Arg	-	-	-	EP-R-NH ₂

Table 3. 1: Inhibitors synthesised at this stage in this study.

Attempts were made to circumvent this problem by normalising the stock concentration to a known concentration using UV-absorption of the peptide bond at 223 nm *via* analytical HPLC. However, this measurement was found to be ambiguous since amino acid side chains also absorb UV light at this wavelength but to differing degrees depending on the amino acid present, leading to a complex absorption spectrum and thus the normalisation strategy

was unreliable. Moreover, potential variation in the amount of sample injected during each LC-MS run was a further limitation for this approach, and thus the freeze-dried mass was used despite its potential inaccuracies. Consequently, the stock concentration determination based on the product weight is reported in this study, and the absolute concentration of the stock solution generated for testing is likely to be somewhat lower. For the purpose of finding specific inhibitors that only label the protease involved in the SlpA cleavage process in the presence of many other bacterial proteases, this potential mass inaccuracy is not a major problem. In summary, a library of 20 compounds was generated for initial *in vivo* testing; the library members are presented in **Table 3. 1**.

3.4 Activity of E-64 analogues

As discussed, the purpose of the study at this stage was to find potent inhibitors that can act as precursors to aid the identification of the target protease(s). The inhibitors presented in **Table 3. 1** were tested *in vivo* using the assay described in section 2.1.1, but before discussing the results it is worth reiterating some of the limitations of the chosen assay technique that should be taken into consideration during interpretation. Firstly, the growth rate of *C. difficile* is rather inconsistent despite the application of some technical improvements (discussed in section 2.1.1). Secondly, because the compounds do not directly inhibit growth, the observed effect may appear relatively insignificant as the samples were normalised by OD. However the behaviour of the bacterium in the presence of the inhibitor could be affected in a subtle way. Secondly, the small inaccuracy in the stock concentration may affect the relative potencies when comparing different inhibitors. Nevertheless, a given compound may be reliably assayed and compared at various concentrations from the same stock solution. Thirdly, a Western blot is not suitable for precise quantitative analysis in the absence of a control band since the intensity of a particular band on the blot varies from one membrane to the next and also depends on the length of the exposure time during imaging. Furthermore, this intensity is judged by eye, which further reduces the accuracy. Despite this, it is usually reliable to compare bands on the same membrane, and a negative and positive control on each membrane acts as a useful reference. As mentioned in the assay section 2.1.1, the band for full length SlpA appears when the target protease is inhibited since there is no longer full cleavage of SlpA by the protease. The extent to which this cleavage is inhibited correlates well with the full

length SlpA band intensity, because the better the inhibitor inactivated the target protease the less it cleaves SlpA and therefore the more intense the band becomes. Thus, within the limitations noted above, this band can be exploited as a semi-quantitative scale for a lower limit of inhibitor potency, where the stronger the band the more potent the inhibitor.

Initially, the inhibitor EP-LR-NH₂ was compared side-by-side with the natural product itself (E-64) at two different concentrations (50 μ M and 250 μ M). The anti-LMW Western blot of this comparison is shown (**Figure 3. 5**, includes a negative control lane in which no inhibitors were added). It is readily noticeable that EP-LR-NH₂, which retains the key stereochemical and functional features of the natural product, is probably somewhat more potent than E-64 since the full length SlpA band is clearly visible at 50 μ M in EP-LR-NH₂ but not present in E-64. This was very encouraging as it implies that to some extent the considerations taken and assumptions made during inhibitor design were justified and with optimisation could potentially lead to more potent inhibitors. The fact that all the LMW-SLP bands are equal indicates that the same amount of sample was loaded on each lane and suggests the results can be interpreted reliably. Comparing inhibition at 250 μ M, there is no significant difference between EP-LR-NH₂ and E-64. It was decided to introduce a concentration of 100 μ M in order to improve the interpretation of results when comparing different inhibitors on Western blot.

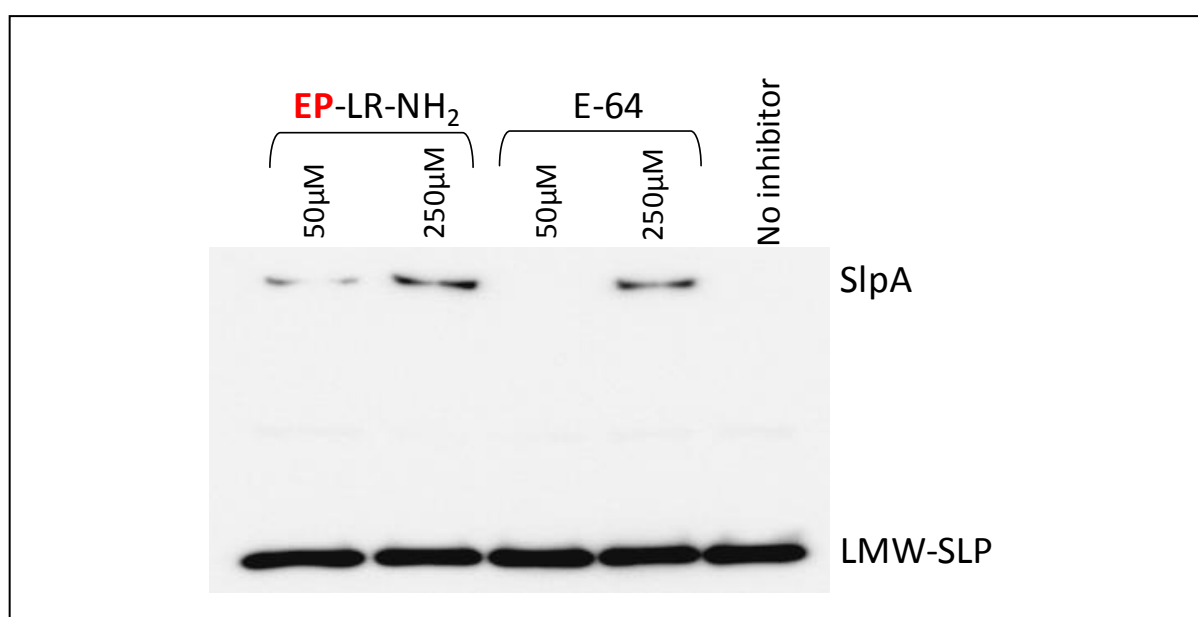


Figure 3. 5: Activity of EP-LR-NH₂ vs. E-64, as measured by Western blot using anti-LMW SLP. *C. difficile* 630 cultures were treated with each compound at 50 μ M and 250 μ M.

The importance of the warhead was next investigated. EtEP-OH, Ac-LR-NH₂, EtEP-LR-NH₂, (*R,R*)EP-LR-NH₂ were tested alongside two positive controls (EP-LR-NH₂ and E-64) and the negative control, with all compounds at 100 μM (**Figure 3. 6**). **Figure 3. 6** demonstrates that both positive controls possess similar activity at 100 μM which is in line with previous observations. It is also evident that the substitution of the N-terminal *trans*-(*S,S*)-epoxysuccinyl warhead with an acetyl group or a *trans*-(*R,R*)-epoxysuccinyl warhead leads to a complete loss of activity. This validates the hypothesis that the epoxide warhead is not only critical for inhibition but its binding to the protease is also highly stereoselective. Furthermore, the warhead on its own without any appended specificity element was determined to be inactive, illustrating that both parts, the warhead and the specificity element, are essential for achieving inhibition.

Finally, no significant difference was identified in the extent of inhibition at this concentration between the free acid and the ester form of the epoxide warhead. It is difficult to speculate on which of the analogues are more likely to accomplish superior cell penetration or reach the target protease more easily and thus it is challenging to differentiate between them at this stage with regards to their relative activities. Another speculation is that the ethyl ester compound might be hydrolysed to the acid form in the culture and therefore it possesses similar potency. Further investigation to answer these intriguing questions fell outside the principal focus of this research and hence this further work will be carried out in subsequent studies.

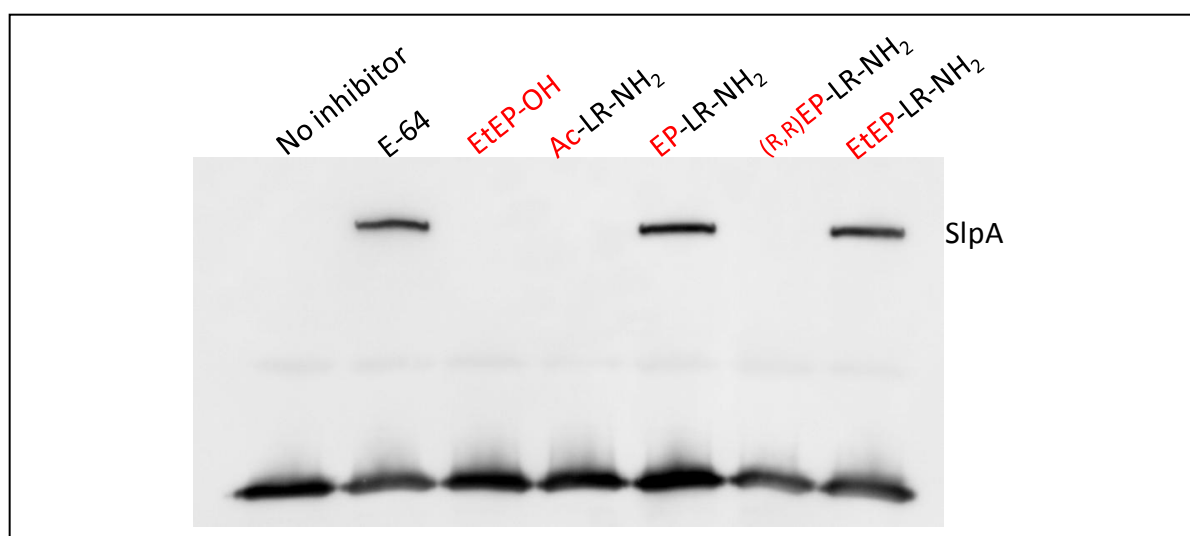


Figure 3. 6: Activity of inhibitors, as measured by Western blot of S-layer fractions using anti-LMW SLP. *C. difficile* 630 cultures were treated with compound at 100 μM. The warhead was modified and a specificity element was kept as LR-NH₂, as present in the parent precursor.

The effects of the variation of the specificity element on inhibition were examined, and the warhead was kept in its acid form (as seen in E-64) for all inhibitors to ease directly comparison (**Figure 3. 7**). The data suggests that increasing the peptide length to mimic P2 up to P5 of the substrate for strain 630 enhanced inhibition (EP-KTEL-NH₂ was more active than EP-KT-NH₂). However, EP-K-NH₂ and EP-KTE-NH₂ were found to be unexpectedly inactive at the concentration tested. The activity of EP-KT-NH₂ and inactivity of EP-K-NH₂ might imply that at least two amino acids are required for binding, but it is difficult to suggest why EP-KTE-NH₂ was inactive whilst EP-KTEL-NH₂ was active. Maybe the glutamic acid is unfavoured in the binding, but this unfavourableness is compensated by adding a leucine in the chain. On the other hand, EP-KA-NH₂, EP-KTT-NH₂ and EP-YT-NH₂ (refer to strain 959 and 167) all show some level of activity. EP-YT-NH₂ is the most potent compound from the entire series as measured by Western blot. The variation of the AA¹ leads to highly intriguing results as can be seen for EP-KT-NH₂, EP-YT-NH₂, EP-LT-NH₂ and EP-KT-NH₂ in **Figure 3. 7**. These inhibitors carry the same AA² residue, threonine, but a different AA¹ residue. An alanine at AA¹ position causes a consistent loss of activity, a phenomenon further confirmed by EP-ATT-NH₂. The activity increases dramatically from leucine to tyrosine. EP-KT-NH₂ and EP-KA-NH₂, where the AA¹ residue is the same but the AA² amino acid is different, are observed to have similar activities. Furthermore, if threonine is shifted to the P2 position (as in EP-T-NH₂), inhibition is lost.

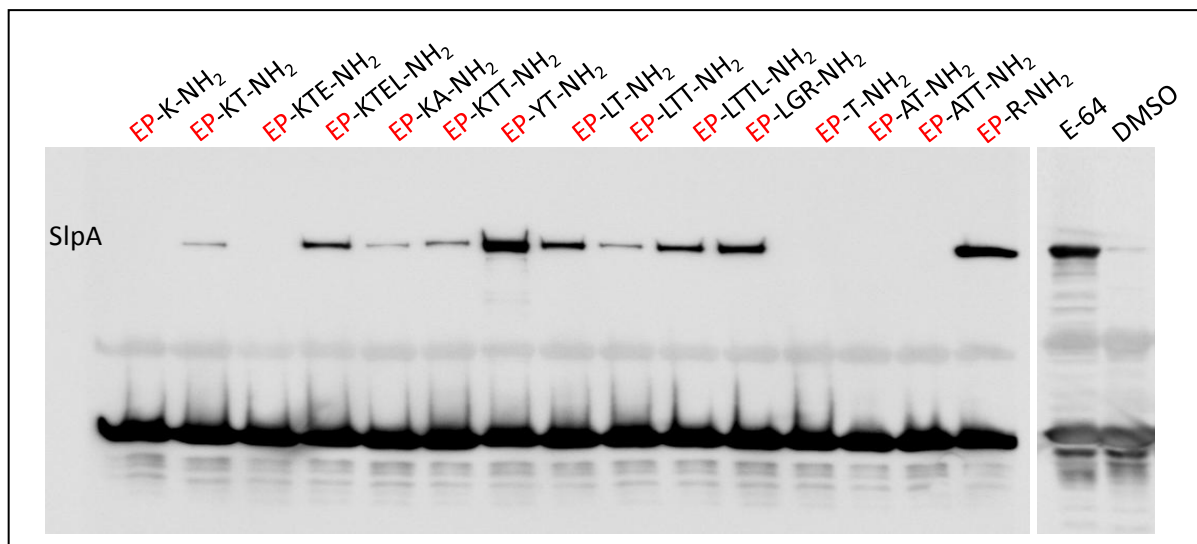


Figure 3. 7: Activity of inhibitors, as measured by Western blot of S-layer fractions using anti-LMW SLP. *C. difficile* 630 cultures were treated either with compound at 250 μ M or 1% DMSO. The specificity elements were varied and the warhead was kept as EP, as present in E-64.

Regarding the inhibitors designed through inspiration from the structure of E-64 (i.e. AA¹= leucine: EP-LT-NH₂, EP-LTT-NH₂, EP-LTTL-NH₂), only moderate activities were observed compared to EP-KTEL-NH₂. The addition of a spacer in the middle of the peptidic sequence of the parent inhibitor did not affect its activity. Unfortunately EP-L-NH₂ was not readily accessible using the routes described above, so a direct comparison with the related EP-R-NH₂ was not possible.

Accumulating the knowledge gained from the comparison between EP-AA¹XX-NH₂ and EP-XXAA²-NH₂ (where XX is the varied amino acid), the inactivity of EP-T-NH₂ and the increased potency of EP-R-NH₂ is compelling evidence that the AA¹ position plays the decisive role in contribution towards binding whereas variation in the AA² position does not affect the activity within the range of compounds explored here. Extending the length of the peptidic chain from AA³ to AA⁴ does not gain substantial additional activity. On the contrary, the potency of EP-R-NH₂ indicates that even one amino acid in combination with the warhead is enough to support a relatively effective inhibition.

Looking at the unanticipated inactivity of EP-K-NH₂ and EP-KTE-NH₂, this could be associated with the limitation of the assay discussed above. The stock concentration of EP-K-NH₂ and EP-KTE-NH₂ might be much lower than determined by mass, consequently their real concentration in this particular experiment could be significantly below 100 μM which may suggest why no activity was detected.

3.5 Determination of the location of the target protease(s)

E-64, discovered as one of the first SlpA cleavage inhibitors, does not have a noticeable effect on bacterial growth in rich BHI media (section 2.1.3). It was found that the inhibitors synthesised in this study also do not have an obvious effect on bacterial growth due to similar OD measured before harvesting. Nevertheless, an impact on cell integrity was found. Boiling *C. difficile* cells in a buffer containing strong anionic detergent (0.7% SDS, 100 °C) does not typically induce lysis because under normal culture conditions (described in section 2.1.1) *C. difficile* possesses an extremely stable cell wall in addition to the S-layer. However, cells treated with inhibitor are less robust under the same conditions and cause a detectible effect on lysis as can be seen in **Figure 3. 8A**. One could speculate, therefore, that inhibition has a marked effect on cell integrity even when bacteria are cultured under optimal

conditions (e.g. BHI medium). From the aspect of future drug development based on this type of inhibitor, S-layer processing might be a potential therapeutic target, if possibly this impact on integrity might increase significantly, when the bacteria suffer under general stress such as antibiotic treatment or attack by the host immune system.

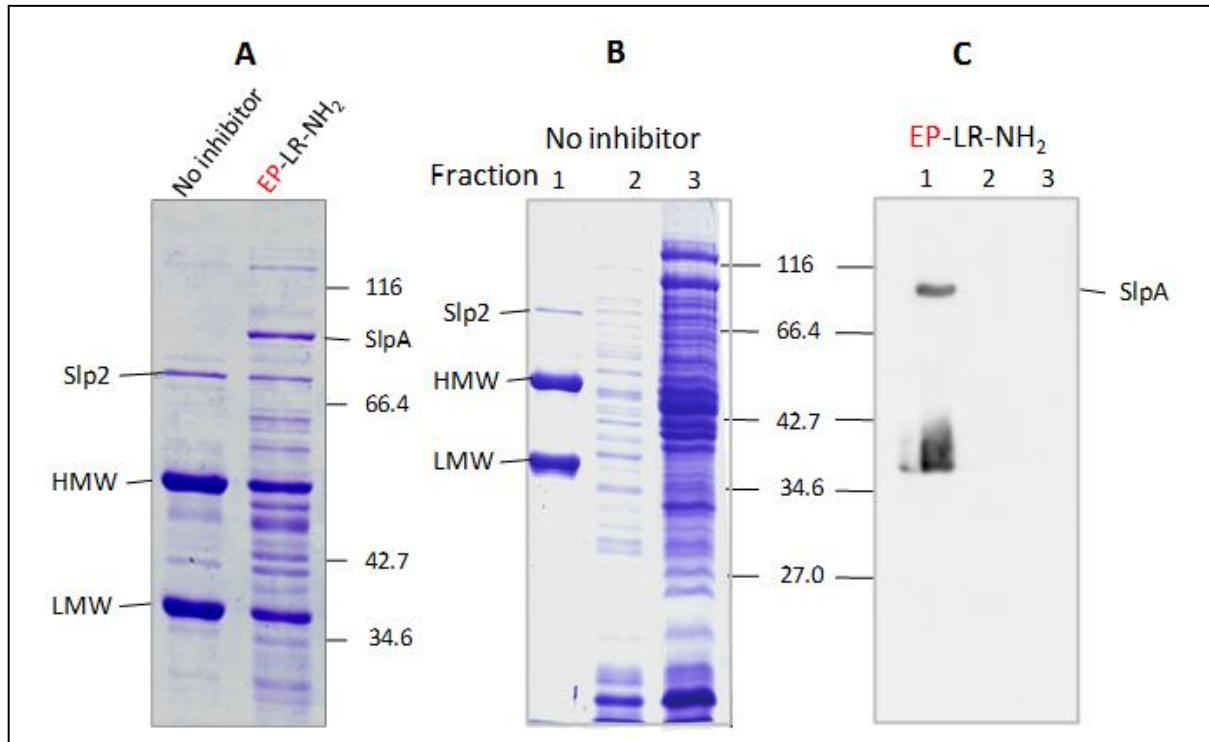


Figure 3. 8: A) Coomassie blue-stained gels of protein released from cells untreated and treated with inhibitor (100 μ M EP-LR-NH₂) after boiling in 0.7% SDS at 100 °C for 5 min. Cells treated with inhibitor release a large quantity of intracellular protein, whilst untreated cells release proteins mainly or exclusively from the S-layer; Coomassie stained gels and anti-LMW-SLP Western blot of protein derived from fractions: 1 = glycine fraction; 2 = soluble fraction; 3 = insoluble fraction. B) *C. difficile* cell without treatment with inhibitor. C) *C. difficile* cell treated with EP-LR-NH₂.

The protein mixture derived from whole cell lysis is highly complex as it contains proteins from the S-layer, the cell wall, cytoplasmic proteins and other cell components. This complexity would present an additional challenge to locate the target enzyme. We next sought to determine the location of the active target protease in live cells (surface layer, membrane or intracellular) with the aim of reducing the complexity of the identification process. The cell was separated into three fractions. Firstly, the S-layer was stripped using low pH glycine following the standard extraction protocol [391]. The remaining cells were treated with lysozyme (to digest the polysaccharide component of bacterial cell walls) and DNase (to reduce the viscosity causing by released nucleic acid material) followed by centrifugation, based on the method of Jonquires [392]. Cell lysis was further enhanced by

additionally treating the cells with glass beads in order to facilitate the breakdown of cell walls after lysozyme treatment. As a result, the stripped cells were collected in two further fractions: the soluble fraction (containing cytoplasmic proteins) and the pelleted insoluble fraction (containing membranes and insoluble cell components). The proof of principle of this technique for achieving effective cell separation was demonstrated in a sample without any inhibitor shown in **Figure 3. 8B**, highlighted by different bands in each fraction (only LMW- and HMW-SLPs, Slp2 in the S-layer fraction, distinctive bands for soluble and insoluble fractions). This experiment was repeated with EP-LR-NH₂ inoculation before the bacteria were grown. **Figure 3. 8C** indicates that the inhibitory effect occurs after (or possibly during) delivery of SlpA to the S-layer fraction as no SlpA band was observed in the soluble and insoluble fractions.

As previously shown and discussed in section 3.4, it was not clear which protease was targeted by this type of inhibitor. However, looking at Figure 3. 8C, it can be concluded that the inhibition most probably follows translocation to the cell surface and that the synthetic inhibitors target the putative protease involved in the internal cleavage to produce mature LMW and HMW-SLPs. This suggests the hypothesis that the putative target protease resides in the S-layer of the bacteria and is probably readily accessible to inhibitors present externally in the cell growth medium. If this is the case then inhibitors do not require intracellular uptake through the dense S-layer/cell wall, lending further support to the potential of this process as a therapeutic target. However, further study is required to provide concrete evidence that this is indeed the case, as described in the next chapter.

3.6 Activity vs. concentration

The dependence of *in vivo* inhibition on inhibitor concentration was investigated. To examine the maximal concentration of inhibitor required for an absolute inhibitory effect the concentration series tested ranged from 250 μ M to 5 mM. Consequently, this experiment required a larger amount of inhibitor. EP-LGR-NH₂ was found to have similar activity to that of the parent inhibitor EP-LR-NH₂, and was relatively high-yielding, therefore it was chosen for this trial experiment. Remarkably, the amount of full length SlpA accumulating in the S-layer increased with increasing EP-LGR-NH₂ concentration until 0.5

mM at which point a plateau was reached where increases in concentration had no further effect (**Figure 3. 9**).

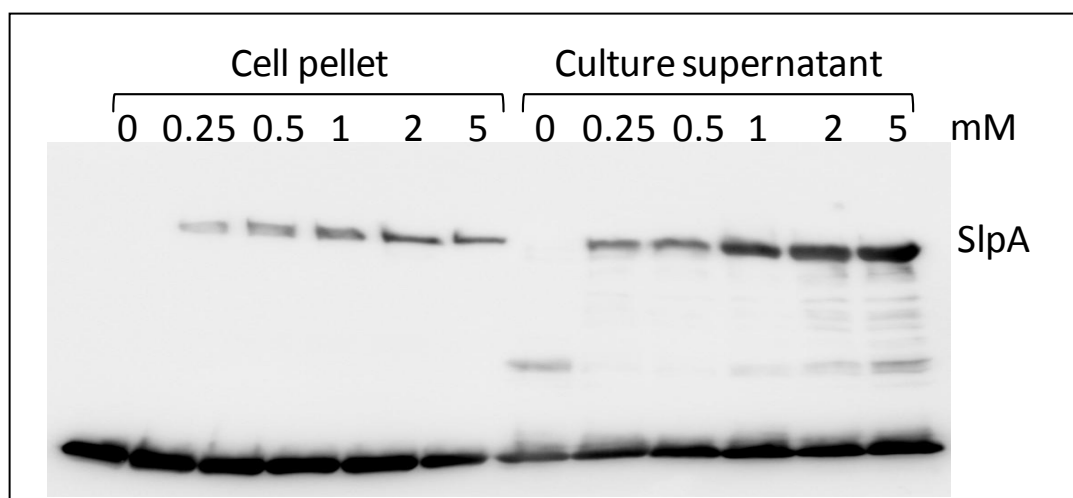


Figure 3. 9: Inhibitor activity vs. concentration, as measured by Western blot of S-layer fractions using anti-LMW SLP. *C. difficile* 630 cultures were treated with EP-LGR-NH₂ at different concentrations.

While examining the supernatant of cultures treated with inhibitor, it was found that SlpA is shed from cells. Although the level of the shedding is relative high, eventually exceeding the total SLPs retained on the cell wall, it does not reach a point where S-layer processing is completely inhibited. This striking result suggests a potency plateau in the cell and suggests that the S-layer probably has an important function for *C. difficile*. Unprocessed SlpA above a certain threshold may be not permit a tight matrix of S-layer and furthermore the production of SlpA appears to have been up-regulated to counteract the inhibitory effect. This response to the inhibitory stress is in agreement with papers that report increased SlpA expression in *C. difficile* exposed to various general stresses [393-394].

3.7 Conclusion

This chapter has detailed the development of improved inhibitors starting from a design based on E-64 through the synthesis and characterisation of activity in live cells. The synthetic route exploited during this study was adapted and optimised from literature methods to achieve high yields and enable facile access to a variety of inhibitor analogues. This can be used as a general platform for generating future generations of inhibitors.

Both efficacy and selectivity of the synthesised inhibitors was explored, showing that the warhead and the specificity element are essential in combination to obtain an inhibitory

effect. Moreover, the inhibition is highly stereoselective with respect to the warhead. The results of this work also point to the important role of the AA¹ position in contributing towards the activity, and that having only one amino acid with the warhead is sufficient for effective inhibition. This work formed the basis for a deeper structure activity relationship (SAR) study described the next chapter.

Finally, resistance to lysis and inhibition dependence on inhibitor concentration were explored. In addition to weakened resistance to lysis, shedding behaviour was found to counteract the inhibitory effect at higher concentration suggesting an important structural role of the S-layer in *C. difficile*.

Chapter 4 Inhibitor analogues and bioorthogonal ligation

To date, little work has been done involving the characterisation of proteases in *C. difficile* and only a few protease activities have been reported. Recent developments in genomics have provided the complete genome sequence of the *C. difficile* strain 630 [277] and proteomic characterisation of its spore [395]. In these previously reported studies, sequence alignment algorithms and transcriptional profiling were applied to create a better understanding of the potential role of specific gene products. However, these technologies do not reveal the functional complexity of an enzyme such as the protease involved in the SlpA cleavage process because they do not provide direct information on protease activity *in vivo*, which is typically regulated at the post-translational level. To address the functional complexities of this protease, activity based probes (ABPs) were applied.

In this chapter, the development of ABPs is described. Epoxysuccinyl inhibitors were appended with a small chemical tag, an alkyne, with the objective of identifying their target. The chemical tag was expected to have a minimal effect on binding to the protease and cell distribution and could be ligated bioorthogonally with a fluorophore to visualise labelled enzymes after gel-based separation. Furthermore, a structure activity relationship (SAR) was generated by correlating AA¹-modified epoxysuccinyl inhibitors with their inhibition potencies. The results of this work are important as they provide a starting point for protease identification (Chapter 6).

4.1 ABP design and synthesis

Having established that the newly synthesised epoxysuccinyl inhibitor possesses improved potency over E-64 and that the AA¹ position contributes significantly to its activity, we explored the SAR for this class of inhibitor with the putative protease. For the sake of easy evaluation, all inhibitors possessed the same warhead and the AA¹ position was varied with different amino acids. The warhead chosen was ethyl epoxysuccinate, eliminating the last hydrolysis step in the synthesis route (see section 3.3). Previous data suggested that there is no difference in inhibitor potency between the ethyl ester and the carboxylic acid at the N-terminus (section 3.4). The amino acids chosen for variation at the AA¹ position featured different side chain motifs for the target binding: lysine, arginine and histidine were selected due to their basicity, while glutamic acid was chosen for its acidic side chain. Hydrophobicity and aromaticity were covered by valine, isoleucine, leucine, phenylalanine and tyrosine, whilst neutral hydrophilicity was covered by threonine and serine. Furthermore, the effect of altering the structural motif of the inhibitor was addressed by proline and glycine.

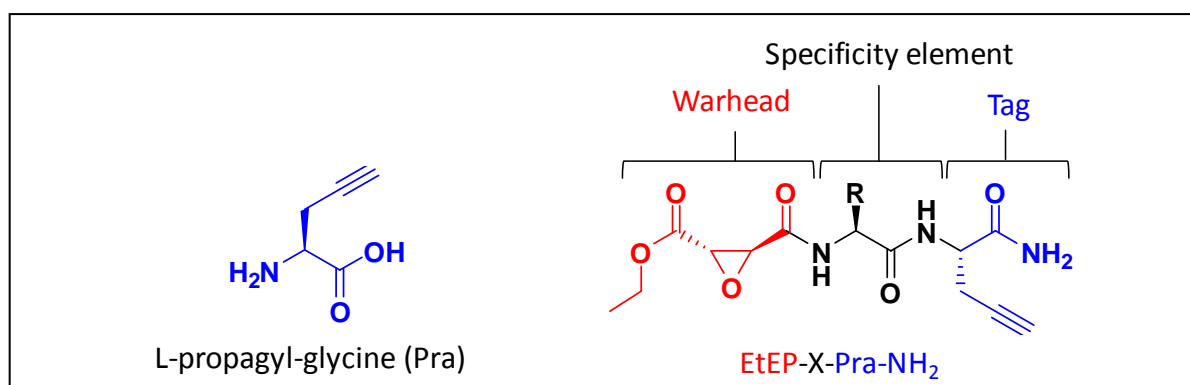
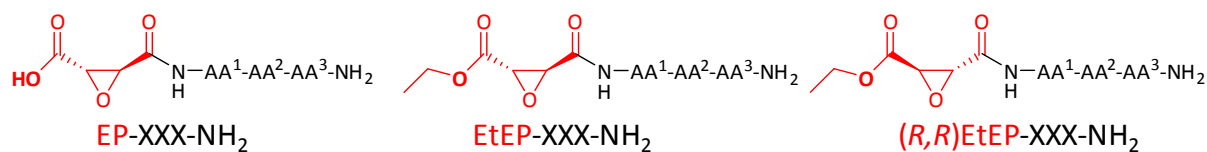


Figure 4. 1: Structures of the chemical tag (Pra) and newly generated ABPs.

As discussed in section 3.3, synthesis with only one amino acid after the warhead was problematic in some cases. Furthermore, the results of Chapter 3 also indicated that the AA² position does not affect potency. In addition, the inhibitors should be readily converted into ABPs with the objective of visualising their labelled target, and thus a chemical tag, an alkyne embedded in L-propargyl-glycine (Pra), was appended to each inhibitor at the C-terminus. The Pra tag motif is used widely in chemical biology and found to be almost completely unreactive toward biological molecules [36]. Its small size, relatively nonpolar character and inability to form significant hydrogen bonds mean it does not change the

properties of the structure of the enzyme it labels. Moreover, it can be inserted straightforwardly into the peptidic chain using SPPS, as with normal amino acids.



Nr.	WH	AA ¹	AA ²	AA ³	Abbr. Name
TD 21	EP	Leu	Arg	Pra	EP-LR-Pra-NH ₂
TD 22	EtEP	Leu	Arg	Pra	EtEP-LR-Pra-NH ₂
TD 23	EtEP	Ala	Pra	-	EtEP-A-Pra-NH ₂
TD 24	EtEP	Phe	Pra	-	EtEP-F-Pra-NH ₂
TD 25	EtEP	Ile	Pra	-	EtEP-I-Pra-NH ₂
TD 26	EtEP	Leu	Pra	-	EtEP-L-Pra-NH ₂
TD 27	EtEP	Val	Pra	-	EtEP-V-Pra-NH ₂
TD 28	EtEP	Tyr	Pra	-	EtEP-Y-Pra-NH ₂
TD 29	EtEP	Ser	Pra	-	EtEP-S-Pra-NH ₂
TD 30	EtEP	His	Pra	-	EtEP-H-Pra-NH ₂
TD 31	EtEP	Lys	Pra	-	EtEP-K-Pra-NH ₂
TD 32	EtEP	Arg	Pra	-	EtEP-R-Pra-NH ₂
TD 33	EtEP	Glu	Pra	-	EtEP-E-Pra-NH ₂
TD 34	EtEP	Gly	Pra	-	EtEP-G-Pra-NH ₂
TD 35	EtEP	Pro	Pra	-	EtEP-P-Pra-NH ₂
TD 36	EtEP	Thr	Pra	-	EtEP-T-Pra-NH ₂
TD 37	(<i>R,R</i>)EtEP	Leu	Arg	Pra	(<i>R,R</i>)EtEP-LR-Pra-NH ₂
TD 38	EtEP	Pra	-	-	EtEP-Pra-NH ₂
TD 39	EtEP	(<i>D</i>)Ala	Pra	-	EtEP-(<i>D</i>)A-Pra-NH ₂
TD 40	EtEP	(<i>D</i>)Lys	Pra	-	EtEP-(<i>D</i>)K-Pra-NH ₂
TD 41	EtEP	(<i>D</i>)Arg	Pra	-	EtEP-(<i>D</i>)R-Pra-NH ₂
TD 42	EtEP	(<i>D</i>)Tyr	Pra	-	EtEP-(<i>D</i>)Y-Pra-NH ₂

Table 4. 1: ABPs synthesised in the work described in this chapter.

To confirm the properties of a Pra-appended inhibitor, EtEP-LR-Pra was synthesised to compare with the parent compound EtEP-LR-NH₂. With regard to the stereochemistry, apart from creating an (*R,R*) configuration in the epoxide warhead, D-amino acid analogues were made of some inhibitors (A, K, R and Y) in the expanded library. Using the synthetic method described in section 3.3, the expanded library of 21 compounds, listed in **Table 4. 1**, was generated with the general structure shown in **Figure 4. 1**.

4.2 Activity of chemical tag appended analogues

To confirm that the Pra group did not affect the potency and that the ethyl ester N-terminal of the epoxide possessed identical activity to the corresponding free carboxylic acid, EtEP-LR-Pra-NH₂ and EP-LR-Pra-NH₂ were compared with the parent compounds EtEP-LR-NH₂, EP-LR-NH₂. As expected, the potencies of all these inhibitors appear identical (**Figure 4. 2**). Next, a panel of ABPs (**Table 4. 1**) was tested *in vivo* against strain 630 at 100 μM. The results from this experiment are striking (**Figure 4. 3**): both tyrosine and arginine as the AA¹ amino acid showed significantly improved potency over EtEP-LR-Pra-NH₂, which is in good agreement with previous results (section 3.4); phenylalanine also showed strong activity compared to tyrosine and arginine. Other analogues such as X = alanine, valine, serine, lysine, glutamic acid, glycine, proline and threonine remained inactive at the concentration tested, while histidine, isoleucine and leucine were slightly active.

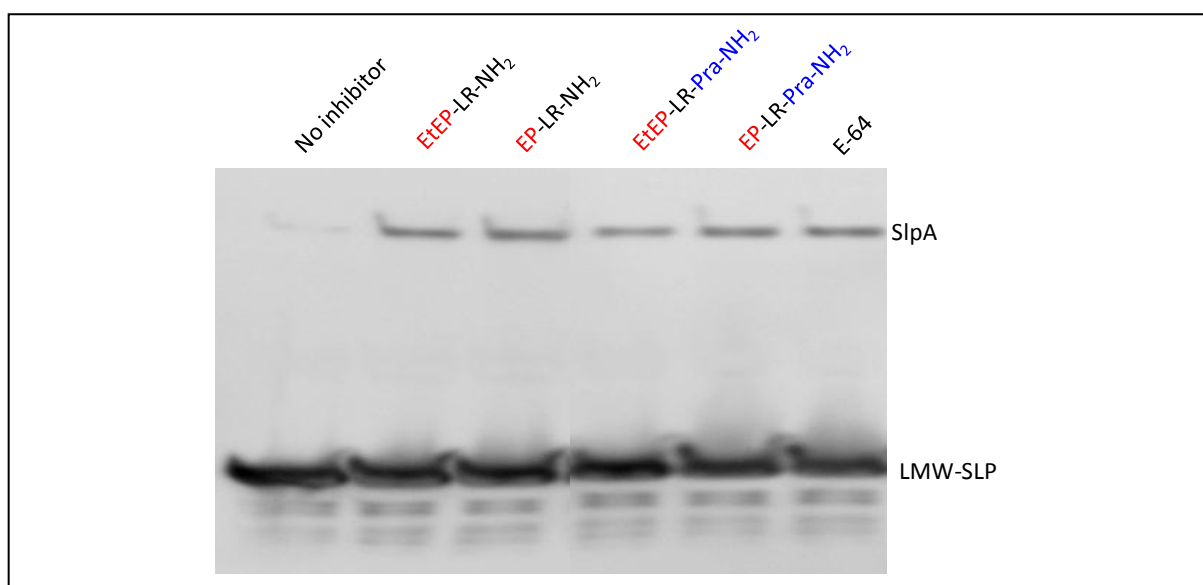


Figure 4. 2: Activity of (Et/)EP-LR-Pra-NH₂ vs (Et/)EP-LR-NH₂, as measured by Western blot against anti-LMW SLP. *C. difficile* 630 were treated overnight with 100 μM inhibitors and whole cells were boiled in loading buffer prior to SDS-PAGE analysis.

Consistent with our previous finding, the opposite stereochemistry of the warhead ((*R,R*) configuration) remained inactive. Interestingly, a switch of the stereochemistry of the amino acid from L- to D-isomers in the specificity element moiety also caused complete loss of potency. The correlation between potency and amino acid structure was divided into two different tendencies. In one of the trends, the large positively charged side chain of arginine was permitted, while in the other large hydrophobic/aromatic side chains (phenylalanine and tyrosine) were tolerated. Aliphatic hydrophobic side chains tended to reduce the potency dramatically as found in the case of isoleucine and leucine. In addition, inhibitors with an (*R,R*)-configured warhead or D-amino acid were not active suggesting a highly stereoselective mode of action on the putative protease. The inhibitor containing only the epoxide warhead and the chemical tag, EtEP-Pra-NH₂, was also inactive. Examining the conserved sequence of the SlpA cleavage locus across *C. difficile* strains (**Table 1. 2**) offers a better understanding of this interesting SAR. According to this, the AA¹ position (corresponding to the second residue upstream of the scissile bond) is either lysine or tyrosine. In the case of EtEP-Y-Pra-NH₂, the result suggested that it bound to the protease in a similar manner as the substrate, in which the epoxide warhead occupied the S1 pocket and tyrosine the S2 pocket. However, this hypothesis was not maintained for EtEP-K-Pra-NH₂. It is possible that EtEP-K-Pra-NH₂ was not active at the concentration tested, but could be active at higher concentration. Indeed, it emerged in later studies (section 4.3), that lysine possessed a higher activity threshold. **Figure 4. 3** also illustrates the correspondence of SlpA bands on the Coomassie stained gel and Western blot as reflected by their intensities.

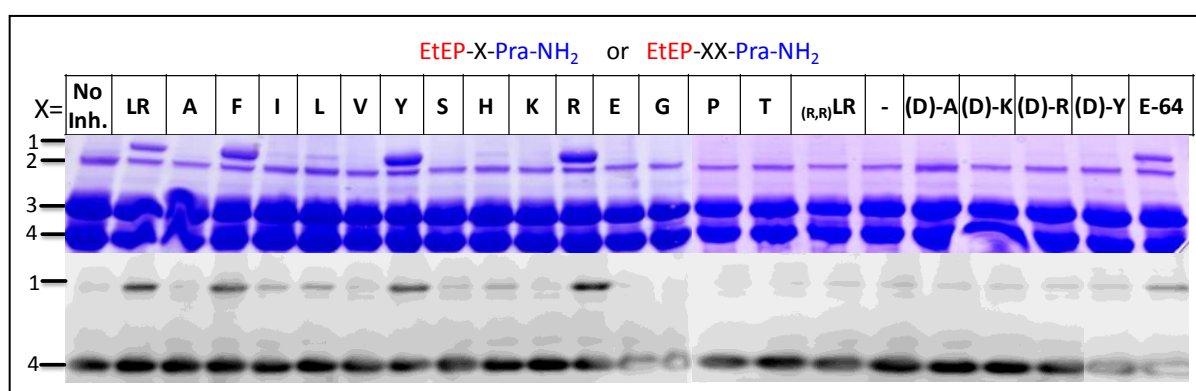


Figure 4. 3: Activity of ABPs, as measured by Coomassie stain and Western blot of S-layer fractions using anti-LMW SLP. *C. difficile* 630 cultures were treated with compounds at 100 μ M. Indicated bands are 1) SlpA; 2) Slp2; 3) HMW-SLP; 4) LMW-SLP.

4.3 Activity vs. Concentration

We next revisited the dependence of *in vivo* inhibition on inhibitor concentration. Comparing the E-64 mimic ABP, EtEP-LR-Pra-NH₂, with the most potent compound, EtEP-Y-Pra-NH₂, revealed remarkable results. The tendency of reaching a plateau at about 500 μ M on the S-layer and the shedding behaviour in the cultured supernatant of the unprocessed SlpA were found as expected (**Figure 4. 4**). However, at 5 mM EtEP-Y-Pra-NH₂ almost exclusively SlpA and almost no LMW-SLP were detected in the cultured supernatant. Furthermore, the ratio of SlpA and LMW-SLP was significantly lower when *C. difficile* was treated with EtEP-LR-Pra-NH₂.

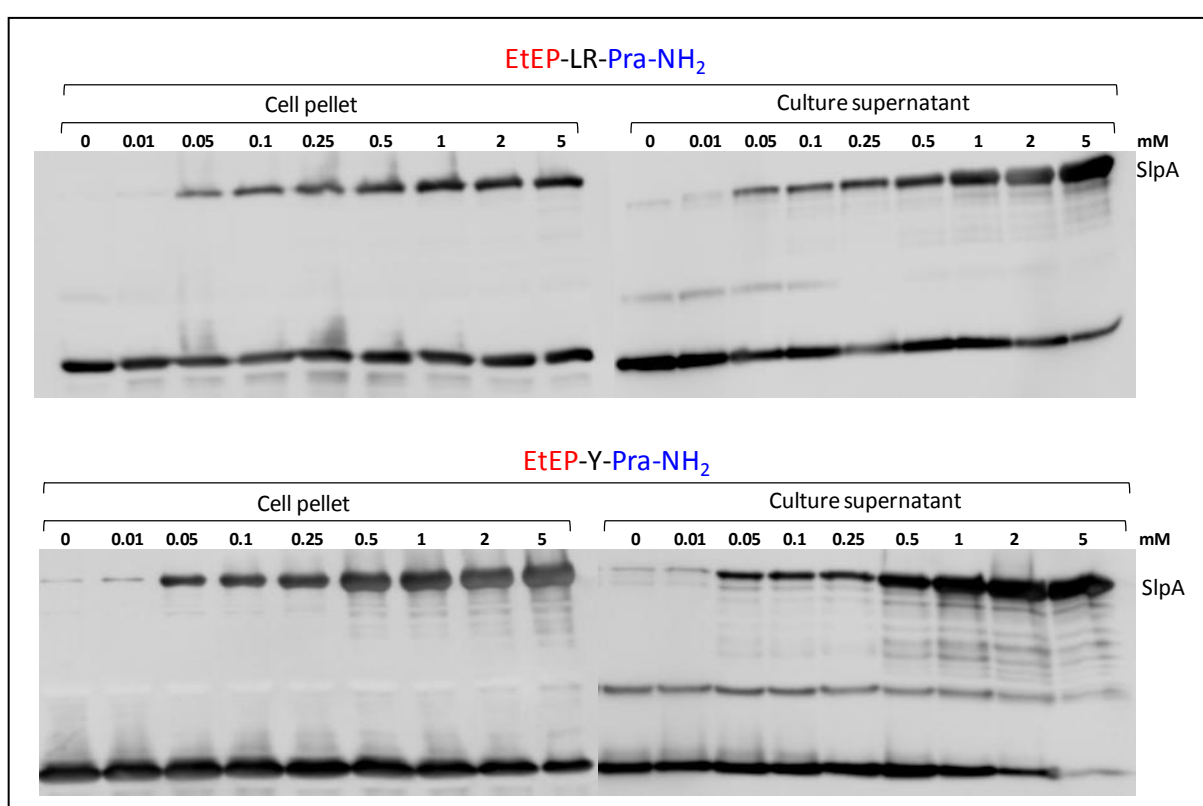


Figure 4. 4: Activity of inhibitors, as measured by Coomassie stain and Western blot of S-layer fractions using anti-LMW SLP. *C. difficile* 630 cultures were treated with compounds at different concentrations, as indicated.

To compare inhibitor potency at lower micromolar concentrations, a panel of ABPs (X=A, L, Y, K, R) was tested at various concentrations, from 10 μ M to 500 μ M. As can be seen in **Figure 4. 5**, the inhibitor containing alanine at the AA¹ position remained inactive, even at 500 μ M. Compared with the previous result shown in **Figure 4. 3**, a slight lack of reproducibility was observed as EtEP-L-NH₂ was active at 100 μ M previously but not in this experiment; this is likely a result of inconsistencies in the Western blot detection of SlpA.

Thus, moderate activity was observed for lysine and leucine at this position above 250 μM . On the other hand, substitution at the AA¹ position with arginine increased the potency to 50 μM and with tyrosine to 10 μM . Further testing of EtEP-Y-Pra-NH₂ at much lower concentrations showed that it retained some activity down to 5 μM . This indicates a significant improvement in *in vivo* activity over the initial hit E-64.

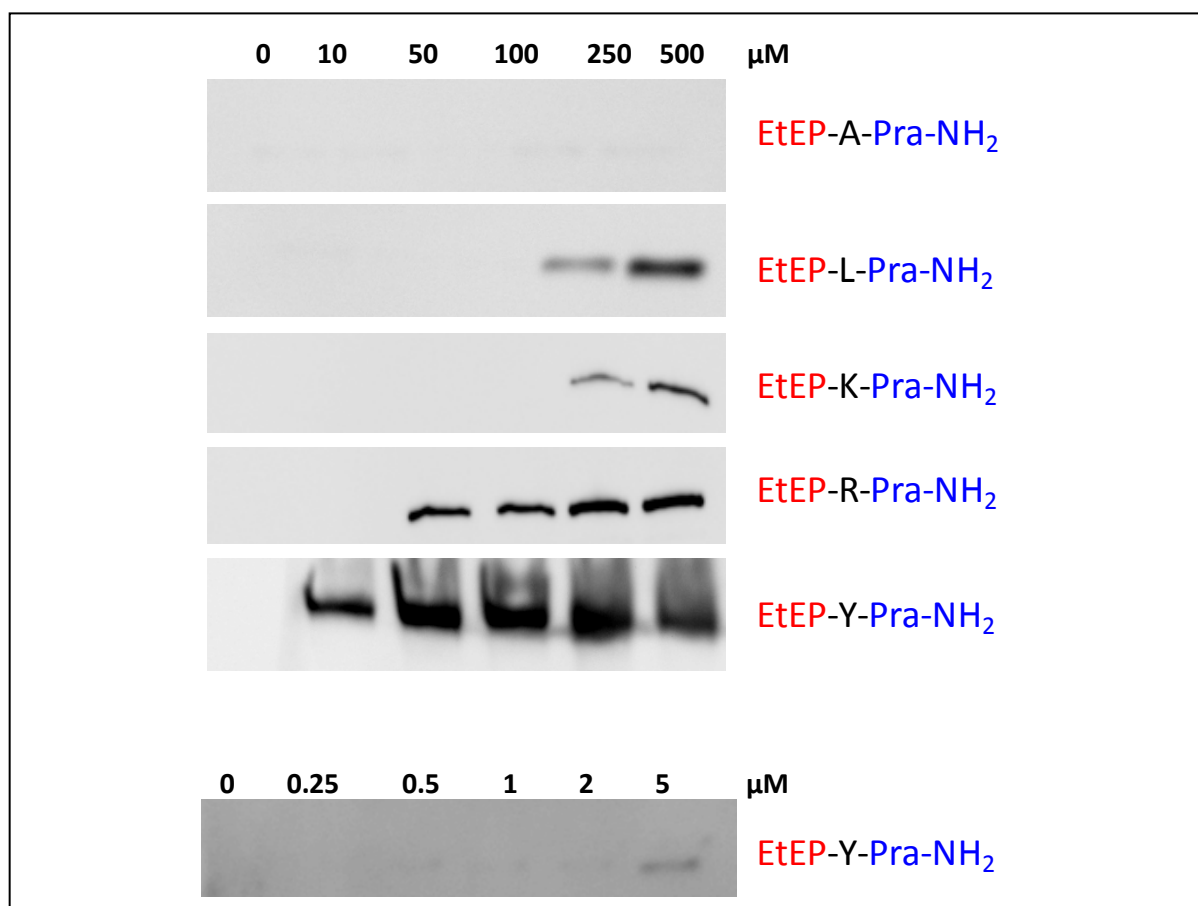


Figure 4. 5: Activity of inhibitors, as measured by Western blot using anti-LMW SLP. SlpA is shown. *C. difficile* 630 cultures were treated overnight with compounds at different concentrations, as indicated and whole cells were boiled in loading buffer prior to SDS-PAGE analysis.

4.4 Chemical labelling

4.4.1 Click ligation in labelled *C. difficile*

Click reactions are often used to ligate a labelling reagent to a bioorthogonally tagged molecule. The details of click reaction and its application in chemical biology have been described in section 1.3.2. The application of CuAAC in *C. difficile* is reported in this study using CuAAC as it was found to be more suitable for detecting alkynes in protein lysates rather than on live cells because of its cytotoxicity [94]. In addition, this method is known to

be relatively sensitive and useful for the detection of low-abundance proteins [94] and its extensive application in the Tate group provided access to strong local expertise and reagents. Moving towards the identification of the target protease involved in SlpA cleavage, although the ABP inactivates (and therefore should tag) this protease *in vivo*, labelling with capture reagent probes can take place *in vitro*.

Conditions suitable for performing CuAAC in different mammalian cell lines (e.g. HeLa and HEK cell lines) and parasite lysates (e.g. *Trypanosoma brucei*) had already been established by members of the Tate group, and the same conditions were employed in the *C. difficile* system. Once the designed ABP (**Figure 4. 1**) has been introduced into the target protease *in vivo* by virtue of its inhibition effect, the capture reagent should label the ABP after lysis using the highly selective CuAAC (**Figure 4. 6**). The capture reagent (AzK(TAMRA)G-PEG₃-Biotin) used in this study was synthesised by Megan Wright (Tate group, Imperial College London, UK) and contains three functionalities: an azide group enabling ligation to the alkyne, a biotin group permitting subsequent affinity pull down for enrichment and a TAMRA group for in-gel fluorescence detection of labelled proteins.

4.4.2 Click assay development

As mentioned above, work in the Tate group has directed considerable effort toward optimisation of click conditions for various systems. These reagents and conditions were employed to develop a protocol for click ligation in *C. difficile*. The bacteria were grown in the presence of an ABP, allowing the tagging of the target protease *in vivo*. Afterwards, the bacteria were harvested and treated with low pH glycine to strip off the S-layer (the 'S-layer fraction'). The remaining cells lacking S-layer proteins were termed the 'cell fraction'. Subsequent ligation was carried out on aliquots of these fractions. The initial click conditions applied were:

1. 10 μl of prepared protein fractions (from a 200 μl of a 5 mL culture)
2. 10 $\mu\text{M}_{\text{final}}$ of azide capture reagent (from a 1 mM stock in DMSO)
3. 2 mM_{final} CuSO_4 (from a 40 mM stock in H_2O)
4. 10 mM_{final} sodium ascorbate (from a 400 mM stock in H_2O)
5. 500 $\mu\text{M}_{\text{final}}$ BCS ligand (**Figure 4. 7**; from a 50 mM stock in H_2O)

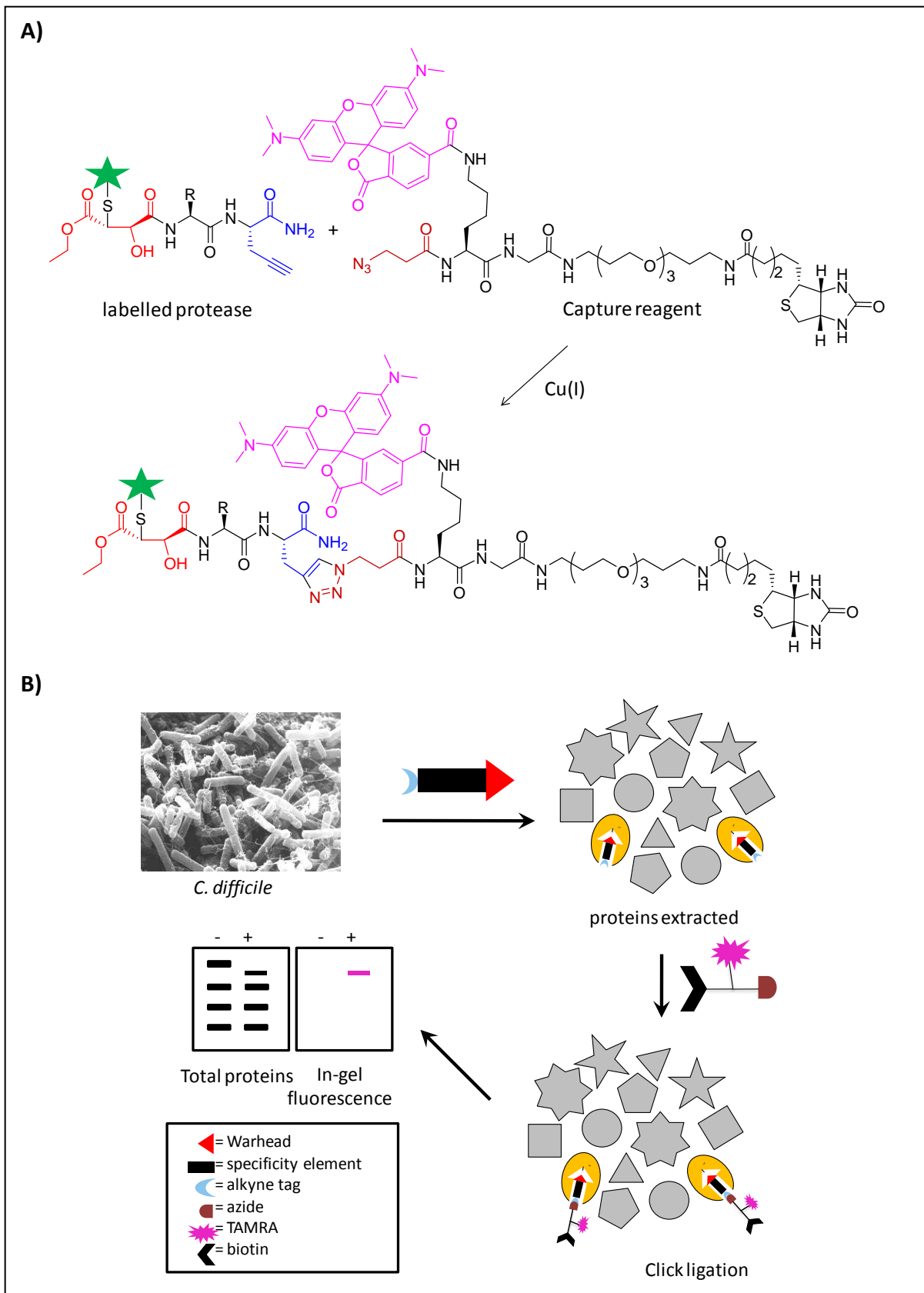


Figure 4. 6: A) Cu(I)-catalysed azide-alkyne cycloaddition reaction of the ABP with AzK(TAMRA)G-PEG₃-Biotin. B) Bioorthogonal labelling in *C. difficile*.

Components were added in the order above and mixed thoroughly at each addition step. The reaction mixture was agitated at room temperature for a further 20 minutes, before quenching with MeOH (0 °C) and storage at -80 °C overnight to promote protein precipitation. The precipitated protein was analysed by SDS-PAGE and in-gel fluorescence imaging. Unfortunately, no bands were observed, apart from excess capture reagent at the bottom of each lane (**Figure 4. 8A**). This could be due to the low concentration of the labelled protein or of the azide capture reagent, or the short reaction time. Therefore, the click reaction was further optimised as follows:

- Increasing the amount of the prepared protein in each fraction from 10 µl to 50 µl
- Increasing the azide concentration from 10 µM to 50 µM
- Increasing the reaction time from 20 minutes to 1 hour

The results remained disappointing as no labelled bands were observed. The click ligation proved to be challenging for *C. difficile* and required further optimisation, perhaps indicating the complexity of the extracted proteome which demands more robust protocols and reagents.

Examining and comparing each component of the click reaction mixture to examples of click ligation in the literature led to the realisation that the prepared protein fractions (S-layer and cell fractions) contained relatively high concentrations of salts (glycine in the S-layer fraction and PBS in the cell fraction). These salts could interfere with the click reaction. Glycine could act as a weak ligand for Cu(I) and outcompete the BCS ligand when it is in large excess. Consequently, the stability of Cu(I) might be sufficient to catalyse cyclisation. Therefore, buffer replacement by ultrafiltration in Ultrafree 0.5 mL centrifugal filter devices (Millipore) with a 10 kDa molecular weight cut-off limit was proposed to try and eliminate the problem.

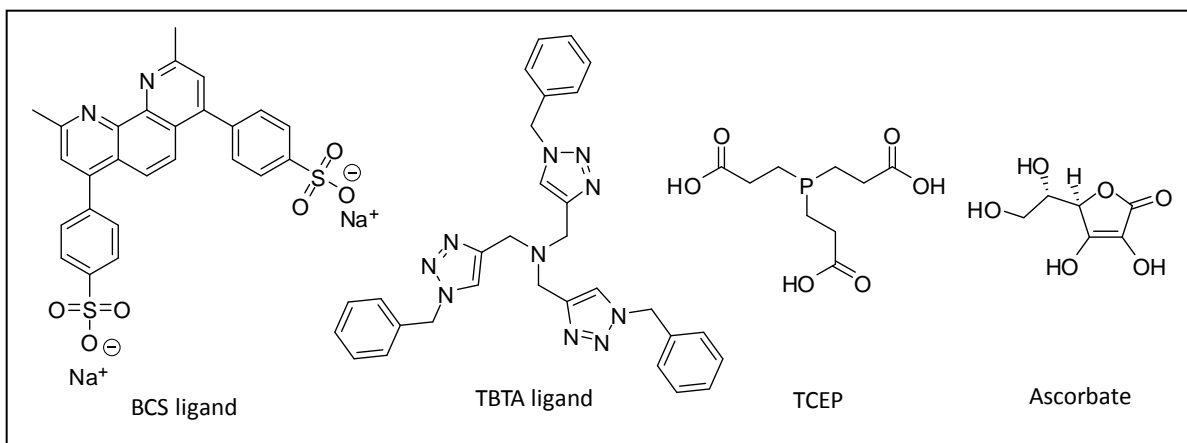


Figure 4. 7: Ligands and reducing agents used in this study.

The reducing agent used to generate the copper (I) catalyst was one of the next concerns. Although only small differences between the effect of ascorbate and TCEP on the ligation in mammalian cell lines and parasite lysates were found [396], sodium ascorbate was replaced by TCEP to avoid addition of further salt to the reaction mixture. Furthermore, the BCS ligand was switched to TBTA. Both ligands were known to stabilise the copper (I) oxidation state, however TBTA better stabilises Cu(I) against oxidation and disproportionation and reduces the background signals better than BCS [397-398]. The optimised click ligation conditions were:

1. 50 μl of prepared protein fractions (from a 200 μl of a 5 mL culture)
2. 50 $\mu\text{M}_{\text{final}}$ of azide capture reagent (from a 1 mM stock in DMSO)
3. 1 mM_{final} CuSO_4 (from a 40 mM stock in H_2O)
4. 1 mM_{final} TCEP (from a 400 mM stock in H_2O)
5. 100 $\mu\text{M}_{\text{final}}$ TBTA ligand (**Figure 4. 7**; from a 50 mM stock in H_2O)
6. 1 hour reaction time at room temperature.

Figure 4. 8B shows the results for the click ligation carried out using these optimised conditions. A positive control (lane L) was added, which was alkynyl myristate-labelled HeLa cell lysate containing 50 μg protein prepared by Megan Wright, a student in the Tate group (Imperial College London, UK). A few bands were observed in lane DS and the control lane (lane L) after a prolonged exposure time. However, the signals were relatively weak and lanes S to C remained blank. Increasing the protein amount from 50 μl to 100 μl led to a striking result (**Figure 4. 8C**): clear labelled bands were observed in all lanes. Comparing lane

DS (S-layer fraction after buffer replacement) and lane S (S-layer fraction without any buffer replacing process) revealed the important role of the buffer replacement step.

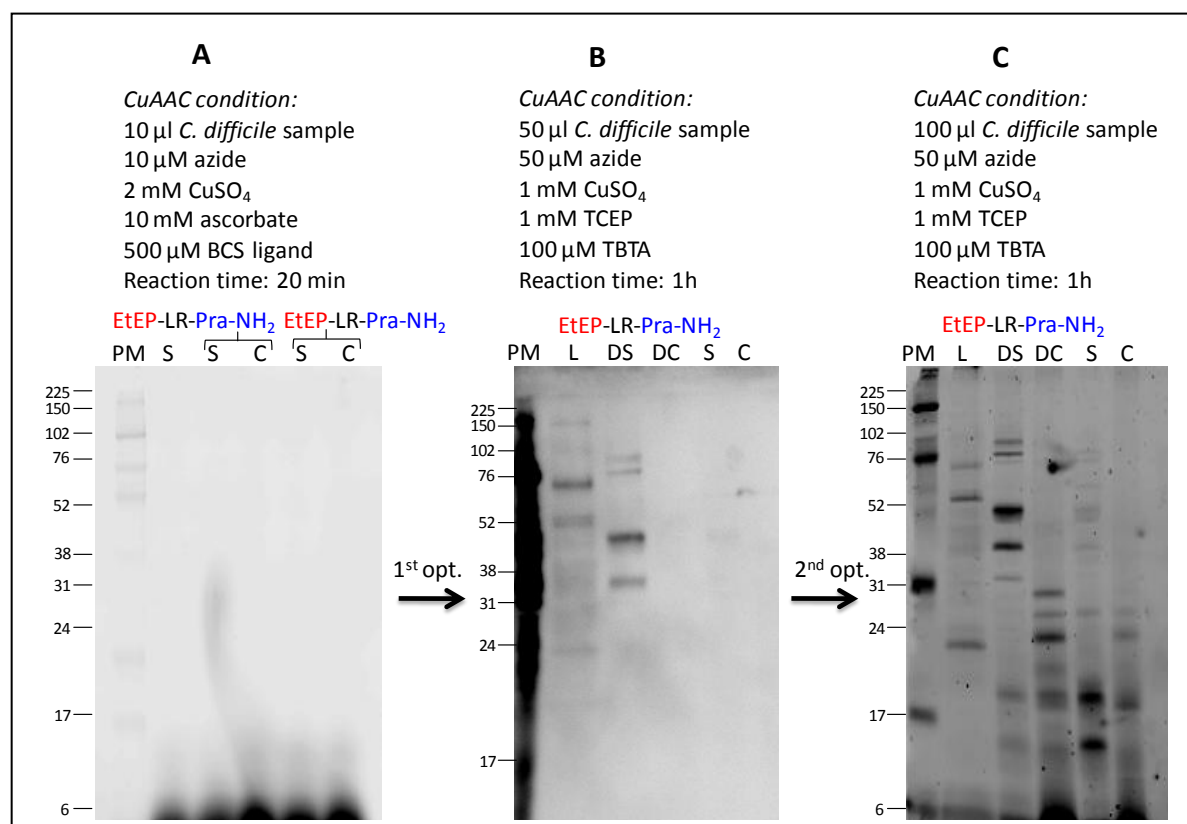


Figure 4. 8: Click optimisation, including the optimised conditions for each stage. Gels visualised by in-gel fluorescence show ligation of protein extract after treatment with 100 μ M of indicated inhibitor. PM = ECL Plex™ Fluorescent Rainbow Markers (kDa), S = S-layer fraction, C = cell fractions, DS = buffer replaced S-layer fraction, DC = buffer replaced cell fraction, L = alkynyl labelled Hella cell lysates as positive control (50 μ g protein lysates, prepared by Megan Wright).

4.4.3 Ligation results

In order to demonstrate the correlation of protein labelling with S-layer inhibition activity, a panel of ABPs was tested using these optimised conditions. The result of this experiment is displayed in **Figure 4. 9**. Generally, labelling was observed in all samples with a more complex pattern than previously found in the last optimisation step (**Figure 4. 8C**). Although the intensities of labelled bands varied strongly between samples, the pattern of these bands was very similar. Comparing bands in no inhibitor lane with other lanes indicated that most of these bands were background labelling. It was difficult to extract meaningful information without further optimisation.

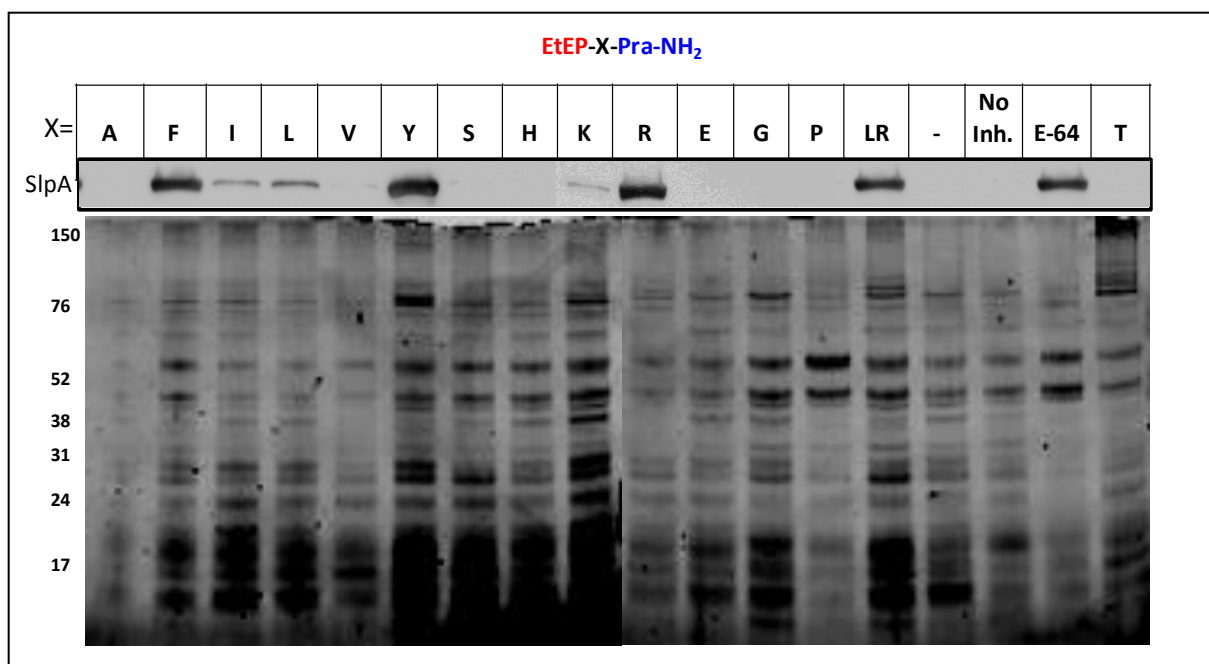


Figure 4. 9: In-gel fluorescence image showing labelling of S-layer fractions across the ABP series (100 μ M) after click ligation to AzK(TAMRA)G-PEG₃-Biotin. SlpA bands of Western blot against LMW-SLP have been added (at top) to facilitate comparison with activity. In the activity assay, cells were treated with ABPs at 250 μ M.

From a practical aspect, there was still scope to improve the click ligation because problems including unpredictable loss of sample resulted in inconsistently loaded protein in each lane (**Figure 4. 9**, lane X=A) and high background labelling made evaluation difficult. Ultrafree centrifugal filter devices are useful to reduce the salt concentration, but have one drawback: a certain amount of protein sticks to the filter membrane, leading to occasional complete or partial loss of samples. In addition, precipitation of protein at -80 °C in MeOH produces fluffy protein pellets that are easy to lose during solvent removal and wash steps. To circumvent the discussed problems, another precipitation method (MeOH/CHCl₃) was applied. The precipitated protein concentrated at the phase interface and was easy to handle.

To reduce the degree of background labelling a pull-down step was added to the protocol prior to SDS-PAGE analysis. ‘Pull-down’ is the term used for a small scale affinity purification technique that utilises the strong affinity binding of two different binding partners, biotin and streptavidin in this case. Commercially-available streptavidin immobilised onto solid support was used (‘Streptavidin Dynabeads’). After resuspending the precipitated protein in 30 mM Tris, 30 μ L Streptavidin was added and agitated for one hour at RT to allow the capture (‘pull-down’) of biotin labelled protein onto the beads. With this method, labelled

protein was enriched and physically separated from unlabelled protein, however there was also nonspecific binding causing a certain amount of undesired co-purification. Stringent washing steps (three times of 50 μ l 1%SDS in 30 mM Tris) were added to elute these non-specific binders, but does not co-elute the biotin labelled protein due to its specific and very strong affinity bond. Finally, the beads were heated at 100° C in loading buffer to break the streptavidin-biotin bond, eluting labelled protein from the beads.

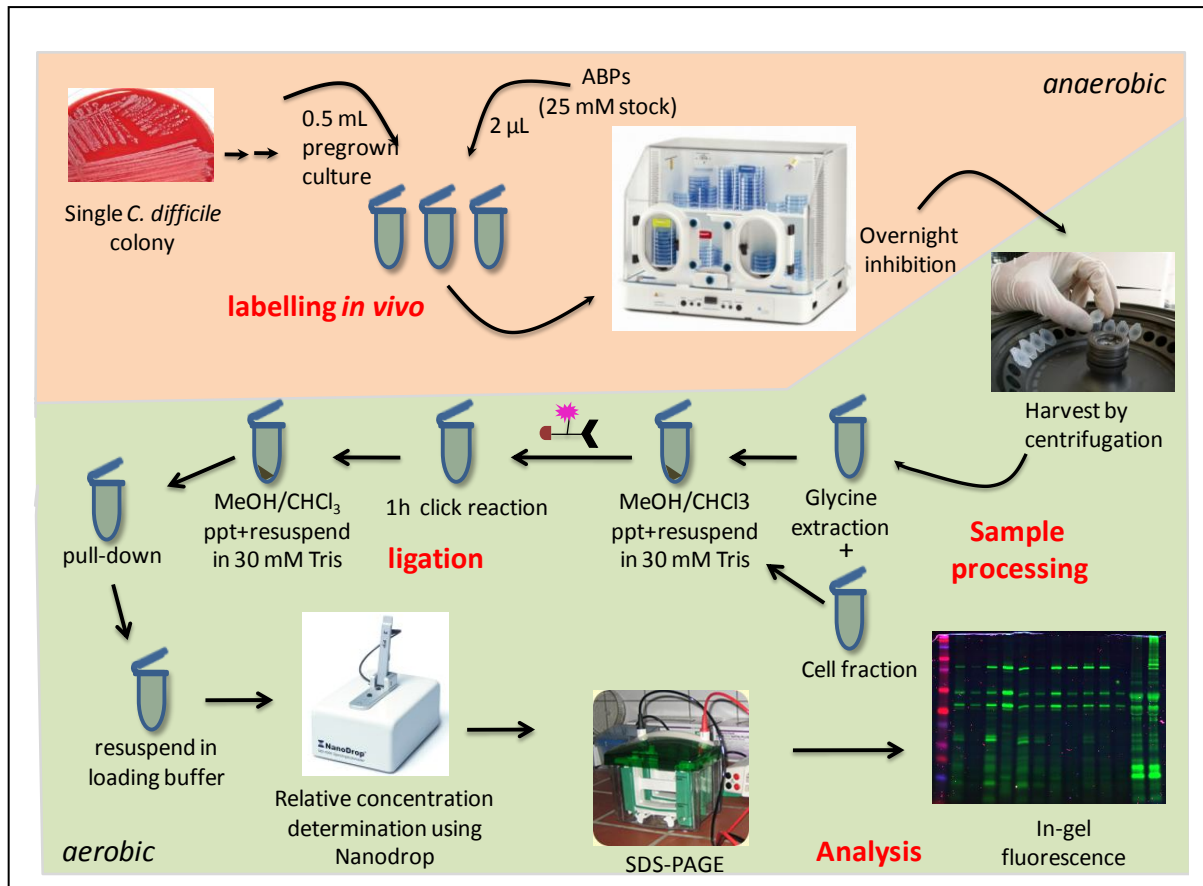


Figure 4. 10: Optimised click ligation protocol for *C. difficile* system used in this study.

To secure relatively equal amounts of protein for each sample in each lane the relative concentration of eluted proteins in loading buffer was measured using a spectrophotometer and normalised prior loading on the gel. A NanoDrop® ND-1000 UV-Vis Spectrophotometer was used as it is designed for quantifying accurately and reproducibly small volumes of sample (0.5-2 μ l), perfect for limited sample volumes such as those used here. The instrument was blanked using loading buffer without any protein, and 1 μ l of sample was measured on the analyser at 230 nm. The final optimised protocol is displayed in **Figure 4.**

10.

The results of the final optimised protocol were clearer (**Figure 4. 11**). In general, the high background was eliminated and the gel looked much cleaner; most of the background labelling remained in the supernatant. However, some of the abundant HMW and LMW-SLPs were retained in the pull-down and showed a very strong labelled band on the S-layer pull-down gel. Intriguingly, strongly labelled bands at approximately 76 kDa were seen in direct correlation with inhibition activity in some cases. However, bands at the same size were also observed with some inactive ABPs including all D- amino acid analogues. Interestingly, EtEP-R-Pra-NH₂, EtEP-Y-Pra-NH₂ and EtEP-F-Pra-NH₂ labelled a second band slightly higher than 76 kDa. Further labelled bands below 38 kDa appeared to be present in most samples but with different intensity, indicating that they were also background labelling. As the cell lysed differently after treatment with different inhibitors due to inhibition of SlpA processing, the level of this background labelling might be depend on the total protein in cell lysate. Cell fractions revealed only background labelling, reinforcing the previous suggestion that the target protease(s) is/are located in the S-layer, and not inside the cell (section 3.5).

4.5 Conclusion

The development of alkyne-tagged ABPs has been described in this chapter. Having investigated more than 40 inhibitors/ABPs (Chapter 3 and 4) covering most aspects of known E-64 SARs and of the SlpA cleavage site, a general SAR for epoxysuccinyl inhibitors was compiled. This SAR is divided in three parts examining the importance of each unit of the designed inhibitors/ABPs, namely the warhead, the specificity element and the appended tag. In the inhibition of proteases involved in SlpA cleavage, the warhead was an essential unit in contribution towards activity, as replacing the epoxysuccinyl warhead with an acetyl group caused a complete loss of activity. Furthermore, the stereochemistry of the epoxide also played an critical role in inhibition. The *S,S*- configuration as found in E-64 was active, while the *R,R*-configuration was not. Although ethyl ester and carboxylic acid functions at the N-terminus of the inhibitor differ in their chemical structure and properties, they possessed similar inhibitor potency. The epoxysuccinyl warhead was indeed essential for inhibition, but not on its own. It required the attachment of the specificity element, with even a single specific amino acid resulting in activity, suggesting equal importance of both these parts of the inhibitors.

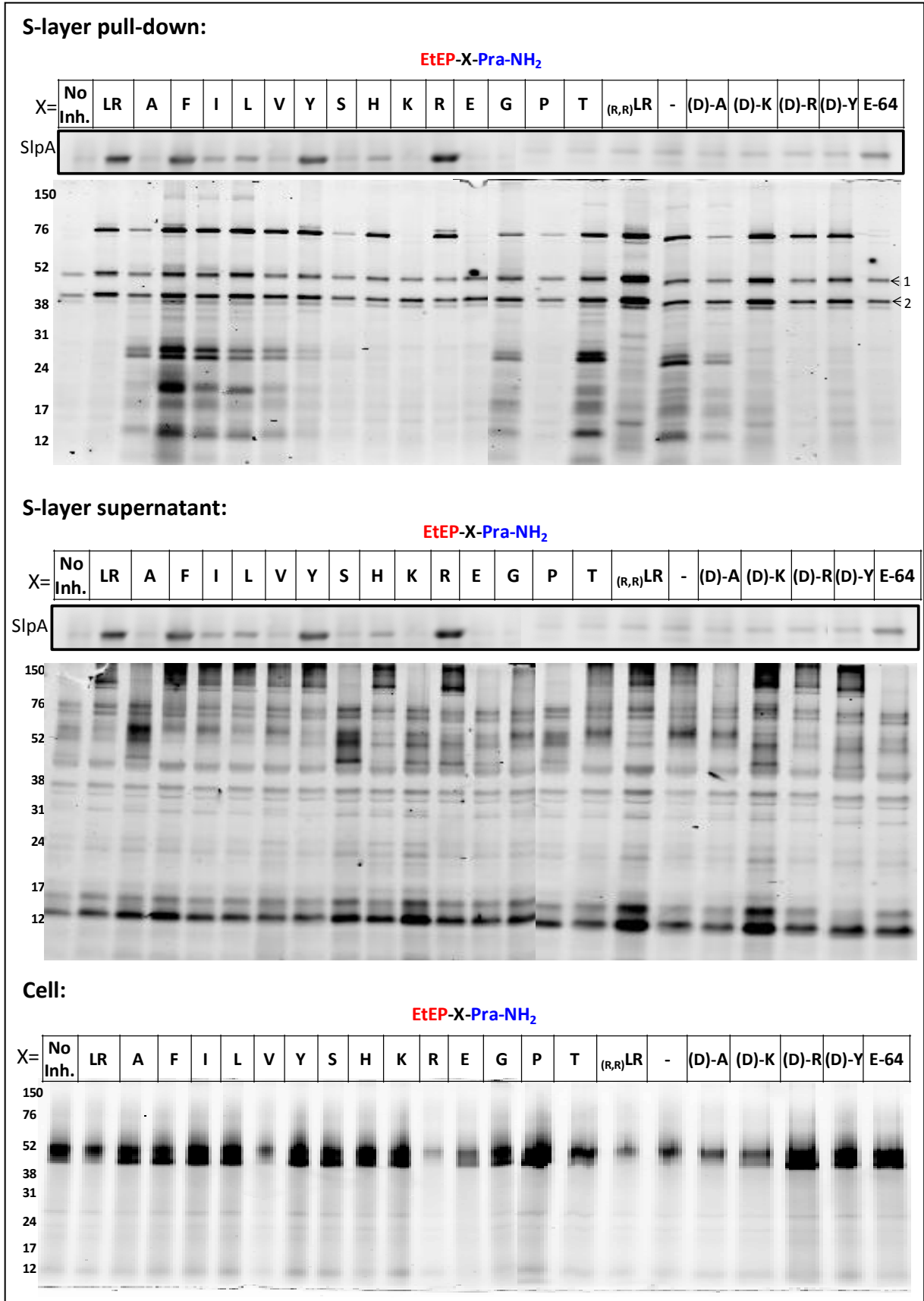


Figure 4. 11: Labelling across ABP series seen by in-gel fluorescence after CuAAC ligation to AzK(TAMRA)G-PEG₃-Biotin. Three different fractions are shown: S-layer pull-down, S-layer supernatant and cell pull-down. Indicated band in S-layer pull-down gel: 1) HMW-SLP; 2) LMW-SLP.

The specificity element was designed to mimic the cleavage locus of SlpA. It was found that the first amino acid after the warhead (AA¹ in the inhibitor nomenclature adopted here) had a significant influence on inhibitor potency. Correlation with activity was divided in two groups. In one group a positively charged side chain was permitted, though it appeared that a larger or more basic group was preferred over smaller group (e.g.: arginine better than lysine, histidine). This effect was observed more clearly in the concentration dependence experiment (**Figure 4. 5**). In the other group, large hydrophobic/aromatic side chains (e.g.: phenylalanine, tyrosine) were accepted. In this group, aliphatic hydrophobic side chains seemed to be disfavoured as they significantly reduced the potency (e.g.: leucine, isoleucine).

Inhibitors bearing a small side chain (alanine), acidic side chain (glutamic acid) or neutral hydrophilic side chain (threonine, serine) as well as structure-altering side chains (glycine, proline) at AA¹ remained inactive. Stereoselectivity was also observed within the specificity element as L-amino acids were active and D-amino acids not.

Examining the importance of the second amino acid after the warhead (AA²) brought more questions rather than answers. If AA¹ was moderately active and kept consistent while AA² was varied, the inhibition activity could be barely distinguished (e.g.: EP-KT-NH₂, EP-KA-NH₂, EP-KPra-NH₂), suggesting that the AA² position does not affect the activity. Indeed, further work in the Tate and Fairweather research groups (Vladimir Turek, Imperial College 2008, unpublished results) building on the data reported here demonstrated that a wide range of amino acids at AA² are tolerated when AA¹ is Tyr or Lys. However, this trend did not appear to extend to leucine. As shown above, leucine possessed a relatively poor activity on its own as AA¹, but in combination with arginine ((Et/)EP-LR-NH₂) at AA² the potency increased significantly, suggesting that, on the contrary, the nature of AA² can have an effect on inhibition. Further extension of the length of the specificity element to AA³ or AA⁴ did not result in substantial additional activity within the limited set of inhibitors investigated here. The chemical tag appended to the C-terminus of the inhibitor did not appear to interfere with inhibition. This was further confirmed by the ABPs described in the next chapter.

Taken together, these data suggest that the epoxysuccinyl inhibitors inactivate a papain class cysteine protease in a canonical manner to that found for E-64, in which the epoxide

warhead probably binds to the S1 pocket and the first amino acid after the warhead (corresponding to the second residue upstream of the scissile bond in the cleavage locus) binds to the S2 pocket. This being the case, binding to the S3 pocket was uncertain, as the effects of AA² on inhibition were ambiguous.

The work of Chapter 4 also describes the optimisation of two-step labelling in the *C. difficile* system. The practical challenge was overcome and very distinct labelling of proteins was seen. Despite the still ambiguous correlation between these labelling with the activity of the corresponding ABP, the labelled proteins could be identified (Chapter 6), and later verified as including the target protease (Chapter 7). Furthermore, the results of protein labelling confirmed that the target protease(s) is/are located in the S-layer and should therefore be readily accessible in the medium. This observation led to the development of directly labelled ABPs, as described in the next chapter (Chapter 5).

Chapter 5 Directly labelled activity-based probes

Verification that the putative proteases are located in the S-layer of *C. difficile* eliminates any cell penetration issues associated with the inhibitors/ABPs, as they can readily reach their targets in the medium. It was then speculated that switching from two-stage ABPs to directly-labelled ABPs would simplify the target identification process due to abolishing the challenged click ligation and its requirements, as well as its interference in the labelling step. This chapter is divided in two parts: the first part explores the work with ABPs bearing a direct affinity label (a biotin tag), while the second part addresses ABPs possessing single and dual direct fluorescent labels (a fluorescent dye and a biotin tag). The results of this chapter served as foundation for the target identification process, discussed in the next chapter.

5.1 Biotinylated ABPs

Biotin is the most widely used reporter unit (see 1.3.1.3). In this study, biotin was attached to the C-terminus of inhibitors via a hydrophilic PEG linker, which improves solubility of the ABP and reduces the chance of interference between warhead and affinity tag. The general structure of these ABPs is displayed in **Figure 5. 1**.

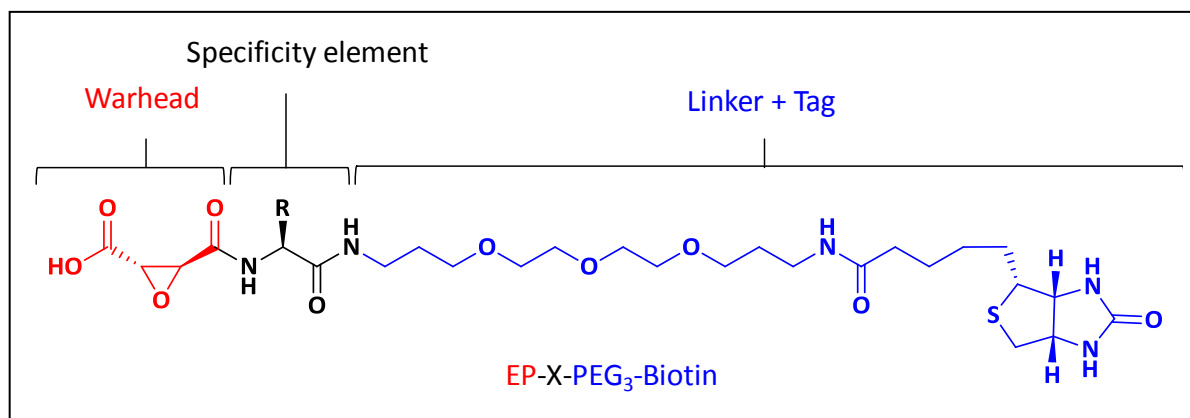
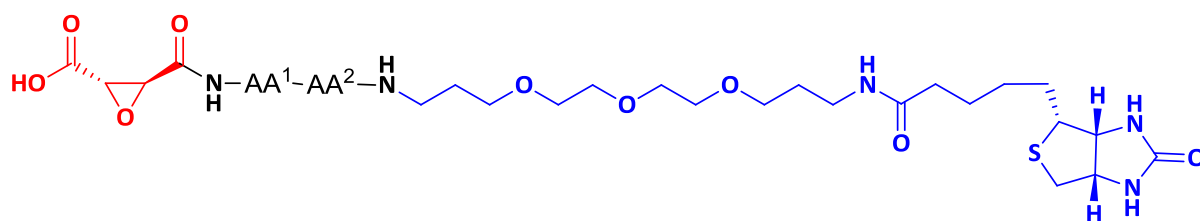


Figure 5. 1: General structure of biotinylated ABPs.

To generate a new series of PEG-biotinylated ABPs, SPPS using the method described in section 3.3 was performed using Biotin-Novatag resin instead of Rink amide resin, permitting the PEG-Biotin to be introduced at the outset of the synthesis. With the aim of confirming the correlation of activity and labelled effect found previously, a similar series of ABPs to those discussed in Chapter 4 were synthesised, and are listed in **Table 5. 1**.



EP-AA¹-AA²-PEG₃-Biotin

No.	WH	AA ¹	AA ²	Abbr. Name
TD 43	EP	Leu	Arg	EP-LR-PEG ₃ -Biotin
TD 44	EtEP	Leu	Arg	EtEP-LR-PEG ₃ -Biotin
TD 45	(<i>R,R</i>)EtEP	Leu	Arg	(<i>R,R</i>)EtEP-LR-PEG ₃ -Biotin
TD 46	EP	Ala	-	EP-A-PEG ₃ -Biotin
TD 47	EP	Phe	-	EP-F-PEG ₃ -Biotin
TD 48	EP	Ile	-	EP-I-PEG ₃ -Biotin
TD 49	EP	Leu	-	EP-L-PEG ₃ -Biotin
TD 50	EP	Val	-	EP-V-PEG ₃ -Biotin
TD 51	EP	Tyr	-	EP-Y-PEG ₃ -Biotin
TD 52	EP	Ser	-	EP-S-PEG ₃ -Biotin
TD 53	EP	His	-	EP-H-PEG ₃ -Biotin
TD 54	EP	Lys	-	EP-K-PEG ₃ -Biotin
TD 55	EP	Arg	-	EP-R-PEG ₃ -Biotin
TD 56	EP	Glu	-	EP-E-PEG ₃ -Biotin
TD 57	EP	Gly	-	EP-G-PEG ₃ -Biotin
TD 58	EP	Pro	-	EP-P-PEG ₃ -Biotin
TD 59	EP	Thr	-	EP-T-PEG ₃ -Biotin
TD 60	EP	-	-	EP-PEG ₃ -Biotin

Table 5. 1: Biotinylated ABPs synthesised in the work described in this chapter.

5.1.1 Activity of biotinylated analogues

Similar to previous work in Chapter 4, a comparison of the parent compounds (Et/)EP-LR-NH₂ with different generations of ABPs (the directly-labelled EtEP-LR-PEG₃-Biotin and the alkyne-tagged EtEP-LR-Pra-NH₂) was performed. As shown in **Figure 5. 2**, these compounds possessed essentially identical potency (at 100 μM) to the parent inhibitors, strongly

suggesting that the tag does not change or interfere with the inhibition activity. This may be due to the fact that the target proteases are located in the S-layer and are readily accessible for inhibitors/ABPs in the medium.

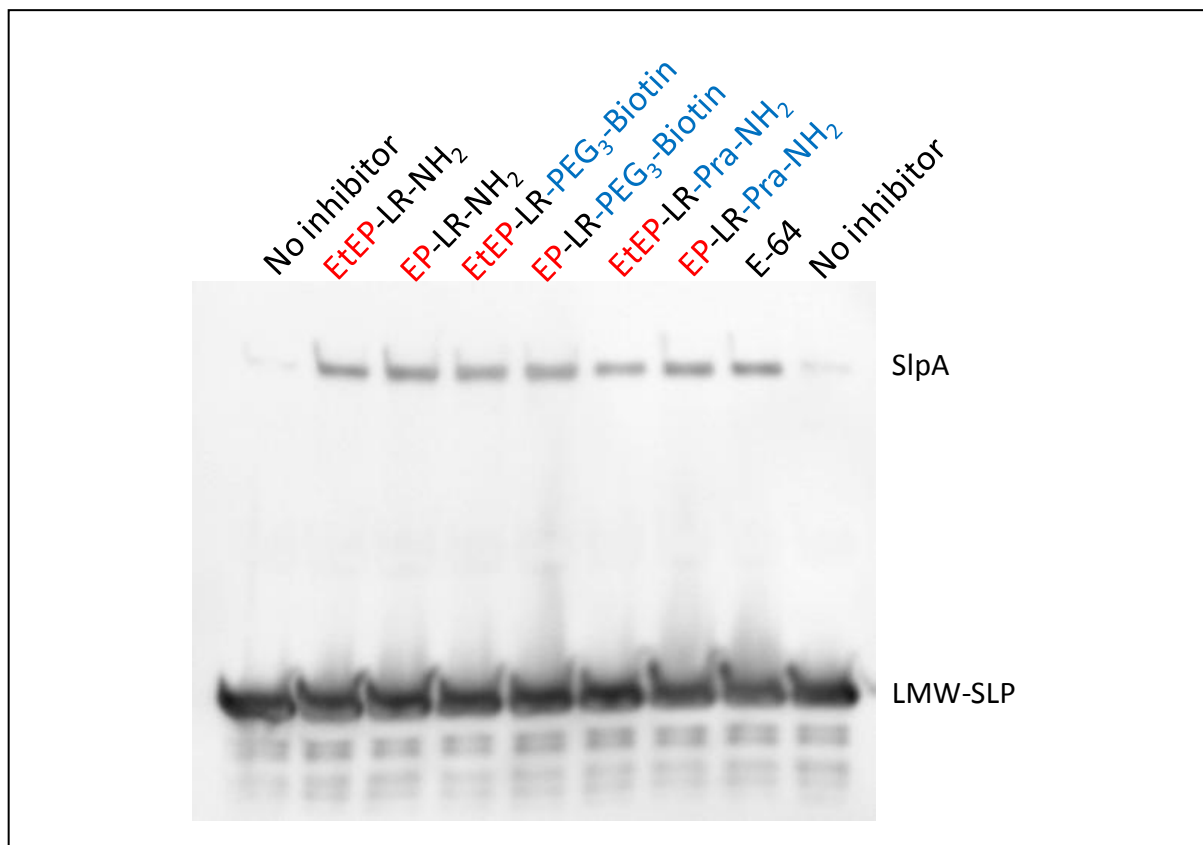


Figure 5. 2: Activity of EP-LR-PEG₃-Biotin and EP-LR-Pra-NH₂ vs. EP-LR-NH₂, as measured by Western blot against LMW SLP. *C. difficile* 630 cultures were treated with compounds at 100 μ M.

Testing selected ABPs against strain 630 *in vivo* at 100 μ M produced similar results as those found in Chapter 4 (**Figure 5. 3**). The tendency for the tolerance of two different characteristic amino acids located in the AA¹ position with improved potency was also observed. Tyrosine (representative of a large aromatic side chain group) or arginine (possessing a large, positively-charged side chain) at the AA¹ position retained higher potency, other ABPs remained moderate or less potent. The appearance of a weak SlpA band in most lanes was also observed, as previously discussed (section 2.1.3).

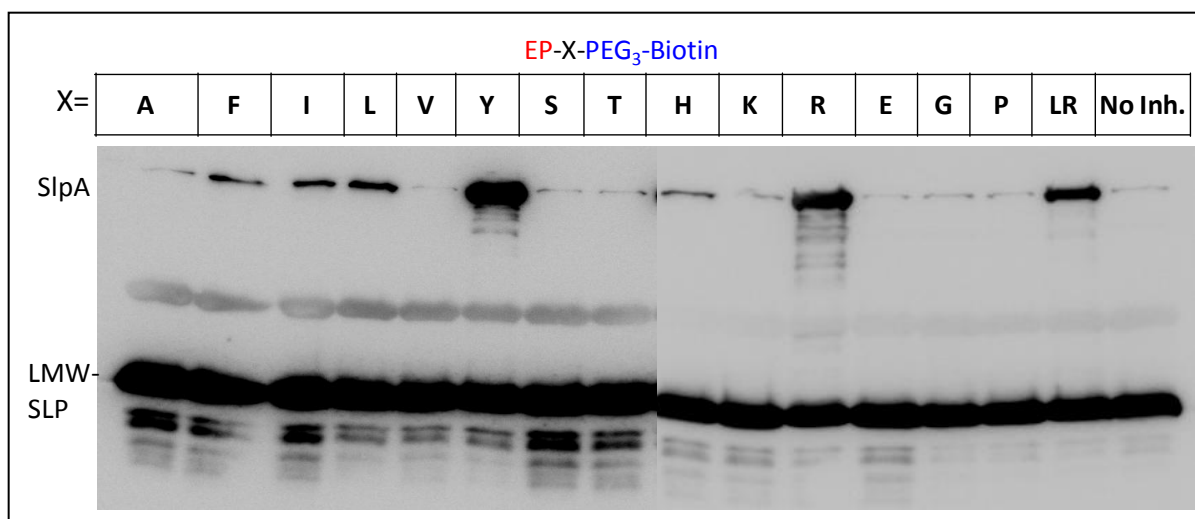


Figure 5. 3: Activity of biotinylated ABPs, as measured by Western blot of whole cell against LMW SLP. *C. difficile* 630 cultures were treated with compounds at 100 μ M.

5.1.2 Labelling effect of biotinylated ABPs

To investigate the utility of directly-labelled ABPs, *C. difficile* overnight cultures were grown in the presence of the panel of ABPs at 100 μ M (**Table 5. 1**) and harvested before the whole cell was boiled in loading buffer. SDS-PAGE analysis (on glycine-tris gel) and Western blotting against NeutrAvidin revealed the interesting results presented in **Figure 5. 4**.

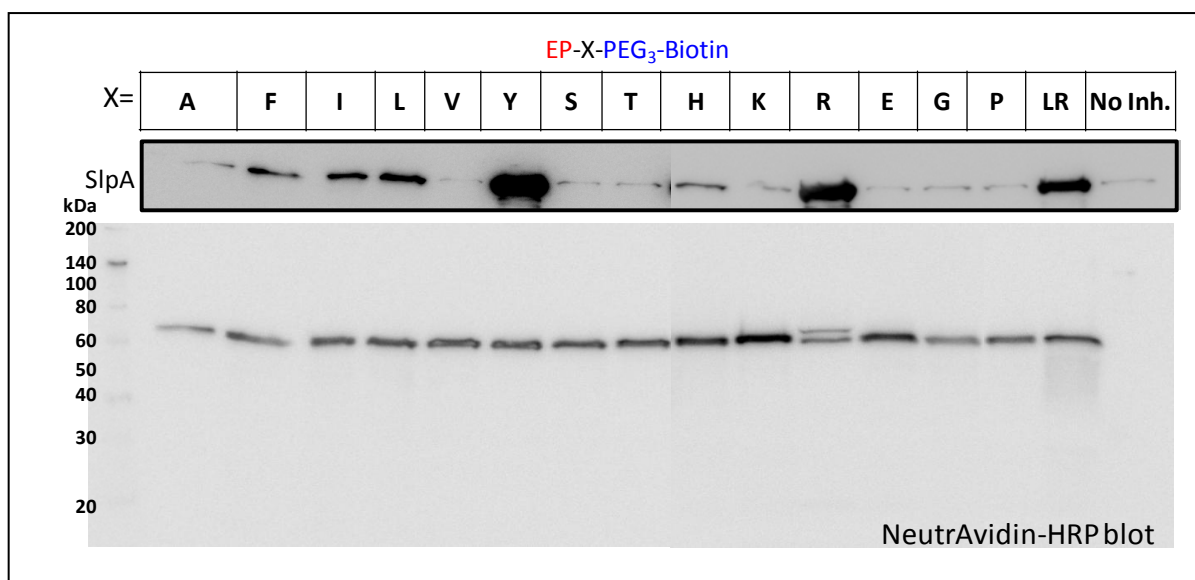


Figure 5. 4: Labelling effect of biotinylated ABPs, as measured by Western blot of whole cell against NeutrAvidin-HRP. *C. difficile* 630 cultures were treated with compounds at 100 μ M. SlpA bands of ABPs from previous section (Figure 5. 3) are shown to aid the evaluation of correlation between activity and labelling.

Intriguingly, the labelling of these probes appeared very specific. A single band, at apparently 70 kDa, was observed for all probes, except EP-R-PEG₃-Biotin that labelled two

bands, one of similar size to the other ABPs and the other located slightly higher on the gel. The band at 70 kDa (apparent m.w.) on the NeutrAvidin blot is probably the same band as was observed with the two-step labelling method at approximately 76 kDa (**Figure 4. 11**). The small difference in size is the result of differing behaviour of the protein on the different gel types (glycine-tris gel and Bis-tris gel (see section 9.1.3) and the less accurate estimation of protein size in the higher molecular weight region by using different protein markers (fluorescent and biotinylated protein markers). The labelling of proteins of apparently the same size with both the inactive and active inhibitor is in agreement with the results of the two-step labelling and is likely to be the same protein labelled. However, all labelled bands were equally intense and consistent throughout the whole series of ABPs, regardless of whether these ABPs are active or not in inhibition of SlpA processing. It was hypothesised that there may be two possibilities explaining this phenomenon. The first possibility is that both active and inactive ABPs label one single protease. However, in this case the inactive ABP must presumably bind to the protease at a site other than the active site (**Figure 5. 5A**). The second possibility is that this type of inhibitor targets at least two proteases, in which case only the active ABPs label the protease involved in the internal SlpA cleavage, whilst the inactive ABPs (and possibly also the active ABPs) label some other protease of *C. difficile* of similar size, running at the same position on the gel after SDS-PAGE analysis (**Figure 5. 5B**). Furthermore, these proteases could be closely related and belong to the same protease family since they are targeted by the same type of inhibitor, but tolerate different binding motifs and so would respond to different specificity elements in the ABPs. The identification of these proteins would reveal which hypothesis is more likely in *C. difficile*.

To examine the minimal concentration of ABP required for a labelling effect, EP-LR-PEG₃-Biotin was tested over a range of 0 to 50 μ M. As shown in **Figure 5. 6**, only slight activity was observed at 50 μ M and at lower concentrations no activity was observed. However, the labelling effect could be detected at concentrations as low as 2 μ M. This result implies that the minimal concentration required for detection of labelling of the target protease(s) is much lower than that required for detectable activity against SlpA processing. It is possible, that *C. difficile* produces an excess of the active protease(s), so that only a fraction of available protease is required for cleaving SlpA. Consequently, inactivating (and labelling) the excess protease with an ABP will not result in inhibition of SlpA cleavage, and

therefore the concentration threshold for labelling/detection is much more sensitive than detection of activity (via detection of SlpA).

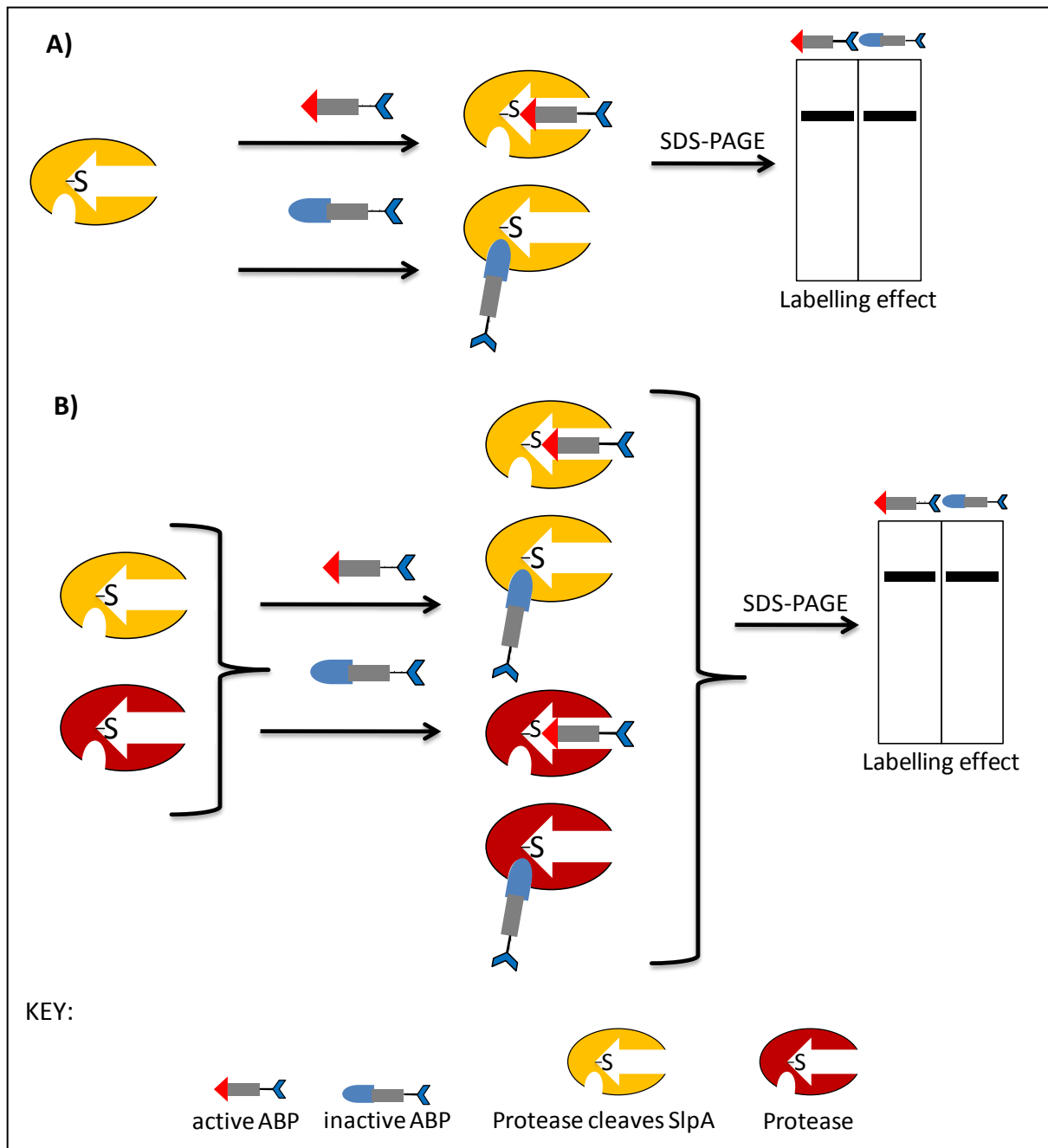


Figure 5. 5: Hypothesis of labelling of epoxysuccinyl ABPs. A) Both active and inactive ABPs target one single protease, but at different binding sites. B) Only active ABPs target label the target protease involved in SlpA cleavage, while inactive and possibly also active ABPs bind to another protease which have similar size to the target protease.

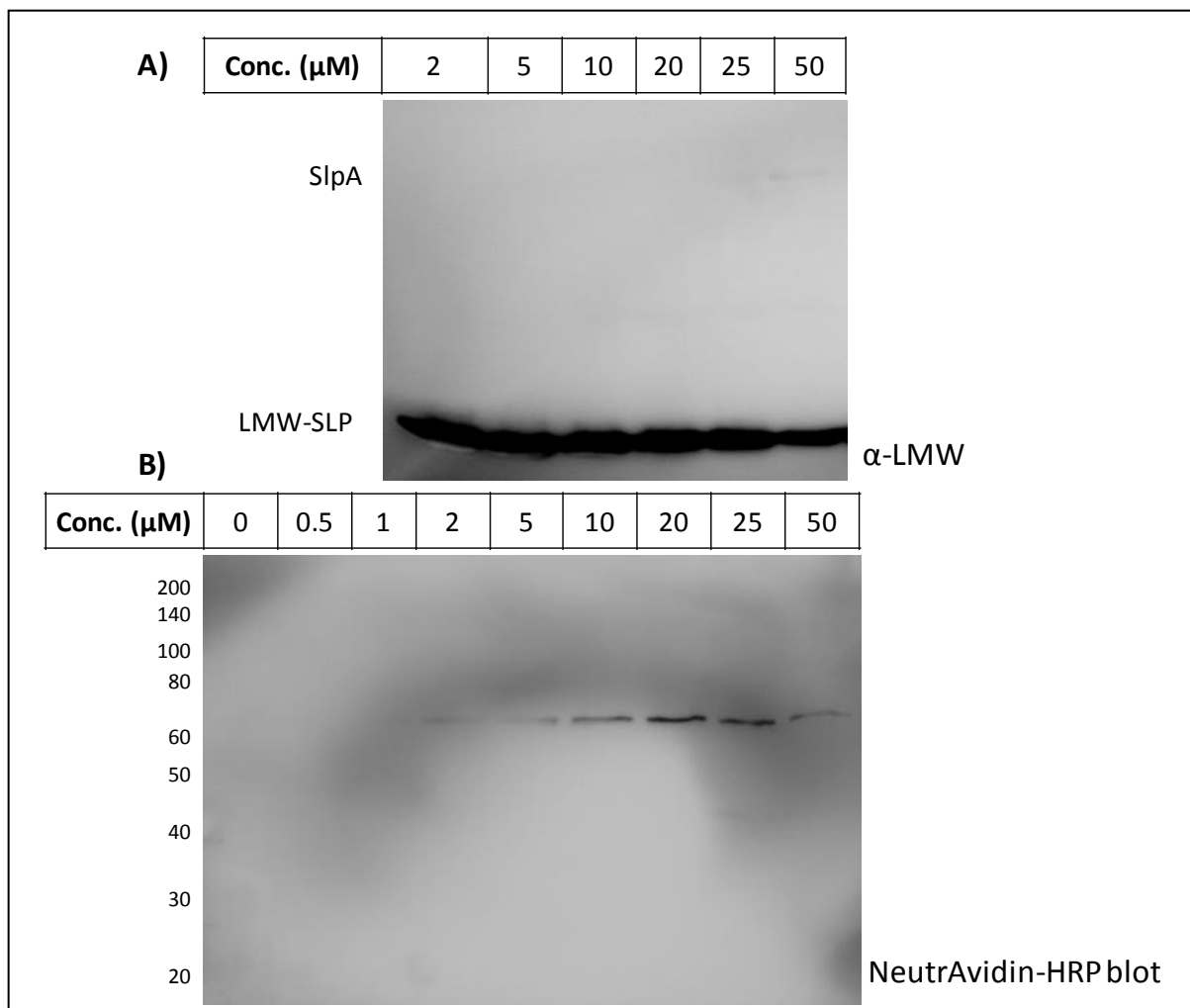


Figure 5. 6: Activity and labelling effect of biotinylated ABPs, as measured by Western blot of whole cell against LMW-SLP (A) and NeutrAvidin-HRP (B). *C. difficile* 630 cultures were treated with EP-LR-PEG₃-Biotin at the indicated concentrations, as per standard inhibition assays (see section 2.1.1).

Presumably the labelling effect correlates directly with the amount of protease(s) bound by the ABPs regardless whether the ABPs inactivate the protease(s). Furthermore, the labelling signals seem to saturate at 20 μM , as the intensity of the labelling from 20 μM to 50 μM is similar. This saturation effect is one drawback of a non-linear visualisation method such as NeutrAvidin-HRP (using chemiluminescence detected with an automated gel reader), which tends to enhance the detection of strong signals and decrease weaker signals, frustrating the detection of proteins at lower concentration. Furthermore, it makes the quantitative evaluation of labelling more difficult, especially at higher concentrations required for detectable potency. Nevertheless, the fact that the designed ABPs tag target protease(s) at relatively low concentration encouraged further investigation in protein enrichment and identification, as discussed in the next chapter.

5.1.3 Stability of inhibitors in culture

Epoxides were reported to be stable under biologically relevant conditions by the Koide group [399]. However, under the condition tested, it might be that denaturated proteins contained in the BHI broth could cross react with the epoxide part of the inhibitors at 37°C in the anaerobic cabinet.

It was assumed that if the epoxide reacts or degrades in BHI broth its effective concentration would decrease significantly and therefore a much lower activity would be expected at later time points. To address this concern, EP-LR-PEG₃-Biotin was pre-incubated in BHI broth for 1.5, 16 or 24 hours prior to inoculating a single *C. difficile* colony for performing the activity assay. The result of this experiment is shown in **Figure 5. 7A**. Both full-length SlpA on Western blot against LMW-SLP and the characteristic labelled band at approximately 70 kDa on a NeutrAvidin-HRP blot were found in all three samples, suggesting that the epoxide inhibitor was stable in BHI broth under the conditions tested. Thus the lifetime of active inhibitor/ABP in BHI broth at 37°C is certainly greater than 24 h. However, it is unclear if the concentrations of inhibitor in the growth medium after different incubation times are identical or slightly lower compared to the initial concentration added, as the similar potency and labelled activity in all samples does not reflect small changes of inhibitor concentration.

It was noted that there was a high molecular weight band at approximate 140 kDa found on the NeutrAvidin blot. This band is presumably either a natively biotinylated protein or is detected non-specifically, as it is also found in negative control which was not treated with any of the biotinylated ABPs.

In addition, a growth curve of active and inactive ABPs (EP-LR-PEG₃-Biotin and EP-A-PEG₃-Biotin) was monitored (**Figure 5. 7B**) showing that these ABPs did not affect the growth of bacteria. Samples OD₆₀₀ were identical at all time points, except the one measured after overnight inhibition which may be due to growth variation rather than any effect of the inhibitors.

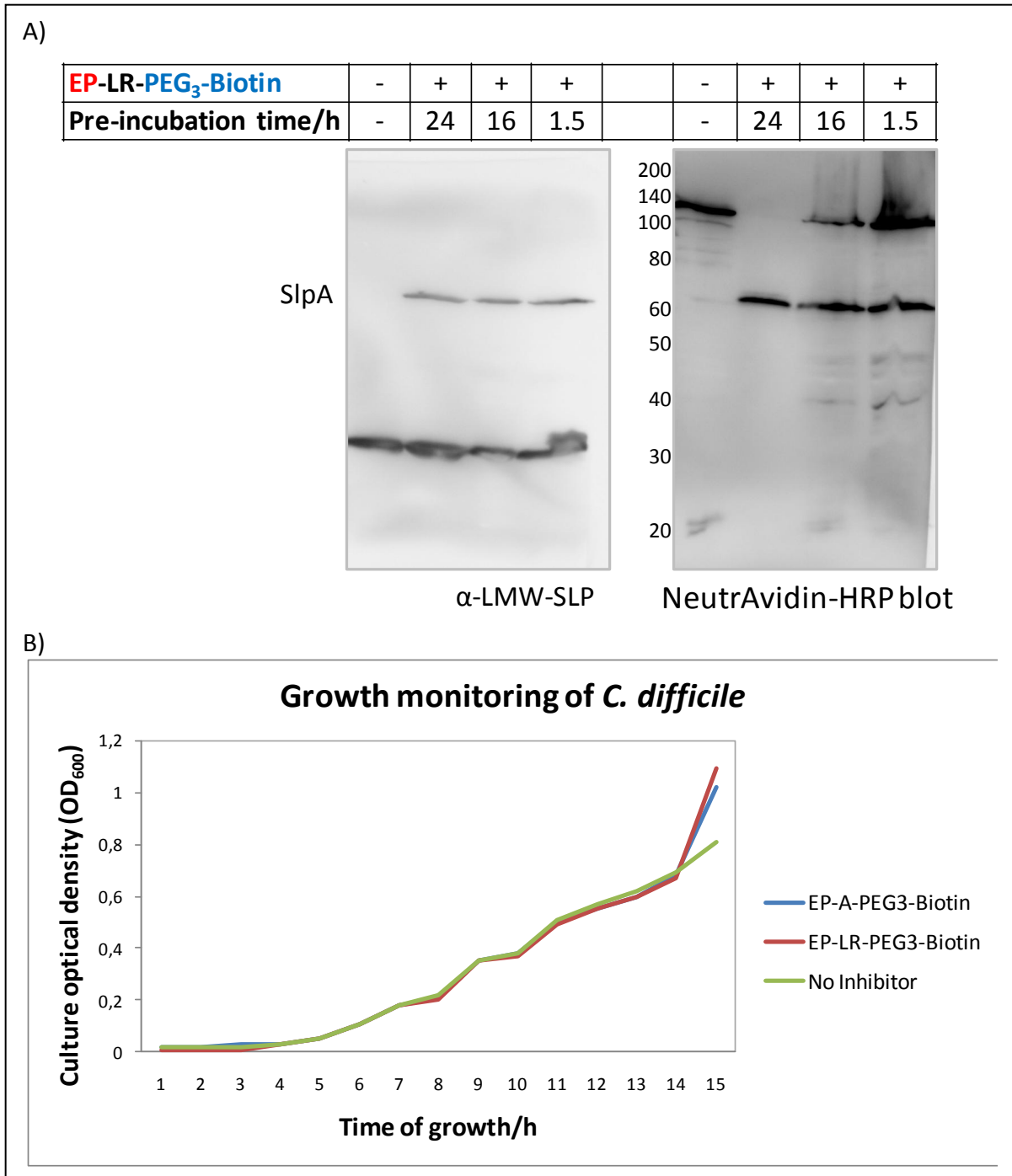


Figure 5. 7: Inhibitor stability test. A) Stability test of epoxide-based inhibitor over time in growth medium. EP-LR-PEG₃-Biotin was pre-incubated in BHI broth for a given time before a *C. difficile* colony was inoculated. The SlpA band on a Western bolt against LMW-SLP antibody reveals the inhibition activity level for each sample. Labelled bands on NeutrAvidin-HRP blot show the labelling effect. B) Measurement of *C. difficile* growth by treating with active ABP (EP-LR-PEG₃-Biotin) and inactive ABP (EP-A-PEG₃-Biotin). Optical density at 600 nm (OD₆₀₀) was measured at each time point.

The whole cell pellet and the concentrate of the culture supernatant were analysed by Western blot against LMW-SLP and NeutrAvidin to reveal the results displayed in **Figure 5.8**. It was found that inhibition of S-layer processing occurred continuously and the longer the inhibitor was present in the culture, the more SlpA was found. Interestingly, the shed SlpA in the culture supernatant obeyed the same tendency as that found in the cell pellet, however unprocessed SlpA was not found in the culture supernatant until at least 4 hours of inhibitor incubation time whilst this was only 2 hours for SlpA found in the cell pellet.

Labelling of protein occurs only in the cell and not in the culture supernatant suggesting that the protease(s) involved in the cleavage remain in the S-layer. Furthermore, a similar tendency was observed as for inhibitory activity; that 2 h of inhibitor incubation time is sufficient for the labelling process to take place.

5.1.5 Strain dependence

Based on the labelling effect and inhibitory activity on SlpA processing, it was hypothesised that epoxysuccinyl inhibitors/ABPs targeted at least two proteases in *C. difficile*. One is involved in SlpA cleavage (inhibited by active inhibitors) and one is not (bound by inactive inhibitors and possibly also active inhibitors). It was decided to investigate this phenomenon further by looking at different *C. difficile* strains.

The cleavage locus of different strains still varies despite the relatively high homology around this locus, as revealed by an alignment of SlpA sequences from 35 strains from different sources (see section 1.6.2.). It is more unlikely that one specific protease can recognise all the different proteins. To address this topic, a collection of ten different strains of *C. difficile* which possess less homology at the cleavage site was investigated.

A panel of non-tagged and biotinylated inhibitors was chosen for this experiment. E-64 and its mimic EP-LR-NH₂ and EP-LR-PEG₃-Biotin were selected as positive controls, while EtEP-OH, EP-PEG₃-Biotin and EP-A-PEG₃-Biotin served as negative controls. EP-LGR-NH₂ and EP-KTEL-NH₂ were chosen to satisfy both design considerations (based on the parent compound EP-LR-NH₂ and cleavage site). EP-K-PEG₃-Biotin and EP-L-PEG₃-Biotin were selected to mimic the P2 position of the majority of substrates (corresponding to the AA¹ position of inhibitors).

Strain	SlpA cleavage sequence ...P ₅ P ₄ P ₃ P ₂ P ₁ P ₁ 'P ₂ 'P ₃ '...	No Inh.	E-64	EtEP- OH	EP-X-PEG ₃ -Biotin				EP-X-NH ₂				
					-	L	K	A	LR	LR	KTEL	LGR	
Cd 1V T T K S A A K....												
Cd 17L E T K S A D I....												
Cd 167F S T Y R A T N....												
Cd 630L E T K S A N D....												
Cd 959L T T K G A T G....												
Cd YF S S Y G K F S....												
SE 528L D T K G A T D....												
R13541A S T Y A S S N....												
R13540L T T K S A T Q....												
R13699L N T L S A S S....												

Figure 5. 9: Strain dependency of inhibitors. Amino acids indicated in red in the cleavage sequence represent the P1 position of the substrate at the cleavage site. Raw data of western blot gels are attached in the Appendix 1.

A summary is illustrated in **Figure 5. 9**, SlpA bands were excised and compiled in the table, while the raw data is attached in the appendix. Overall, E-64 and its mimics showed moderate activity, whereas EtEP-OH, EP-PEG₃-Biotin and EP-A-PEG₃-Biotin remained essentially inactive in all strains tested. Furthermore, some different levels of inhibition were observed for certain inhibitors across different strains. In particular, EP-KTEL-NH₂ and EP-K-PEG₃-Biotin possessed reduced or no inhibitory effects with strains including 167, E13540 and SE 538, but good activity with other strains.

As found in the reference strain 630, all biotinylated ABPs labelled a strong band at approximate 70 kDa. In some strains such as CD17, CD167 and Y, additional bands at lower molecular weights were found with lower intensity (**Figure 5. 10**). These bands showed similar behaviour as found with the 70 kDa bands; they were present in all samples regardless of the activity of the inhibitors. This trend might relate to a degradation process of the above mentioned proteins, but it could also result from non-specific labelling of cytoplasmic proteins released by cell lysis, as different strains might possess varied sensitivity to induce cell lysis in response to inhibitor treatment.

Results of activity and labelling indicated comparable inhibition behaviour to that found in 630. Epoxysuccinyl inhibitors may target multiple proteases, but the protease involved in SlpA cleavage might also vary between strains. Similarly to SlpA, these proteases might possess high homology among each other, but could possess subtly different active sites

appropriate for the corresponding SlpA cleavage locus. Due to the possible high homology and similarity of these proteases across all strains, a potent inhibitor could be active in most of the strains, and hypothetically a single drug could treat all strains related to CDAD if S-layer processing were to be selected as a therapeutic target. Fortunately, this ambiguity regarding the conservation of the target protease(s) was resolved in the later study described in Chapters 6 and 7.

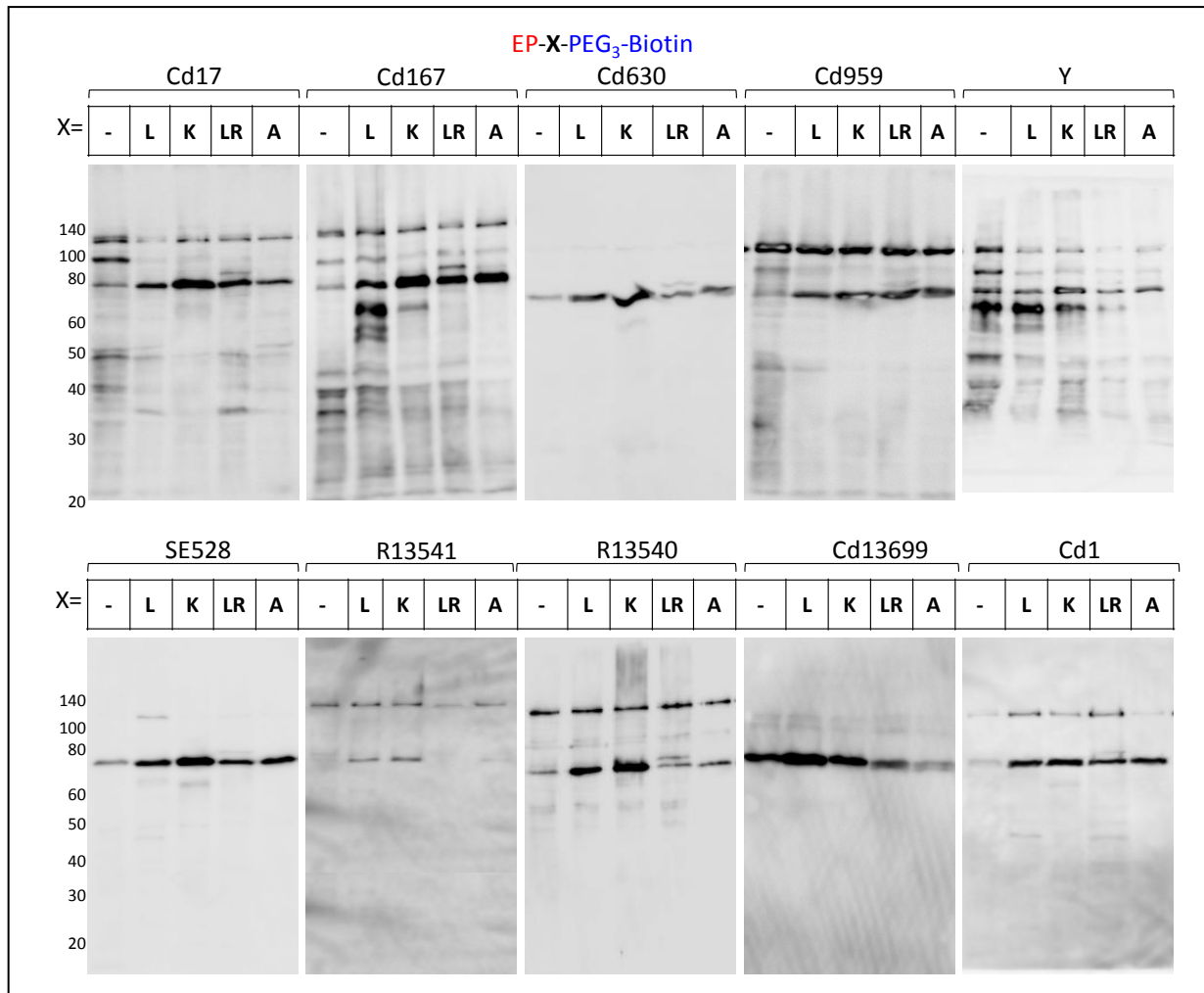


Figure 5. 10: Labelling of targets across different strains, visualised by NeutrAvidin blot. Bacteria were grown overnight with given compounds at 100 μ M. Whole cell pellet was boiled in loading buffer prior SDS-PAGE analysis.

5.2 Fluorescent ABPs

5.2.1 Design of fluorescent ABPs

Although biotinylated ABPs proved to be practical due to their elimination of two-step labelling by accessing the target protease directly in the medium, they brought a new

problem to the assay. Proteins labelled by biotinylated ABPs can only be visualised by Western blot e.g. against biotin antibody, streptavidin, avidin, NeutrAvidin, which could not be used for precise or semi-quantitative analysis. In addition to the common limitations of typical Western blots discussed in section 2.1.3, NeutrAvidin blot signals appeared to saturate at very low inhibitor concentrations and therefore the biotin label may not be optimal for exploring any correlation between activity and labelling. For this reason, another better visualisation label such as a fluorescent dye (Carboxytetramethylrhodamine called TAMRA) was considered to replace biotin.

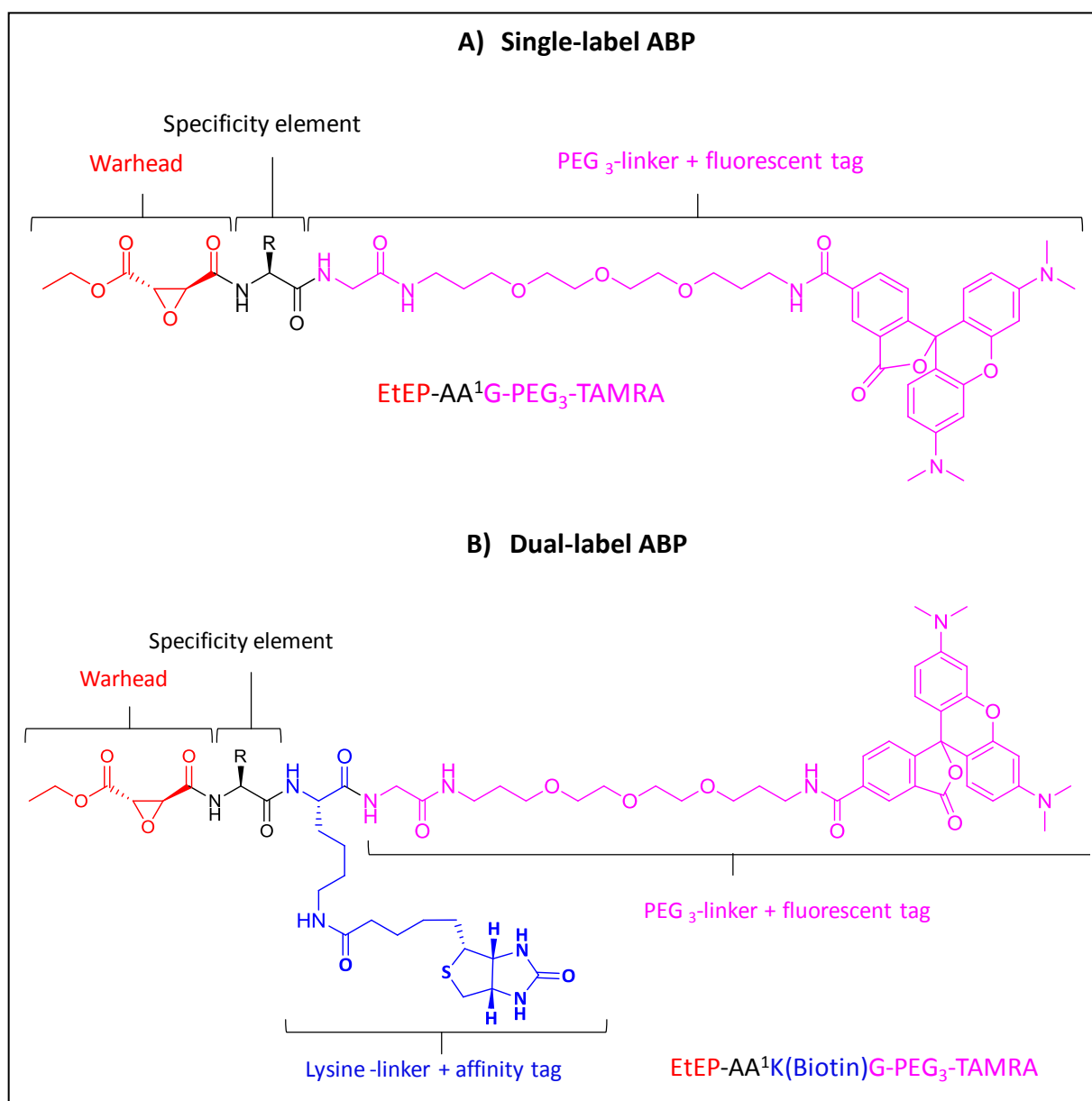
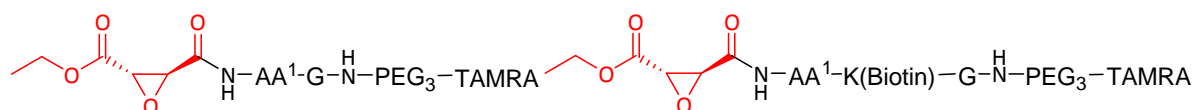


Figure 5. 11: Anatomy of the fluorescent ABPs. A) Singly labelled ABPs warhead and specificity element were separated from the fluorescent tag (TAMRA) by a PEG linker. B) Dually labelled ABPs contain an affinity label appended on the side chain of a lysine and a fluorescent label appended via a PEG linker.

ABPs are not only useful for monitoring labelling and activity, but also for facilitating the enrichment of labelled proteins required for the target identification process. The enrichment process usually involves affinity or antibody pull-down [103]. The TAMRA-antibody used for pull down of TAMRA-appended proteins was found to be less effective than the NeutrAvidin-biotin system due to weaker binding which did not survive the stringent wash step for removing unspecific labelling [400]. Therefore, it was decided to keep the biotin tag for subsequent protein enrichment. Consequently, the new generation of ABPs described here possess two labels, a TAMRA group for fluorescent detection and a biotin group for affinity pull down.

Combination of the two different labels compensated for their individual limitations, but significantly increased the size of the ABPs over the singly-labelled ABPs. It has been reported that bulky reporter tags could have a detrimental effect on enzyme/labelling [37, 75]. To investigate the potential effect of increasing the size of the ABP, singly labelled (fluorescent only) ABPs were also synthesised in the present study.



No.	WH	AA ¹	Affinity tag	Fluorescent tag	Abbr. Name
TD 61	EtEP	Ala	-	TAMRA	EtEP-AG-PEG ₃ -TAMRA
TD 62	EtEP	Arg	-	TAMRA	EtEP-RG-PEG ₃ -TAMRA
TD 63	EtEP	Tyr	-	TAMRA	EtEP-YG-PEG ₃ -TAMRA
TD 64	EtEP	Ala	Biotin	TAMRA	EtEP-AK(Biotin)G-PEG ₃ -TAMRA
TD 65	EtEP	Arg	Biotin	TAMRA	EtEP-RK(Biotin)G-PEG ₃ -TAMRA
TD 66	EtEP	Tyr	Biotin	TAMRA	EtEP-YK(Biotin)G-PEG ₃ -TAMRA

Table 5. 2: Table of fluorescent ABPs.

The design of epoxysuccinyl inhibitors/ABPs with respect to activity has been described extensively in Chapters 3, 4, and 5.1, therefore only the labelling effect is discussed in this section. For this purpose, three representative examples which contain different AA¹ amino acids were chosen: alanine (representing inactive inhibitors), arginine (representative of one of the two alternate substrate/inhibitor tendencies for S2 tolerance and the only inhibitor

which caused the double band labelling effect) and tyrosine (representing the remaining activity trend and the most potent inhibitor so far). The general structure of these ABPs is illustrated in **Figure 5. 11** and the newly synthesised ABPs are summarised in **Table 5. 2**.

5.2.2 Synthesis of fluorescent ABPs

Similar to the previous generation of inhibitors/ABPs, the synthesis was performed on solid phase, however it was more challenging due to the need to incorporate two different labels (biotin and TAMRA) on a single molecule and the instability of the TAMRA group to the conditions employed in Fmoc/^tBu SPPS [68]. For this reason, an orthogonal protection strategy was applied on Universal PEG Novatag resin (**9**) (Novachem, UK). This resin enables peptide chain extension on a secondary Fmoc protected amine under standard conditions and the coupling of TAMRA occurs following orthogonal deprotection of the monomethoxytrityl (Mmt) protected primary amine [68]. The general synthesis route employed for fluorescent ABPs is displayed in **Scheme 5. 1**.

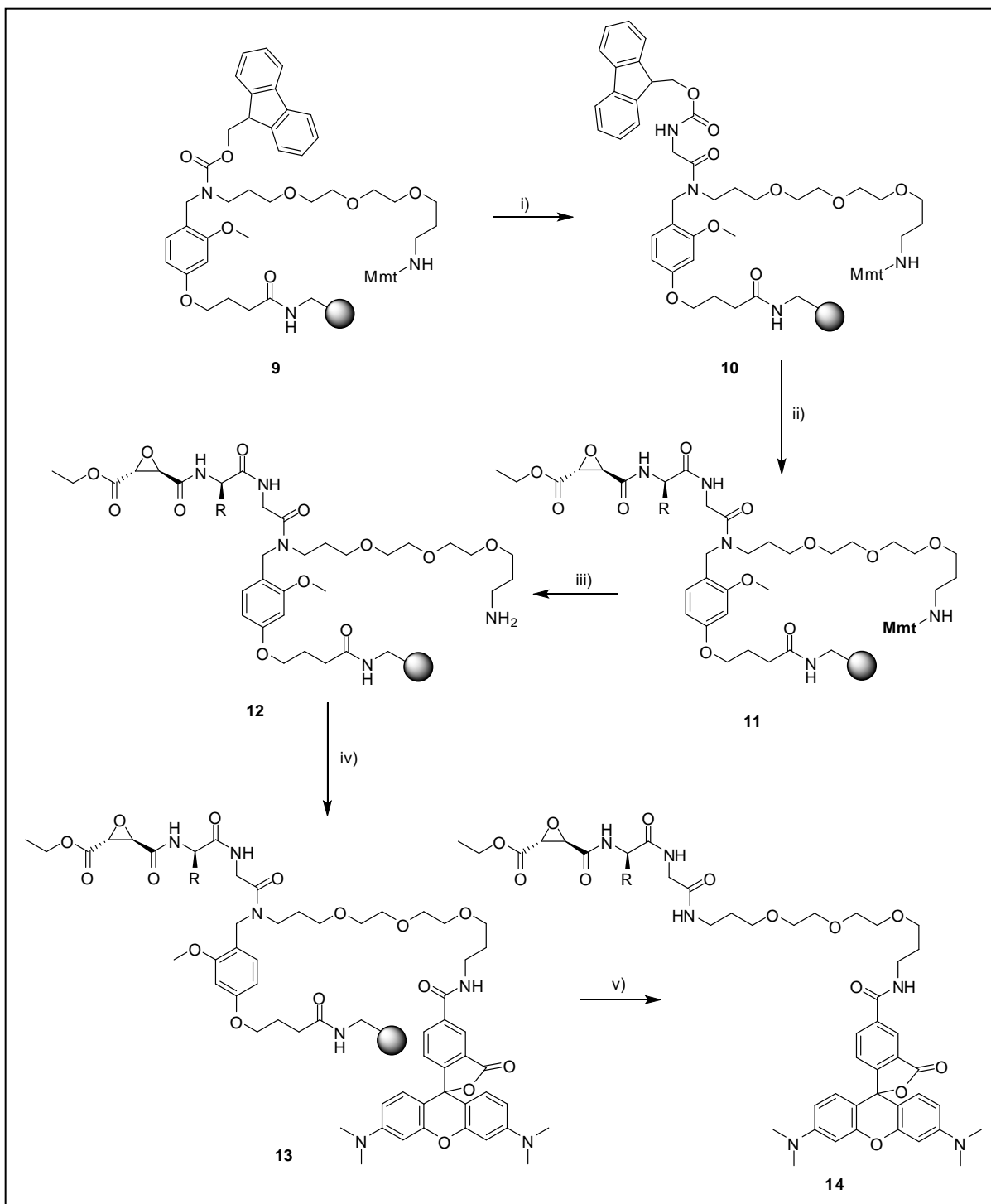
The synthesis was performed as follows: the Fmoc protecting group was removed using 20% piperidine in DMF. Afterwards, the resin was washed thoroughly before the first amino acid was loaded onto the resin. As mentioned previously, the coupling involved is acylation of a secondary amine and therefore stronger coupling conditions were needed. Several activation strategies for the first amino acid coupling were explored including the utilisation of HBTU, HCTU, HATU or HOBt/DIC, but only activation using HOBt/DIC was efficient. Furthermore, HOBt/DIC preactivation is a harsh condition and could possibly cause racemisation of a coupled chiral amino acid. To prevent this potential problem, glycine was chosen as the first amino acid for initial loading on the resin. Glycine is small and achiral, and could act as additional spacer extension to the PEG linker; in itself, it should not interfere with inhibition/labelling. As glycine was the initial loading for all ABPs, the resin was loaded on a large scale and aliquoted subsequently for the next coupling.

Much effort was put into optimisation of the first loading, because it is very important to occupy as much as free amine as possible. Unreacted amines were capped by acetic anhydride; without capping, short-peptidic by-products e.g.: EtEP-PEG₃-TAMRA, TAMRA-PEG₃-TAMRA, caused dramatic purification problems.

Secondly, the desired peptide was extended on the preloaded resin using standard conditions with HBTU and DiPEA: alanine/tyrosine/arginine was the second loaded amino acid in the case of single label ABPs, whilst in the case of dual label ABPs biotin appended lysine was inserted prior in order to introduce the biotin label. Chain extension was completed by coupling of the warhead using HCTU and DiPEA, followed by an intensive resin wash with DFM and DCM.

Lastly, the Mmt group, stable during the chain extension under basic conditions but unstable in mildly acidic media [401], was removed with 1%TFA in DCM. This deprotection procedure was repeated twice and was relatively quick (2 min) to prevent possible cleavage of the pre-synthesised peptide still attached on the resin. TAMRA pre-activated with HOAt/DIC was coupled to the resin via the PEG linker. It was important to get rid of the excess TAMRA by thoroughly washing with DCM-DMF-DCM-MeOH-Et₂O. The resin was dried and cleaved to obtain the final products. **14** was purified using LC-MS and lyophilised to yield the fluorescent inhibitors. It is worth noting that the more by-products the synthesis contained, the more challenging the purification of the crude material was. In some cases the purification had to be repeated several times with appropriate gradient settings to obtain the final pure products. For this reason, the whole synthesis was implemented manually by hand with extra care in each step rather than using an automated synthesiser. Unfortunately, the yield of each inhibitor was very poor due to substantial losses during the purification process, and no attempt was made to calculate the overall yield.

Similar to other inhibitors/ABPs, stock solutions of 10 mM were prepared in DMSO. The stock solution calculation was based on the mass of the freeze-dried product. An attempt to minimise the limited accuracy of the stock concentration determination discussed in section 3.3 was made by normalising the stock solution against a standard solution of TAMRA. It should be noted that the standard solution was prepared by the same manner as the inhibitors (i.e. by mass), but at greater dimension (ca. 5 mg) to minimise the inaccuracy of weighing, followed a dilution to 10 mM. After normalising, the concentration of most inhibitors dropped from 10 mM (by mass) to 4 mM or 2 mM, thus it was estimated that the true inhibitor concentrations were four or five fold lower.



Scheme 5. 1: Synthesis of fluorescent ABPs using Universal PEG Novatag™ resin (I). Reagents and conditions: i) Fmoc deprotection with 20% piperidine/DMF, 5 min; 3 × coupling of glycine (5 eq) preactivated with DIC, HOBT and DiPEA, 30 min, capping the unreacted amine with acetyl anhydride/DiPEA/DMF (2:1:7); ii) Fmoc deprotection with 20% piperidine /DMF, 5 min; 2 × coupling of either AA¹ (5 eq) or K(Biotin) (2.5 eq) then AA¹ (5 eq) preactivated with HBTU and DiPEA, each coupling 30 min; coupling of EtEP-OH (3 eq) preactivated with HCTU and DiPEA, 1h, capping the unreacted amine with acetyl anhydride/DiPEA/DMF (2:1:7) after each coupling step; iii) Mmt deprotection with DCM/TFA/TIS (94:1:5), 2 min; iv) 2 × coupling of TAMRA (2 eq) preactivated with DIC, HOAt and DiPEA, 1h; v) cleavage from the resin with TFA/TIS/H₂O (95:2.5:2.5), 2h.

5.2.3 Activity and labelling effect of fluorescence appended ABPs

Although the focus of this section is not the potency of inhibitors, newly synthesised fluorescent ABPs were tested at 100 μ M to ensure that they were still active despite the expanded size. As can be seen in **Figure 5. 12A**, the activity of this new ABP generation remained, though at a modified degree: alanine in AA¹ position was inactive as previously found; tyrosine and arginine at the same position were active, but arginine appeared to be significantly more active than tyrosine (in both type of ABPs), which is contradictory to the previous findings with other ABPs/inhibitors (sections 4.2). Counting all possible limitations in each step, starting from the synthesis to the stock solution determination, *in vivo* assay and western blotting, the observed inverted potency of tyrosine and arginine in AA¹ position is certainly subject to some uncertainty. This difference could arise from the altered structure of TAMRA labelled ABPs vs. previously described ABPs, but on the basis of the observations to date this seems unlikely.

Looking at the in-gel fluorescence (**Figure 5. 12B**), one single band at approximately 76 kDa was found in all samples which is in agreement with the labelling effect of biotinylated ABPs. However, the degree of labelling differed significantly among these ABPs. For singly-labelled fluorescent ABPs, EtEP-AG-PEG₃-TAMRA gave the distinct impression that it labelled the protein less strongly than EtEP-RG-PEG₃-TAMRA and EtEP-YG-PEG₃-TAMRA. The labelling of EtEP-RG-PEG₃-TAMRA and EtEP-YG-PEG₃-TAMRA were equally strong, though any 'double band' effect of EtEP-RG-PEG₃-TAMRA (as previously seen with analogous Arg-containing ABPs) was apparently minor, as only a faint band on top of the 76 kDa band was observed.

In the case of dually labelled fluorescent ABPs, all three ABPs possess similar labelling effects, which might arise from the interfering effect of biotinylated lysine adjacent to the AA¹ amino acid. The 'double band' effect of was also missing for EtEP-RG-PEG₃-TAMRA.

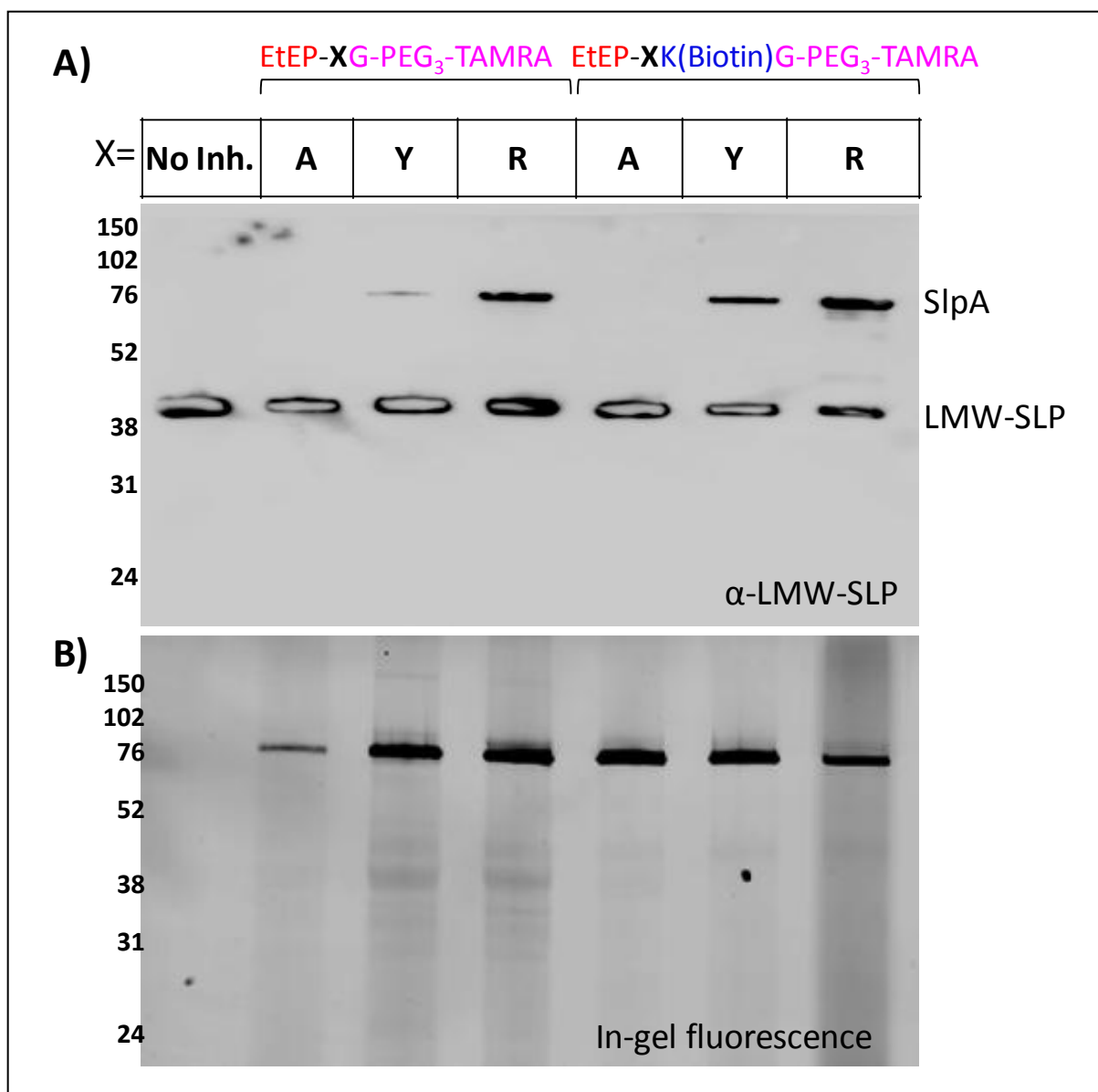


Figure 5. 12: Activity and labelling of fluorescent ABPs, as measured by A) Western blot using anti-LMW SLP and B) in-gel fluorescence. *C. difficile* cultures were treated with compound at 100 μ M.

Given that the size of 'first' labelled bands with fluorescent ABPs were similar to the size of labelled bands with alkyne-tagged ABPs (both on Bis-Tris gel at ca. 76 kDa) and the tendency of all bands being labelled at one size in all three generation of ABPs, it could be supposed that these labelled bands were the same protein(s) in each case. Therefore, enrichment and proteomics using one type of ABPs may be representative of the other types.

Combining the changed behaviour of dual label ABPs in activity (inverted potency of EtEP-RG-PEG₃-TAMRA and EtEP-YG-PEG₃-TAMRA) and labelling (all ABPs seemed to labelled double bands equally), it was supposed that the expansion in size could have an effect on

activity and labelling of the ABPs. Therefore, it was decided to apply the previous (single label) biotinylated ABPs for initial target identification, rather than the fluorescent ABPs.

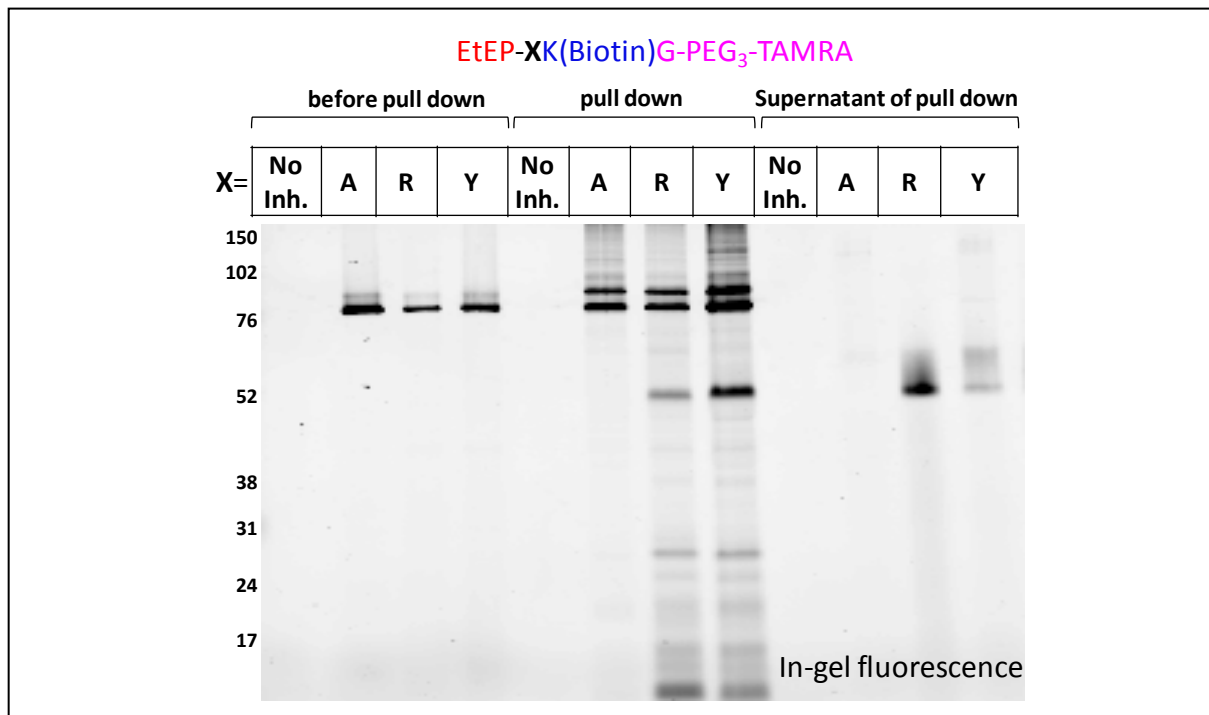


Figure 5. 13: Labelling of dual label fluorescent ABPs (S-layer fractions), visualised by in-gel fluorescence. *C. difficile* 630 was grown overnight in the presence of 50 μ M ABPs. S-layer fraction was extracted using glycine at low pH.

5.2.4 Strain dependence

Investigation of strain dependence with biotinylated ABPs provided useful information (section 5.1.5), but the results were rather difficult to analyse. The cell contains a certain amount of protein that carries biotin, and these are also visualised by streptavidin blotting causing a certain level of background signals. These background signals increased where the cell lysed and released more cytoplasmic protein. The visualisation method itself also varied between blots causing different intensity of bands, and made comparison between membranes problematic. Fluorescence detection does not suffer from these problems and therefore fluorescent ABPs might provide clearer results.

Because most of the stocks of dual label fluorescent ABPs were consumed in previous pull down experiments only single label ABPs were tested in this case. Bacteria were treated with ABP at 100 μ M overnight, then were harvested and boiled in loading buffer. Whole cell was loaded on the gel for SDS-PAGE analysis and in-gel fluorescence imaging. All ABPs

labelled the 76 kDa band very strongly across all strains tested. EtEP-RG-PEG₃-TAMRA showed clearly the double band effect in all strains. While the 76 kDa bands appeared to have the same size and same intensity in all strains, the upper band (of the double band) varied in size slightly and in intensity more so. Background labelling remained relatively weak and more consistent in all three ABP treatments.

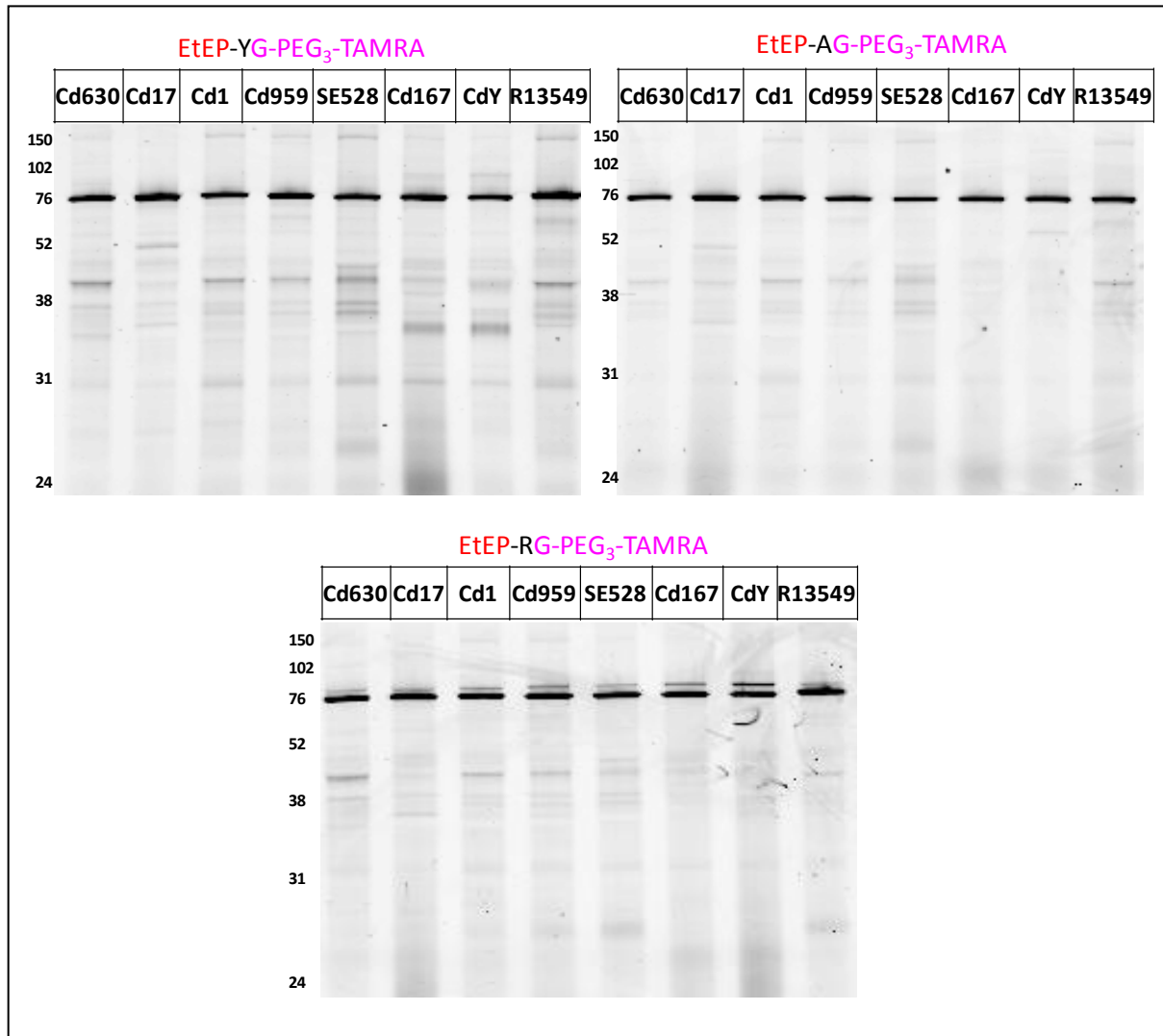


Figure 5. 14: Labelling across different strains visualised by in-gel fluorescence. *C. difficile* was treated with 100 μ M overnight. Whole cell was boiled in loading buffer and loaded on the gel.

5.3 Conclusion

This chapter has described the development of directly-labelled ABPs that carry either a biotin or a fluorescent group (TAMRA), or both. The synthesis of these ABPs was similar to the synthesis described in section 3.3 with some important modifications. It was more challenging to synthesise fluorescent ABPs than biotinylated ABPs because of the application of an orthogonal protection strategy and more challenging purification.

Efficacy and selectivity of biotinylated ABPs were explored and the result was in line with previous observed SAR (section 4.5), while the dual label ABPs showed some variance. Labelling of directly-labelled ABPs was clean and specific, as generally only one band was labelled in all ABPs, except ABPs bearing an arginine in AA¹ position.

As part of this work, the labelling effect across different strains that possess alternative homology at the cleavage locus was also investigated. It showed that the potency varied somewhat among strains but that labelling was consistent across all strains. These results suggest that the target enzyme(s) in different strains have both a similar mode of action (mechanism) and similar size.

The clean and mostly specific labelling given by biotinylated ABPs, coupled with ease of synthesis, resulted in their selection for testing in the enrichment and target identification experiments described in the next chapter.

Chapter 6 Target enrichment, identification and proteomic study

Whilst previous chapters described in detail the synthesis of chemical probes and their biological effects on *C. difficile*, this chapter focuses on the methods to identify the target proteases labelled by the developed probes, starting with affinity enrichment and moving on to different mass spectrometric methods. In the first reported application of ABPP to Clostridia, two putative cysteine proteases were identified in the S-layer of *C. difficile*, Cwp84 and Cwp13; further study was then performed on the potential functions of these proteases, seeking the candidate responsible for internal cleavage of SlpA to mature the two surface layer proteins (HMW- and LMW-SLPs). Furthermore, the effect on other cell wall protein members was investigated to increase our understanding of the role of SlpA proteolysis in S-layer integrity.

6.1 Enrichment of labelled proteins in *C. difficile*

Having established that the target proteases were located in the S-layer and were readily accessible for direct labelled ABPs, the labelling experiment was scaled-up to provide sufficient protein for subsequent mass spectrometry analysis. The probes used in this experiment contained a biotin label allowing for the affinity purification, in which the highly specific interaction of biotin with avidin and its modified forms (NeutrAvidin and streptavidin) was exploited.

6.1.1 Pull down of labelled proteins with NeutrAvidin agarose resin

Aiming for a sufficient protein amount for mass spectrometry analysis, a 5 ml *C. difficile* culture was grown overnight in the presence of EP-LR-PEG₃-Biotin or EP-A-PEG₃-Biotin at 100 µM. After treatment of the bacterial cell pellet and dividing in three fractions (S-layer, soluble and insoluble) as described in section 3.5, NeutrAvidin agarose suspension was added to each fraction and the mixtures agitated gently at 4 °C for 1 hour to allow immobilisation of biotin-labelled protein onto the beads. Afterwards, the mixture was centrifuged and the supernatant removed. The beads were washed three times with PBS-T buffer to aid in removing potential non-specifically bound proteins and then boiled in the loading buffer at 100 °C. At these conditions (0.7% of SDS in loading buffer and high temperature) NeutrAvidin is denatured, thereby breaking the interaction with biotin, eluting the labelled proteins in the loading buffer readily for SDS-PAGE analysis.

The result of this experiment is shown in **Figure 6. 1**. Coomassie stained gels and NeutrAvidin-HRP blotting of all three fractions are shown to demonstrate the correlation of the pull down bands and the labelling effects. The Coomassie stained gel of the supernatant fraction after S-layer pull-down also confirmed the activity of the ABPs through the presence of the SlpA band (band 1 in S-layer fraction).

In the S-layer fraction, several bands were observed on the Coomassie stained gels in the pull down samples. These bands were enriched, since they were not readily detectable before (cf. the supernatant gels). Comparing bands from samples treated with active and inactive probe (EP-LR-PEG-Biotin and EP-A-PEG-Biotin respectively), it can be seen that the pattern of the pulled down proteins in both cases was similar. However, the intensity of the active sample was far greater than the inactive one. The same effect was observed in the

NeutrAvidin-HRP blot. There were about 6-7 bands enriched at different levels, but only the band at approximately 70 kDa was detected in the NeutrAvidin-HRP blot of both samples. Two further bands similar in size were observed in the EP-LR-PEG₃-Biotin sample, but with a much lower intensity.

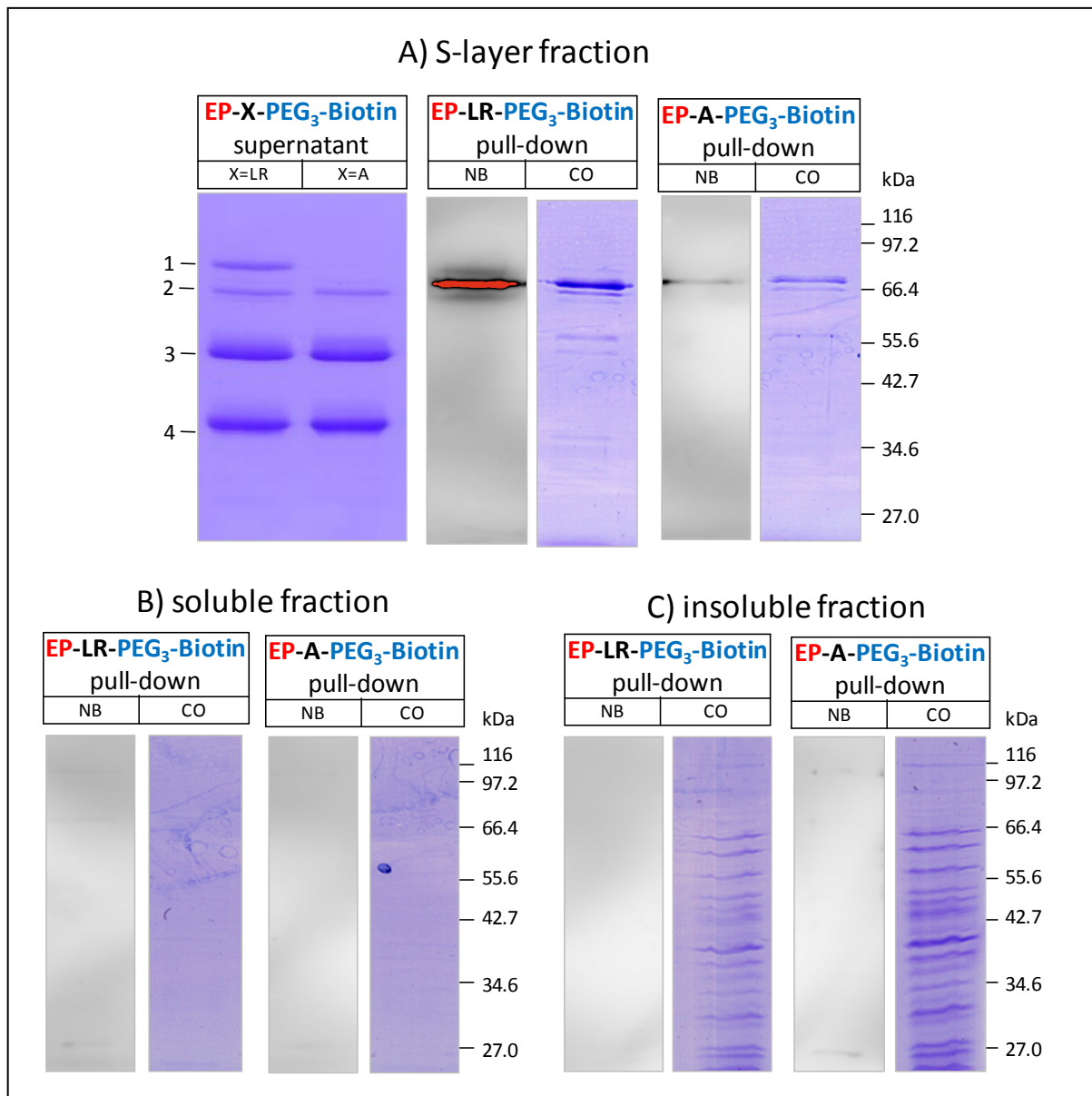


Figure 6. 1: Pull down experiment with biotinylated ABPs using NeutrAvidin agarose resin. *C. difficile* was treated with 100 μ M indicated ABP overnight, and then divided in subcellular fractions. Coomassie stained gels (CO) and NeutrAvidin-HRP blotting (NB) are shown for each fraction. The supernatant of the pull down of the S-layer fraction is shown to demonstrate the activity through the present of SlpA band. The lanes are: 1= SlpA; 2= Slp2; 3= HMW-SLP; 4= LMW-SLP.

Examination of the soluble fraction shows that no real enrichment occurred; bands were very faint and hardly detectable. The same tendency was found on the NeutrAvidin-HRP

blot. In contrast, many more bands were observed in the sample deriving from the insoluble fraction. However, in terms of the bands detected, the two samples were barely distinguishable. Furthermore, no bands on the Coomassie stained gel were found to correspond to those on the NeutrAvidin-HRP blot, suggesting that these bands might represent proteins 'sticking' on the beads non-specifically and therefore not related to the labelling.

Combining all the observations discussed above, the labelling only occurred in the S-layer, which is in agreement with previous observations (Chapters 4 and 5). Only a few enriched bands were correlated with the labelling, suggesting that co-precipitation occurred. These bands could be bound either to the beads or to the labelled proteins. Because no real correlation between enriched proteins and labelling was found in the soluble and insoluble fractions, only bands from the S-layer fraction were excised and prepared for mass spectrometry analysis. As a result, only the S-layer fraction was considered in the subsequent experiments.

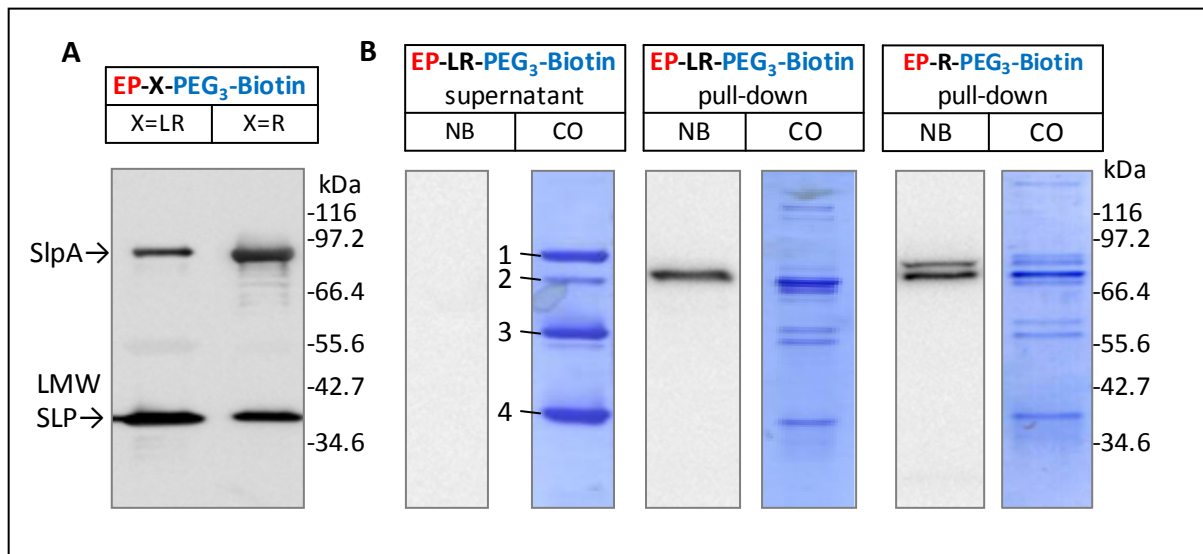


Figure 6. 2: Pull down experiment of S-layer fractions with biotinylated ABPs using NeutrAvidin agarose resin. *C. difficile* was treated with 100 μ M of indicated ABPs overnight. A) Western blot against LMW-SLP. B) Coomassie stained gels (CO) and NeutrAvidin-HRP blotting (NB) are shown. Supernatant of the pull down of the S-layer fraction was added to demonstrate the activity through the presence of the SlpA band. The lanes are: 1= SlpA; 2= Slp2; 3= HMW-SLP; 4= LMW-SLP.

Revisiting the labelling results described in Chapters 4 and 5, it was noted that the labelling effects of epoxysuccinyl ABPs was very specific. In most cases, a strongly labelled band at approximately 70 kDa was observed with both active and inactive ABPs. However, in some

cases a further labelled band above this 70 kDa band was found. The intensity of this band varied between samples, but it appeared significantly stronger when AA¹ was an arginine residue. In particular, in the case of biotinylated ABPs, only EP-R-PEG₃-Biotin gave rise to this 'double labelled band' effect, while all other ABPs labelled (mainly) one single band. In this experiment, the focus shifted to identifying this second band. Therefore, the pull down experiment with the S-fraction described above was repeated with EP-R-PEG₃-Biotin, and including EP-LR-PEG₃-Biotin as a positive control. As can be seen in **Figure 6. 2**, the double labelled band characteristic of EP-R-PEG₃-Biotin was observed in line with previous results. The upper band was less intense, but easily distinguishable compared with EP-LR-PEG₃-Biotin. A similar pattern of nonspecific co-purification was also observed, as in **Figure 6. 1**.

6.1.2 Pull down of labelled proteins with Streptavidin Dynabeads

It was noticed during implementation of the pull down experiment that it was practically difficult to remove all the supernatant from the NeutrAvidin agarose resin without accidentally withdrawing some resin. If a small volume of supernatant was left to ensure that the resin was not touched, then the final volume of the samples after adding loading buffer varied slightly. Consequently, there was a degree of inconsistency among the prepared samples, which might affect any quantitative analysis. To avert this potential problem, NeutrAvidin was replaced by streptavidin Dynabeads, which provide several advantages over the agarose resin, such as:

- Easy separation of the protein-bead complex from the supernatant with a Dyna magnet by collecting the beads on the tube wall; this provides a quick and gentle method, as centrifugation was avoided.
- According to the supplier's product data [402], batch consistency and reproducibility of results were aided by the presence of a streptavidin monolayer (covalently coupled on the hydrophobic bead surface) whose leakage was negligible and which contained no excess adsorbed streptavidin

The beads were first trialled with EP-R-PEG₃-Biotin labelled sample. Compared with the previous method of using NeutrAvidin agarose resin, the level of background pull down was very high, which indicated a higher level of non-specific binding to the streptavidin beads (**Figure 6. 3A**). Matching up bands in the not treated sample with those in the treated

sample, only a few additional bands at high molecular weight were observed in the treated sample. The experiment was repeated with EP-X-PEG₃-Biotin (X=A, Y or R), and more stringent wash conditions (1% SDS in PBS) were used to successfully remove some of the non-specifically binding proteins (**Figure 6. 3B**).

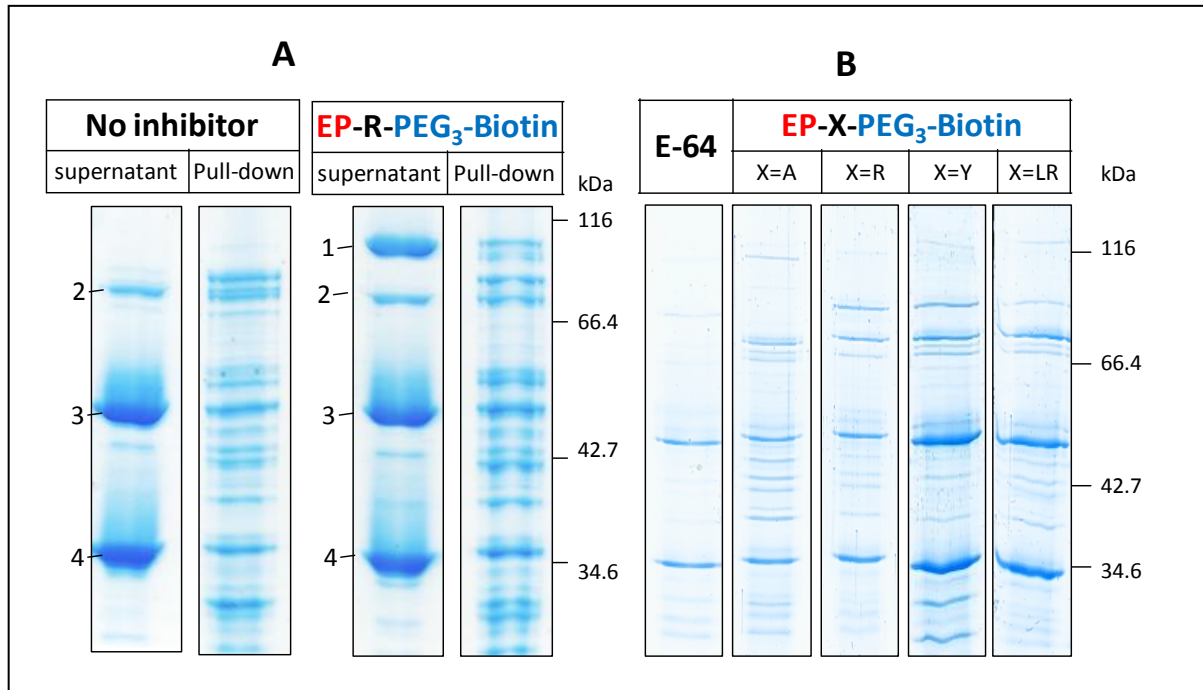


Figure 6. 3: Pull down experiment of S-layer fractions using super-paramagnetic Streptavidin Dynabeads, visualised by Coomassie staining. *C. difficile* was treated with 100 μ M ABP overnight. Prepared S-layer fractions were pulled down with 30 μ L of beads. A) The bands in the supernatant are: 1= SlpA; 2= Slp2; 3= HMW-SLP; 4= LMW-SLP. Beads were washed three times with PBS-T. B) Only the pull down samples are shown. Beads were washed three times with PBS containing 1% SDS.

6.2 Identification of protease(s) by MALDI-TOF-MS and PMF

Having successfully enriched the labelled protein(s) from *C. difficile*, a MS-based proteomic analysis was carried out using a bottom-up, sort-then-break approach to identify these enriched protein(s) (see section 1.5.3.2). As this study was focused on the identification of the labelled proteins in a relatively less complex peptide mixture (only one band per sample), MALDI-TOF-MS with subsequent PMF analysis was chosen as the least complex and most matured proteomic approach, simplifying the data analysis and saving time and cost.

The enriched bands visualised by Coomassie staining prepared in section 6.1.1 were excised from the 1D-gel and digested with trypsin. The tryptic digest was then spotted on the MALDI-plate readily for MS analysis. This experiment was first implemented by Dr. Len

Packman (Protein & Nucleic Acid Chemistry Facility of Biochemistry department (PNAC), Cambridge, UK) as an external service (**Figure 6. 4A**) and then repeated by the author using facilities at Imperial College London (Division of Molecular Biosciences) (**Figure 6. 4B**). Because the experiments were practiced using different facilities (although the same type of instrument and approach was used in both cases) by different analysers, besides the biological variation of samples discrepancies in performance might be expected. The resulted peptide masses were processed with the online database search engine MASCOT [403] with bacteria (Eubacteria) as selected Taxonomy and the NCBI non-redundant database selected for the search. Protein identification was validated when the MOWSE score was significant ($p < 0.05$) according to MASCOT [403].

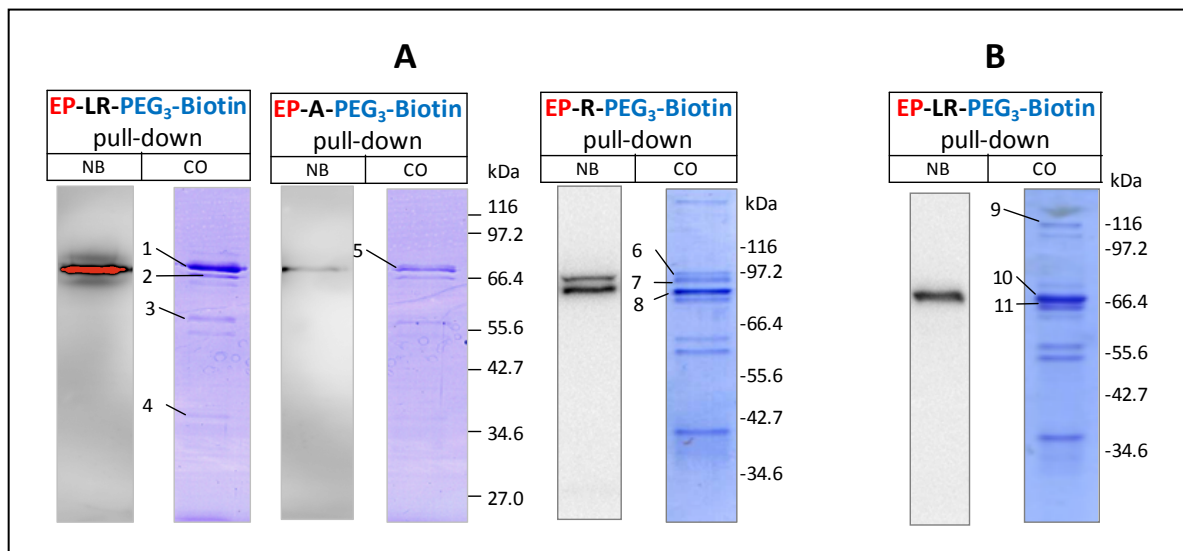


Figure 6. 4: Identified bands on Coomassie stained gels. Samples were prepared and enriched as described in section 2.1.1. In-gel digestion with trypsin and subsequently MALDI-TOF-MS and PMF were implemented by A) Dr. Len Packman, PNAC Cambridge, UK, B) the author, Imperial College London, UK. Identified bands are: 1, 5, 7, 8 and 10) Cwp84; 2 and 11) Cwp19 and Cwp66; 3) Cwp6; 4) LMW-SLP; 6) SlpA; 9) CD2797.

The PMF of labelled bands successfully obtained and their MASCOT analysis results are listed in **Table 6. 1**. The protein scores (probability scores in Mascot program) were relatively high (above 110), except the score for Cwp66 and CD2767 (around 60-70). Higher scores indicate more confidence in correct protein identification. Generally, all enriched proteins were identified as members of the cell wall protein family. Intriguingly, all the bands corresponding to labelling (bands 1,5,7, 8 and 10 in Figure 6. 4) were revealed to be the putative cysteine protease Cwp84, regardless of whether the probe was active (EP-LR-PEG₃-Biotin, EP-R-PEG₃-Biotin) or inactive (EP-A-PEG₃-Biotin). **Figure 6. 5** shows an example

of the final full spectra and the sequence coverage of the matched protein identified by peptide mass fingerprinting (PMF), while raw data can be found in the appendix. That this protein gave the highest score in all PMF analyses suggests very strongly that the identity of this protein is correct, especially given the relatively high sequence coverage. Variation in scores and sequence coverage of Cwp84 reflected possible discrepancies in the biological and technical reproducibility; nevertheless, this variation seemed to be minor and in all cases the protein identification was high confidence.

Assuming that biological and technical variation can be neglected by virtue of the near-identical sample treatment (starting from the same culture, same growth time, running on the same gel, identical chemical treatment, practising by the same analyst, processing with the same instruments), certain aspects of the results from treatment with active EP-LR-PEG₃-Biotin (band 1 in Figure 6. 4) can be compared directly with the results from treatment with inactive EP-A-PEG₃-Biotin (band 5 in Figure 6. 4). According to this, although both probes labelled Cwp84, the active probe seemed to label Cwp84 in a much more effective manner than the inactive probe, reflected in the significantly stronger band on NeurAvidin-HRP blot and Coomassie stained gel, associated with a larger amount of enriched protein. However, no quantitative comparison using MALDI-TOF-MS acquired data could be performed, and this will be discussed in detail in section 6.8.

More interestingly, both labelled bands of EP-R-PEG₃-Biotin treated samples (bands 7 and 8 in Figure 6. 4) were identified as Cwp84, suggesting a specific structure-dependent activity of this probe. Comparing the biological replicates with EP-LR-PEG₃-Biotin, the identified bands were identical, despite the variation in score and sequence coverage which maybe the result of technical variation. Bands 1 and 10 were revealed as Cwp84, whereas bands 2 and 11 were Cwp19 and Cwp66, further increasing confidence in the identification results.

Co-purified bands that did not correspond to the labelling were identified as Cwp19, Cwp66, Cwp6, LMW-SlpA, SlpA and CD2797. LMW-SlpA and SlpA were probably co-purified due to their high abundance in the sample. Cwp19, Cwp66, Cwp6 and CD2797 are part of the same gene cluster as *slpA*. However, only Cwp66 has been functionally annotated, as a temperature-dependent adhesion protein [256], whilst the other proteins have no known function.

Band No.	Accession No.	Protein Name	Protein score	Calc. Protein MW	Sequence coverage
1	gi 145953268	Cwp84	297	87456	45(%)
2	gi 126700383 gi 11066029	Cwp19 (also known as CD2767) Cwp66	263		30(%) 19(%)
3	gi 168708027	Cwp6 (also known as CD2784)	206	72964	35(%)
4	gi 94966244	LMW-SlpA	192	36607	53(%)
5	gi 145953268	Cwp84	246	87456	38(%)
6	gi 126700409	SlpA	205		29(%)
7	gi 126700403	Cwp84	215	87455	35(%)
8	gi 126700403	Cwp84	314	87455	41(%)
9*	gi 126700413	CD2797	70	221748	9(%)
10*	gi 126700403	Cwp84	657	87227	45(%)
11*	gi 126700383 gi 11066029	Cwp19 (also known as CD2767) Cwp66	110 60	77782 66724	26(%) 21(%)

Table 6. 1: Identified proteins using MALDI-TOF-MS and PMF with band number as shown in Figure 6. 4. Protein name and accession number (AC) are reported according to the NCBI data base, percentage of sequence coverage is generated from experimental PMF; protein score is probability score in Mascot program. Proteins in bands 1 to 8 were identified by Dr. Len Packman (PNAC Cambridge, UK), while proteins in bands 9 to 11 (indicated with a *) were identified by the author using facilities of Imperial College London (Division of Molecular Biosciences).

The activity and labelling data (from the previous chapters 3, 4 and 5) as well as the identification data for the labelled bands combine to provide convincing evidence that Cwp84 is a key target of this type of inhibitor and the ABPs described in this study. However, attempts to identify the tryptic peptide bearing the probe (covalently bound to the cysteine of the active side) failed in all samples. It is possible that either this modified peptide did not ionise well and therefore could not be detected, or the bond between the probe and the peptide was broken by the impact of the high ionisation energy, or during the sample preparation process.

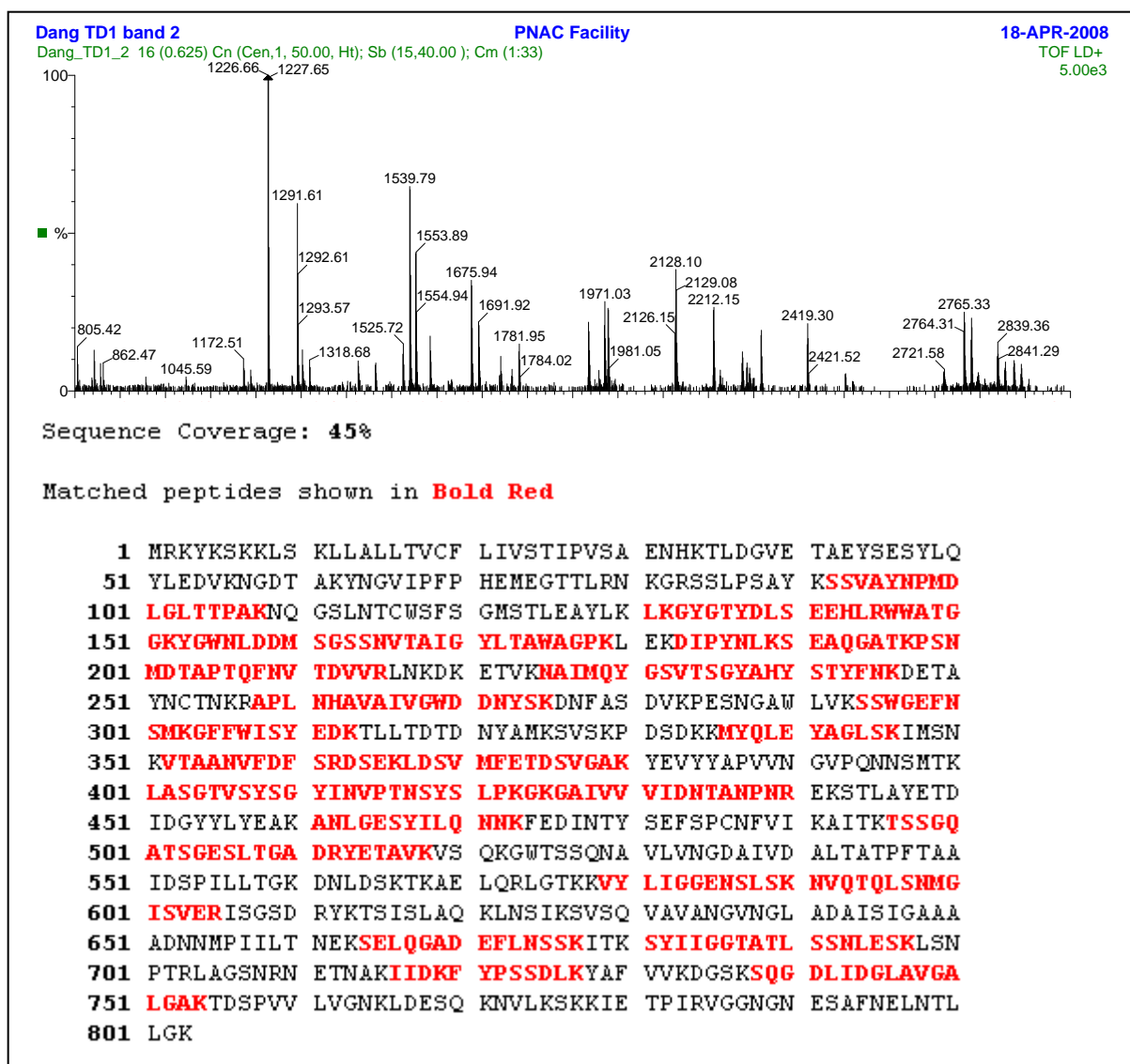


Figure 6. 5: MALDI-TOF-MS analysis of band 1 in Figure 6. 4, identified as Cwp84. The MALDI-TOF mass spectrum is shown on top and the Cwp84 sequence is shown at the bottom. Matched peptides are indicated in red.

6.3 Identification of proteases by LC-MS/MS

Although MALDI-TOF-MS is the most cost-efficient approach in proteomics today [175, 404] it suffers from some fundamental and practical limitations. The identification of a mixture of two proteins (Cwp66 and Cwp19 in bands 2 and 11) observed in the work described in the previous section may be due to a very distinct protein sequence, however, in general a MALDI-TOF-MS based method cannot identify multiple components in a mixture, especially if these components are highly homologous or low abundance [216]. This results from lower search scoring, which occurs when experimentally derived peptide masses for a given mass accuracy tolerance match to more than one protein, especially when these masses are small

and the peptide signal intensity is weak [216]. Furthermore, if proteins in the mixture are highly homologous, their derived peptide masses are very similar and so the software cannot differentiate the identification hits. This limitation may affect the results of this study, since in 1D gels a band often contains more than one protein [405]. This missed information could be decisive in understanding the similar labelling by active and inactive probes. However, this information can be obtained by applying tandem-MS, because during MS/MS a peptide can be trapped and sequenced giving unique information about its identity.

Another drawback of MALDI-TOF-MS is encountered when any degree of quantitative analysis is required [216, 404, 406-407]. This restriction is the result of multiple contributing factors, including variation in the ionisation efficiency due to non-homogeneous application of the peptide sample to the MALDI plate by spotting, the quality of the sample crystal, initially crystal form of samples and therefore only a diluted amount was ionised, ion suppression effects due to the presence of residual surfactant and salts, resulting in selective ionisation of certain peptide species over others depending on their amino acid sequence. These peptide species exhibit typically poor ionisation potential and small peptide masses are beyond the optimum mass range of the MALDI-TOF instrument due to interference from matrix. Unfortunately, the limitations discussed above restrict not only the quantitative analysis but also the ability to trace PTM modification, which is required for detection the probe-modified peptide fragment.

To circumvent these limitations and aim for a more sensitive method, identification of labelled bands with different probes was reproduced using the most recently developed technology, LC-ESI-MS/MS with an LTQ-Orbitrap-Velos as mass analyser and with ETD as the fragmentation method. The enriched bands visualised by Coomassie staining prepared in section 6.1.2 were excised from the 1D-gel, covered with 50 mM ammonium bicarbonate solution and frozen on dry ice by the author before sending to the University of Leicester Proteomics Facility, UK. There, these gels pieces were digested with trypsin and prepared for MS/MS analysis by Dr. Andrew Bottrill, a senior experimental officer at the University of Leicester, UK.

The resulting tandem mass spectra were analysed using Mascot (Matrix Science, London, UK) with *Clostridium difficile* as selected Taxonomy and the Trembl_39.6 database selected for the search. Furthermore, modification of cysteine by the probes was specified in Mascot as variable modifications. Scaffold (version Scaffold_3_00_03, Proteome Software Inc., Portland, OR) was used to validate MS/MS based peptide and protein identifications. Peptide identifications were accepted if they could be established at greater than 95,0% probability as specified by the Peptide Prophet algorithm [408] and contained at least 2 identified peptides.

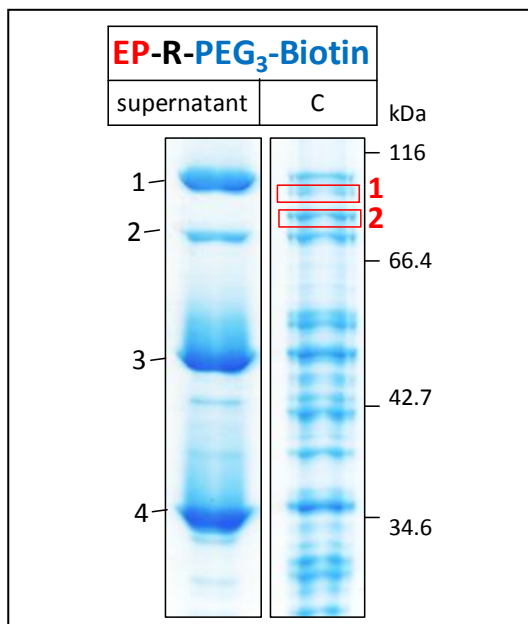


Figure 6. 6: Identified bands on Coomassie stained gels. Samples were prepared and enriched as described in section 6.1.2. Bands 1 and 2 indicated as red boxes were excised. In-gel digestion with trypsin and subsequent LC-MS/MS analysis were carried out by Dr. Andrew Bottrill, University of Leicester, UK.

The first attempt re-explored the labelled bands of EP-R-PEG₃-Biotin, therefore two bands (indicated in red box, in **Figure 6. 6**) were excised based on the knowledge about the labelling of this ABP acquired in various stages of this study. There were 13 proteins identified for band 1

(**Figure 6. 6**) and 15 proteins for band 2; however, by spectral count only one or two proteins were major species, while other proteins were more likely trace species that may co-migrate in the gel with the prominent proteins. A summary of the spectral counts of both bands can be found in **Table 6. 2**.

Intriguingly, in the first band two major proteins were found: Cwp84 and Cwp13 (also known as CD1751), both of which are putative cysteine proteases. A few examples of the proteolytic activity of Cwp84 had been previously reported in the literature and the function of both Cwp84 and Cwp13 were unknown at the time this work was performed. According to the spectral count Cwp13 seemed less abundant than Cwp84 in band 1, while in band 2 Cwp84 was the only prominent species and other proteins including Cwp13 were insignificant. Comparing this result with the previous results generated by Dr. Packman

using MALDI-TOF-MS, it was clear that Cwp84 is present in both bands, indicating the consistency of results regardless of the approach used. However, a few important differences were found. Besides the increased amount of Cwp13 in the first band, the sequence coverage of the proteins identified using LC-MS/MS – LTQ-Orbitrap was greater (almost double) compared with MALDI-TOF-MS; sequence coverage increased from 35% to 64% for the top band and from 41% to 77% for the bottom band in the case of Cwp84. This improvement was expected, since LC-MS/MS by LTQ-Orbitrap is more sensitive and has better dynamic range for data acquisition than MALDI-TOF-MS (see section 1.5.1). Unfortunately, this improvement was not enough to detect the peptide modified with the probe. Generally, at least 80% sequence coverage is desirable for a good chance of identifying modified sites [409], if the location of the modification is not known. In our case, the location of the probe modification is known (peptide containing the active cysteine residue C116), but this peptide was not found in the identified sequences, which might be due to less well ionisation property and consequently this peptide is not detectable in MS.

Identified proteins	Sequence coverage (%)		Quantitative Value		Unique peptides		Unique spectra		Assigned spectra	
	B 1	B 2	B 1	B 2	B 1	B 2	B 1	B 2	B 1	B 2
Cwp84	64	77	83	134	36	49	43	59	65	164
Cwp13	37	13	40	7	23	6	24	6	30	6
SlpA	30	18	17	7	13	8	13	9	13	9
Cpw66	15	27	8	7	6	9	6	9	6	9
CD2767	27	24	14	9	11	11	11	11	11	11
CD2784	7	19	5	7	4	9	4	9	4	9

Table 6. 2: Spectral counts of identified proteins of bands in Coomassie stain gel Figure 6. 6A. B 1= band 1, B 2 = band 2.

6.4 In-depth proteomic study using LC-MS/MS with the wild type 630

The appearance of Cwp84 in the top band and the presence of Cwp13 raised further questions. Did epoxysuccinyl ABPs intercept these putative cysteine proteases at multiple stages of their self-processing, assuming that they undergo a self-processing like that of recombinant Cwp84 [410], and therefore label different sizes of Cwp84 (and maybe Cwp13 as well)? If so, does inactivation occur in an ABP structure-dependent fashion, as only ABPs with arginine at the AA¹ position seemed to grant the strong double-labelled band effect? Furthermore, why do inactive ABPs also label proteins in the lower band?

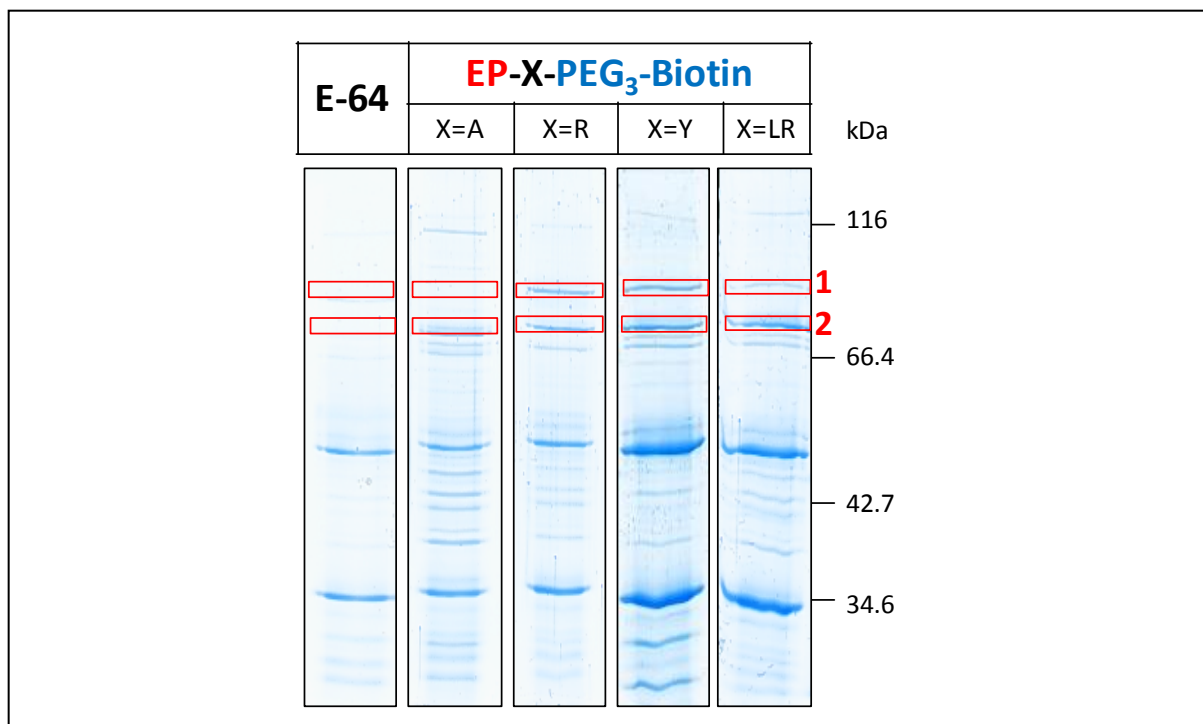


Figure 6. 7: Identified bands on Coomassie stained gels. Samples were prepared and enriched as described in section 6.1.2. Bands 1 and 2 indicated as red boxes were excised. In-gel digestion with trypsin and subsequent LC-MS/MS analysis were carried out by Dr. Andrew Bottrill, University of Leicester, UK

Seeking an answer to those questions, the proteomic experiment was re-run with a panel of ABPs. EP-R-PEG₃-Biotin was chosen as an interesting candidate due to its ability to generate two strongly labelled bands, whilst EP-Y-PEG₃-Biotin was selected as the most potent ABP and EP-A-PEG₃-Biotin as an inactive ABP. EP-LR-PEG₃-Biotin was added to compare the labelling between moderately active and potent ABPs. E-64 treated samples served as controls. The experiment was implemented as described above with one additional step: before loading onto the MS instrument, the tryptic digested samples were mixed with a standard peptide digest (ADH- alcohol dehydrogenase from yeast – with Uniprot accession P00330) as an internal standard to aid with quantitative analysis, and to correct for run-to-run variations in the instrumentation. Two bands at the same gel location (labelled as 1 and 2 and corresponding to the two labelled bands in samples treated with EP-R-PEG₃-Biotin) were excised for each sample. Unfortunately, protein amount loaded on the Coomassie stained gel was not equal between samples (**Figure 6. 7**). Judging by the stained protein bands, one could speculate that sample X=Y and LR had significantly more total protein than other samples. This phenomenon could be a result of variability in enrichment ability of each ABP or cell lysis susceptibility, and non-specific co-pull down could increase in the case

of X=Y and LR. To prevent this problem, more stringent wash and loading normalisation prior SDS-PAGE could be applied in future experiments. Because of the unequal loading, quantitative evaluation of the identified proteins could not follow. Importantly, the quantitative aspects of the data illustrated below are only indicative, and require future repetition and refinement. However, identification of proteins appears secure.

The quantitative values generated by Scaffold are given in **Table 6. 3**. The relatively similar quantitative value of ADH in most samples (excepting the E-64 treated sample) demonstrated the effective injection of an equal amount of sample by the auto sampler in each run, variation was likely related with the unequal amount of protein in each lane. Assuming that this variation is small and negligible , in general Cwp84 seemed to be enriched in all samples treated with ABPs.

Identified Proteins	Molecular Weight	E-64		EP-R-PEG ₃ -Biotin		EP-Y-PEG ₃ -Biotin		EP-A-PEG ₃ -Biotin		EP-LR-PEG ₃ -Biotin	
		band 1	band 2	band 1	band 2	band 1	band 2	band 1	band 2	band 1	band 2
Cwp84	87 kDa	13	50	64	159	74	241	65	130	87	222
Cwp13	87 kDa	4	8	46	1	40	48	57	80	41	37
SlpA	76 kDa	26	32	46	36	75	35	26	17	41	25
Slp2	66 kDa	11	40	11	19	22	19	17	14	16	20
Cwp66	67 kDa	1	40	1	14	19	27	15	30	14	28
CD2784	73 kDa	1	1	1	8	6	3	4	1	3	1
ADH	37 kDa	34	51	32	24	27	24	33	24	27	39
Trypsin	24 kDa	66	100	50	67	75	65	70	59	64	63

Table 6. 3: Quantitative value of identified proteins. ADH = yeast alcohol dehydrogenase standard.

Identified Proteins	E-64		EP-R-PEG ₃ -Biotin		EP-Y-PEG ₃ -Biotin		EP-A-PEG ₃ -Biotin		EP-LR-PEG ₃ -Biotin	
	band 1	band 2	band 1	band 2	band 1	band 2	band 1	band 2	band 1	band 2
Cwp84	0.38	0.98	2.00	6.63	2.74	10.04	1.97	5.42	3.22	5.69
Cwp13	0.12	0.16	1.44	0.04	1.48	2.00	1.73	3.33	1.52	0.95
SlpA	0.76	0.63	1.44	1.50	2.78	1.46	0.79	0.71	1.52	0.64
Slp2	0.32	0.78	0.34	0.79	0.81	0.79	0.52	0.58	0.59	0.51
Cwp66	0.03	0.78	0.03	0.58	0.70	1.13	0.45	1.25	0.52	0.72
CD2784	0.03	0.02	0.03	0.33	0.22	0.13	0.12	0.04	0.11	0.03
ADH	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Trypsin	1.94	1.96	1.56	2.79	2.78	2.71	2.12	2.46	2.37	1.62

Table 6. 4: Relative abundance of identified proteins to the ADH control. ADH = yeast alcohol dehydrogenase standard.

6.5 Cwp13 may be an off-target of epoxysuccinyl ABPs

Given that Cwp84 is the target protease mediating SlpA cleavage, it is clear why active ABPs possess a corresponding strong labelled band at apparently 76 kDa on the Bis-Tris gel. However, it was unclear why inactive ABPs seemed to label the same band even though they did not functionally inhibit Cwp84 (see labelling effect in Chapter 5). Proteomic analysis described in section 6.3 also revealed the presence of Cwp13 in the identification after enrichment, leading to the hypothesis that the cysteine protease Cwp13 could be an off-target of epoxysuccinyl inhibitors/ABPs targeted by both inactive and active ABPs; because both proteases are very similar in size they appear at the same position on the gel.

To verify this hypothesis, we employed antibody against recombinant Cwp84 which was raised in mice. Recombinant Cwp84 was obtained by transforming plasmid pET-28a orf 4H-1, containing the cysteine protease domain of Cwp84, which was kindly donated by the Fairweather's laboratory, into *E. coli* Rosetta cells. These cells were grown and induced for protein expression using a pre-optimised protocol. After purification using nickel affinity chromatography, the concentrated recombinant Cwp84 was sent to School of Biological Sciences (Royal Holloway University of London, Surrey, UK) for raising of antibodies in mice.

C. difficile was treated with dual fluorescent ABPs overnight, and a glycine extraction performed as previously described to obtain S-layer fractions. The labelled proteins were enriched via pull-down on streptavidin Dynabeads and labelling was demonstrated using in-gel fluorescence, while the presence of Cwp84 was revealed using Western blot. As seen in **Figure 6. 8**, strong labelling was observed in all ABPs before and after enrichment. However, Cwp84 was only significantly enriched by the active ABPs and not by the inactive probes. Furthermore, multiple bands of Cwp84 observed in the enrichment suggested the intercepting of self-processing intermediates of this protease. Interestingly, the apparent labelled Cwp84 band by the EtEP-RK(Biotin)G-PEG₃-TAMRA, is slightly shifted in higher molecular weight compared to other bands. This observation will be discussed in a later session of this chapter.

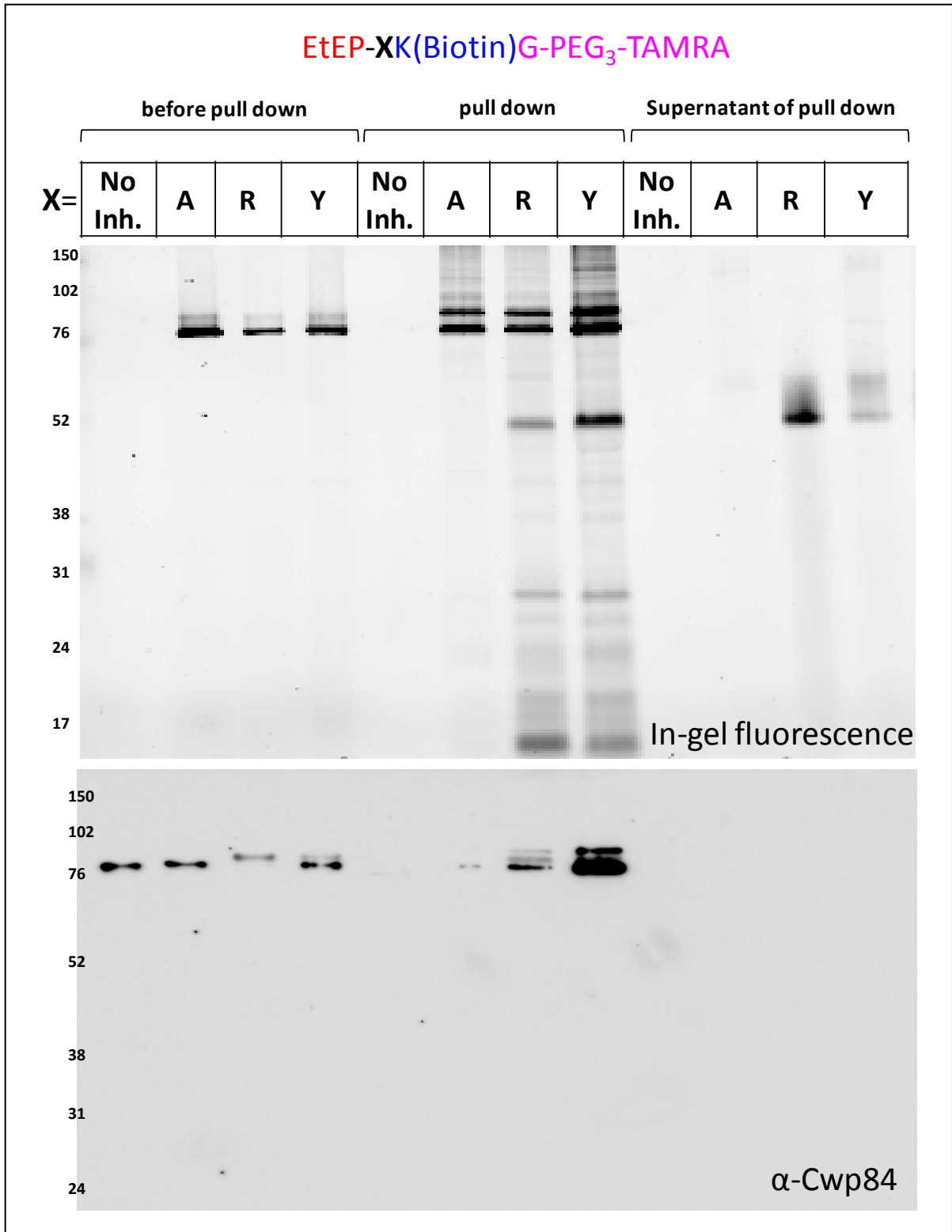


Figure 6. 8: Labelling effect and identification of Cwp84 from epoxysuccinyl ABPs. *C. difficile* 630 was treated with 50 μ M of ABPs overnight, followed by isolation of the S-layer fraction as described in section 3.5. A pull-down step was added to enrich the labelled proteins prior SDS-PAGE analysis. Visualisation was performed using in-gel fluorescence and Western blot against Cwp84.

6.6 Effect of Cwp84 on other cell wall proteins

At the time this work was performed, Kirby *et al.* [411] had reported the genetic knock-out of *cwp84* in *C. difficile* using the Clostron system. This work was repeated and further developed by Dr. de la Riva Perez, a postdoctoral researcher in Neil Fairweather's laboratory, Imperial College London, UK, providing access to several mutants using in the work of this session. Colony morphology changes and effect in cell integrity in the 630 Δ *cwp84* mutant [411] indicated a significant change in the S-layer and cell wall composition. Aside from the major effect on SlpA, there are other cell wall proteins present as minor components in the S-layer which could be also affected. Building on the initial work of Dr. de la Riva Perez investigating the 630 Δ *cwp84* strain, the effect of inhibitor on other cell wall proteins was investigated. Interestingly, treatment of wild type strain 630 with inhibitors possessing different levels of potency also resulted in shedding of Cwp66 and Cwp2 in the culture supernatant. Furthermore, shedding behaviour seemed to depend on the potency of the inhibitors (**Figure 6. 9**).

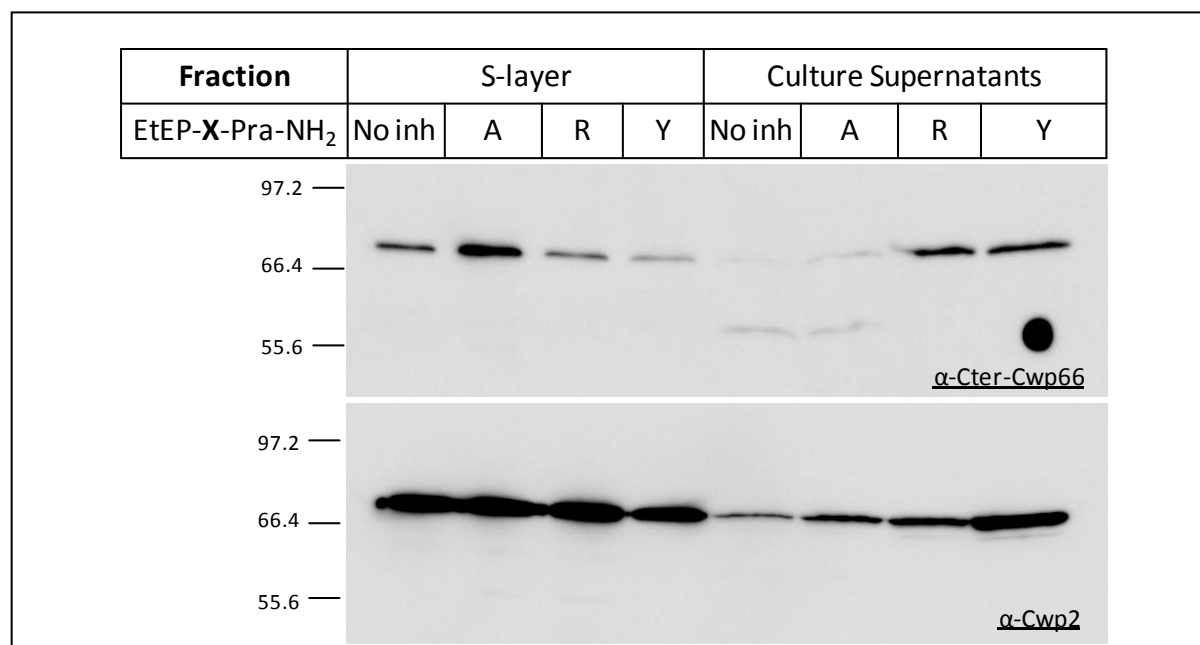


Figure 6. 9: Effect of Cwp84 on Cwp66 and Cwp2. *C. difficile* 630 were treated with compound at 100 μ M overnight, followed by isolation of the S-layer fraction as described in section 3.5 and culture supernatant concentration. Visualisation was using Western blot against Cwp66 and Cwp2 as indicated. Note: Black spot on Western blot against Cwp66 is a random signal.

6.7 Proteomic studies with a *C. difficile* 630 Δ *cwp84* mutant

Revisiting the labelling effect of epoxysuccinyl ABPs in wild type 630, it was noticed that the labelled Cwp84 signal was very strong. Consequently, this signal might prevent the

detection of possible other less strongly labelled targets. In this section, a hunt for those potential hiding targets was attempted by treatment of 630 Δ cwp84 with a series of biotinylated ABPs. However, the NeutrAvidin-HRP blot of the overnight treated samples revealed no further labelling, except the strong bands of the positive control treated with the wild type 630.

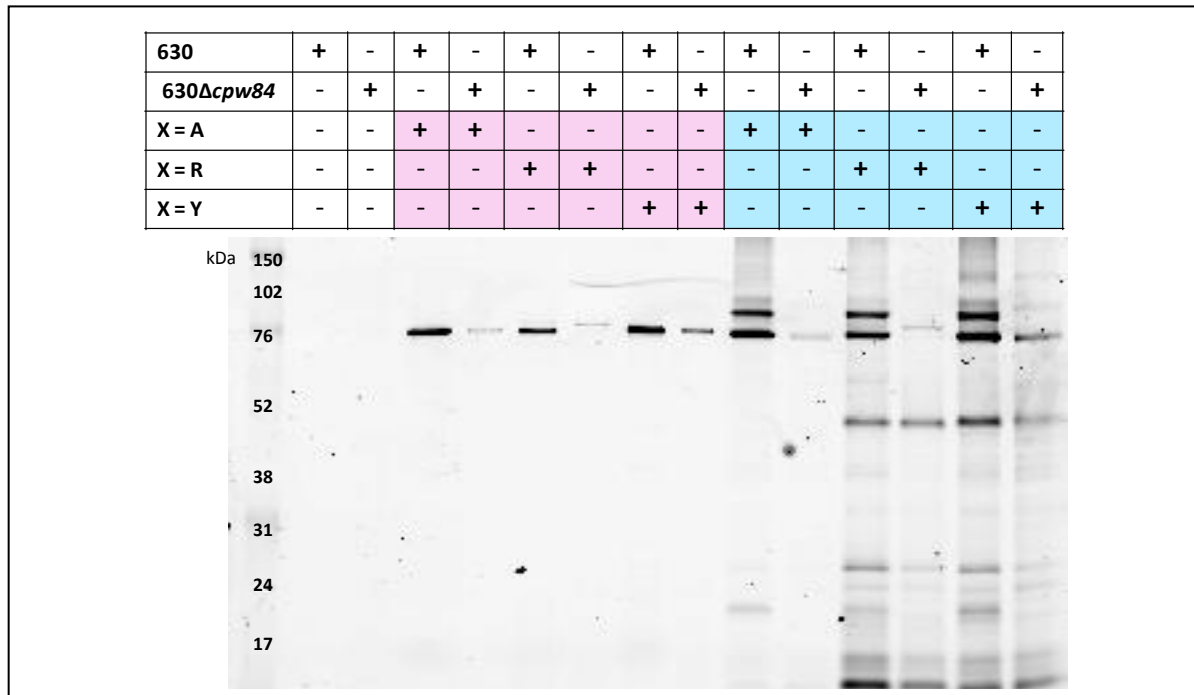


Figure 6. 10: Labelling of fluorescent ABPs in *C. difficile* 630 and 630 Δ cwp84 mutant visualised by in-gel fluorescence. 630 was treated with 50 μ M ABPs overnight, while 630 Δ cwp84 mutant was treated with 25 μ M ABPs. Whole cell bacteria were boiled in loading buffer for SDS-PAGE analysis. Single-label ABPs EtEP-X-PEG₃-TAMRA were shown in pink and dual-label ABPs EtEP-XK(Biotin)G-PEG₃-TAMRA were shown in blue.

It was difficult to locate the exact position of the labelled bands on the Coomassie stained gel, so gel slices were cut to include all proteins between the bands for SlpA and Cwp2, as shown in **Figure 6. 12A**. Each gel slice was divided into three further smaller bands and treated separately as independent samples to enhance coverage [412]. LC-MS/MS services were provided by Dr Andrew Bottrill as described in the previous section. A pre-digested standard, alcohol dehydrogenase from yeast ADH, was added post-digest to allow later quantitative evaluation. Each sample had three technical replicates and the spectral counts of each run were added together to give a final value minimising the impact of run-to-run variation on subsequent quantification.

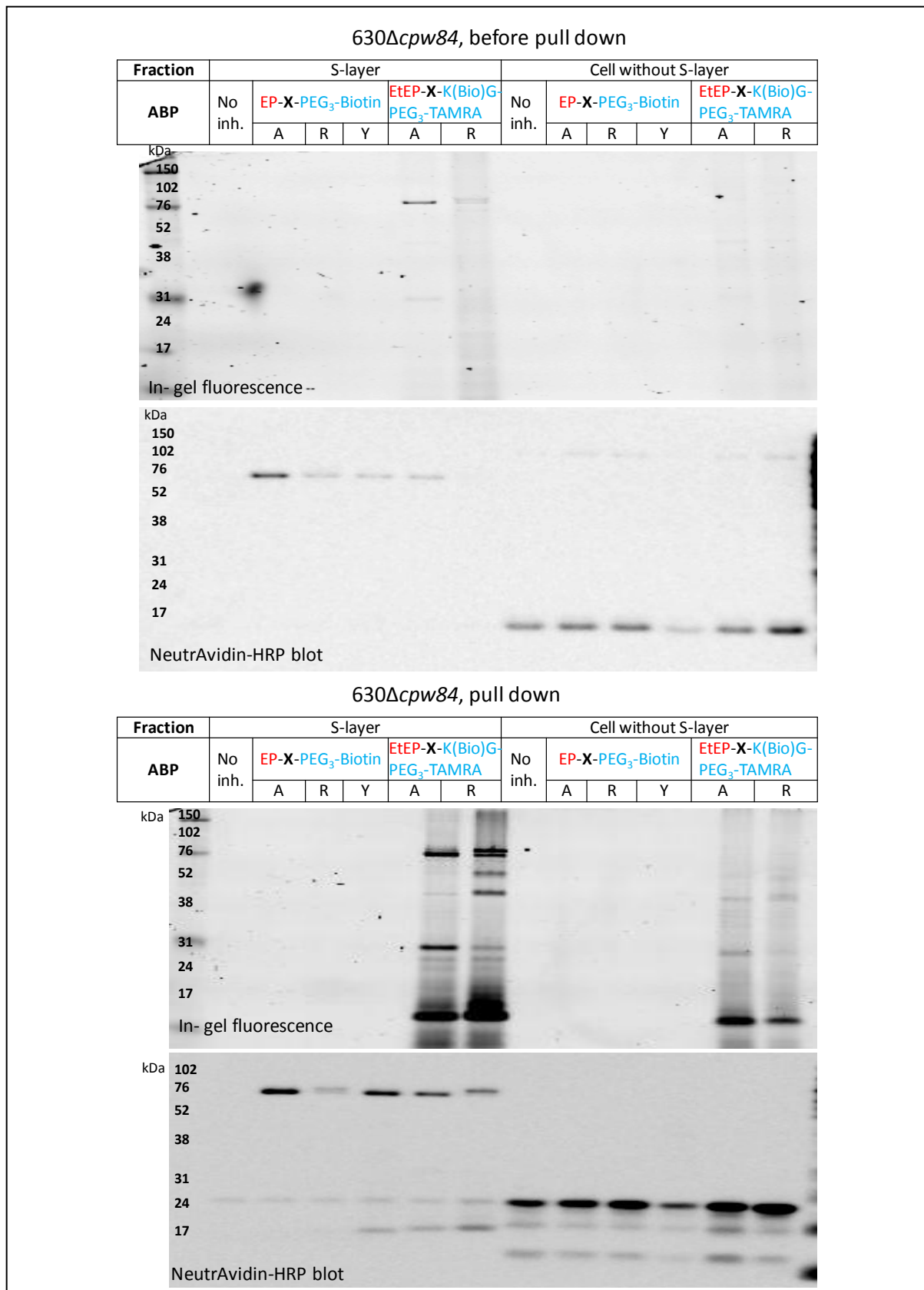


Figure 6. 11: Enrichment of proteins labelled by biotinylated and fluorescent ABPs in 630Δcpw84 mutant, visualised by in-gel fluorescence and NeutrAvidin blot as indicated. Knock-out mutant was treated with ABP at 25 μM overnight, followed S-layer fraction isolation and cell fraction preparation as described in section 3.5. Labelling is shown before and after the enrichment.

Identification of peptides against the database revealed the presence of SlpA, Cwp2, Cwp66, Cwp6 and Cwp13 as the most abundant cell wall proteins across the excised gel slices. Relative abundance of the identified cell wall proteins was calculated in correlation first to the control ADH, and then to the uninhibited control sample, as summarised in **Figure 6. 12B**.

As expected, SlpA and Cwp2 were by some margin the most abundant proteins identified, although there is some variation between runs. Cwp13 was enriched significantly (ca. 3-fold) more than other proteins in EtEP-**AK**(Biotin)G-PEG₃-TAMRA and EtEP-**RK**(Biotin)G-PEG₃-TAMRA relative to control. Furthermore, other cell wall proteins were found to co-purified as well including Cwp6 and Cwp66, which tend to be significantly enriched by EtEP-**RK**(Biotin)G-PEG₃-TAMRA compared to EtEP-**AK**(Biotin)G-PEG₃-TAMRA.

Because the NeutrAvidin-HRP blot might not be sufficiently sensitive due to the practical issues discussed in section 5.1.2, fluorescent ABPs were tested again in both wild type 630 and 630Δ*cwp84*, and run side by side for comparison. A relatively weakly labelled band was found in all 630Δ*cwp84* samples at the same position as Cwp84 in wild type 630. Furthermore, a small shift of this band to higher apparent molecular weight was observed when 630Δ*cwp84* treated with EtEP-**RK**(Biotin)G-PEG₃-TAMRA in agreement with all labelling observations with ABPs bearing arginine at the AA¹ position. It should be noted that the loading on the gel was not normalised between wild type and mutant samples, and therefore samples with wild type 630 contained more protein due to better growth. The effect of non-normalised loading is reflected in the different intensities of labelled bands in both strains. However, the presence of the labelled bands in 630Δ*cwp84* was confirmed in a pull down experiment repeat. The same phenomenon was seen with biotinylated ABPs suggesting that separating the cell in fractions resolved the practical problem mentioned at the beginning of this section.

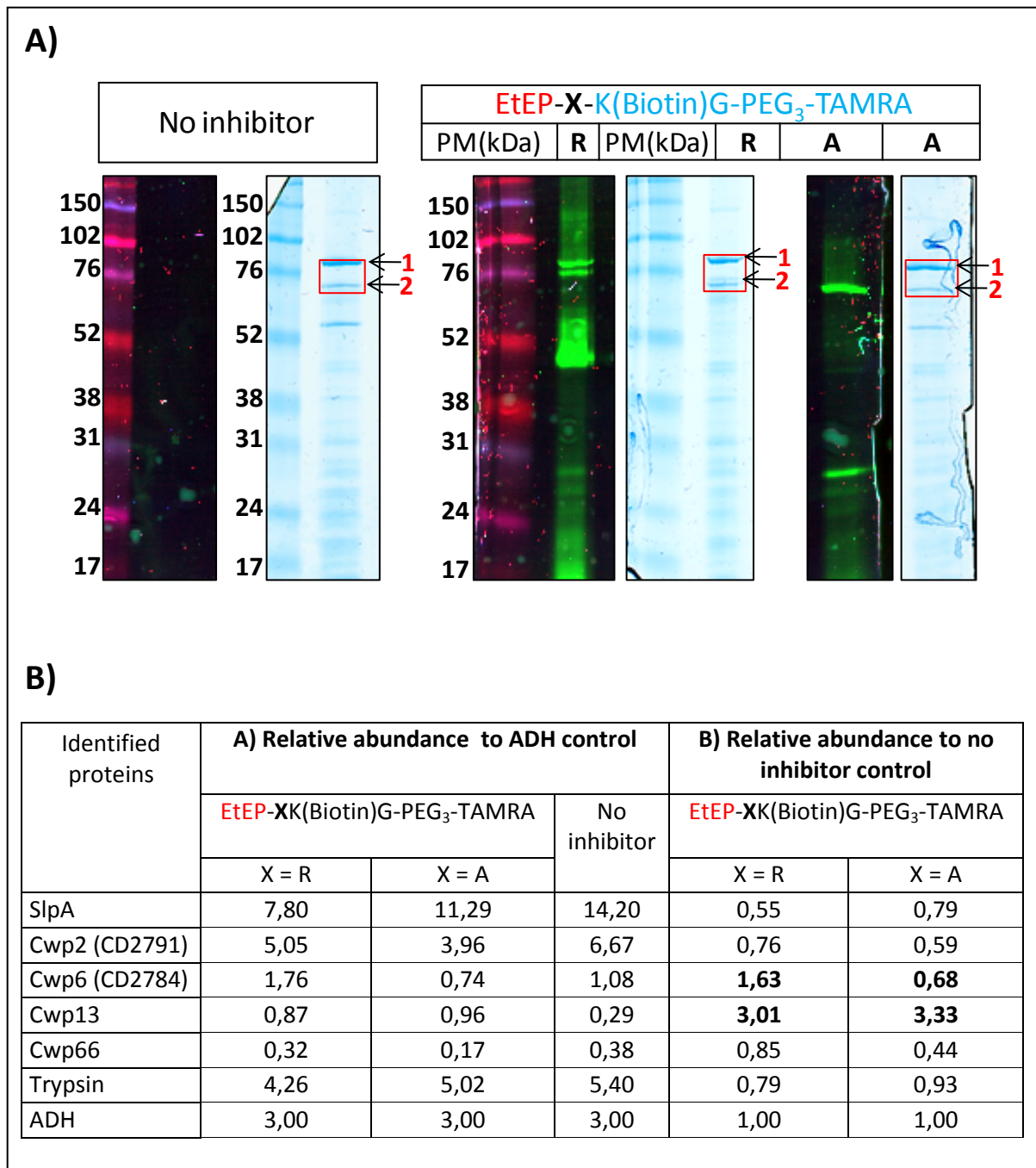


Figure 6. 12: A) Coomassie stained gel and in-gel fluorescent of glycine extracts labelled by dual fluorescent ABPs. *630Δcwp84* mutant was treated with ABPs at 25 μ M overnight, followed by glycine extraction as described in section 2.1.1. Red box indicated the part of the gel lane excised for proteomic analysis and labelled bands are: 1) SlpA, 2) Cwp2. B) Table of proteomic analysis results. Relative abundances were calculated based on the total spectra counts of each proteins.

6.8 Conclusion

This chapter has described the MS-based proteomics of labelled proteins by epoxysuccinyl ABPs. Starting from *in vitro* enrichment of labelled proteins after *in vivo* labelling, the pull down protocol was improved by replacing NeutrAvidin agarose resin with Streptavidin

Dynabeads. In general, the enrichment showed that active ABPs enriched more proteins than inactive ABPs. This tendency was also found for certain co-purified proteins.

Proteomic analysis after enrichment was first applied using MALDI-TOF-MS and PMF, where only one putative cysteine protease, Cwp84, was found, corresponding to the labelled band. Furthermore, a number of proteins were co-purified during the enrichment process, including Cwp66, SlpA, Cwp19 (CD2767), CD2797 and Cwp6 (CD2784). LC-MS/MS was found to be more sensitive and useful in terms of identifying multiple proteins in a mixture and providing data for relative quantification. Re-exploration of proteomic experiments using LC-MS/MS reveal the identity of another cysteine protease, the Cwp13, a homologue of Cwp84. Some of the co-purified protein species were also identified as traces in the Cwp84 and Cwp13 bands by LC-MS/MS. This observation could be properly related to many reasons including distribution of proteins over lane of the gel, sensibility of LC-MS/MS system.

In-depth proteomics analysis using the spectral counting method was applied and improved during the work of this chapter. However, due to some inherent problems such as inconsistencies in gel loading, these experiments require repetition and optimisation. For example, the enrichment process could be optimised with more stringent washes to remove as much of the non-specific binding as possible, and by using on-bead digest for 'gel-free' proteomic experiments rather than in-gel digest to avoid inconsistency related to gel loading, protein migration on gel, altered behaviour of proteins due to modification, or the excised band not including all proteins of interest. Above all, the on-bead digest method would be very useful for quantification of S-layer proteins that tend to undergo auto-processing and distribute over several molecular weight ranges on the gel.

Despite the limitations of the applied method, the results of this chapter still enrich our understanding of the influence of proteases in S-layer processing of *C. difficile*. First, the identification of the two putative cysteine proteases confirmed the success of ABPP application to *C. difficile* system. Second, the results described in this chapter contribute further evidence for potential functions of the identified proteases, which will be discussed in the chapter 8.2.

One interesting point is the SAR across inhibitors/ABPs (described in chapter 4) and the genetic study (described in the work of Dr. de la Riva Perez): both studies demonstrated the

importance of the AA¹ position of inhibitors/ABPs or the corresponding P2 position of the substrate. Combining the significant enrichment of Cwp84 by active ABPs and less effective enrichment by inactive ABPs (see section 6.1) with the proteomics data showing Cwp84 enriched when wild type 630 is treated with active ABPs and significant enrichment of Cwp13 in 630Δ*cwp84* strain, these results suggest that active ABPs label both proteases.

Another interesting finding is the mysterious property of ABPs bearing an arginine in the AA¹ position that causes “the double band effect”. In the proteomic study with wild type 630, the upper labelled band contained almost as much Cwp13 as Cwp84, but the amount of Cwp13 in the bottom band was insignificant for EP-R-PEG₃-Biotin (see section 6.4). In addition, the similar relative abundance of Cwp13 when 630Δ*cwp84* was treated with ABPs having AA¹ as arginine and alanine (section 6.7) indicated that arginyl ABPs labelled Cwp13 as strongly as inactive alanyl ABPs. There appears to be a structure-activity effect for the arginyl ABPs that is not yet fully explicable.

Although proteomic data strongly indicate specific labelling of Cwp13 by both active and inactive ABPs, the function of Cwp13 and any the potential impact of this protease in S-layer processing (e.g.: Cwp13 might process Cwp84 or it might be involved in the host-pathogen interactions) remain unclear. Future work could focus on investigation of Cwp13 to provide better understanding of S-layer biogenesis of *C. difficile*.

As part of this work, the effect of Cwp84 on other cell wall proteins was also investigated. Affecting the activity of this protease either by chemical inhibition or gene knock-out tended to increase the shedding of Cwp2 and Cwp66 into the culture supernatant. This increased the shedding could be an indication of up-regulation of these proteins or somehow related with the stabilisation of these proteins causing their detection on Western blot. In addition, proteomic analysis of 630Δ*cwp84* strain also indicated a larger enrichment of Cwp66 and Cwp6. However, much of these data is preliminary and further work will be required to provide a clearer picture of the effects of deleting *cwp84* and the effects of inhibitors on Cwp84 and Cwp13.

Chapter 7 Exploring the scope and potential antibacterial activity of irreversible cysteine protease inhibitors in *C. difficile*

The discovery of the first inhibitors of SlpA processing, and their development into epoxysuccinyl warhead-containing inhibitors/ABPs, led to the identification of two cysteine proteases: Cwp84 and Cwp13. It was subsequently verified that only Cwp84 was responsible for SlpA cleavage, while its homologue Cwp13 was unable to complement its functions. Furthermore, in addition to the genomic studies by various groups [276, 413], the SAR of epoxysuccinyl inhibitors with their targets developed in this study suggested strongly that Cwp84 and presumably also Cwp13 belong to the papain subclass of cysteine protease.

Cysteine protease inhibitors/probes were widely studied and applied to identify novel cysteine proteases [50, 60-61, 150, 377]. Among those selective cysteine protease inhibitors/probes, various types of electrophilic warhead appended to a specific peptidyl moiety were found to target different sub-classes of cysteine proteases [414]. In this chapter, a number of inhibitor classes bearing different electrophilic warheads were designed and investigated with the objective of exploring their scope and potential antibacterial activity as well as identifying further potentially unknown cysteine proteases in *C. difficile*. These inhibitors were categorised according to both the chemical properties of the warhead as well as where it is positioned in the molecule (N-terminus or C-terminus of the peptidic chain).

7.1 Warheads against cysteine protease inhibitors

Cysteine proteases can be differentiated into three major structurally distinct classes: papain-like (e.g. cathepsins), interleukin-1 β converting enzyme (ICE)-like (e.g. caspases) and picorna-viral like (similar to serine proteases with cysteine substituting serine) proteases [415]. Having developed epoxysuccinyl inhibitors, the next major aim of this study was to design novel inhibitors/probes in order to explore the potential S-layer processing inhibition and antibacterial activity of these irreversible inhibitors. Furthermore, variation of the warhead might lead to labelling of additional potential cysteine proteases as well as serine proteases in *C. difficile* that can accommodate and react with such compounds. A strategy commonly used for designing specific protease inhibitors is the substrate mimetic approach, which is based on cleavage site of the protease substrate and appending a reactive warhead to that substrate moiety. Similar to epoxysuccinyl inhibitors/probes, the design was centred on the specificity element bearing a peptidyl mimic of a particular length so as to fit in the protease's active site and a reactive electrophilic warhead to trap irreversibly the nucleophilic active centre of the proteases. Common warheads include aldehydes, halomethyl ketones, diazomethyl ketones, acyloxymethyl ketones, epoxides and Michael acceptors. These warheads, with the exception of epoxides (which have been explored in detail in previous chapters) and aldehydes (discussed in Chapter 2), will be discussed in the following section.

Halomethyl ketone (HMK)

Peptidyl chloromethyl ketones (CMK) were first developed for serine proteases [416-417], however, it was found subsequently that this type of inhibitor was also potent against cysteine proteases [417-418]. Various studies have established the different mechanisms of inhibition by HMK to serine and cysteine proteases (**Figure 7. 1**). In the case of serine proteases, HMKs form two covalent bonds with the protease: the first is formed with the active site histidine residue [419-420] while the second is generated between the nucleophile catalytic serine oxygen and the ketone carbonyl carbon, leading to the formation of a tetrahedral adduct [421]. In the case of cysteine proteases, HMKs alkylate the active site cysteine residue, forming a covalent thioether bond [422-423]. Altering the peptidyl moiety of the inhibitors can lead to enhanced specificity against individual serine or

cysteine proteases, while the variation of the strength of HMK has been shown to improve both inhibitor potency (F>I>Br>Cl) and selectivity [129]. Fluoromethyl ketones (FMKs) are usually more specific for cysteine proteases, whereas CMKs inhibit both protease classes [129].

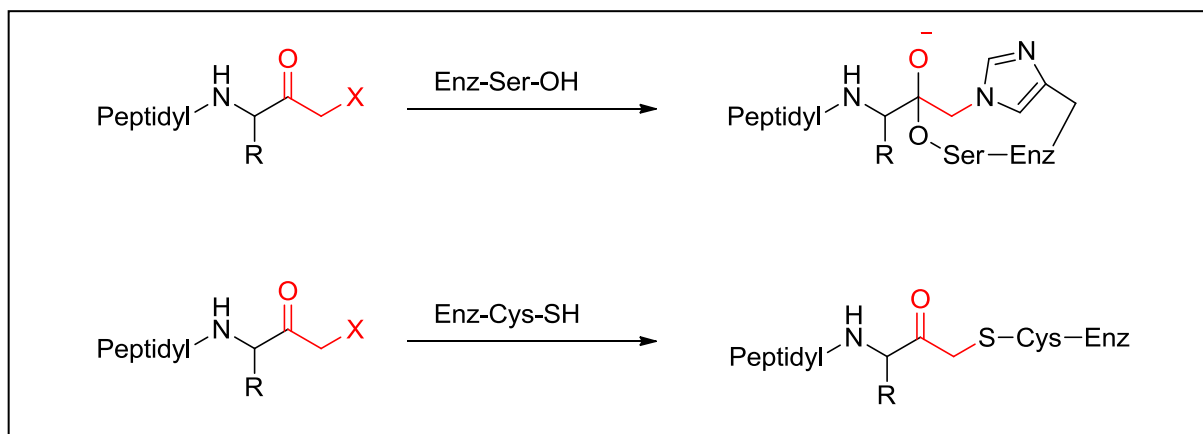


Figure 7. 1: Inactivation of serine and cysteine proteases by HMK inhibitors. (X = F, Br, Cl)

Diazomethyl ketone

Peptidyl diazomethyl ketones irreversibly inactivate cysteine proteases by alkylating the active site cysteine residue. They have been developed for some sub-classes of cysteine proteases including papain [424-425], cathepsins [426-428], and calpain [427, 429]. The proposed inhibition mechanism of diazomethyl ketones involves a proton transfer from the active site histidine residue to the methylene carbon of the inhibitor, and the alkylation of the active cysteine residue eliminates N₂. It is still unclear which of these processes occurs first.



Figure 7. 2: Inactivation of cysteine proteases by diazomethyl ketone inhibitors.

Acyloxymethyl ketone (AOMK)

AOMKs are analogues of HMKs that bear an ester group instead of a halogen atom, and were first designed by Krantz *et al.* [153, 430]. The major drawback of using HMKs is that

they possess an inherently reactive halomethyl functional group, which can lead to issues with broad reactivity. The exploitation of AOMKs in inhibitor design overcomes this design flaw by having a less reactive functional group than HMKs, and thus these compounds suffer from non-specific reactivity to a lesser extent. Similar to other types of inhibitors, AOMKs alkylate the active site cysteine residue to form a thioether ketone. AOMKs were found to inhibit cathepsins (B, L, S) [431-432], calpain [154], caspases [433] and other cysteine proteases [434]. The potency of an AOMK depends on the nature of the leaving group as well as the affinity of the peptidyl region.

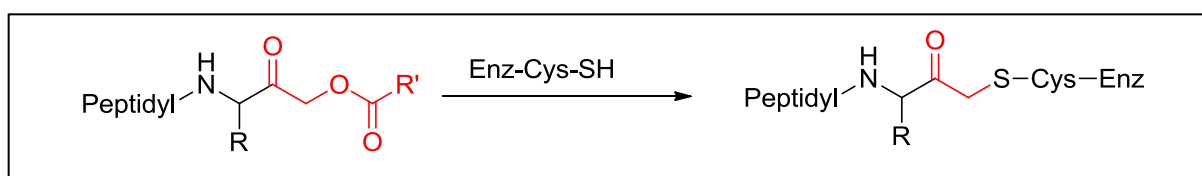


Figure 7. 3: Inactivation of cysteine proteases by AOMK inhibitors.

Michael acceptor (MA)

Peptidyl derivatives bearing an MA are generally specific cysteine protease inhibitors, found to inhibit cathepsins [149, 435-436], papain [436-438], calpains [149], cruzain [149, 439] and rhinovirus 3C proteases [62, 440]. They remain unreactive towards other nucleophiles, and require the catalytic system of cysteine proteases for activation. A number of studies have shown that the alkylation process takes place on the β -carbon of the α,β -unsaturated carbonyl moiety by the active site cysteine residue, followed by protonation of the α -carbon to form the thioester adduct [149, 436-437] (**Figure 7. 4**). Furthermore, the strength of the electron-withdrawing groups on the MA correlates linearly with the potency of the inhibitors.

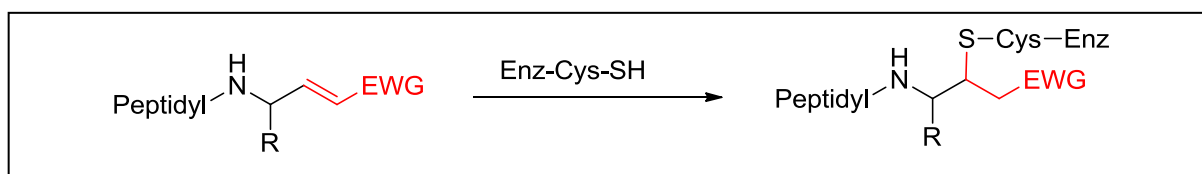


Figure 7. 4: Inactivation of cysteine proteases by MA inhibitors. R is the amino acid side chain and EWG can be varied with different electron withdrawing groups e.g.: vinyl sulfone ($-\text{SO}_2\text{R}$) to various carbonyl derivatives (COOR).

7.2 Inhibitor design

As detailed in previous chapters, an ABP usually consists of three moieties: the warhead, the specificity element and the tag. These three elements were considered separately in the design as follows.

The reporter unit

It was necessary to use some kind of reporter unit in order to detect the labelling of the probes, and this could occur either *via* direct attachment of a label to the inhibitors or indirectly *via* small chemical tag accessible for subsequent bioorthogonal ligation *in vitro*. However, when the *in vivo* assay was taken into consideration, a biotin direct label was chosen as the reporter for all inhibitors. There are two major advantages gained from using biotin in this manner: it is smaller than the fluorescent label, and attaching biotin has the potential to enhance the uptake of inhibitor into the cell *via* the biotin-mediated vitamin uptake pathways [441]. The biotin group was linked either via a PEG linker, as described in Chapter 5, or via incorporated on the side-chain of a lysine residue in the peptidic chain.

The specificity element

The aim of this study is to explore the S-layer inhibitory and antibacterial activity of irreversible inhibitors in *C. difficile*. Considering the substrate mimetic strategy, a panel of probes containing different peptide components were chosen based on their diverse physical properties and the knowledge of SlpA processing inhibition gained thus far: alanine, serine and threonine were chosen for the 'inactive' compounds whereas arginine, tyrosine and arginine-leucine dipeptide were chosen for the 'active' compounds.

The warhead

The warheads were attached either on the N-terminus or on the C-terminus of the peptidic specificity element (**Figure 7. 5**). Inhibitors with the warhead at the N-terminus were derived using a similar route to that exploited for the epoxysuccinyl inhibitors (see section 3.3). In contrast to introducing a ketone group upon warhead attachment (as for HMKs), the warhead was coupled onto the N-terminal amine hence forming haloacetamide unit. Since the electron withdrawing effect of amines is much weaker than that of ketones, these newly designed probes would be expected to have reduced potency in comparison to the ketones.

Nevertheless, these compounds would still retain moderate potency, and since inhibitor potency was not the main focus of this study haloacetamides and acyloxyacetamides were included as warheads.

Introducing the warhead onto the C-terminus of the peptidyl inhibitor during SPPS is synthetically more challenging. One useful method of achieving this however involves having the warhead preloaded on the solid support before peptide elongation. For this reason, the building block containing the warhead should comprise of at least two functional groups: one handle for attachment to the resin and another for appending the specificity element. In the case of the MA warhead, Fmoc-Ser was modified on its C-terminus to provide the corresponding MA. Once this building block was obtained, attachment onto the solid support could follow *via* the serine side chain while the peptide elongation could occur on the Fmoc protected amine as usual. In the case of the CMK and AOMK warheads, Z-alanine served as the starting point to generate Z-alanine-CMK or Z-alanine-AOMK, with the aim of introducing these compounds onto a solid support. Details of the synthetic route taken will be described in the following sections.

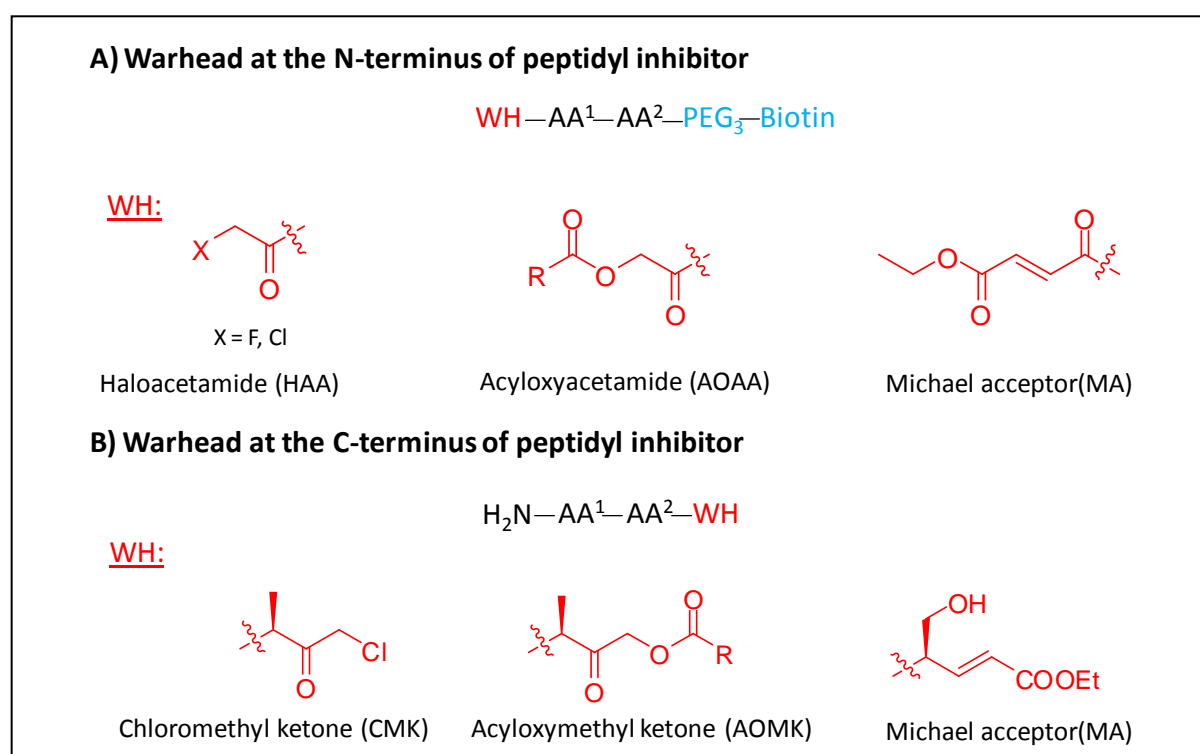
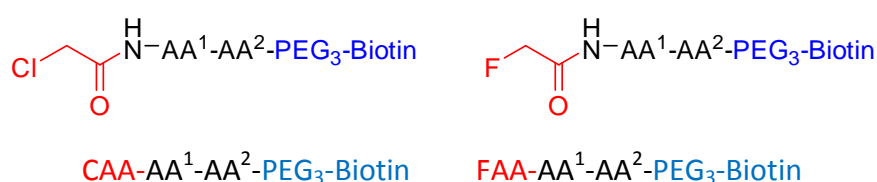


Figure 7. 5: Inhibitor design. The warhead was attached either on the N-terminus (**A**) or the C-terminus (**B**) of the peptidyl specificity element. The biotin label was introduced at the C-terminus *via* use of PEG-Biotin-NovatagTM resin in **A** or at the side chain of a lysine residue in the peptidyl specificity element in **B**.

7.3 Inhibitors with the warhead appended to the N-terminus

7.3.1 Synthesis

The generic route employed for the preparation of inhibitors with the warhead appended to the N-terminus of the specificity element is similar to that employed for the synthesis of the epoxysuccinyl inhibitors described in Chapters 3 and 5. The specificity element was prepared on the PEG-Biotin-NovatagTM resin utilising standard Fmoc/^tBu SPPS. Once the desired peptide chain had been made and still attached to the resin, the warheads were incorporated in a straightforward manner onto the peptide N-terminus (**Figure 7. 6**).



No.	WH	AA ¹	AA ²	Abbr. Name
TD 67	CAA	Leu	Arg	CAA-LR-PEG ₃ -Biotin
TD 68	CAA	Ala	-	CAA -A-PEG ₃ -Biotin
TD 69	CAA	Ser	-	CAA-S-PEG ₃ -Biotin
TD 70	CAA	Thr	-	CAA-T-PEG ₃ -Biotin
TD 71	CAA	Arg	-	CAA-R-PEG ₃ -Biotin
TD 72	CAA	Tyr	-	CAA-Y-PEG ₃ -Biotin
TD 73	FAA	Leu	Arg	FAA -LR-PEG ₃ -Biotin
TD 74	FAA	Ala	-	FAA -A-PEG ₃ -Biotin
TD 75	FAA	Ser	-	FAA -S-PEG ₃ -Biotin
TD 76	FAA	Tyr	-	FAA -Y-PEG ₃ -Biotin
TD 77	FAA	Arg	-	FAA -R-PEG ₃ -Biotin

Table 7. 1: Inhibitors with the haloacetamide warhead appended to the N-terminus of the specificity element.

To obtain chloroacetamide (CAA) warhead, the resin-bound peptide chain was treated with chloroacetyl chloride in the presence of *N*-methyl morpholine (NMM). Alternatively, the fluoroacetamide warhead (FAA) was obtained from the coupling of preactivated sodium fluoroacetate using HCTU and DiPEA, while the Michael acceptor (MA) warhead was derived from coupling with preactivated fumaric acid mono ethyl ester also using HCTU and DiPEA.

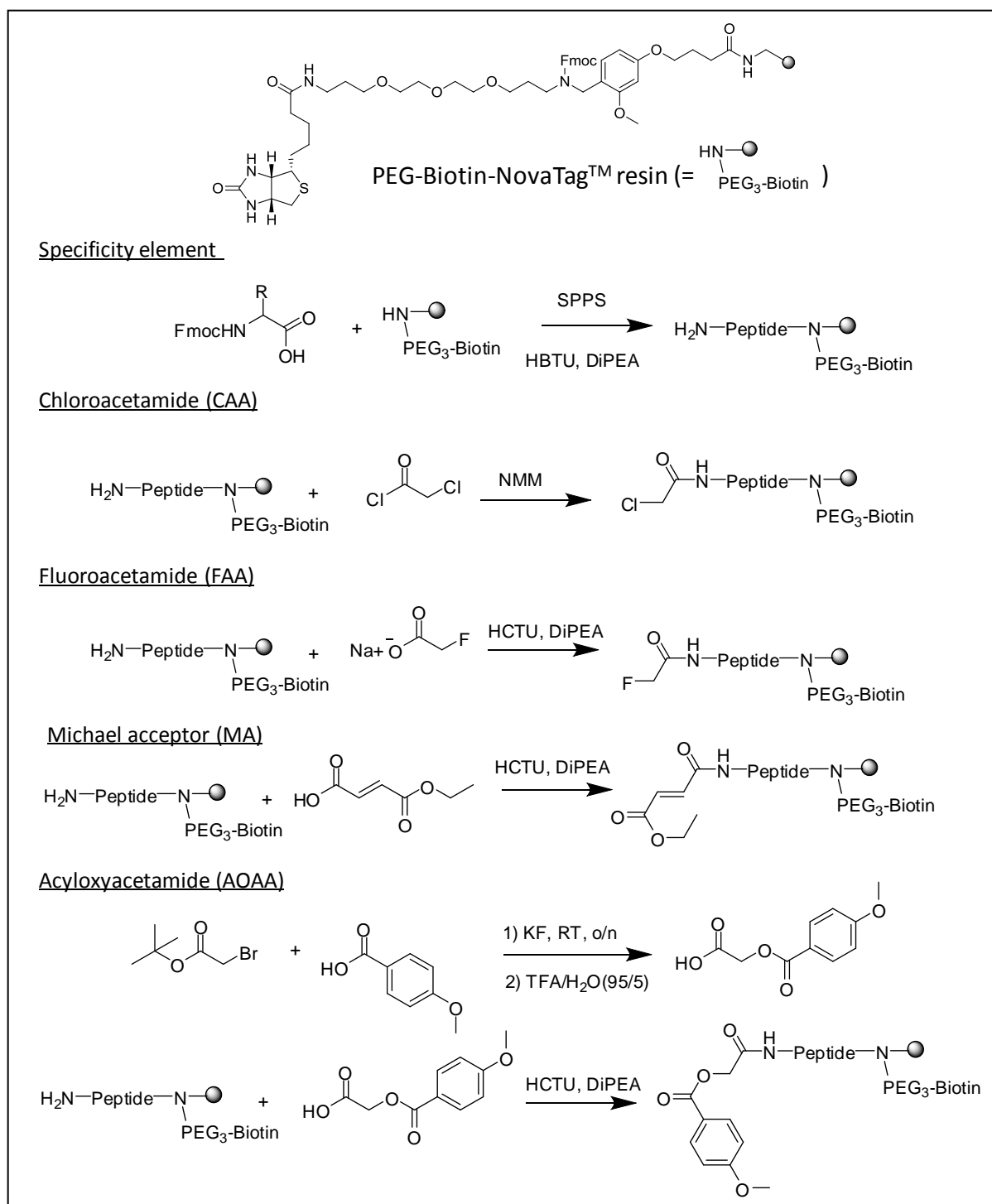
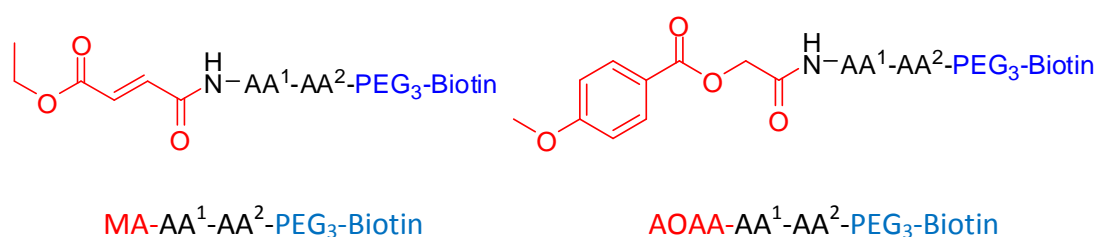


Figure 7. 6: Synthesis of inhibitors with warhead appended to the N-terminus. The specificity element was synthesised on the PEG-Biotin-NovaTag™ resin using standard SPPS with subsequent coupling of the desired warhead. *CAA warhead:* ClCOCH_2Cl (1.2eq.), NMM (2.5eq.), 1h, RT, yield 4-37%. *FAA warhead:* $\text{NaCOOCH}_2\text{F}$ (5eq.), HCTU (5eq.) and DiPEA (6eq.), 1h, RT, yield 6-15%. *MA warhead:* HOOCCHCHCOOEt (5eq.), HCTU (5eq.) and DiPEA (6eq.), 1h, RT, yield 10-39%. *AOAA warhead:* tBuAcBr (1eq.), 4-MeOBnCOOH (2eq.), KF (4eq.), RT, o/n; coupling was used 4-MeOBnCOOCH₂COOH (5eq.), HCTU (5eq.) and DiPEA (6eq.), 1h, RT, yield 18-87%.

The advantage of using this method is that all three starting materials (chloroacetyl chloride, sodium fluoroacetate and fumaric acid ethyl ester) are commercially available. In contrast,

the AOAA warhead building block was synthesised from *tert*-butyl bromoacetate and methoxybenzoic acid in the presence of potassium fluoride, followed by deprotection with TFA. The resulting methoxy benzoyloxy acetic acid was then activated with HCTU and DiPEA before coupling to the resin-bound peptide chain. Finally, all inhibitors were cleaved from the resin using standard conditions with 95% TFA. After purification *via* LC-MS, inhibitors stocks were prepared as described in section 3.3 ready for the inhibition assay. In summary, 23 inhibitors were generated as shown in **Table 7. 1** and **Table 7. 2**.



No.	WH	AA ¹	AA ²	Abbr. Name
TD 78	MA	Leu	Arg	MA-LR-PEG ₃ -Biotin
TD 79	MA	Ala	-	MA -A-PEG ₃ -Biotin
TD 80	MA	Ser	-	MA-S-PEG ₃ -Biotin
TD 81	MA	Thr	-	MA-T-PEG ₃ -Biotin
TD 82	MA	Arg	-	MA-R-PEG ₃ -Biotin
TD 83	MA	Tyr	-	MA-Y-PEG ₃ -Biotin
TD 84	AOAA	Leu	Arg	AOAA-LR-PEG ₃ -Biotin
TD 85	AOAA	Ala	-	AOAA-A-PEG ₃ -Biotin
TD 86	AOAA	Ser	-	AOAA-S-PEG ₃ -Biotin
TD 87	AOAA	Tyr	-	AOAA-Y-PEG ₃ -Biotin
TD 88	AOAA	Arg	-	AOAA-R-PEG ₃ -Biotin
TD 89	AOAA	Thr	-	AOAA-T-PEG ₃ -Biotin

Table 7. 2: Inhibitors with Michael acceptor or acyloxyacetamide warhead appended to the N-terminus of the specificity element.

7.3.2 Activity and labelling effect

In order to explore the scope and activity of the newly synthesised inhibitors as well as enable a comparison between different types of inhibitors, the inhibitors were tested at 100 μ M with Cd630 and whole cells boiled in loading buffer analysed using SDS-PAGE. For

the sake of more practical sample handle (e.g.: sample processing, loading on one gel instead of several gels, etc.), the library was divided into two sets: the haloacetamides were classed as one set and the rest of the compounds made up the other set.

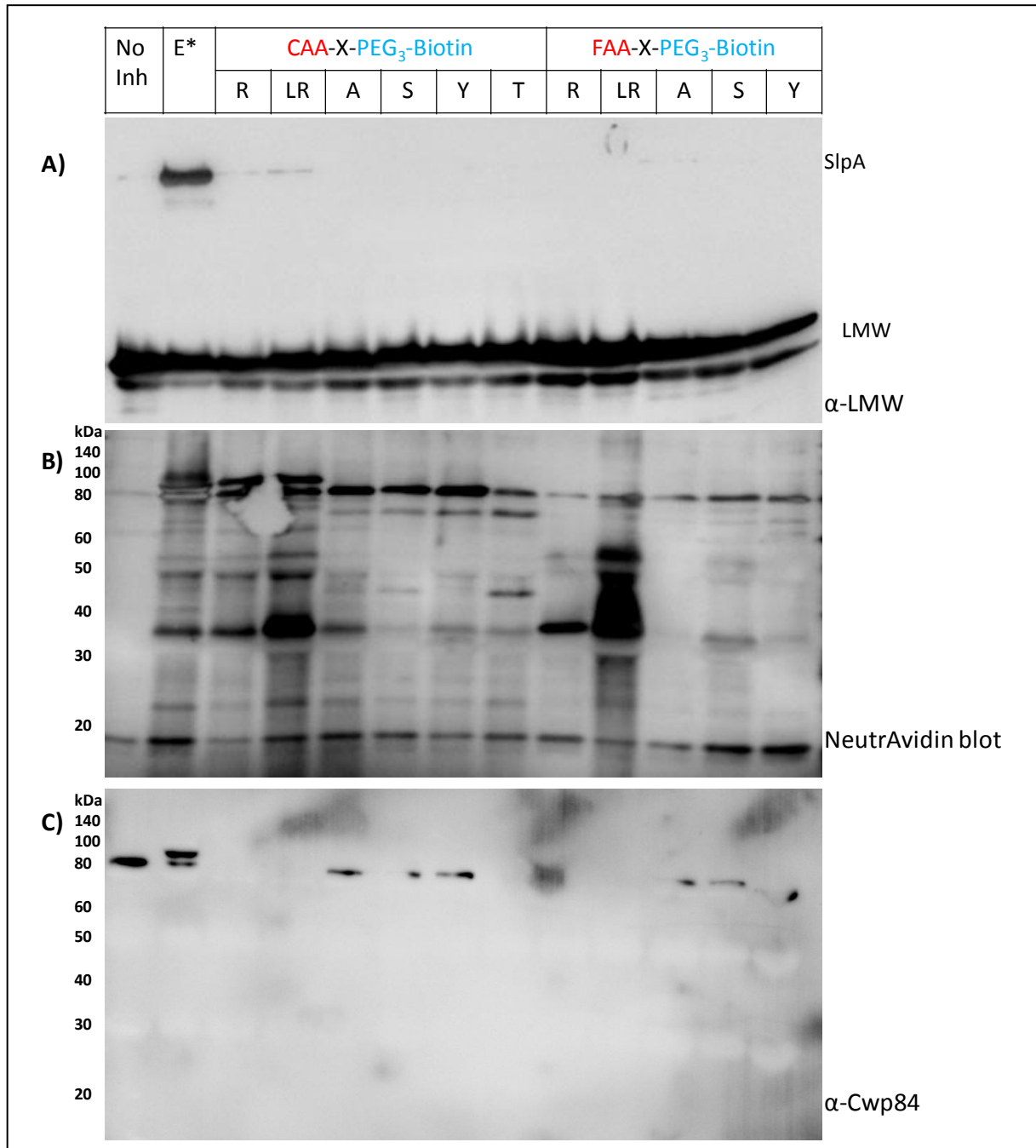


Figure 7. 7: Activity and labelling effect of haloacetamide inhibitors. *C. difficile* 630 were treated with 100 μ M inhibitors and whole cells were boiled in loading buffer prior to SDS-PAGE analysis. A) Western blot against LMW-SLP is shown to demonstrate the activity. B) NeutrAvidin blot is shown to demonstrate the activity the labelling effects. C) Western blot membrane of A was stripped and re-probed against Cwp84. E*=EtEP-R-PEG₃-Biotin.

As can be seen in **Figure 7. 7**, the inhibitory effect of the haloacetamides against S-layer processing at the tested concentration varied between samples. CAA-R-PEG₃-Biotin and CAA-LR-PEG₃-Biotin appeared to be slightly active, but at a much lower level than found with EtEP-R-PEG₃-Biotin, while the remaining inhibitors did not show any activity. The labelling showed similarity to one of the epoxysuccinyl inhibitors, EtEP-R-PEG₃-Biotin probe. Besides the “double band effect” as seen in active compounds CAA-R-PEG₃-Biotin and CAA-LR-PEG₃-Biotin, strong bands were observed at approximately 80 kDa, where Cwp84 was identified with EtEP-R-PEG₃-Biotin probe. There were a few additional signals on the NeutrAvidin blot compared to the epoxysuccinyl probe control, although it is hard to claim which signals are labelling related and which are biotinylated native proteins in the cell lysates. It should be noted that the differential lysis of the cells also affects the detected labelling signal of biotinylated native proteins (e.g.: fewer bands were found in the no inhibitor control sample). After probing against LMW-SLP, the membrane was stripped using low pH buffer and re-probed against Cwp84 (membrane C in **Figure 7. 7**). Unfortunately, the quality of the blot was poor and some signals were missing, which may be due to the nature of the stripping process. Despite this, the few strong bands that were observed on the blot indicated that Cwp84 was among the proteins labelled by these HAAs.

The set of MA and AOAA inhibitors were tested in the same manner as the HAA inhibitors (**Figure 7. 8**). No inhibition activity against SlpA processing could be detected under the tested conditions, which may also relate to the low quality of the Western blot. The labelled protein pattern was almost identical to that of the control probe (EtEP-R-PEG₃-Biotin), except for MA-LR-PEG₃-Biotin. Despite the noticeably stronger bands in the Western blot against Cwp84 for the MA and AOAA probes, there was still a visible reduction in signal detection due to the stripping and re-probing process as observed in the corresponding blot for the HAAs. As for the HAAs, it was observed that the strongly labelled bands in the higher molecular weight range did indeed contain Cwp84.

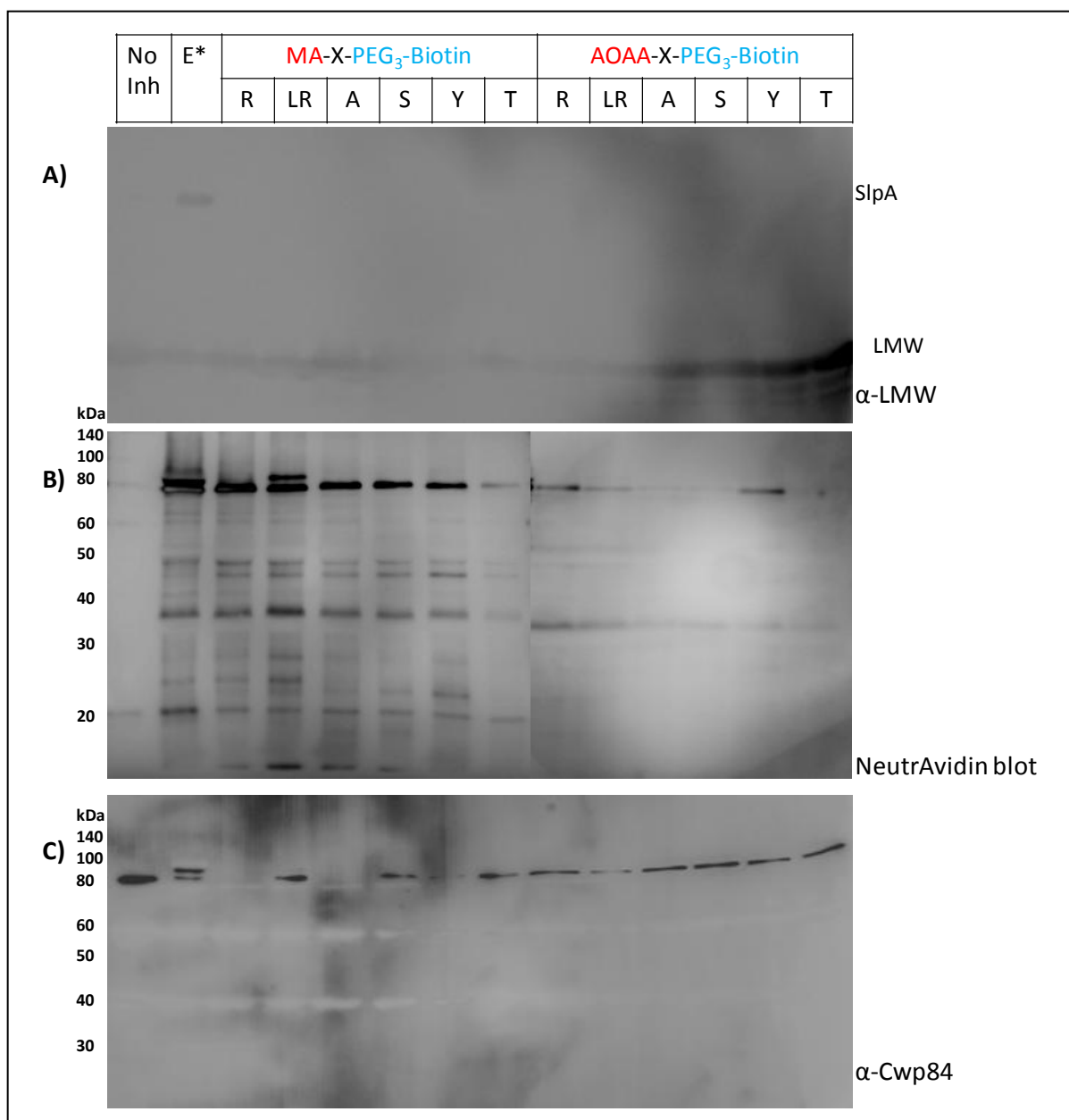


Figure 7. 8: Activity and labelling effect of MA and acyloxyacteamide inhibitors. *C. difficile* 630 were treated with 100 μ M inhibitors and the whole cells were boiled in loading buffer prior to SDS-PAGE analysis. A) Western blot against LMW-SLP is shown to demonstrate the activity. B) NeutrAvidin blot is shown to demonstrate the labelling effects. C) Western blot membrane was probed against Cwp84. E*=EtEP-R-PEG₃-Biotin.

Because NeutrAvidin blots of both inhibitor sets were unclear with many bands, analysis proved difficult and thus further experiments were needed to be conducted in order to make concrete conclusions. It was thought that dividing the bacterial cells into sub-factions should simplify the proteome and provide clearer and more interpretable results, as observed previously (see section 5.1.2). One representative compound from each of the four classes of inhibitors (CAA, MA, FAA, AOAA) was chosen for *in vivo* testing at 100 μ M,

each possessing the same specificity element (LR), in addition to the positive control EtEP-R-PEG₃-Biotin. *C. difficile* was divided into two fractions: the S-layer fraction and the cell fraction as previously described in section 3.5. In addition, a pull down step was added to remove non-specific labelling. **Figure 7. 9** depicts the Western blots obtained from this experiment, and it is seen that the results, as expected, are indeed much cleaner than those conducted without splitting the cell into two fractions (**Figure 7. 7** and **Figure 7. 8**). On the NeutrAvidin blot, one major band at approximately 80 kDa was observed for all of biotinylated inhibitors in the S-layer fraction except AOAA-LR-PEG₃-Biotin. The "double labelled band" effect of EtEP-R-PEG₃-Biotin was also observed for CAA-LR-PEG₃-Biotin. Once again, the Western blot against Cwp84 strongly indicated that those labelled bands contained Cwp84, which is also present in the lane containing the S-layer fraction treated with the AOAA compound. On the Coomassie-stained gel, which showed the enriched proteins, it was difficult to determine a distinct band in comparison to the positive and negative controls (EtEP-R-PEG₃-Biotin and no inhibitor lanes). Extrapolating from the results seen in Western blot against Cwp84 and the Coomassie-stained gel, the missed labelling effect of AOAA-LR-PEG₃-Biotin on the NeutrAvidin blot is most likely due to incomplete transferring or unexpected blotting effect during the sample processing.

In the cell fraction, no additional labelling was observed, as only biotinylated proteins of the bacteria were seen, which were also present in the no inhibitor control. Similar to S-layer fraction, no distinct bands were observed when compared to the enriched proteins on Coomassie stained gel. As a result, no further attempt was made to identify any enriched proteins in either fraction.

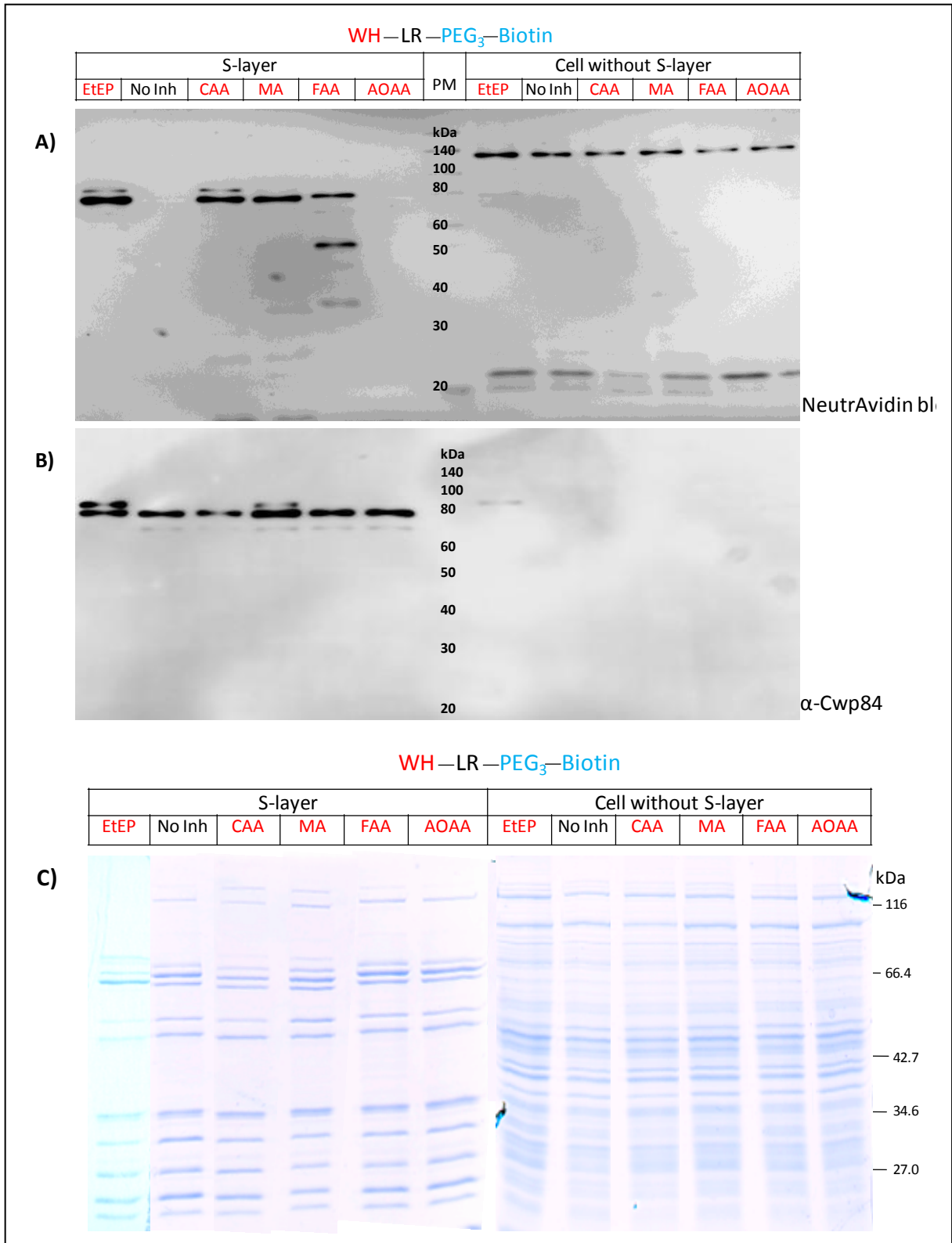


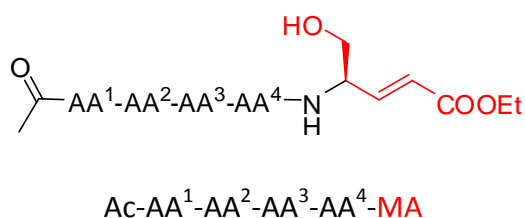
Figure 7. 9: Activity and labelling effect of inhibitors with the warhead appended to the N-terminus. *C. difficile* 630 were treated with 100 μ M Inhibitors and sub-fractionated into the S-layer fraction and the cell fraction. A pull-down was conducted prior to SDS-PAGE analysis to remove the majority of non-specific labelling. A) NeutrAvidin blot shows the labelling effects. B) Western blot against Cwp84. C) Coomassie stained gel visualised the enriched proteins for proteomic evaluation. All inhibitors have the LR specificity element, except the control EtEP=EtEP-R-PEG₃-Biotin.

7.4 Inhibitors with the warhead appended to the C-terminus

As mentioned in section 7.2, the strategy for synthesising inhibitors with the warhead appended to the C-terminus commences with the synthesis of the warhead building block using solution phase chemistry before attachment to the solid support for further peptide elongation. The details for each warhead type will be treated separately in the following sections.

7.4.1 Synthesis of Michael acceptor inhibitors

Revisiting the cleavage site of SlpA, it was noted that P1 of the substrate is either an alanine or a serine in most of the strains. The synthesis of inhibitors with warhead appended to the C-terminus using SPPS proved difficult to adapt for a number of reasons, including restriction on the number of amino acids with the properties required for transformation into the corresponding warhead functionality and/or loading onto resin. For example: a functionalised side chain is required to enable the attachment on the resin, and acid and/or base stability is typically essential to survive the harsh condition during solution phase warhead synthesis.



No.	AA ¹	AA ²	AA ³	AA ⁴	Abbr. Name
TD 90	Leu	Glu	Thr	Lys	Ac-LETK-MA
TD 91	-	-	Lys(Biotin)	Tyr	Ac-K(Biotin)-Y-MA
TD 92	-	-	Lys(Biotin)	Leu	Ac-K(Biotin)-L-MA
TD 93	-	-	Lys(Biotin)	Glu	Ac-K(Biotin)-E-MA
TD 94	-	-	Lys(Biotin)	Met	Ac-K(Biotin)-M-MA

Table 7. 3: Inhibitors with Michael acceptor warhead appended to the C-terminus of the specificity element.

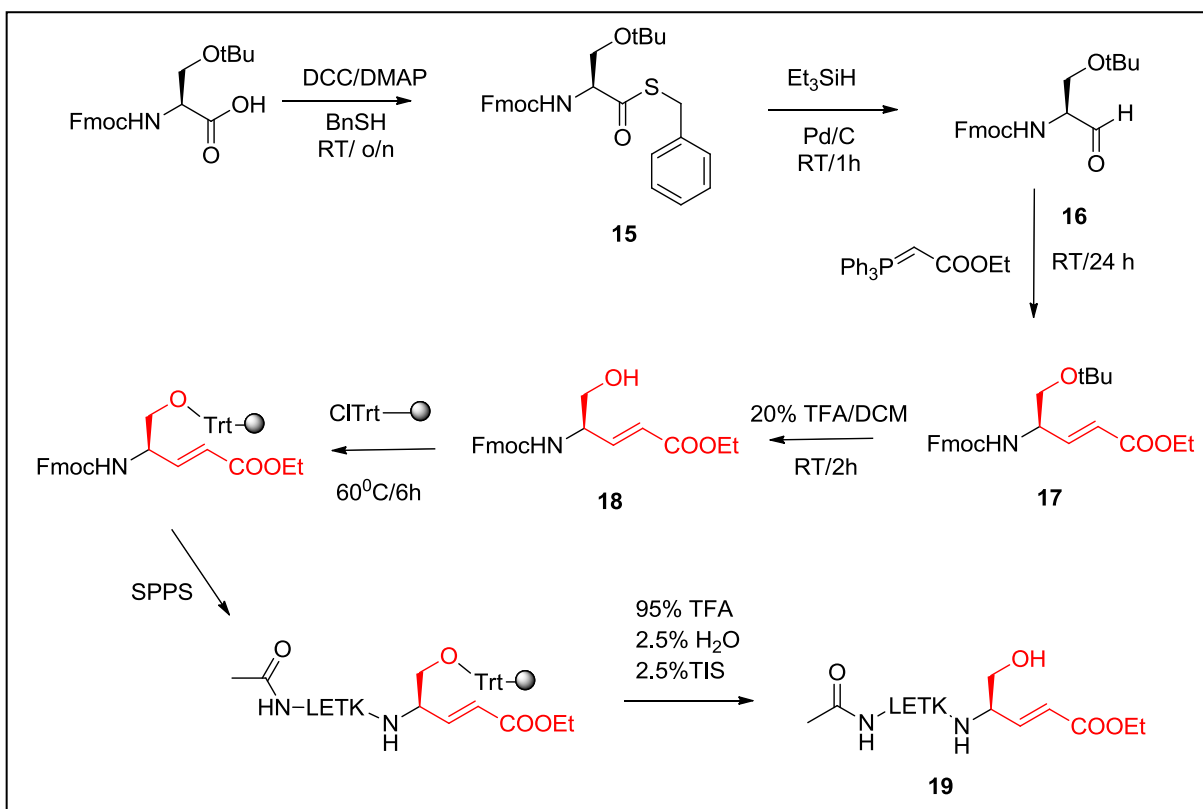


Figure 7. 10: Synthesis of inhibitors with an MA warhead appended on the C-terminus. Reagents and conditions: IX) DCC (1.05eq.), DMAP (0.10eq.), BnSH (2eq.), RT, o/n; X) Et₃SiH (5eq.), Pd/C (10%), in acetone RT,1h; XI) Ph₃PCHCOOEt (1.5eq.) in THF, RT, 24h; XII) 20%TFA in DCM, RT, 2h; loading on resin in pyridine and THF, 60°C, 6h; peptide elongation using standard SPPS; cleavage with 95% TFA, 2.5%H₂O, 2.5% TIS, 2h, RT.

The synthesis of MA inhibitors commenced with Fmoc-Ser(^tBu)-OH, which consists of three functional groups utilised for conversion into the MA warhead building block: the carboxyl group for modifying to the corresponding MA, the alcohol side chain to enable attachment onto the solid support and the Fmoc protected amine for subsequent peptide elongation. The reaction conditions were adapted from the literature procedure [62] (**Figure 7. 10**). Firstly, serine was converted into S-benzyl thioester (**15**) by coupling to benzyl mercaptan. Reduction of this thioester (**15**) using palladium over carbon and a source of hydrogen (Et₃SiH in this case) provided the corresponding aldehyde intermediate (**16**). Crude **16** was then condensed with the carbethoxymethylene triphenylphosphorane Wittig reagent to afford the α,β -unsaturated ester (**17**). Removal of the ^tBu protective group using 20% TFA supplied the corresponding side chain deprotected ester (**18**) allowing the attachment of the modified serine onto the solid support, which is a trityl resin in this case. Subsequently, the peptidyl specificity element was built up on the loaded resin by exploiting standard Fmoc/^tBu SPPS. The biotin label was introduced to the inhibitor by incorporating a side

chain biotinylated lysine in the peptidic specificity element, and the free amine at the N-terminus was acetylated to prevent additional charges in the molecule. Under standard TFA cleavage conditions as used previously, MA inhibitors were obtained and presented in **Table 7. 3**. The warhead of MA inhibitors is defined as the combination of the remaining serine side chain and the α,β -unsaturated ester (indicated in red).

7.4.2 Synthesis of AOMK- and CMK- inhibitors

A general method for synthesising AOMK and CMK probes was developed by Bogyo and co-workers [53], and this involved the attachment of Fmoc-protected CMK derivatives on hydrazine preloaded resin. Based on this work, an adapted synthetic route was proposed for the production of CMK inhibitors (**Figure 7. 11**).

Like the synthesis of the MA inhibitors, the initial amino acid building block was modified to contain the CMK functionality using solution chemistry. CMK derivative syntheses have been widely reported in the literature [53, 442-443]. The conventional route involves multiple transformations of the corresponding amino acid via diazomethyl ketones as intermediates [444]. These methods utilised diazomethane as the key reagent to generate the diazomethyl ketone moiety. However, diazomethane suffers from numerous safety issues due to its high toxicity, shock sensitivity and explosiveness [444-446]. Alternative routes avoiding the use of diazomethane and employing other reagents of a significantly reduced toxicity were explored, such as the use of chloromethyl lithium enolate, silyl enol ether, or a bromination step followed by decarboxylation, and a sulfur ylide. However, only the sulfur ylide route, which was based on the work of Wang *et al.* [444], proved to be simple and reliable with only three steps in total to obtain the CMK derivative. This procedure was adapted to synthesise the CMK building block used in the present study. Firstly, Z-protected alanine, where the 'Z' group is stable under acidic conditions, was converted into the corresponding methyl ester (**20**) **Figure 7. 12**. Treatment of this ester (**20**) with trimethylsulfoxonium iodide and potassium *tert*-butoxide afforded the trimethylsulfoxonium ylide (**21**), which, without further purification, was subsequently converted into Z-Ala-CMK (**22**) utilising HCl generated *in situ via* heating a mixture of lithium chloride and methanesulfonic acid in THF. One part of **22** was transformed into the AOMK derivative (**23**) by treatment with methoxybenzoic acid

in the presence of potassium fluoride in DMF; a protocol adapted from the work of Wood et al. [443].

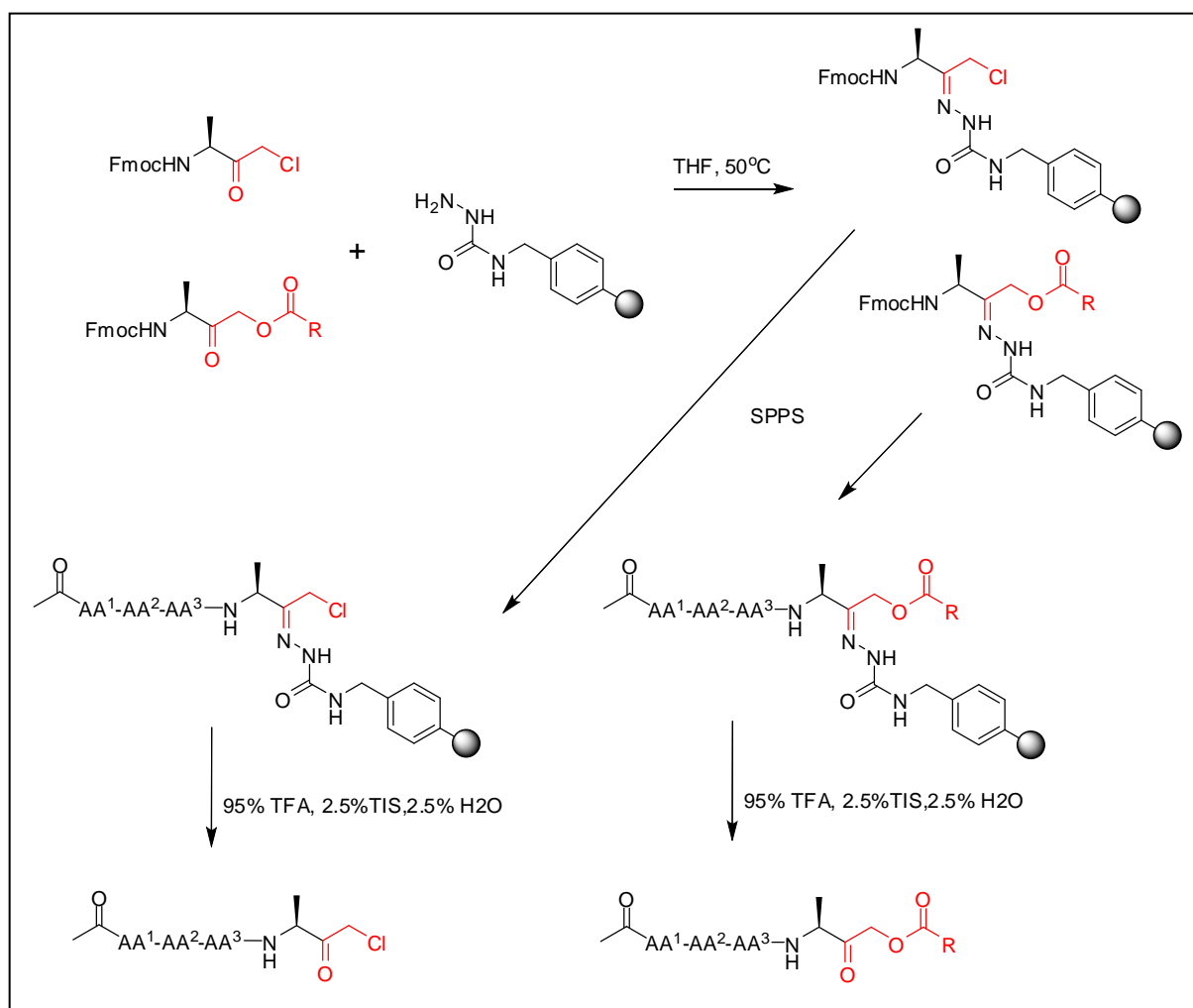


Figure 7. 11: Proposed synthetic route for CMK and AOMK inhibitors [53]. The Fmoc-protected building blocks are loaded onto hydrazine-functionalised resin in THF at 50°C . Standard SPPS was used for peptide elongation and final cleavage was performed with 95% TFA, 2.5% TIS, 2.5% H_2O , 2h at RT.

Prior to loading these building blocks onto the solid support, *via* a hydrazone linkage as depicted in **Figure 7. 11**, removal of the Z protecting group was attempted in order to enable subsequent substitution by an Fmoc group. Unfortunately, classical Z-deprotection methods such as catalytic (Pd/C) hydrogenolysis led to decomposition for both the CMK and AOMK building blocks, and therefore replacement with an Fmoc group was not possible. Due to time constraints, persistent decomposition upon deprotection meant that these building blocks could not be coupled onto the solid support and hence the desired final inhibitors could not be synthesised. The Z-Ala-CMK and Z-Ala-AOMK compounds were

nevertheless taken through to *in vivo* testing as representatives for these classes of inhibitors.

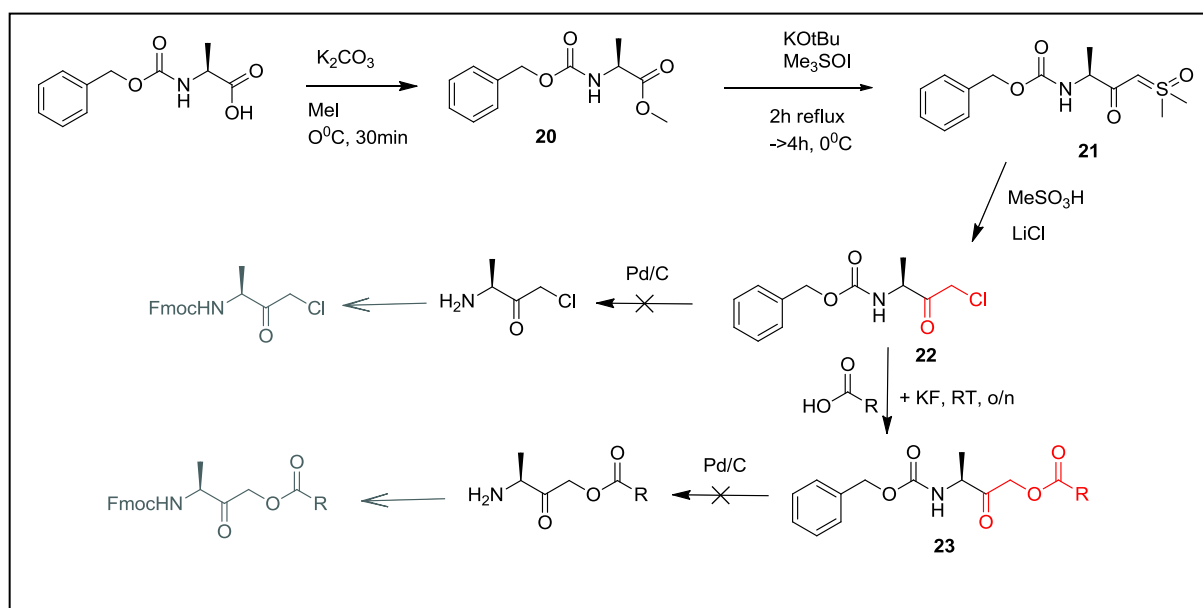


Figure 7. 12: Synthesis of CMK and AOMK building blocks. Reagents and conditions: XIV) K_2CO_3 (1eq.), MeI (2eq.) in DMF, 0°C , 30min; XV) KO^tBu (3eq.), Me_3SOI (3eq.) in THF, 2h reflux, then cooled down to 0°C before adding XIV (1eq.), 4h; XVI) MeSO_3H (1.1eq.), LiCl (1.14 eq.) in THF, 0°C -> 70°C , 2h, and quenched with H_2O ; XVII) 4-MeOBnCOOH (2eq.), KF (4eq.) in DMF, RT, o/n.

7.4.3 Activity and labelling effect of inhibitors with the warhead appended to C-terminus

Testing all inhibitors with the warhead appended at the C-terminus at $100\ \mu\text{M}$ against *C. difficile* 630 revealed that all inhibitors possessed moderate inhibition activity towards SlpA processing with the exception of Z-Ala-CMK (**Figure 7. 13**). This suggests that the strength of the electrophilic trap may play an important role since the corresponding amides were not active at the concentration tested. In addition, the resultant significant change in structure may also contribute to activity variation between inhibitors with the warhead appended to different ends of the peptidic chain. Intriguingly, although Z-Ala-CMK did not possess activity against SlpA processing, it did prove effective at killing *C. difficile*. Further investigation of this compound is discussed in a later section.

The labelling effect was observed on the NeutrAvidin blot; however, it was observed that there was a similar tendency to that found for other inhibitor types. Generally, the labelled protein pattern of Michael acceptor inhibitors (Ac-XXS-MA) was almost identical to the one of the control sample EtEP-R-PEG₃-Biotin though less intense. Since no distinct bands were

observed in comparison with the epoxysuccinyl control sample, no post proteomics study was carried out on this type of inhibitor.

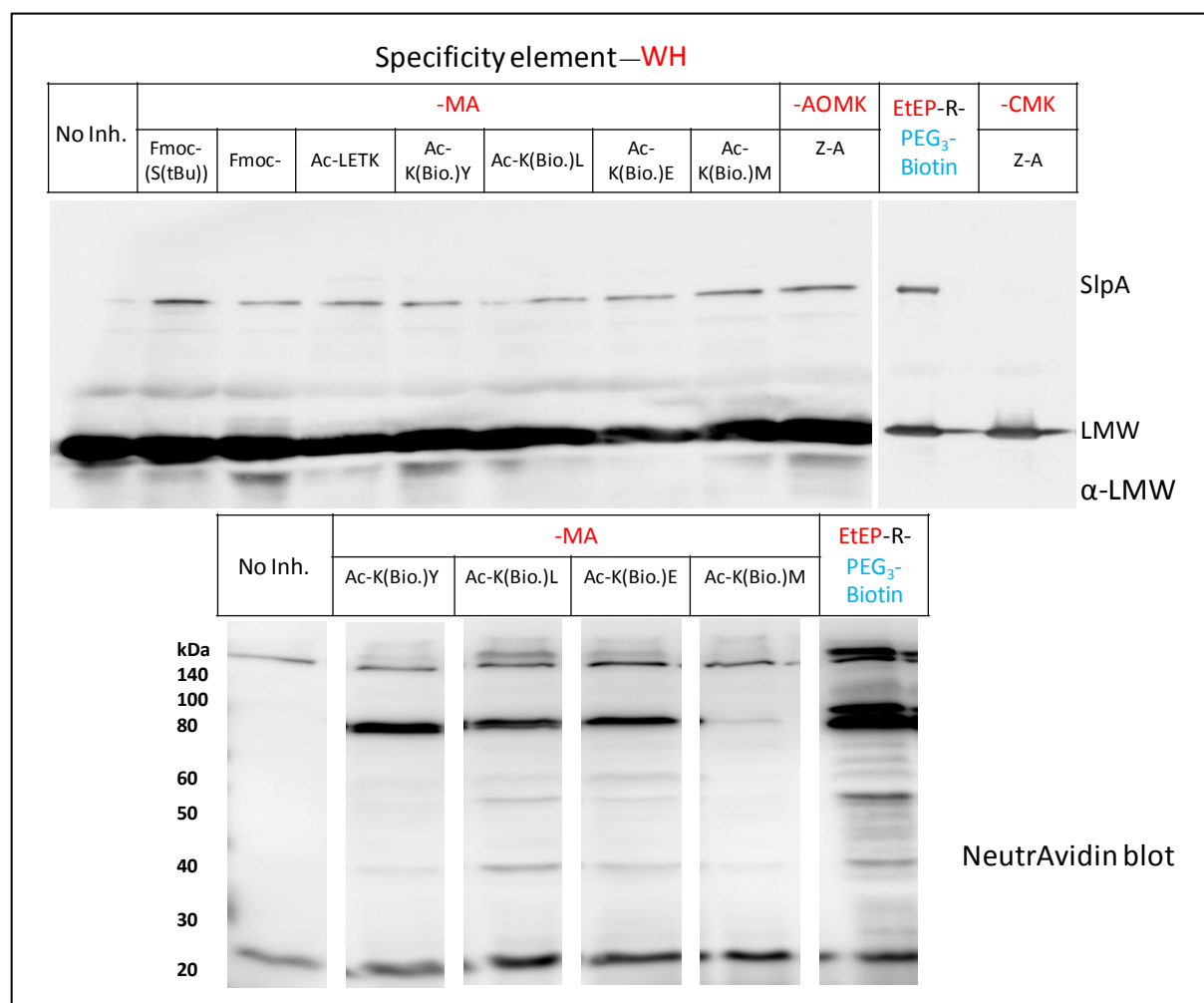


Figure 7. 13: Activity and labelling effect of inhibitors with the warhead appended to the C-terminus. *C. difficile* 630 cultures were treated with indicated inhibitors/probes at 100 μ M, as described standard inhibition assays (see section 2.1.1).

7.5 Growth inhibition of *C. difficile*

7.5.1 Growth inhibition by Z-Ala-CMK and EtEP-R-Pra in solution culture

As mentioned in previous section, Z-Ala-CMK tends to kill *C. difficile* at 100 μ M, as judged by the failure of many growth attempts in the presence of this compound. Only moderate bacterial growth was achieved at 10 μ M concentration (with OD of 0.5 after overnight culturing) while normal growth was attained at 1 μ M (with OD of 0.85 after overnight culturing). Following this finding, the concentration required for 50% effect (EC_{50}) was investigated by growing *C. difficile* in the presence of Z-Ala-CMK at different concentrations

ranging from 1 μM to 50 μM in BHI medium. Considering that the S-layer might play an important role in the growth inhibition, two mutants were chosen for this experiment: 630 Δerm (possessing a correctly-assembled S-layer as found in the wild type) and 630 Δcwp84 (possessing a less stable S-layer containing full length SlpA). Interestingly, the result revealed that the EC_{50} of Z-Ala-CMK was about 10 μM and small inhibition effect was observed already at 1 μM (**Figure 7. 14**). However, it should be noted that the small inhibition effect could be related with the growth variation of *C. difficile*.

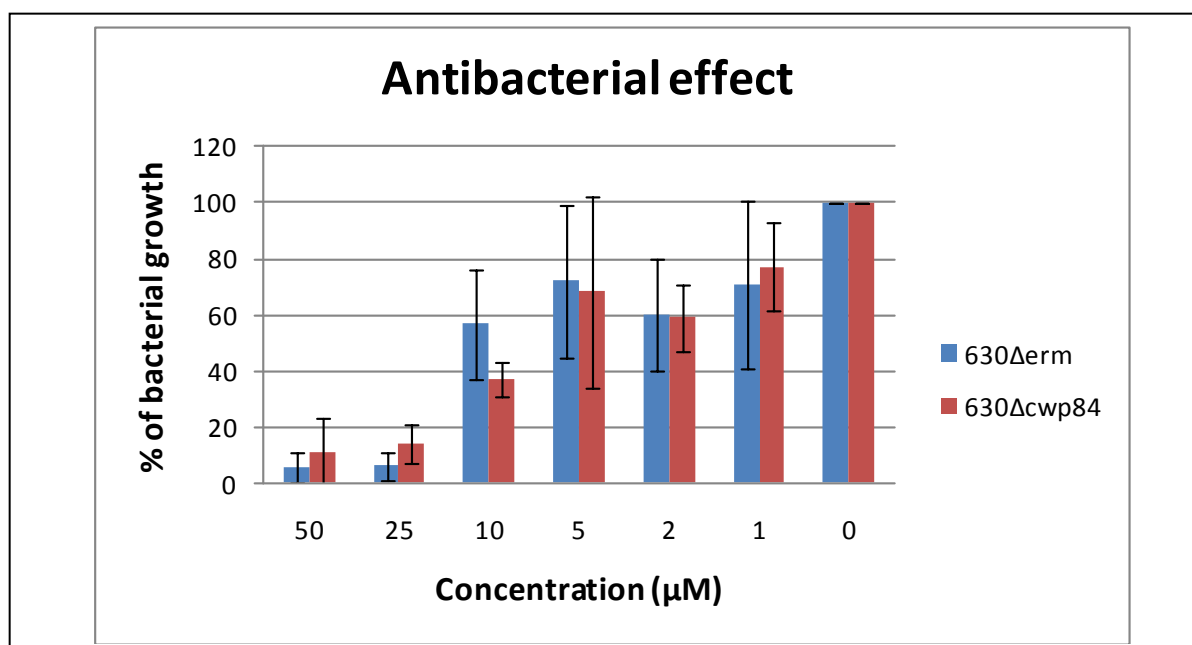


Figure 7. 14: Antibacterial effect of Z-Ala-CMK. *C. difficile* was grown overnight with different concentrations of Z-Ala-CMK. The percentages of bacterial growth were calculated based upon the OD values of the cultures.

Discovering the growth inhibition effect of Z-Ala-CMK and revisiting the inhibition results of epoxysuccinyl inhibitors, it was noticed that growing 630 Δcwp84 in the presence of inhibitors bearing arginine in the AA¹ position appeared to be difficult at high concentration (see section 6.7). The growth inhibition experiment was repeated using EtEP-R-Pra. Generally, inhibition effect was only observed for the 630 Δcwp84 mutant and not for the 630 Δerm mutant. Monitoring the growth over a concentration range from 0 μM to 50 μM it appeared that the EC_{50} is about 10 μM , similar to that of Z-Ala-CMK (**Figure 7. 15**).

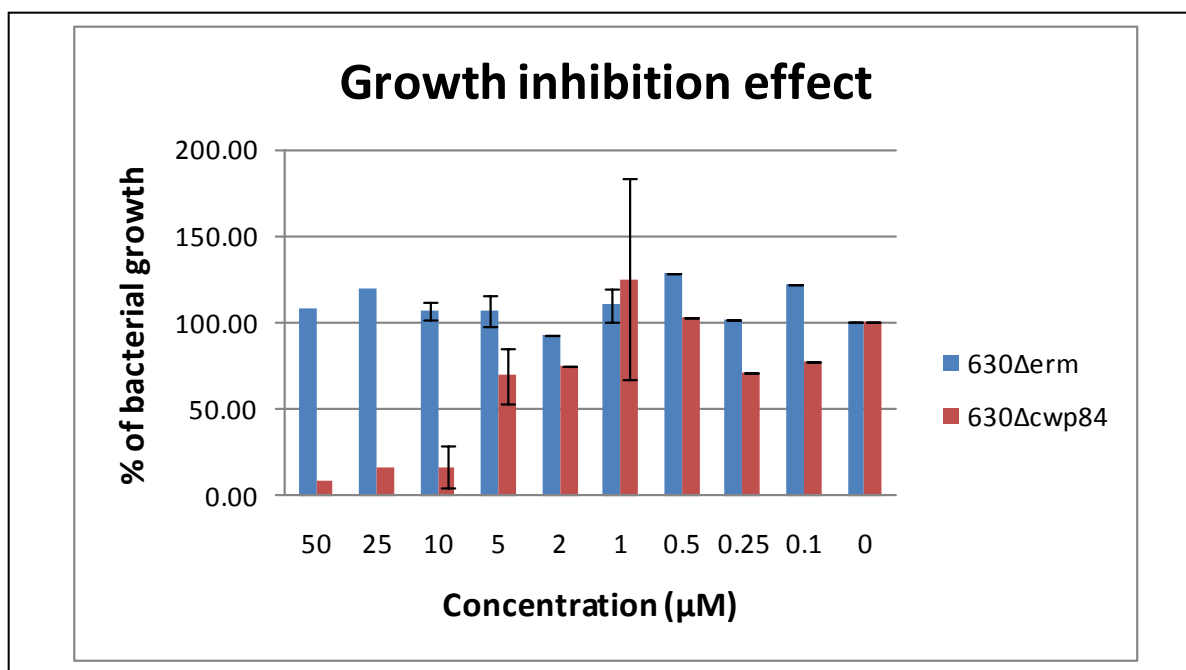


Figure 7. 15: Growth inhibition effect of EtEP-R-Pra. *C. difficile* mutants was grown overnight with different concentrations of EtEP-R-Pra. The percentages of bacterial growth were calculated based upon the OD values of the cultures.

7.5.2 Growth inhibition of 630 and 630Δcwp84 by antibiotics

Both full deletion of Cwp84 by gene knockout mutation and partial Cwp84 inactivation using inhibitors lead to a marked effect on cell integrity as judged by lysis sensitivity of the cell (see section 6.6), and by growth rate in rich medium in the case of the knockout mutant (see section 7.5.1). It was speculated that this effect might have a domino effect on viability under stress conditions such as antibiotic treatment or attack by the host immune system, presenting S-layer biogenesis as a potential therapeutic target. To examine this hypothesis the effect of antibiotics on wild type *C. difficile* 630 and 630Δcwp84 are described in this section. Both strains were streaked evenly on agar plates and subsequently E-test strips (Biomérieux, Marcy l’Etoile, France) containing antibiotics of interest were placed on the agar surface for 24 hours growth. Afterward, the minimum inhibitory concentrations (MICs) to chloramphenicol (CL), clindamycin (CM), erythromycin (EM), metronidazole (MZ), moxifloxacin (MX) and tetracycline (TC) were determined, a summary of the results is illustrated in **Figure 7. 16**. Higher MIC of the mutants compared to the wild type means the mutant strains are less sensitive to the antibiotics, which was observed when 630Δcwp84 mutant treated with chloramphenicol, moxifloxacin and tetracycline. Only metronidazole

caused an opposite effect. This result is interesting, however further experiments are required to ensure any conclusive antibiotic effect of the knockout mutants.

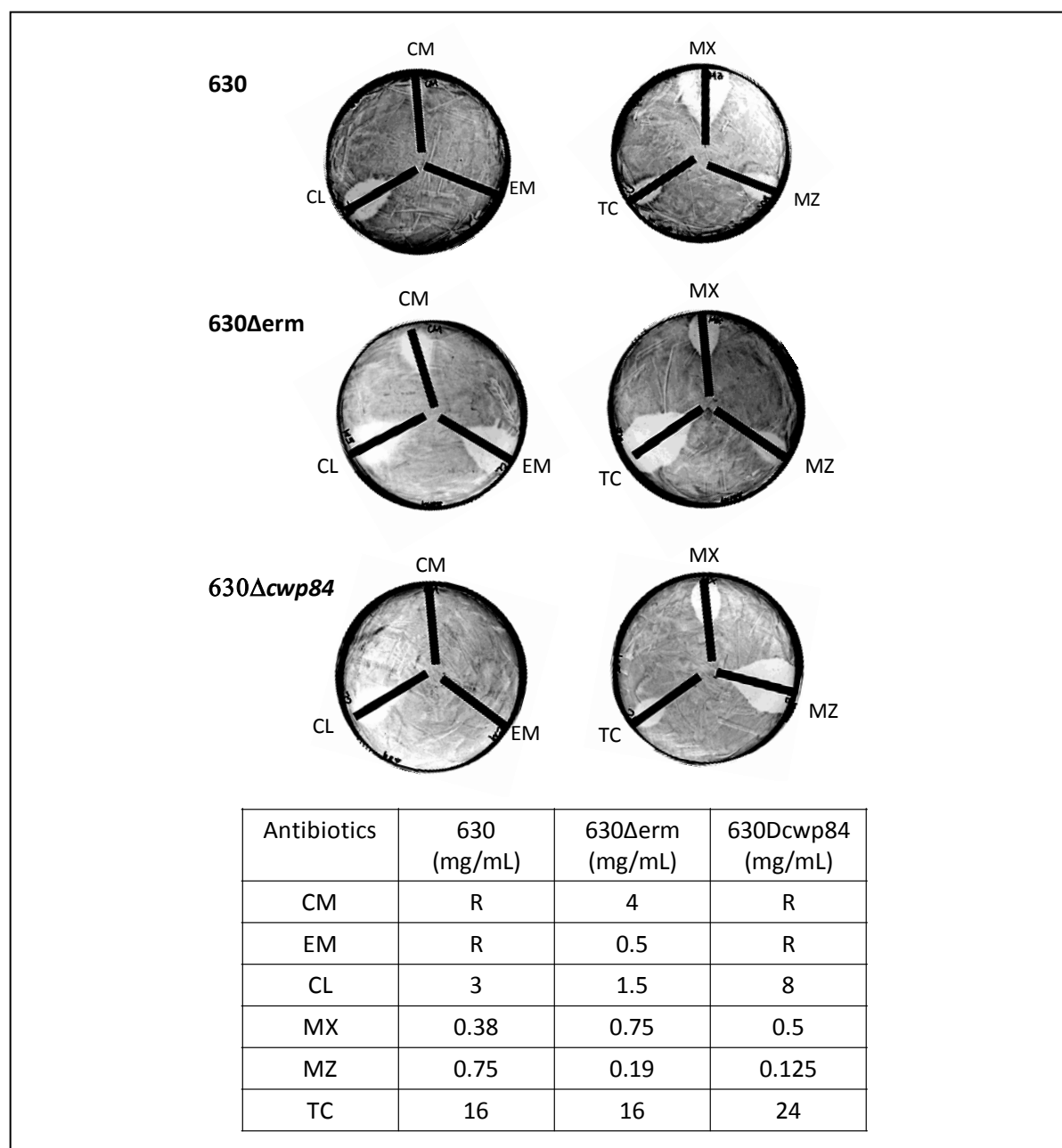


Figure 7. 16: Antibiotic effects. E-test strips containing antibiotic was placed on even streaked agar plated with *C. difficile* strains. MICs were determined after 24 hour growth according to manufacturer’s instructions. Antibiotics were chloramphenicol (CL), clindamycin (CM), erythromycin (EM), metronidazole (MZ), moxifloxacin (MX) and tetracycline (TC). MIC is given in $\mu\text{g/mL}$ while R means resistant.

7.6 Conclusion

This chapter has described the exploration of putative irreversible cysteine protease inhibitors bearing diverse types of warhead attached either side of the peptidyl specificity element. The synthesis of inhibitors with the warhead attached to the N-terminus was

adapted successfully from the previous synthesis used to generate epoxysuccinyl inhibitors. This success confirmed the adaptability of the synthetic route. On the other hand, the synthesis of inhibitors with the warhead attached on the C-terminus using SPPS proved difficult. Although five MA inhibitors/probes were successfully synthesised, modification of the warhead was still restricted to only a few amino acids possessing the side chain functionality required for preloading on solid support (e.g.: serine, threonine, tyrosine). Serine was used in this study because of its alcohol side chain and being the most common P1 amino acid at the cleavage site of SlpA in many *C. difficile* strains (**Table 1. 2**). Future studies could apply the synthesis route to other functionalised amino acid side chains such as the primary amine in lysine, carbonate in aspartic acid and glutamic acid, or the amide in asparagine and glutamine, which would require an alternate choice of resin than the trityl resin used for loading the alcohol side chain in this work.

In contrast to the synthesis of MA inhibitors, amino acids bearing a side chain without polar functionality (alanine, leucine, isoleucine etc.) are more useful for the synthesis of CMK/AOMK inhibitors because these amino acids are base and acid stable, which is essential to survive the harsh conditions of solution phase synthesis. Unfortunately, removal of the Z-protecting group of the intermediate suffered from decomposition of the whole compound. Facile reduction of the CMK/AMOK warheads may be the cause of the decomposition when using palladium on carbon to reduce the Z-group. Future studies could vary the protecting group of the amino acid to one that is still acid stable but requires a non-reductive deprotection method to remove (e.g.: acetamide)

Activity of the newly generated inhibitors was variable, as only weak inhibitory effect was observed in two of the CAA inhibitors (CAA-R-PEG₃-Biotin and CAA-LR-PEG₃-Biotin), while the rest of the inhibitors having the N-terminus attached warhead were not active. Inhibitors bearing the warhead at the C-terminus appeared to have increased inhibitory effect, as all of these inhibitors were moderately active, except Z-Ala-CMK. Interestingly, the specificity element does not tend to affect the inhibitory effect as strongly as was found in epoxysuccinyl inhibitors. This observation is reflected in the similar activity level of these inhibitors. In the case of MA inhibitors, it could be that the alcohol side chain of the serine included in the warhead occupies the S1 pocket of the protease, which was found to be important in epoxysuccinyl inhibitors. Therefore, the adjacent amino acid like lysine,

tyrosine, leucine, glutamic acid and methionine might not affect the activity significantly by occupying the S3 pocket (which was found to be relatively unimportant in epoxysuccinyl inhibitors). However, if this hypothesis is true for MA inhibitors, it is then difficult to understand the activity of Z-Ala-AOMK, as the alanine being adjacent to the warhead was not active in epoxysuccinyl inhibitors. If future studies could provide a crystal structure of Cwp84 for docking experiments with all of these inhibitors in the active site further insight and better explanation for these observations might be obtained.

Intriguingly, although Z-Ala-CMK did not show activity in the inhibition of SlpA processing, it did inhibit bacterial growth at concentrations above 10 μ M. One hypothesis is that this CMK inhibitor might target protease(s) (e.g.: sortase, serine proteases) that may be involved in posttranslational processing of surface proteins. This hypothesis is supported by the known inactivation of cysteine and serine proteases by CMK inhibitors [417] and the predicted serine protease serving as the signal peptidase responsible for posttranslational signal peptide cleavage of secreted proteins (discussed in section 3.1). It is possible that Z-Ala-CMK targets this protease and blocks all proteins possessing a signal peptide inside the cell, which may lead to the growth inhibition observed in the work of this chapter. Further work is required in order to answer these questions, including the development of a new synthetic strategy for synthesising the corresponding probes to enable target identification.

Although activity of the new inhibitors was observed in some cases, labelling was relatively similar to that of the positive control (EtEP-R-PEG₃-Biotin). Only a few bands were found to be distinct to the control, however, these bands were not found on the Coomassie stained gel after fractionating the bacteria. Therefore, post-proteomic studies were not conducted.

One general observation of labelling with the new compounds is the appearance of a band at ca. 80 kDa in NeutrAvidin blot. In addition, Western blot against Cwp84 also reveals a band at the same region. Previous proteomic studies described in Chapter 6 with both wild type and 630 Δ *cwp84* excising gel slices at a similar region showed the dominant presence of Cwp84, Cwp13 and traces of SlpA, Cwp6, Cwp66 and Cwp2. Considering the combined observations, it suggested that the new inhibitors in the work of this chapter might also label and enrich these proteins.

In addition, growth inhibition of 630 Δ cwp84 by EtEp-R-Pra, representative of epoxysuccinyl inhibitors bearing an arginine at AA¹ position, occurs readily at low concentration raising further questions about the effect of these compounds against *C. difficile* generally and against the S-layer specifically. As the Arg-containing compounds resulted in several unexpected observations across the whole study, further detailed discussion summarising all observed effects is continued in the next chapter.

Chapter 8 Conclusion and Perspectives

The main aim of this thesis was to apply ABPP to the pathogen *C. difficile* to identify potential therapeutic targets, with the ultimate objective of aiding in the battle against this multi-drug resistant bacterium. In this chapter, results of this study are evaluated from two perspectives: a technology development aspect and a biological aspect. Placing this work in the wider picture of chemical biology and current understanding of *C. difficile*, future perspectives are discussed, and suggested directions for future study.

8.1 Technology development

Chemical proteomics is a multidisciplinary research area that makes use of synthetic organic chemistry, biochemistry, cell biology and mass spectrometry. It is a powerful tool for the identification of small molecule-protein interactions within a complex proteome. As the aim of the project is to find a potential therapeutic target, the process of this study has similarities with drug development. Starting from initial discovery of a compound from a screen, in this case the natural product E-64, a cycle of design, synthesis and testing led to the identification of the target, and insights into its mode of action.

8.1.1 Design

Design of inhibitors/ABPs considered all three components of a typical ABP: the warhead, the specificity element and the reporter unit. Creation of a specificity element was based on the known cleavage locus of SlpA, whilst the choice of reporter units depended on the labelling method (e.g.: two-step labelling or direct labelling). Designing the warhead was difficult in terms of finding an optimal electrophilic warhead strong enough to be potent towards the target protease, but not too reactive to attack any nucleophile in the cell. Following a screen (see section 2.1), an epoxysuccinyl warhead was chosen as the electrophilic trap to react with the thio-nucleophile of cysteine residue of cysteine proteases; both the type and position (on C- or N-terminus of the specific element) of the warhead were based on E-64. This type of warhead has been investigated most intensively, but other cysteine protease-targeted warheads such as MA, CMK or AOMK were also developed.

As the electrophilic strength of the warhead contributes directly to the potency of the compounds, future design of the warhead could put more effort in improving this property. While the HAA, AOAA and MA warheads at the N-terminus of the specificity element benefit from ease of synthesis compared to the HMK, AOMK and MA warheads at the C-terminus, these compounds suffer from lower electrophilicity due to the presence of a weakly electron-withdrawing amide. In addition, inhibitors with a warhead at the N-terminus possess a different chemical composition, which makes side-by-side comparison with C-terminal warhead inhibitors more difficult, especially the putative orientation and fitting of

these compounds into the protease active site. Future work could make more effort in establishing more comparable compound sets Figure 8. 1.

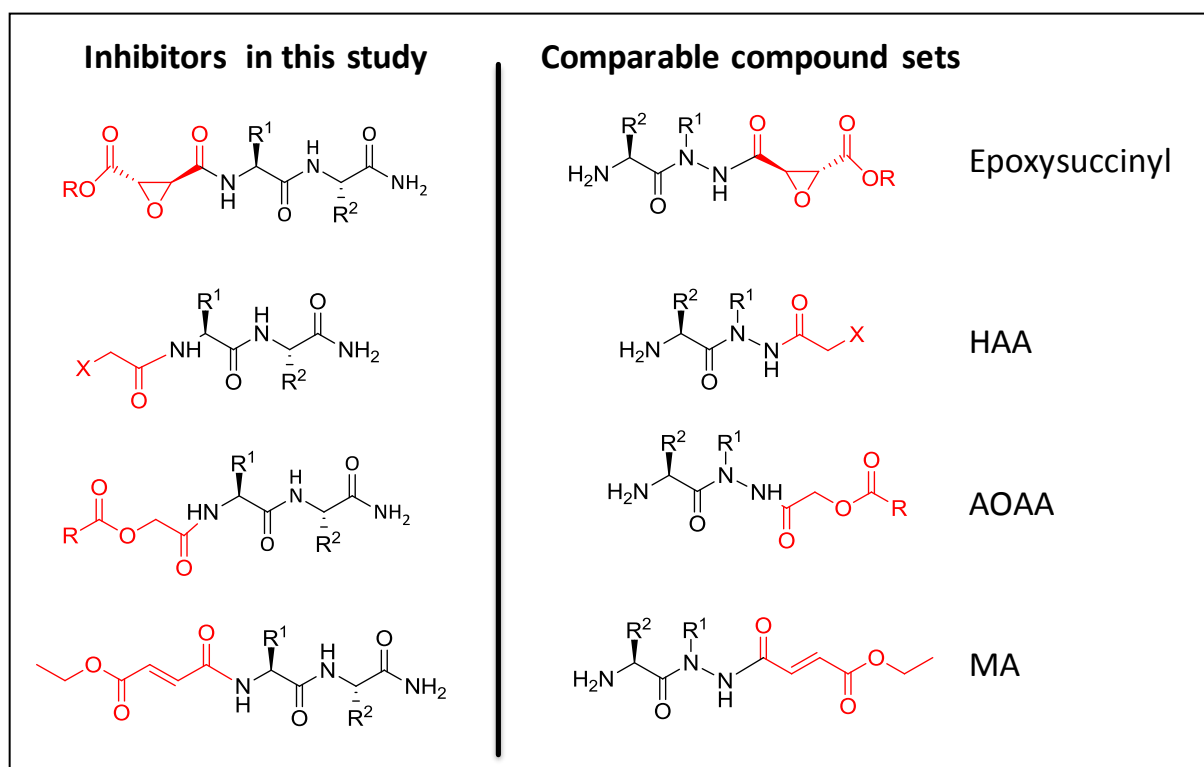


Figure 8. 1: Suggested comparable compound sets. The warheads have similar electrophilicity, but are attached at different end of the peptide.

8.1.2 Synthesis of inhibitors/ABPs

8.1.2.1 Inhibitors/ABPs bearing a warhead at the N-terminus of the specificity element

Based on numerous literature approaches to epoxysuccinyl peptidic inhibitors, the synthesis of inhibitors/ABPs of this group was adapted and improved [50, 72, 375]. Low-yielding reactions and even failure to couple the warhead to the peptidyl resin-bound specificity element arose from defects in reported literature procedures (e.g.: using crude warhead building block directly from solution phase synthesis). These were prevented by including additional purification of the warhead prior to coupling, as it turned out that warhead purity is very important for successful coupling. This synthesis route was further integrated for other types of warhead (see chapter 7).

One advantage of the optimised synthesis route is flexibility, as both warhead and specificity element can be easily varied. Therefore this method has the potential for a wide range of applications, and provides access to different inhibitors; a few proposed examples are

displayed in Figure 8. 2. Furthermore, the synthesis is relatively straightforward. Because the coupling of both components is a peptide coupling, the choice of the warhead building block is restricted to those possessing a free carboxylate group available for coupling. Some building blocks are hydrophilic and dissolve poorly in the organic solvent used in SPPS. Consequently, the yield for those couplings was also limited.

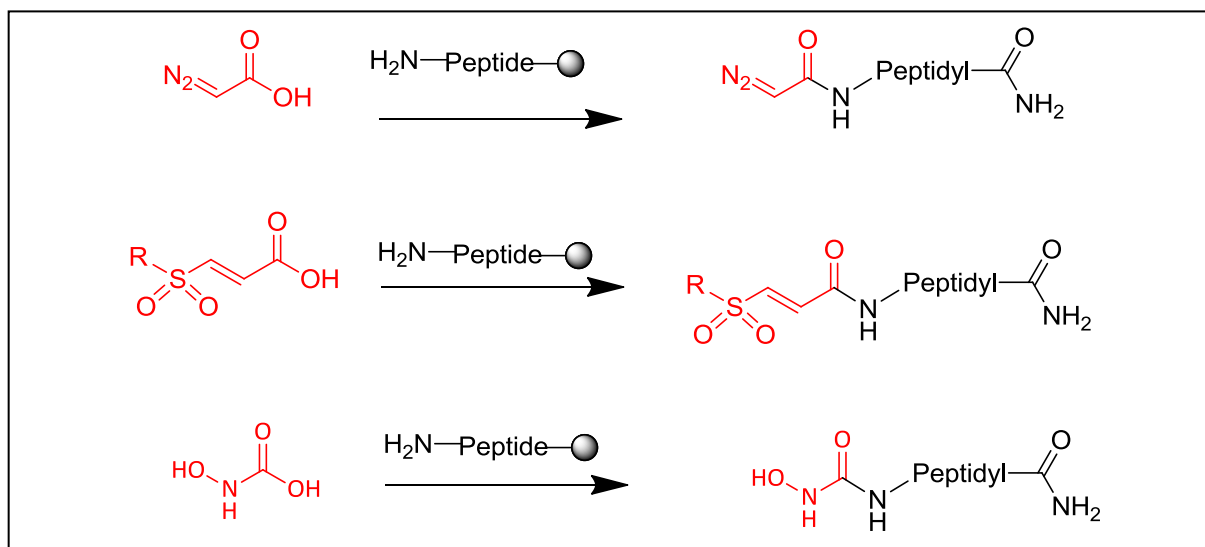


Figure 8. 2: Examples of potential application of the synthesis of inhibitors bearing a warhead at the N-terminus for access to other inhibitor types.

8.1.2.2 Inhibitors/ABPs bearing warhead at the C-terminus of the specificity element

Most common syntheses of inhibitors/ABPs bearing the warhead at the C-terminus involve either preloading of the warhead building block on the solid support with subsequent on-resin peptidyl elongation, or coupling of the warhead on the C-terminus when the elongation was already performed (e.g.: coupling to the secondary amine of an aza peptide).[61, 150]

The first approach was chosen for this study, where MA inhibitors/ABPs were successfully generated, while CMK and AOMK inhibitors could not be transformed into the corresponding ABPs. Although the MA inhibitor/ABP synthesis was well-adapted, the loading of the warhead on the resin was not easy to control under the conditions tested. As a result, the yield of these compounds might be adversely affected (see section 7.4.1). Similar observations were found in attempted loading of CMK or AOMK on resin. Furthermore, in order to be accessible for subsequent peptidic elongation, the protecting group of the amine is preferably an Fmoc-group, which is non-trivial to achieve in some

cases. Another drawback of this approach is the requirement of dual functionality in addition to the warhead itself: one for the loading on resin and one for peptide extension. Consequently, the choice of the warhead building block is also limited. Despite the synthesis hurdle, this group of inhibitors/ABPs is worth consideration, as they present the same orientation of the peptide backbone as the substrate (i.e. do not have the backbone reversed). Future work could build up from an aza peptide to prevent the inefficient preloading step (Figure 8. 3).

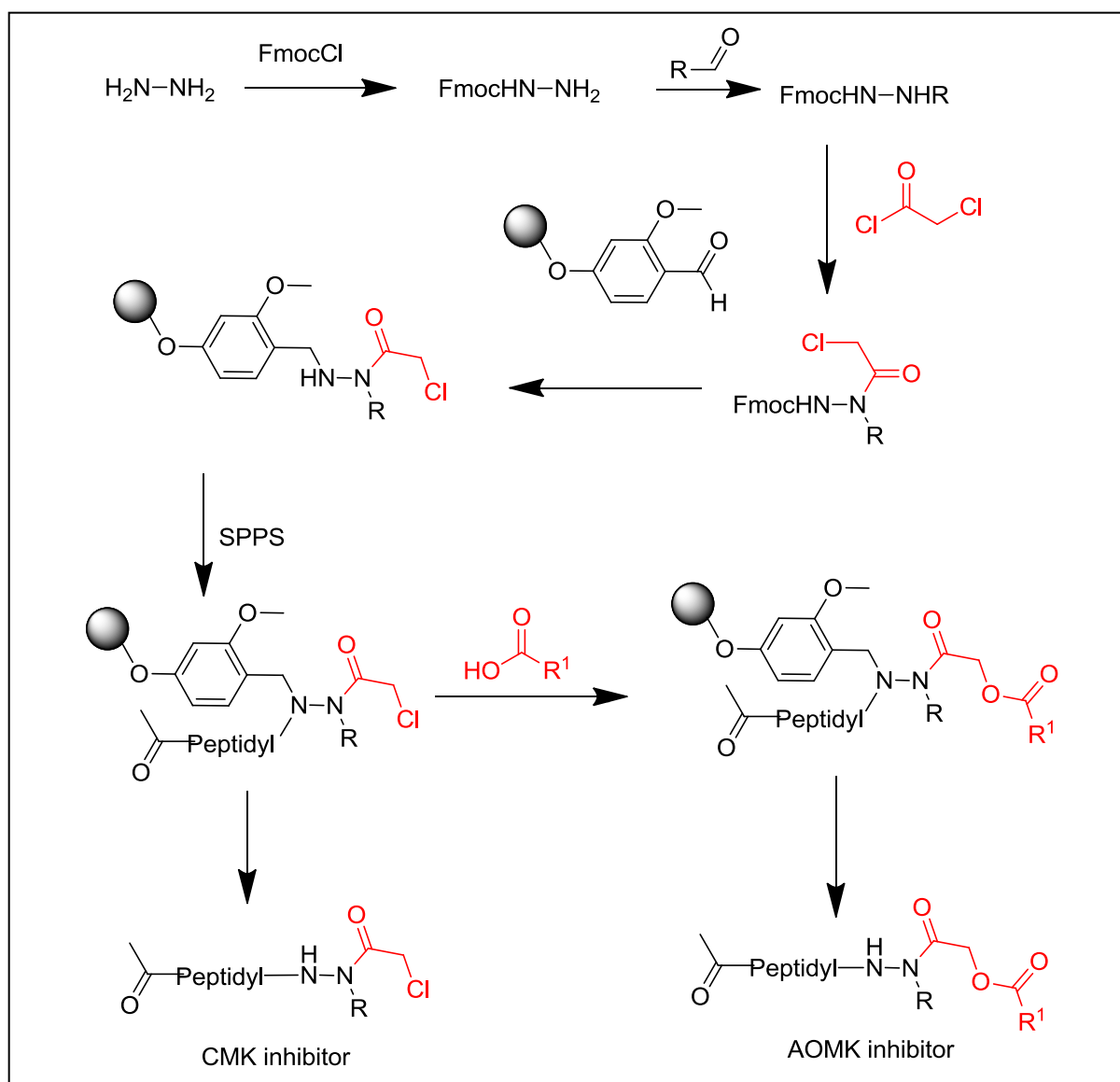


Figure 8. 3: Proposed synthesis of aza-peptidyl inhibitors to access CMK and AOMK inhibitors.

8.1.3 Activity assay

The assay for SlpA cleavage was very successfully implemented *in vivo* with live bacteria, however, working with *C. difficile* is very challenging. Variation in bacterial growth might affect the result, despite normalisation using OD₆₀₀. Normalisation confirms equal total loading between samples for comparison on a gel, but does not reflect any effects of inhibitors on bacteria at different stages of growth, such as inducing stress response or up-regulation of SlpA synthesis. Although the inhibitors were carefully designed to target SlpA processing by mimicking the substrate, the potency of generated inhibitors in general was not very high, except EP-Y-Pra-NH₂ and EP-R-NH₂. This could result from a number of factors, including:

- *Inhibitor concentration.* There are inherent problems in determining the absolute concentration of inhibitor stock (see section 3.3). Therefore, the actual effective concentration of inhibitor given in each assay likely differs to some degree from the calculated concentration. This is more obvious in attempted concentration determination using a fluorescent standard (see section 5.2.2), which revealed that this stock concentration may be up to four to five fold lower than that determined by the weighing method.
- *Substrate competition.* Imagining that each second ca. 400 units of SlpA are synthesised and transported to the surface per cell, the amount of SlpA substrate against which the inhibitor has to compete is huge. This may decrease the effect of the inhibitors towards the target protease.
- *Inhibitor composition.* Both orientations of the peptidic moiety were investigated by integrating the warhead at different ends of the specificity element. The potency differences of inhibitors of both groups were big enough to clarify a structure-activity trend. However, it is impossible to prove that these differences derive mainly from better fitting of the specificity element into the active pocket of the protease or from warhead reactivity, as each inhibitor possesses a slightly different warhead and might also have a different binding mode.

To improve the potency on the target (Cwp84), an *in vitro* assay could be designed, which could introduce high throughput screening of a library for covalent and non-covalent inhibitors. Based on the hits of this screening, future compounds could be designed and optimised.

Although an SAR was developed (for epoxysuccinyl inhibitors/ABPs), several aspects remained unclear, particularly questions regarding the binding mode of inhibitors into the active pocket of the target protease. A crystal structure of the target protease or co-crystal structure with inhibitors would provide further insight and aid our understanding of the binding mode, as well as offer a way to improve the potency.

Homology models of a protease based on crystal structures of proteases in the same family do not give as much information as a co-crystal structure with inhibitor, but they can give hints for possible binding modes. Accordingly, a Cwp84 homology model to a cathepsin C (1JQP) [288] was built using the Open Source Swiss-Model [447]. Aligning this with the co-crystal of Ervatamin-C – E-64 (2PRE) [372], it can be seen that E-64 fits nicely in the active pocket of the Cwp84 homology model (**Figure 8. 4**). According to this model, the S₂ pocket can tolerate long charged side chains at AA¹ that interact with Arg215/Asp165, or hydrophobic side chains that could bind to Leu260. Furthermore, the S₂ pocket seems relatively large and offers sufficient space for optimisation of the AA¹ to achieve optimal fit. The S' side is free and offers the possibility for development of “double headed inhibitors”[72]. For example, a large aromatic group could be introduced to form π - π interactions with Trp296. Although this model might not reflect the real binding mode of epoxysuccinyl inhibitors, it clearly suggests that there is scope available for optimisation to reach higher potency.

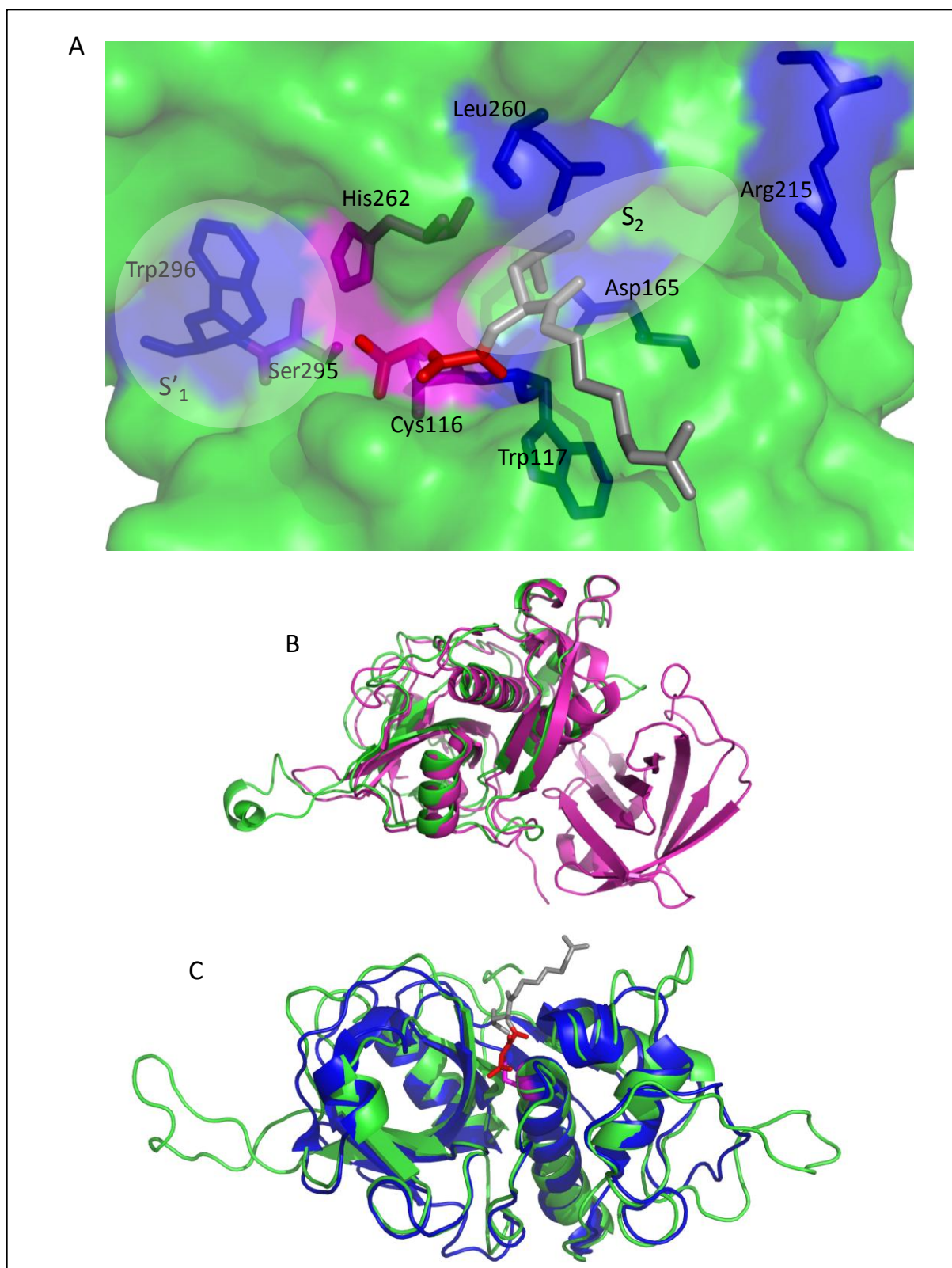


Figure 8. 4: Fitting of E-64 in Cwp84 homology model, based on the crystal structure of a cathepsin C (1JQP) [288]. A) Surface view of the active centre. Catalytic triad is shown in purple, some specific side chain of amino acid presented for possible involving in the binding are shown in blue. B) Overlay of Cwp84 model (green) and 1JQP (purple) (RSMD = 3.459 (171 to 171 atoms). C) Overlay of Cwp84 model (green) and 2PRE (blue) (RSMD = 2.647 (131 to 131 atoms). E-64 is presented as sticks, the warhead is labelled in red and the peptidyl moiety in gray.

8.1.4 Labelling effect

Different types of probes were developed during this thesis. Two-step labelling using alkyne-tagged probes was optimised and achieved a certain amount of success, considering that click ligation conditions tend to vary strongly between biological systems. However, the results were still biased, and in some cases not explainable at this stage (see section 4.4.3). The difficulties lay in the decision of whether a labelled band is real and related to the probe, and when it is just an effect of background labelling resulting from unspecific binding or endogenous biotinylated proteins. Further optimisation might lead to clearer and unbiased results, but this might require an extensive investment of time.

To circumvent the above mentioned problem, direct label probes were generated. Single label probes such as biotinylated or fluorescent ABPs tended to provide more consistent results than dual-label probes. This observation could be a result of detrimental effects of significantly increasing the size and hydrophobicity of the probes. Although direct labelling was very clean and aided the identification of the target proteases in the S-layer fraction, the labelling of the cell/soluble fractions still contained too much background signal. It was very difficult to differentiate between background labelling and probe related labelling. Similar results were also found with alkyne-tagged probes, and there could be several reasons for this phenomenon.

- *Permeability of probes.* Probes generated in this study possessing different size (e.g.: alkyne-tagged probes are small, whilst the dual label probes are very big) and groups that could potentially aid up-take (e.g.: biotin [112, 115]), the possibility that at least one of the probes can get inside the cell seems high. However, none of the probes provided concrete evidence of cell permeability. If the probe cannot penetrate inside the cell, intracellular probe-related labelling can be excluded.
- *Target specificity of the probes.* Despite the high number of predicted proteases in *C. difficile* (see section 1.6.3), only two cysteine proteases were identified in the S-layer fraction. Either the probes designed in this study targeted very specifically the cysteine protease(s) involved in SlpA processing and not other cysteine proteases, or there are no other proteases assessable to the tested probes. In both cases, probe related labelling in the cell fraction can also be excluded. It should be noted that the

manner in which the proteomic experiments were conducted probably also led to bias, potentially missing other targets that were not included in the excised bands.

- *High degree of background.* It has been observed that the reporter units give a high level of background signal, potentially from biotinylated proteins of the cell co-visualised in the streptavidin blot or hydrophobic sticking of the TAMRA group to hydrophobic proteins causing fluorescent signals. Optimisation of washing conditions could minimise the problem, however, due to the small working volume of samples, intensive and harsh wash conditions might cause loss of sample material during each wash step. Finding optimal conditions is often non-trivial.

To summarise, direct label probes were more useful in this particular study, however they also have some drawbacks and there is still scope to improve e.g.: reducing the size of the probes, optimising washing steps to reduce background signal, trying other reporters (e.g. other fluorescent groups) or introducing cleavable linkers for further investigation. It might be worth optimising click ligation in *C. difficile*, as this could be very useful in other applications to hunt for new targets.

8.1.5 Proteomics

Different proteomic techniques were used in this study, starting from classic MALDI-TOF to more modern LTQ-Orbitrap LC-MS/MS. As the primary interest of the study was the identification of the target protease, MALDI-TOF was the most simple and quickest method to access protein identification by PMF. In addition, this approach offers cost efficiency and flexibility by using in-house facilities. However, the data generated using MALDI-TOF were not very sensitive, less informative and not suitable for quantification analysis, whilst Orbitrap LC-MS/MS can overcome these limitations. With Orbitrap LC-MS/MS, the second protease (Cwp13) was revealed, whereas the MALDI-TOF only identified the presence of the cysteine protease Cwp84. Interest in quantitative analysis was evoked only later in the study by the desire to compare specific labelling of Cwp84 and Cwp13 between probes. Therefore, experimental design was not considered to cover quantification issues at the outset, such as adding standard peptide controls, producing biological and technical replicates, minimising variation through in-gel digestion, etc. Consequently, the data generated were very informative but insufficient for absolute quantification analysis (see section 6.8). Future

work could optimise the proteomic experiments by using on-bead digest rather than running SDS-PAGE and in-gel digestion to avoid missing target proteins either through absence in the excised bands or misbehaviour of proteins during SDS-PAGE, and introducing normalisation methods in each step to avoid variation in relative protein concentration. Despite imperfections in implementation, the primary objective of the study was fulfilled, revealing the identity of cysteine protease Cwp84, and a putative cysteine protease Cwp13.

8.2 Impacts of this study on current understanding of *C. difficile* biology.

Although ABPP has proven to be valuable for identifying new proteins and facilitating studies on a wide range of basic biological processes, it has primarily been used in eukaryotic systems. There are only a handful of examples demonstrating the application of ABPP in bacteria generally, and in pathogens specifically [448-449]. This is somewhat surprising, because many pathogens lack genetic tools or the necessary machinery for genetic manipulating, and have a complex biology (e.g. host pathogen interactions), which would make ABPP even more useful. On the other hand, application of ABPP on bacteria possessing advanced genetic tools such as *E. coli* is less unique due to the more convenient access to information via genetic experiments; however, in combination with genetics, ABPP can provide more striking insights into dynamic posttranslational phenomena than genetics alone.

At the outset of this study, proteases involved in SlpA processing were unknown. The identification of the target protease, Cwp84, has resulted in the first reported successful application of ABPP in live *C. difficile*. This achievement offers an additional chemical tool to study this challenging pathogen and related bacteria with similar limitations. Recently, Puri *et al.* reported the use of ABPs to identifying inhibitors of *C. difficile* cysteine protease domain in the toxins, and intend to use these probes as tools for studying inactivation mechanism of these protease domains [450]. However, it should be noted that this work was applied to recombinant toxin, and not live bacteria.

8.2.1 Cwp84 is the target protease responsible for SlpA internal cleavage

In vitro protease activity of Cwp84 had been previously reported against extracellular matrix proteins but not S-layer processing [280]. After the identification of Cwp84 using different proteomic techniques, a series of subsequent experiments confirmed our revised functional

assignment of this protease. Firstly, the identification of Cwp84 coincides with the ABP labelling signal, and this result was consistent, no matter which types of probes or which type of proteomic techniques were used. Furthermore, similar labelling was found in multiple strains of *C. difficile*. Secondly, *cwp84* locates in the same gene cluster as *slpA* and shows normal expression in the genetic analysis (see section 1.6.3.1). Thirdly, the knockout-strain 630 Δ *cwp84* suggested strongly that Cwp84 is responsible for SlpA cleavage due to the morphology change, the presence of only full-length SlpA and absence of its cleavage products in the S-layer [411], and this mutant was later generated for further experiments by Lucia de la Riva Perez (Fairweather and Tate, unpublished work). In addition, the presence of SlpA cleavage products was found only in mutant strains complemented with active Cwp84 (unpublished work of Dr. de la Riva Perez, ICL). Fourth, a reconstituted cleavage assay in *E. coli* provided further conclusive proof that Cwp84 is the only cysteine protease responsible for internal cleavage of SlpA into mature HMW- and LMW SLPs, at least under the conditions of laboratory culture used here [451]. Furthermore, this cleavage assay additionally confirmed aspects of the substrate specificity, which was in line with the SAR found in this study (see section 4.5).

While the function of Cwp84 was confirmed, further properties of the protease were also revealed. Although a crystal structure of Cwp84 is not yet available, our SAR suggests that Cwp84 acts as a papain-like cysteine protease. Cys116 is essential for activity, as the mutation of this Cys residue to Ala led to inactivation of the protease [451].

Several results suggested strongly that Cwp84 undergoes further (possibly autocatalytic) processing to form the active form (**Figure 8. 5**). The first observation was the ability of certain inhibitors such as EP-R-PEG₃-Biotin to intercept the processing, apparently labelling both forms of the protease. In one of the ABP experiments (**Figure 6. 8**), only the upper band of Cwp84 was observed by EtEP-RK(Biotin)G-PEG₃-TAMRA before pull-down, while the derivative containing tyrosine in the AA¹ position labelled both apparent 'forms' of Cwp84, as confirmed by Western blot against Cwp84, but favouring the lower band. Furthermore, several Cwp84 bands appeared in the pull down, also supporting the hypothesis of (auto-) processing.

In parallel work in the Fairweather laboratory, two different bands of Cwp84 have been confirmed genetically with mutants over-expressing active Cwp84, whilst only one band was shown in mutants expressing inactive Cwp84 (unpublished work of Dr de la Riva Perez, ICL). In addition, degradation of this protease was observed in the culture supernatant. Interestingly, degradation tended to be stronger if the activity of Cwp84 was neutralised, as judged by the amount of degraded product in culture supernatant. However, the mechanism of (auto-) processing remains unclear as well as other components that may be involved in this maturation process. This offers a new direction for future study, especially if Cwp84 is considered to be a potentially useful therapeutic target.

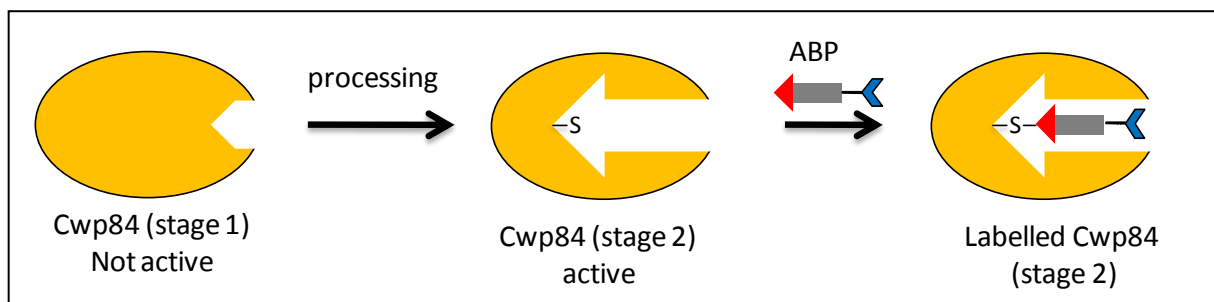


Figure 8. 5: A proposal for Cwp84 processing.

8.2.2 Cwp13 and other cell wall proteins

The discovery of the second putative cysteine protease, Cwp13, was perhaps surprising, and confirmed the sensitivity of the Orbitrap LC-MS/MS. Lying outside the SlpA locus at ca. 1.2 Mb distance, Cwp13 was labelled by epoxysuccinyl ABPs, but does not seem to target SlpA as Cwp84 does. The results of section Cwp13 may be an off-target of epoxysuccinyl ABPs suggest that Cwp13 is a target for many epoxysuccinyl inhibitors/ABPs. Despite the tendency of greater enrichment of Cwp13 by ABPs in the knockout strain $630\Delta cwp84$, Cwp13 was unable to complement Cwp84 function, as only SlpA was found in the S-layer fraction of $630\Delta cwp84(pCwp13)$ mutant, i.e. a mutant lacking *cwp84* but over-expressing *cwp13*. Constitutive over-expression of the protein was achieved by placing the gene under the control of P_{cwp2} , a promoter responsible for constitutive expression of a cell wall protein in *C. difficile* [279] (unpublished work of Dr. Riva de la Perez, ICL). Apart from being identified as a cell wall protein containing a putative cysteine protease domain [452], no studies on Cwp13 are reported as yet. The function of Cwp13 remains to be elucidated, but its co-purification together with Cwp84 and the effect with ABPs possessing arginine at the

AA¹ position (see next section) lead to speculation that Cwp13 may play a role in Cwp84 maturing process.

A number of other cell wall proteins were co-purified along with the two cysteine proteases (see section 6.2). Cwp6 is an interesting case, as it was identified at lower molecular weight range in the wild type strain, but at high molecular weight in the 630Δ*cwp84* strain. This might be related to a protein processing effect involving Cwp84, in which case unprocessed Cwp6 could be found at higher molecular weight regions in 630Δ*cwp84*. Cwp6 shows distant homology to Cwp13, but its function remains unknown. This tendency was also found for Cwp66, suggesting that the co-purification of these proteins might be related to a complex formed with the labelled target proteases rather than non-specific binding to the beads.

8.2.3 Epoxysuccinyl inhibitors/ABPs with Arg as AA¹ is a special case

A general SAR was developed for epoxysuccinyl inhibitors/ABPs providing better understanding of the mode of action of Cwp84 (see section 4.5). However, inhibitors/ABPs bearing arginine at the AA¹ position (arginyl ABPs) did not fit in the observed SAR. For example, although the inhibitory activity of EP-R-PEG₃-Biotin was better than that of EP-LR-PEG₃-Biotin, this compound seemed to label Cwp84 less efficiently (see section 6.4). Arginyl ABPs are also unique in provoking the Cwp84 “double band effect” clearly and reasonably consistently across many experiments. Both of these bands contain Cwp84 giving rise to the hypothesis of Cwp84 auto-processing. Naming the top band as a putative ‘stage one’ of Cwp84 processing and the bottom band ‘stage two’, a larger amount of Cwp84 was intercepted at stage two than at stage one by these ABPs. Furthermore, arginyl ABPs appeared to label significantly more Cwp13 than others ABPs judging by the proteomic experiments. However, it should be noted that the data are in most cases preliminary and require further reproduction. Unfortunately, an antibody against Cwp13 is not yet available, and quantitative proteomic analysis requires further optimisation. Making firm conclusions about the relationship between arginyl ABPs, Cwp84 and Cwp13 would therefore be premature at this point.

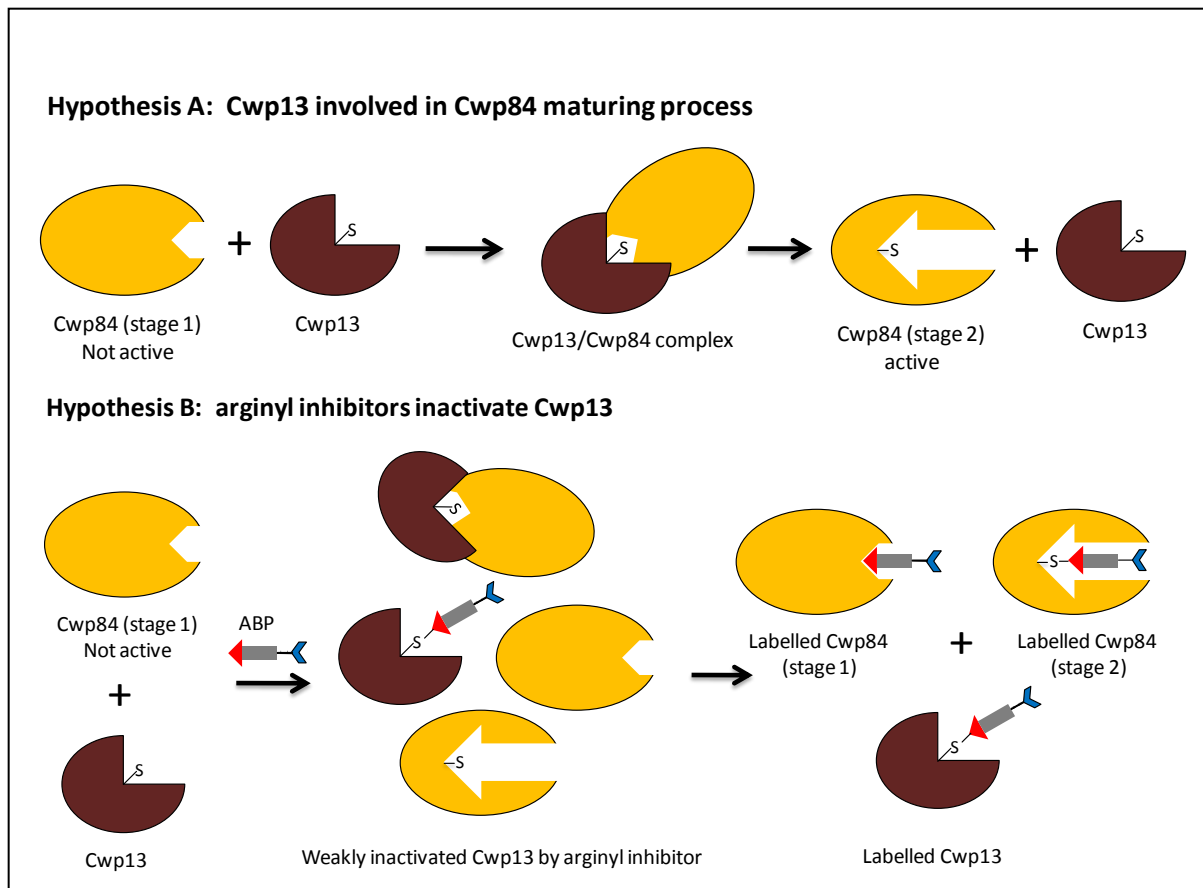


Figure 8. 6: Hypotheses on the role of Cwp13 and the interception of active enzymes by arginyl inhibitors/ABPs in the Cwp84 maturation process.

8.2.4 Could Cwp84 be a potential therapeutic target for CDI treatment?

Although the knockout strain 630Δ*cwp84* is very sick and difficult to culture, as well as very prone to lysis, it was reportedly able to cause CDI in a hamster model with similar pathology to the wild type strain [411]. Generally, animal models are used to mimic the pathological features and immunological responses observed in humans. However, different animal models have different specific features, particularly in the response to pathogens and toxins, leading to slightly varied results or sometimes to completely different results to those observed in humans. This is especially the case when comparing the survival time after infection for hamsters, which are extremely susceptible to *C. difficile* toxins and die after a few days to that of humans, who typically survive for many weeks, or can remain asymptomatic. Therefore, one cannot conclude that 630Δ*cwp84* will also induce CID in humans; testing of this strain in other animals less prone to the toxic effects of CDI is ongoing and may provide interesting results.

Targeting Cwp84 means targeting SlpA specifically or the S-layer in general. The question is what role does the S-layer play for *C. difficile*, particularly during infection? From an evolutionary perspective, the huge effort the bacteria make in synthesising an enormous number of SlpA molecules in each cell each second to build up the S-layer has to provide a payoff. Notably, the expression of SlpA and Cwp84 is up-regulated when the bacteria are under stress conditions such as antibiotic treatment [394], or Cwp84 inhibitor treatment (see section 4.3 and 6.4). Still, the question raised above remains open for future study, and therapeutic aspects of Cwp84 inhibition could be investigated through establishing a method of infection human enteric tissue cells with *C. difficile* to mimic the infection in real human body as closely as possible. The colonisation ability of 630 Δ cwp84 on human enteric tissue cell could then be compared with the wild type strain, as well as the wild type strain treated with inhibitors against Cwp84.

8.2.5 Other potential targets in *C. difficile*

Other targets in *C. difficile* might also be interesting in the battle against this bacterium, with the objective of finding new methods to cure CDI. It is fascinating that Z-Ala-CMK can prevent *C. difficile* from growing at early stage or at 10 μ M, similar to the effective concentration of antibiotics such as chloramphenicol. This concentration is also much lower than that of the epoxysuccinyl inhibitors, and Z-Ala-CMK does not inhibit SlpA processing. Unfortunately, the conversion of this compound into an ABP failed in this study, but future studies could invent a more effective synthesis for CMK probes (see suggestions in section 7.6) and open up new applications for ABPP in hunting down the next potential targets in *C. difficile*.

Chapter 9 Materials and methods

This chapter provides details about materials and methods including the characterisation of the inhibitors synthesised in this study. It is divided in three parts :

- Microbiology
- Proteomics
- Chemistry

9.1 Microbiology

9.1.1 Buffers

Name of buffer	Components in buffer
Running buffer	25 mM Tris, 192 mM glycine, 0.1% SDS
Sample loading buffer	45 mM Tris-Cl pH 6.8, 0.7% SDS, 0.7 mg/mL sucrose, 3.5% β -mercaptoethanol, 0.01 mM bromophenol blue
NuPAGE® MES SDS running buffer	Tris, SDS, 4-morpholinoethane sulfonic acid
NuPAGE® LDS 4x sample loading buffer	141 mM Tris base, 2% LDS, 10% glycerol, 0.51 mM EDTA, 0.22 mM SERVA Blue G250, 0.175 mM Phenol Red, pH 8.5
Staining buffer	45% methanol, 10% glacial acetic acid, 1 mg/mL Coomassie blue
Destaining buffer	45% methanol, 10% glacial acetic acid
Transfer buffer	25 mM Tris, 192 mM glycine, 20% methanol
10 × PBS	2 g KCl, 80 g NaCl, 17.8 g $\text{Na}_2\text{HPO}_4(2\text{H}_2\text{O})$ and 2.4 g KH_2PO_4 in 800 mL dH_2O
1 × PBS or PBS	100 mL of 10 × PBS in 900 mL dH_2O
FPLC wash buffer	50 mM NaH_2PO_4 , 300 mM NaCl pH 7.8, 20 mM imidazole
FPLC elution buffer	50 mM NaH_2PO_4 , 300 mM NaCl, 250 mM imidazole
Ambic buffer	50 mM NH_4HCO_3 pH 8.4
CHCA solution	20 mM α -cyano-4-hydroxycinnamic acid in 70% ACN, 0.1% TFA (v/v)
BCA reagent A	sodium carbonate, sodium bicarbonate, bicinchoninic acid and sodium tartrate in 0.1 M sodium hydroxide
BCA reagent B	4% cupric sulphate

Table 9. 1: Buffers used in this study.

9.1.2 Bacteria strains and culture conditions

C. difficile strains

C. difficile 630 [453] and the erythromycin sensitive derivative of 630 (630 Δ erm) [454] were provided by Dr Peter Mullany, Eastman Dental Institute, London, and has been fully

sequenced by the Wellcome Trust Sanger Institute [277]. CD1, CD17[455], and Y [456] were obtained from St. Bartholomew's Hospital, London. CD167 and CD959 are clinical isolates that were provided by Professor C. P. Kelly, Harvard Medical School, Boston, USA. CD167 was isolated from a patient with a single episode of diarrhoea that responded well to metronidazole treatment; CD959 was obtained from an asymptomatic carrier. SE528 is a toxin negative strain and was provided by Professor Michel Demee, Université Catholique de Louvain, Brussels, Belgium. R13549, R13540, R13699 were obtained from Dr Deirdre Ni Eidhin, Univeristy of Limerick, Limerick, IRELAND. 630 Δ cwp84, 630 Δ cwp84(pCwp84+) and 630 Δ cwp84(pCwp84_{C116A}) were generated by Dr. Lucia de la Riva Perez, ICL, UK (**Table 9. 2**).

C. difficile culture conditions

C. difficile was routinely cultured either on blood agar base II (Oxoid Ltd, Basingstoke, UK) supplemented with 7% horse blood (TCS Biosciences, Botolph Claydon, UK); or in brain-heart infusion (BHI) broth (Oxoid). Culture was undertaken in an anaerobic cabinet (Don Whitley Scientific, Shipley, UK) at 37 °C in a reducing anaerobic atmosphere (10% CO₂, 10% H₂, 80% N₂).

To obtain cells during exponential phase, an overnight (ca. 16-18 h) liquid culture was inoculated 1:20 v/v into BHI that had been pre-equilibrated (at least 2 h) to the temperature and atmosphere of the anaerobic cabinet, and the culture was shaken vigorously for 15 sec to mix thoroughly. During growth, the culture liquid was agitated gently using the shaker at 200 rpm to allow evenly growth. In order to monitor the growth of *C. difficile*, the OD₆₀₀ of a sample was determined using either a Spectronic™ Heλios™ ε spectrometer (Thermo Scientific, Waltham, MA) or a WPA biowave CD8000 cell density meter (Isogen Life Science, UK).

Strain name	Alternative names	Toxin status	Ribotype ^a	Serotype ^b	Source
630*		A ⁺ B ⁺ CDT ⁻	012	C	P. Mullany, UCL, UK
CDKK167	CD167	A ⁺ B ⁺ CDT ⁻	016	nk	C. Kelly, Boston, USA
CDKK959	CD959	A ⁺ B ⁺ CDT ⁻	053	K	C. Kelly, Boston, USA
Y	Y	A ⁻ B ⁻ CDT ⁻	010	nk	P. Mullany, UCL, UK
R8366	CD1	A ⁺ B ⁺ CDT ⁻	001	G	S. Ward, London, UK
R7404	CD17	A ⁻ B ⁺ CDT ⁻	017	F	S. Ward, London, UK
SE528		A ⁻ B ⁻	nk	C	M. Delmee, Brussels, Belgium
R13549		A ⁺ B ⁺	002	A ₂	D. Eidhin, Dublin, Ireland
R13540		A ⁺ B ⁺	078	nk	D. Eidhin, Dublin, Ireland
R13699		A ⁻ B ⁻	066	A ₉	D. Eidhin, Dublin, Ireland
630 Δ erm		A ⁺ B ⁺ CDT ⁻	012	C	P. Mullany, UCL, UK
630 Δ cwp84		A ⁺ B ⁺ CDT ⁻	012	C	L. de la Riva Perez, ICL,UK
630 Δ cwp84 (pCwp84+)		A ⁺ B ⁺ CDT ⁻	012	C	L. de la Riva Perez, ICL,UK
630 Δ cwp84 (pCwp84 _{C116A})		A ⁺ B ⁺ CDT ⁻	012	C	L. de la Riva Perez, ICL,UK

Table 9. 2: Table of *C. difficile* strains used in this study.

The table presents known characteristics of *C. difficile* strains used in this study. The data in italics is predicted from the known characteristics of the ribotype. ^a as determined by the method of Stubbs [455] ^b according to Delmee [457-458]; nk, not known. * genome sequence is available.

9.1.3 SDS-PAGE

Glycine-Tris gels:

Standard glycine-Tris gels at 10% or 12% acrylamide were prepared using mini-protean kits (Bio-Rad Laboratories). The separating gel (10 or 12% 39:1 acrylamide: bisacrylamide, 375 mM Tris pH 8.8, 0.1% SDS, 0.5% APS, 0.07% TEMED) was poured and ethanol carefully layered on top. Once the gel was set, the ethanol was discarded and the stacking gel (5%

39:1 acrylamide:bisacrylamide, 62.5 mM Tris pH 6.8, 0.1% SDS, 0.1% APS, 0.24% TEMED) was poured and the comb put in place to form the wells.

Once set, the comb was removed and the gel was immersed in the running buffer. Protein samples were diluted in sample loading buffer. These, along with a 10 µL aliquot of Broad Range Protein Marker (NEB), and a 3.5 µL aliquot of biotinylated protein ladder (Cell Signalling Technology) were heated to 100 °C for 5 minutes. Samples were loaded into the wells of the gel and a constant voltage of 200 V applied for 40 minutes.

Bis-Tris gels:

Bis-Tris gels at 12% acrylamide were prepared using mini-protean kits (Bio-Rad Laboratories). The separating gel (4 mL 30% acrylamide, 2.9 mL Bis-tris (1.25 M, pH 6.7), 3 mL dH₂O, 0.1 mL 10% w/v ammonium persulfate and 4 µL TEMED) was poured and ethanol carefully layered on top. Once the gel was set, the ethanol was discarded and the stacking gel (0.333 mL 30% acrylamide, 0.741 mL Bis-tris (1.25 M, pH 6.7), 1.453 mL dH₂O, 0.0125 mL 10% w/v ammonium persulfate and 2 µL TEMED) was poured and the comb put in place to form the wells.

Once set, the comb was removed and the gel was immersed in NuPAGE® MES SDS running buffer from Invitrogen. Protein samples were diluted in NuPAGE® LDS 4x sample loading buffer containing 16% 2-mercaptoethanol and were heated to 100 °C for 5 minutes. These, along with a 10 µL aliquot of Broad Range Protein Marker (NEB), a 3.5 µL aliquot of biotinylated protein ladder (Cell Signalling Technology), and a 2.5 µL aliquot of Amersham ECL Plex™ Fluorescent Rainbow Markers were used where appropriate to allow molecular weight comparison. Samples were loaded into the wells of the gel and a constant voltage of 200 V applied for 40 minutes.

Visualisation

For general visualisation of proteins, gels were stained in Coomassie blue staining buffer for 4–16 hours, followed by destaining using destaining buffer. Gels were image scanned and then rehydrated in 10% glycerol and dried between two cellulose membranes. Alternatively, gels were transferred to PVDF membrane for immunoblotting.

In-gel fluorescence visualisation was applied with Bis-Tris gels. Generally gels were soaked in water for 5 min and visualisation was carried out using an Ettan™ DIGE Imager (Amersham Biosciences, Cy3 channel used to detect TAMRA-labelled proteins) and results analysed with ImageQuant™ TL software.

9.1.4 Western immunoblotting

SDS-polyacrylamide gels were transferred to Trans-blot™ transfer medium (Bio-Rad Laboratories) PVDF membrane using a semi-dry protocol. The membrane was prepared for transfer by incubation sequentially in methanol (1-2 sec), H₂O (1-2 min) and finally in transfer buffer (2 min). It is important to keep the membrane wet at all times; if it dries out, the transfer can be adversely affected. Gel and filter paper were pre-soaked in transfer buffer. Filter paper, membrane and gel were stacked in the Trans-blot™ SD semi-dry transfer cell (Bio-Rad) as a sandwich in the following order: filter paper-membrane-gel-filter paper. The transfer was performed at 15 V for 15 minutes.

After transfer, the membrane was soaked in methanol (5 mL) and dried for 20 minutes, which avoids the additional blocking step typically required for minimising non-specific binding or back ground signal. The membrane was incubated with primary antibody diluted to the specified ratio v/v (Table 9. 3) in 10 mL PBS-T +3% (w/v) non-fat powdered milk for at least 1 hour with gentle rocking. After three washes in 10 mL PBS-T, secondary antibody diluted to the specified ratio v/v (Table 9. 3) in 10 mL PBS-T + 3% (w/v) non-fat powdered milk was added and the membrane rocked gently for at least 30 minutes. The membrane was washed three times in PBS-T. Signal was detected using ECL reagents (Amersham Biosciences) or SuperSignal® West Pico Chemiluminescent Substrate (Pierce) followed by image capture using a LAS-3000 Image Reader (Fujifilm), with appropriate sizing carried out using Microsoft Office PowerPoint 2007.

9.1.5 NeutrAvidin™-HRP immunoblotting

SDS-polyacrylamide gels were transferred using the method describe above, and the membranes were blocked for 1 hour in 10 mL PBS-T containing 5% BSA (w/v) before washing for 3 × 10 min in 10 mL PBS-T. The blots were probed using NeutrAvidin™-HRP at 1:5000 dilution in 10 mL PBS-T for 1 hour at room temperature. The membranes were then washed with 10 mL PBS-T for 3 × 5 min before detection as described above.

Antibody	Species	Conjugate or Label	Specificity	Antigen Purification	Source	Dilutions used:
α -Cwp84	mouse	none	Cd79685 Cwp84	Ni-Affinity Purification	This study	1/4000
α -HMW-SLP	Rabbit	none	630 HMW SLP	Ni-Affinity Purification	E. Calabi, Fairweather laboratory	1/200,000
α -LMW-SLP	Rabbit	none	630 LMW SLP	Ni-Affinity Purification	R. Fagan, Fairweather laboratory	1/200,000
α -Cwp66	mouse	none	630 Cwp66	Ni-Affinity Purification	J. Emerson, Fairweather laboratory	1/20,000
α -Cwp2	mouse	none	630 Cwp2	Ni-Affinity Purification	J. Emerson, Fairweather laboratory	1/50,000
α -rabbit-HRP	goat	horse radish peroxidase	rabbit immunoglobulins	Commercial	Dako Cytomation	1/2000
α -mouse-HRP	goat	horse radish peroxidase	mouse immunoglobulins	Commercial	Sigma Adrich	1/2000
NeutrAvidin TM -HRP	(synthetic)	horse radish peroxidase	biotin	Commercial	Sigma Aldrich	1/5000

Table 9. 3: Primary and secondary antibodies and immune sera used in this study, including the species in which they were raised, and the antigen against which they were raised. The antibody dilution used in Western blotting is also given.

9.1.6 Cell fractionation

Glycine extraction of SLPs

Standard SLP extraction was performed essentially as previously described in Dubreuil *et al.* [459]. *C. difficile* bacteria from a 16–18 hour culture in BHI were harvested by centrifugation (15 min at $3,500 \times g$) and resuspended in 0.04 volumes (relative to culture volume) of 0.2 M glycine-HCl pH 2.2. After rotating incubation for 30 min at room temperature, the mixture was centrifuged to remove bacteria (5 min at $10,000 \times g$) and the supernatant neutralised by adding 2 M Tris portion-wise (5 μ L at a time) until the desired pH was obtained (monitoring the pH with pH strips (Whatman)).

Preparations of soluble and insoluble fractions

After removing the SLPs, the pellet was washed two times with PBS and resuspended in 1 mL PBS supplemented with 30 μ L of a 50 mg/mL solution of lysozyme in water. The suspension was incubated for 20 minutes at room temperature, and the cells were then sonicated (Sonoplus sonicator, Bandelin) on ice for 4 minutes in pulses of 10 s on, 10 s off. After centrifugation at $17,900 \times g$ at 4 °C for 20 minutes the supernatant was separated as the soluble fraction. The pellet was resuspended in 100 μ L of 10% SDS solution; this was defined as the insoluble fraction.

Preparations of only soluble fraction using glass bead lysis

After removing the SLPs, the pellet was washed two times with PBS and resuspended in 1 mL PBS and 30 μ L of a 50 mg/mL solution of lysozyme in water. The suspension was incubated for 20 minutes at room temperature and then added to a vial containing ca. 0.5 mL of 150-212 μ m diameter glass beads (Sigma). The sample and the glass beads were vigorously agitated in a closed vial in a BIO101 Fast Prep FP120 bead beater (BIO101 Savant) at setting 6 for 45 s to lyse the cells. The supernatant was separated as the soluble fraction. The glass beads were subsequently washed with 200 μ L PBS. Wash supernatant was removed using centrifugation at 14000 rpm for 2 min and combined with the soluble fraction.

9.1.7 CuAAC ligation

CuAAC reagents were prepared and stored as stocks in DMSO or water (**Table 9. 4**). The reagents were premixed in the given order as shown in Table 3 from top to bottom. CuAAC

ligation was carried out on protein fractions (e.g.: glycine extraction, soluble fraction) by vortexing the sample for 1 hour (unless otherwise specified) at RT and then quenched by the addition of at least 10 volumes of ice-cold MeOH. Samples were left at -80°C overnight for precipitation and then protein pelleted by centrifugation at 13300 rpm for 20 min and at 4°C. Pellets were washed with another ~0.5 mL ice-cold MeOH and pelleted once more, then air-dried for ~15 min. Protein was resuspended in buffer (usually 4 % SDS in PBS).

Reagents	Concentration of stock	Typical final concentration
Capture reagents	10 mM in DMSO	50 µM
CuSO ₄	40 mM in dH ₂ O	1 mM
TCEP	50 mM in dH ₂ O	1 mM
TBTA	10 mM in DMSO	100 µM

Table 9. 4: Reagents and their stock concentration using for click ligation.

9.1.8 Enrichment of captured proteins

Using NeutrAvidin™ agarose beads

Typically, 30 µL NeutrAvidin™ agarose beads (supplied as a 50% slurry from Thermo Scientific) were added to the bacterial sample fraction (starting from a 5 mL culture) containing the captured proteins and incubated with rotation for 2 hours at 4 °C. The beads were then pelleted and the supernatant removed, followed by washing of the beads 3 times with PBS. Then the beads were boiled in ca. 20 µL sample loading buffer for 10 min for subsequent SDS-PAGE analysis.

Using Dynabeads® MyOne™ Streptavidin C1

Typically, 20 µL Dynabeads® MyOne™ Streptavidin C1 (10 mg/mL suspension from Invitrogen) were added to the bacterial sample fraction (starting from a 5 mL culture) containing the captured proteins and incubated with rotation for 2 hours at 4 °C. The supernatant was removed, followed by washing of the beads 3 times with PBS (unless otherwise specified). Then the beads were boiled in ~20 µL sample loading buffer for 10 min for subsequent SDS-PAGE analysis.

9.1.9 Expression and purification of recombinant protein

E. coli strains used in this study

Protein expression was carried out in Rosetta™ (Novagen) cells. Rosetta™ are BL21(DE3) cells carrying a plasmid encoding tRNAs specific to codons that are rare in *E. coli* but common in other organisms. Genotype: F⁻ *ompT hsdS_B (r_B⁻, m_B⁻) gal dcm* (DE3) pRARE (Cm^R).

Plasmid purification

Plasmids were purified using QIAprep miniprep kit (Qiagen), following the procedure recommended by the manufacturer. The Qiaprep column was transferred to a fresh 1.5 mL microfuge tube and 50 µL of dH₂O or buffer EB (elution buffer; 10 mM Tris-Cl, pH 8.5) was added to the centre of the QIAprep membrane and incubated at room temperature for 1 minute. The column was centrifuged for 1 minute at 17,900 × g to elute the DNA into the microfuge tube. Purified plasmid DNA was then stored at -20 °C.

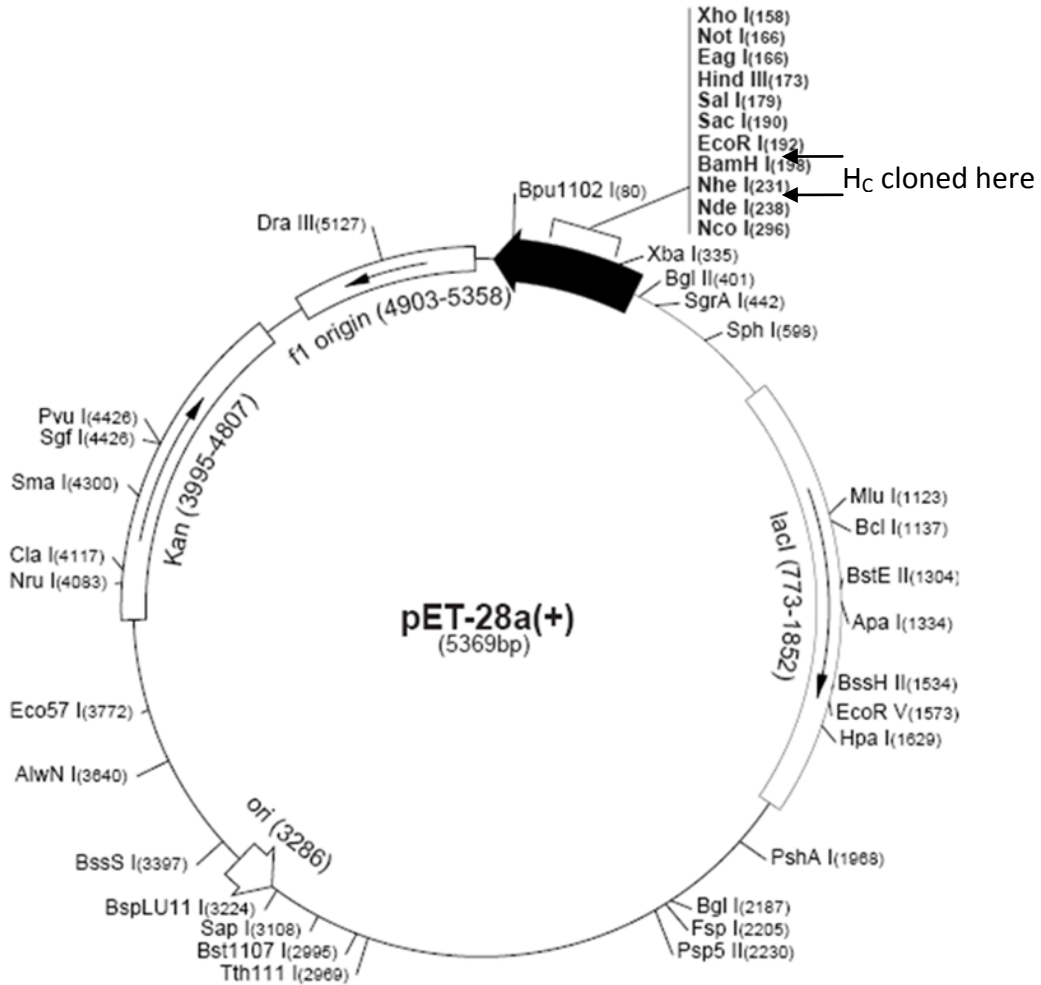
Plasmid DNA sequencing

30 µL of a plasmid after purification was subjected to sequencing at GATC, Germany, where the samples were subjected to dideoxy chain terminator DNA sequencing using BigDye™ technology.

E. coli heat-shock transformation

A 100 µL aliquot of CaCl₂ competent *E. coli* Rosetta cells in a 1.5 mL microfuge tube was thawed on ice. 2 µL of p-DNA purified as described above was added, and the cells chilled on ice for 15 minutes. The mixture was heat-shocked at 42 °C for 30 seconds and then chilled on ice for 1 minute. To the cells, 250 µL of S.O.C medium (Invitrogen, Paisley, UK) was added and the mixture incubated at 37 °C for 1 hour. The culture was then plated on LB agar plates containing kanamycin and chloramphenicol antibiotic to facilitate the selection of transformants.

A)



B)

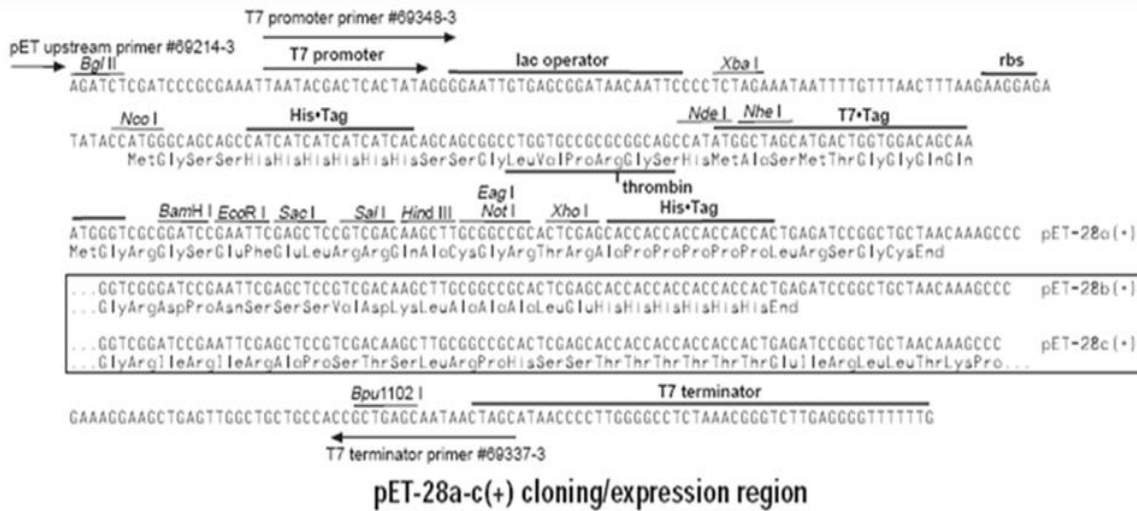


Figure 9. 1: The pET-28a (+) plasmid vector. A) The circular vector map of the pET-28a (+) plasmid is shown; the H_c DNA sequence is cloned between *Bam* HI and *Nde* I. The plasmid is 5369 bp in size (6695 bp with H_c sequence) and contains a kanamycin (Km) resistance gene for selection. The plasmid also contains the gene for the lac repressor (*lac* I) preventing expression of the H_c DNA until induced. Panel B) shows in more detail the multiple cloning site and the red arrows indicate the location of the H_c DNA sequence. Images from the Novagen catalogue (Novagen, TB074 12/98, Cat. No. 69337-3).

Growth and induction

Overnight cultures of *E. coli* Rosetta/pEt orf 4H-1, kindly donated by Neil Fairweather's laboratory, ICL, UK, harbouring plasmids based on the pET vector system (**Figure 9. 1**) were used to inoculate LB (CM and KM 50 $\mu\text{g mL}^{-1}$). Cultures were incubated at 37 °C on a shaking incubator until the OD₆₀₀ was 0.3–0.6. A 1 mL sample was taken, pelleted and stored at -20 °C. Expression of the cloned genes was chemically induced to alleviate repression by the *lac* repressor using 0.1–0.2 mM isopropyl- β -D-thiogalactoside (IPTG). Protein over-expression was induced by the addition of IPTG at the appropriate concentration (37 °C/30 °C). The culture was incubated for a further 3–4 hours, and the bacterial cells pelleted by centrifugation at 16,700 $\times g$ for 15 min. Bacterial cell pellets were stored at -20 °C. Samples of culture were taken prior to and after IPTG induction and SDS-PAGE was used to compare the protein content of these samples prior to protein purification.

Affinity purification of Cwp84

The Cwp84 mutant protein has a 6 \times His affinity tag at the N-terminus and was therefore purified by nickel affinity chromatography.

E. coli Rosetta/pEt orf 4H1 carrying the Cwp84 mutant plasmid was grown in 5 mL Overnight Express™ Instant TB medium containing 50 $\mu\text{g mL}^{-1}$ of kanamycin and 50 $\mu\text{g mL}^{-1}$ chloramphenicol overnight at 37 °C. The 5 mL culture was pelleted by centrifugation at 5000 $\times g$ for 15 minutes and the supernatant discarded. The bacterial pellet was then stored at -20 °C.

The bacterial pellet from the 5 mL cultures was defrosted and was resuspended in 500 μL of PBS, with 10% of 10 \times BugBuster Protein Extraction Reagent (Novagen) as well as 5 μL lysozyme (50 mg/mL stock) and 5 μL DNase (50 mg/mL stock) added to a final concentration of 1 mg mL^{-1} . The lysed cell pellet was centrifuged for 20 minutes at 17,000 $\times g$, 4 °C and the supernatant containing the soluble protein was removed. The pellet was resuspended in 100 μL of PBS, with 10% of 10 \times BugBuster solution. After vortexing to obtain an even suspension, 600 μL 0.1 \times BugBuster solution was added and the suspension was centrifuged at 5000 $\times g$, 4 °C. The supernatant containing the soluble protein was removed, and the pellet was washed three times by resuspending in 50 μL of

0.1 × BugBuster solution, centrifuging at 5000 × g, 4 °C, for 15 minutes. Finally, the pellet was resuspended in 1 mL of FPLC wash buffer, incubated for 1 hour at room temperature and diluted to 3 mL in a syringe ready for purification by FPLC with an affinity chromatography Ni-NTA agarose column (Qiagen). The column was then washed with 5 column volumes of FPLC wash buffer. Wash fractions were monitored at A_{280} to determine the contamination level of other proteins. Finally His-tagged recombinant protein was eluted and collected as fractions with 5 column volumes of FPLC elution buffer. $A_{280\text{nm}}$ was used to monitor elution and if necessary more elution buffer was added to elute all protein from the column. Estimation of protein purity was based on the stained SDS-PAGE gel containing fractions of interest. The pure protein fractions were combined and dialysed at 4 °C. Slide-A Lyzer Dialysis Cassette (Thermo Scientific) with a 10 kDa molecular weight cut-off limit and 3 mL sample volume were used and three changes of the dialysis buffer (2 L of 1 × PBS) were made. The protein was then aliquoted (150 µL/aliquot) and stored at –70 °C.

Concentration of Cwp84 proteins

Proteins were concentrated using ultrafree 0.5 mL centrifugal filter devices (Millipore) with a 10 kDa molecular weight cut-off limit. The membrane in the centrifugal device allows the passage of the solvent but retains molecules larger than 10 kDa. The separation membrane was first pre-rinsed; 0.5 mL of dH₂O was added to the filter device and centrifuged for 5 minutes at 3,264 × g. The flow through was discarded and the protein solution added. The filter device was centrifuged at 3,264 × g until most of the solution had flowed through. Additional protein solution was added and repeated rounds of centrifugation and protein addition were carried out until an appropriate protein concentration (judged by volume change) was achieved. At this point the concentrated protein was removed for concentration determination.

Determination of protein concentrations

The absorbance maximum of the samples were determined at 280 nm ($A_{279} - A_{350}$) using a Perkin-Elmer Lambda 800 UV-Vis spectrophotometer. The protein concentrations were calculated using theoretically determined extinction coefficients (determined from amino acid content). Solutions of 0.5 mL were scanned in 1 cm path length quartz cuvettes from 500 nm to 240 nm. Because the protein concentrations determined by BCA assay (see

below) and by absorbance using the theoretical and calculated extinction coefficients (from amino acid content) agreed reasonably well, the theoretical extinction coefficients were used for the purposes of the optical spectroscopy. Samples of approximately 2 mg ml^{-1} were diluted to an $A_{280\text{nm}}$ of approximately 1.0, and the concentrations were determined for these solutions.

BCA protein concentration assay

The Pierce BCA assay was used according to the manufacturer's guidelines. This system relies on the fact that Cu^{2+} ions are reduced to Cu^+ by protein in an alkaline environment, and reduction is proportional to protein concentration. Bicinchoninic acid (BCA) detects Cu^+ ions by chelating them to form a soluble complex (purple) with a strong absorbance at 562 nm (A_{562}). Because this assay is reliant on a colour reaction proportional to the number of exposed peptide bonds (which will be different for different proteins), the standard protein used should ideally be a known concentration of the same protein. However, a known concentration of the protein may be unavailable and bovine serum albumin (BSA) is commonly used instead. A series of seven concentrations of BSA were made starting at 1 mg mL^{-1} (made from a 2 mg mL^{-1} stock solution), doubling the dilution at each step. $20 \mu\text{L}$ of each standard solution was added to a 96 well micro-titre plate in triplicate. Test proteins were diluted (1:5, 1:10, 1:15 or 1:20) and $20 \mu\text{L}$ of each sample was added to the plate in duplicate. Pierce working reagent (WR) was made up in a 50:1 ratio of BCA reagent A: BCA reagent B and $180 \mu\text{L}$ of this mixture were added to each sample well and the plate incubated for 30 min covered in Clingfilm at 37°C . Absorbance was measured on a plate reader at 595 nm and the averaged, reference corrected standards used to plot a protein concentration standard curve, allowing unknown concentrations of protein to be calculated.

9.1.10 Antibiotic inhibition using Etest

The minimum inhibitory concentrations (MICs) of chloramphenicol, clindamycin, erythromycin, metronidazole, moxifloxacin and tetracycline were determined using Etest strips (Biomérieux, Marcy l'Etoile, France). $150 \mu\text{L}$ of an overnight *C. difficile* culture in BHI at OD 0.7 (equivalent to no. 3 McFarland standard) was spread evenly onto the surface of a BHI agar plate and was then dried openly in the anaerobic cabinet for 5 minutes. Etest strips

were placed onto the agar surface. Agar plates were incubated for 24 hours and MICs were determined following the manufacturer's instructions.

9.2 Proteomics

9.2.1 Proteomics using Imperial College facilities

Gel Preparation

Protein sample was separated using SDS-PAGE. The gel was stained with coomassie stain and the bands of interest were excised. 200 μ L of ambic buffer and 200 μ L acetonitrile (ACN) were added to the excised gel pieces and incubated at RT (ca. 25 °C) for 5 min. The mixture including gel pieces was mixed well in order to disperse the stain at the bottom of the Eppendorf tube and the supernatant carefully removed followed by drying the gel pieces on SpeedVac.

Reduction/carboxymethylation

Gel pieces were swelled in 200 μ L solution of 10 mM dithiothreitol (DTT) in ambic buffer and incubated at 56 °C for 30 min. After removing the DTT solution, gel pieces were washed briefly with 200 μ L ACN and dried on SpeedVac. Gel pieces were then re-swelled in 200 μ L 55 mM iodoacetic acid dissolved in ambic buffer and incubated at RT for 30 min in the dark. The iodoacetic acid solution was removed and gel pieces were washed with 500 μ L ambic buffer for 15 min. The supernatant was removed by pipette and discarded followed by addition of 200 μ L ACN. After 5 min ACN was removed and gel pieces were dried by SpeedVac.

Tryptic digestion

Gel pieces were re-swelled in 20 μ L sequencing grade modified trypsin solution (Promega cat V5111; 0.5 μ g trypsin in 195 μ L ambic buffer) and incubated at RT for 15 min. Further ambic buffer was added to cover gel pieces and the digestion was allowed to proceed overnight in a water bath at 37 °C.

Elution of peptides from the gel pieces

The supernatant containing hydrophilic peptides was removed from the gel pieces, and placed in a clean, labelled 1.5 mL Eppendorf tube. 50 μ L 0.1% TFA was added to each gel

piece (to halt digestion) and incubated at 37 °C for 10 min. Afterwards 100 µL ACN was added and incubated further at 37 °C for 15 min. The supernatant was removed and pooled with the hydrophilic peptide solution. These procedures (adding ACN) were repeated once before reducing the volume to 10-30 µL by SpeedVac.

MALDI-TOF Tandem Mass Spectrometric Analysis

The eluent was mixed 1:1 with CHCA solution (Agilent) and spotted on the MALDI target. Peptides were identified by matrix-assisted laser desorption ionisation time-of-flight (MALDI-TOF) mass spectrometry on a 4800 MALDI tandem time-of-flight mass spectrometer (Applied Biosystems). Briefly, peptide mass fingerprints were acquired with 1000 shots in reflector positive mode from m/z 800–4000. Numerous peptide m/z signals were identified and MS peak masses were compared with the peptide fingerprint database by Mascot 2.1 (Matrix Science) using Global Proteomics Server 3.1 (Applied Biosystems).

9.2.2 Proteomics using the facilities of the Proteomics Centre of the University of Leicester

Sample processing

Protein sample was separated using SDS-PAGE. The gel was stained with Coomassie stain and the band contained the interested protein will be cut out in pieces. Gel pieces were covered in 100 µL of ambic buffer and were frozen with liquid nitrogen. Samples were sent to University of Leicester and were processed by Dr Andrew Bottrill as follows: after digestion with trypsin the samples were spun-down and the supernatant (containing peptides) was removed. The gel pieces were extracted with acetonitrile for 30 min and this was then added to the supernatant. The combined sample was evaporated in a SpeedVac until dry. Each sample was resuspended in 20 µL of 0.1% trifluoroacetic acid containing 10 fmol/µL of a standard protein digest (ADH—Alcohol dehydrogenase from yeast—Uniprot P00330). 6 µL of this sample was put into a vial for each replicate and 5 µL was injected for LC-MS/MS analysis. Thus each replicate consisted of 25% of the sample and contained 50 fmol of the ADH standard. Mass spectra were analysed using Scaffold.

Database searching

Tandem mass spectra were extracted. Charge state deconvolution and deisotoping were not performed. All MS/MS samples were analysed using Mascot (Matrix Science, London, UK;

version Mascot). Mascot was set up to search the Trembl_39.6 database (selected for *Clostridium difficile*, 3896 entries) assuming the digestion enzyme trypsin. Mascot was searched with a fragment ion mass tolerance of 0.60 Da and a parent ion tolerance of 5.0 PPM. Oxidation of methionine, iodoacetamide derivative of cysteine and inhibitors (e.g.: EtEP-R-PEG₃-Biotin) of cysteine were specified in Mascot as variable modifications.

Criteria for protein identification

Scaffold (version Scaffold_2_06_02, Proteome Software Inc., Portland, OR) was used to validate MS/MS based peptide and protein identifications. Peptide identifications were accepted if they could be established at greater than 95.0% probability as specified by the Peptide Prophet algorithm [408]. Protein identifications were accepted if they could be established at greater than 95.0% probability and contained at least two identified peptides. Protein probabilities were assigned by the Protein Prophet algorithm [460]. Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony.

9.3 Chemistry

9.3.1 Reagents and Equipment

All general laboratory chemicals obtained from chemical suppliers (Novabiochem UK or Aldrich Chemical Co.) were used without further purification.

Flash chromatography was performed on Screening Devices silica gel 60 (0.04–0.063 mm). TLC-analysis was conducted on DC-alufolien (Merck, Kieselgel60, F254) with detection by UV-absorption (254 nm).

¹H and ¹³C NMR spectra were recorded on a Brüker AV-400 (400/100 MHz) spectrometer. Chemical shifts (δ) are given in ppm relative to tetramethylsilane and chloroform, H₂O or DMSO residual solvent peak was used as internal standard. Coupling constants are given in Hz.

Optical rotations were measured at 25 °C on a Perkin-Elmer 241 polarimeter with sodium lamp, and are given in 10⁻¹ deg cm² g⁻¹; concentration (*c*) is in g per 100 mL.

Infrared spectra were obtained on PerkinElmer Spectrum 100 FT-IR spectrophotometer, from a thin film deposited onto a sodium chloride plate from dichloromethane. The IR spectra were then analyzed with Spectrum Express, Version 1.01.00 (PerkinElmer).

Mass spectra and accurate mass data were obtained on an AUTOSPEC P673 spectrometer by J. Barton at the Chemistry Department Mass Spectrometry Service (Imperial College London) by electrospray ionisation or chemical ionisation techniques.

Solid Phase Resins: Rink-amide resin and Biotin-PEG-Novatag resin were purchased from Novabiochem UK. The resin substitution ratios were as follows: Rink-amide resin: 0.71 mmol/g; Biotin-PEG-Novatag resin: 0.48 mmol/g; Universal-PEG-Novatag resin: 0.33 mmol/g.

Amino Acids: *N*- α -9-Fluorenylmethoxycarbonyl (*N*- α -Fmoc) protected amino acids were obtained from Novabiochem UK with the following side chain protecting groups: Arg(Pbf), Lys(Boc), Tyr(*t*Bu), Thr(*t*Bu), Glu(*t*Bu), Ser(*t*Bu).

Reagents: Peptide synthesis grade dimethylformamide (DMF) was purchased from National Diagnostics, UK. Benzotriazole-1-yloxy-tris-pyrrolidino-phosphonium hexafluorophosphate (PyBop), *N*-hydroxybenzotriazole (HOBt), diisopropylethylamine (DIPEA) and 2-(1*H*-benzotriazole-1-yl)-1,1,3,3-tetramethylaminium hexafluorophosphate (HBTU) were purchased from Novabiochem UK and Sigma Aldrich. Dry tetrahydrofuran (THF) and DMF, chloroform, dimethylsulfoxide (DMSO), dichloromethane (DCM), methanol, piperidine, thioanisole, *t*butylmethylether (TBME), trifluoroacetic acid (TFA), ethanedithiol (EDT), iodoacetonitrile, 3-mercaptopethyl propionate, sodium thiophenolate, monobasic and dibasic sodium phosphate, trimethylsilylisopropane (TIPS) were reagent grade from Sigma Aldrich. A 'deprotection mixture' of TFA–H₂O–TIPS (97:2.5:1) was used for cleavage and deprotection reactions.

9.3.2 General procedures

Solid Phase Peptide Synthesis (SPPS)

Automated solid phase synthesis was performed using an Advanced ChemTech Apex 396 multiple peptide synthesizer (Advanced ChemTech Europe, Cambridge, UK). Peptide synthesis was carried out in peptide synthesis grade DMF. Resin (25 μ mol per well) was

swelled in DMF for 60 min before coupling each cycle, *N*- α -amino Fmoc group was treated with 20% (v/v) piperidine in DMF for 15 min, with two further repetitions. A fivefold excess of amino acid over resin reactive groups (125 μ mol, 250 μ L of 0.5 M solution) was used for each coupling. *In situ* activation and coupling was carried out for 45 min with a mixture of HBTU/HOBt (125 μ mol, 250 μ L of 0.5 M solution) and DIPEA (125 μ mol, 250 μ L of 0.5 M solution). The peptide was washed with DMF (3 \times 1 mL) between each deprotection and coupling step. A typical cycle consisted of deprotection, DMF wash, coupling and a further DMF wash. The *N*- α -Fmoc protection at the final residue was removed at the end of the synthesis under the usual conditions. After synthesis was complete a preactivated solution of (2*S*,3*S*)-3-(ethoxycarbonyl)oxirane-2-carboxylic acid (12.01 mg, 3 eq), HCTU (51.7 mg, 3 eq) and DiPEA (19.3 mg, 6 eq) in DMF (1 mL) was added to each well and agitated for one hour. The peptidyl resin was removed from the synthesiser, washed several times (3 \times 2 mL DMF, 3 \times 2 mL DCM, 3 \times 2 mL methanol, 3 \times 2 mL diethylether) and dried *in vacuo*.

Deprotection and cleavage from resin

Cleavage and deprotection of peptides was achieved by adding 1.5 mL of deprotection mixture to dry peptidyl resin (25 μ mol). The mixture was agitated on an orbital shaker for up to 3 hr and then filtered. The resin was washed twice with a small volume of TFA and the combined washings and filtrate were precipitated with 10 mL ice cold *tert*-butylmethylether (tBME). The mixture was centrifuged at 5000 rpm for 15 min at 0 $^{\circ}$ C, the supernatant discarded and the remaining peptide washed with a fresh aliquot of tBME. The process was repeated three times to ensure complete removal of all organic impurities. The crude peptide was dried in a desiccator over silica gel to yield an off-white solid.

Peptide purification

Peptide analysis and purification was performed either on Gilson RP-HPLC system or Waters LC-MS system.

Purification of crude peptides was performed on a Gilson semi-preparative RP-HPLC system (Anachem Ltd., Luton, UK) equipped with 306 pumps and a Gilson 155 UV/Vis detector. Analytical RP-HPLC was performed on a Gilson analytical HPLC system (Anachem Ltd., Luton, UK) equipped with a Gilson 151 UV/Vis detector and Gilson 234 auto injector. For both HPLC systems, the peptide bond absorption was detected at 223 nm. The following elution

methods were used: *Method A*: 2% to 98% MeCN in H₂O over 40 min; *Method B*: 2% MeCN in H₂O for 10 min then 2% to 98% MeCN in H₂O over 40 min; all solvents used were supplemented with 0.1% TFA and were degassed with helium.

LC-MS analysis was performed on Waters HPLC system (Waters 2767 autosampler for samples injection and collection; Waters 515 HPLC pump to deliver the mobile phase to the source; XBridge C18 column (Waters, 4.6 mm D × 100 mm L for analytical and 19 mm D × 100 mmL for preparative); Waters 3100 mass spectrometer with ESI and Waters 2998 Photodiode Array (detection at 200–600 nm)). The following elution methods were used: *Method C*: 5% to 95% MeOH in H₂O over 15 min; *Method D*: 5% to 45% MeOH in H₂O over 10 min, then 45% to 55% MeOH in H₂O over further 10 min and finishing with 55% to 95% MeOH in H₂O over the remain 5 min. All solvents used were supplemented with 0.1% formic acid and were degassed with helium.

9.3.3 Synthesis and characterisation

Key data are summarised in **Table 9. 5**, experimental procedures and further characterisation data for selected compounds is provided below.

Nr.	Abbr. Name	ESI-MS [M+H ⁺]		R _t	PM	Yield (%)
		Calculated (Chem. formular)	found			
TD 1	EP-LR-NH ₂	401.2149 (C ₁₆ H ₂₉ N ₆ O ₆)	401.2153	9.15	A	22
TD 3	Ac-LR-NH ₂	329.2301 (C ₁₄ H ₂₉ N ₆ O ₃)	329.2286	5.95	A	43
TD 4	(<i>R,R</i>)EP-LR-NH ₂	401.2143 (C ₁₆ H ₂₉ N ₆ O ₆)	401.2153	10.72	A	9
TD 5	EtEP-LR-NH ₂	429.2462 (C ₁₈ H ₃₂ N ₆ O ₆)	429.2463	10.88	A	22
TD 6	EtEP-K-NH ₂	288.1559 (C ₁₂ H ₂₂ N ₃ O ₅)	288.1588	4.12	A	n.d
TD 7	EP-KT-NH ₂	389.2050 (C ₁₆ H ₂₉ N ₄ O ₇)	389.2046	4.15	A	n.d
TD 8	EP-KTE-NH ₂	490.2513 (C ₁₉ H ₃₂ N ₅ O ₁₀)	490.2506	4.15	A	n.d
TD 9	EP-KTEL-NH ₂	603.2990 (C ₂₅ H ₄₃ N ₆ O ₁₁)	603.2997	8.62	A	25
TD 9b	EtEP-KTEL-NH ₂	631.3320 (C ₂₇ H ₄₇ N ₆ O ₁₁)	631.3297	10.11	A	25
TD 10	EP-KA-NH ₂	331.1612 (C ₁₃ H ₂₃ N ₄ O ₆)	331.1606	3.83	A	8
TD 11	EP-KTT-NH ₂	462.2200 (C ₁₈ H ₃₂ N ₅ O ₉)	462.2212	4.08	A	2

TD 12	EP-YT-NH ₂	396.1401 (C ₁₇ H ₂₂ N ₃ O ₈)	396.1414	6.72	A	17
TD 13	EP-LT-NH ₂	346.1609 (C ₁₄ H ₂₄ N ₃ O ₇)	346.1618	9.39	A	39
TD 14	EP-LTT-NH ₂	447.2091 (C ₁₈ H ₃₁ N ₄ O ₉)	447.2084	10.36	A	16
TD 15	EP-LTTT-NH ₂	560.2932 (C ₂₄ H ₄₁ N ₅ O ₁₀)	560.2953	10.78	B	1
TD 16	EP-LGR-NH ₂	458.2363 (C ₁₈ H ₃₂ N ₇ O ₇)	458.2375	9.08	A	33
TD 17	EP-T-NH ₂	233.0729 (C ₈ H ₁₃ N ₂ O ₆)	233.0768	8.91	A	62
TD 18	EtEP-AT-NH ₂	332.1438 (C ₁₃ H ₂₂ N ₃ O ₇)	332.1461	5.81	A	87
TD 19	EtEP-ATT-NH ₂	433.1935 (C ₁₇ H ₂₉ N ₄ O ₉)	433.1930	3.39	A	n.d
TD 20	EtEP-R-NH ₂	316.1621 (C ₁₂ H ₂₂ N ₅ O ₅)	316.1609	9.44	A	62
TD 21	EP-LR-Pra-NH ₂	496.2520 (C ₂₁ H ₃₃ N ₇ O ₇)	496.2539	8.02	C	60
TD 22	EtEP-LR-Pra-NH ₂	524.2833 (C ₂₃ H ₃₇ N ₇ O ₇)	524.2811	9.08	C	44
TD 23	EtEP-A-Pra-NH ₂	326.1352 (C ₁₄ H ₁₉ N ₃ O ₆)	326.1339	7.68	C	7
TD 24	EtEP-F-Pra-NH ₂	402.1665 (C ₂₀ H ₂₃ N ₃ O ₆)	402.1679	10.56	C	1
TD 25	EtEP-I-Pra-NH ₂	368.1822 (C ₁₇ H ₂₅ N ₃ O ₆)	368.1830	9.85	C	4
TD 26	EtEP-L-Pra-NH ₂	368.1822 (C ₁₇ H ₂₅ N ₃ O ₆)	368.1808	10.00	C	8
TD 27	EtEP-V-Pra-NH ₂	354.1660 (C ₁₆ H ₂₃ N ₃ O ₆)	354.9669	9.01	C	3
TD 28	EtEP-Y-Pra-NH ₂	418.1599 (C ₂₀ H ₂₃ N ₃ O ₇)	418.1614	8.98	C	8
TD 29	EtEP-S-Pra-NH ₂	342.1301 (C ₁₄ H ₁₉ N ₃ O ₇)	342.1293	7.55	C	7
TD 30	EtEP-H-Pra-NH ₂	392.1568 (C ₁₇ H ₂₂ N ₅ O ₆)	392.1570	6.44	C	2
TD 31	EtEP-K-Pra-NH ₂	383.1931 (C ₁₇ H ₂₇ N ₄ O ₆)	383.1920	6.48	C	4
TD 32	EtEP-R-Pra-NH ₂	411.1992 (C ₁₇ H ₂₇ N ₆ O ₆)	411.2005	6.92	C	1
TD 33	EtEP-E-Pra-NH ₂	384.1407 (C ₁₆ H ₂₂ N ₃ O ₈)	384.1394	5.44	C	4
TD 34	EtEP-G-Pra-NH ₂	312.1196 (C ₁₃ H ₁₈ N ₃ O ₆)	312.1206	7.09	C	7
TD 35	EtEP-P-Pra-NH ₂	352.1509 (C ₁₆ H ₂₂ N ₃ O ₆)	352.1494	7.90	C	4.6
TD 36	EtEP-T-Pra-NH ₂	356.1458 (C ₁₅ H ₂₂ N ₃ O ₇)	356.1464	7.82	C	n.d
TD 37	(<i>R,R</i>)EtEP-LR-Pra-NH ₂	524.2833 (C ₂₃ H ₃₇ N ₇ O ₇)	524.2833	9.85	C	n.d
TD 38	EtEP-Pra-NH ₂	255.0981 (C ₁₁ H ₁₅ N ₂ O ₅)	255.0978	7.67	C	5
TD 39	EtEP-(<i>D</i>)-A-Pra-NH ₂	326.1352 (C ₁₄ H ₁₉ N ₃ O ₆)	326.1354	7.70	C	29
TD 40	EtEP-(<i>D</i>)-K-Pra-NH ₂	383.1931 (C ₁₇ H ₂₇ N ₄ O ₆)	383.1931	6.18	C	53
TD 41	EtEP-(<i>D</i>)-R-Pra-NH ₂	411.1992 (C ₁₇ H ₂₇ N ₆ O ₆)	411.1992	7.08	C	62
TD 42	EtEP-(<i>D</i>)-Y-Pra-NH ₂	418.1614 (C ₂₀ H ₂₃ N ₃ O ₇)	418.1619	10.55	C	64

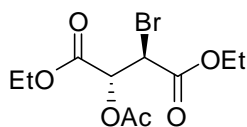
TD 43	EP-LR-PEG ₃ -Biotin	830.4468 (C ₃₆ H ₆₄ N ₉ O ₁₁ S)	830.4470	10.68	A	46
TD 44	EtEP-LR-PEG ₃ -Biotin	858.4781 (C ₃₈ H ₆₈ N ₉ O ₁₁ S)	858.4778	11.09	A	46
TD 45	(<i>R,R</i>)EtEP-LR-PEG ₃ - Biotin	858.4759 (C ₃₈ H ₆₈ N ₉ O ₁₁ S)	858.4757	11.12	A	20
TD 46	EtEP-A-PEG ₃ -Biotin	660.3278 (C ₂₉ H ₅₀ N ₅ O ₁₀ S)	660.3270	10.01	A	54
TD 47	EtEP-F-PEG ₃ -Biotin	736.3645 (C ₃₅ H ₅₄ N ₅ O ₁₀ S)	736.3611	12.53	B	14
TD 48	EtEP-I-PEG ₃ -Biotin	702.3751 (C ₃₂ H ₅₆ N ₅ O ₁₀ S)	702.3751	11.66	B	8
TD 49	EtEP-L-PEG ₃ -Biotin	702.3748 (C ₃₂ H ₅₆ N ₅ O ₁₀ S)	702.3747	11.94	B	46
TD 50	EtEP-V-PEG ₃ -Biotin	688.3573 (C ₃₁ H ₅₄ N ₅ O ₁₀ S)	688.3601	11.52	B	54
TD 51	EtEP-Y-PEG ₃ -Biotin	752.3452 (C ₃₅ H ₅₄ N ₅ O ₁₁ S)	752.3539	10.40	B	8
TD 52	EtEP-S-PEG ₃ -Biotin	676.3228 (C ₂₉ H ₅₀ N ₅ O ₁₁ S)	676.3232	9.55	A	20
TD 53	EtEP-H-PEG ₃ -Biotin	726.3577 (C ₃₂ H ₅₂ N ₇ O ₁₀ S)	726.3529	9.81	A	30
TD 54	EP-K-PEG ₃ -Biotin	689.3557 (C ₃₀ H ₅₃ N ₆ O ₁₀ S)	689.3557	11.45	A	29
TD 55	EtEP-R-PEG ₃ -Biotin	745.3918 (C ₃₂ H ₅₇ N ₈ O ₁₀ S)	745.3939	9.78	A	16
TD 56	EtEP-E-PEG ₃ -Biotin	718.3255 (C ₃₁ H ₅₂ N ₅ O ₁₂ S)	718.3333	9.73	A	13
TD 57	EtEP-G-PEG ₃ -Biotin	646.3103 (C ₂₈ H ₄₈ N ₅ O ₁₀ S)	646.3131	10.70	B	62
TD 58	EtEP-P-PEG ₃ -Biotin	686.3475 (C ₃₁ H ₅₂ N ₅ O ₁₀ S)	686.3447	11.73	A	43
TD 59	EtEP-T-PEG ₃ -Biotin	690.3424 (C ₃₀ H ₅₂ N ₅ O ₁₁ S)	690.3389	9.94	A	33
TD 60	EtEP-PEG ₃ -Biotin	589.2907 (C ₂₆ H ₄₅ N ₄ O ₉ S)	589.2911	11.81	A	25
TD 61	EtEP-AG-PEG ₃ - TAMRA	903.4122 (C ₄₆ H ₅₉ N ₆ O ₁₃)	903.4135	11.22	D	n.d
TD 62	EtEP-YG-PEG ₃ - TAMRA	995.4442 (C ₅₂ H ₆₃ N ₆ O ₁₄)	995.4438	11.22	D	n.d
TD 63	EtEP-R-PEG ₃ -TAMRA	988.4861 (C ₄₉ H ₆₆ N ₉ O ₁₃)	988.4819	10.84	D	n.d
TD 64	EtEP-AK(Biotin)G- PEG ₃ -TAMRA	1256.5787 (C ₆₂ H ₈₅ N ₁₀ O ₁₆ S)	1256.5787	9.39	D	n.d
TD 65	EtEP-RK(Biotin)G- PEG ₃ -TAMRA	1341.6427 (C ₆₅ H ₉₂ N ₁₃ O ₁₆ S)	1341.64	9.37	D	n.d
TD 66	EtEP-YK(Biotin)G- PEG ₃ -TAMRA	1348.6050 (C ₆₈ H ₈₉ N ₁₀ O ₁₇ S)	1348.6050	10.76	D	n.d

TD 67	CAA-LR-PEG ₃ -Biotin	792.4209 (C ₃₄ H ₆₂ ClN ₉ O ₈ S)	792.4240	9.87	C	28
TD 68	CAA-A-PEG ₃ -Biotin	594.2728 (C ₂₅ H ₄₅ ClN ₅ O ₇ S)	594.2730	9.45	C	5
TD 69	CAA-Y-PEG ₃ -Biotin	686.2990 (C ₃₁ H ₄₉ ClN ₅ O ₈ S)	686.2992	10.69	C	3
TD 70	CAA-R-PEG ₃ -Biotin	679.3337 (C ₂₈ H ₅₂ ClN ₈ O ₇ S)	679.3362	8.71	C	37
TD 71	CAA-T-PEG ₃ -Biotin	624.2834 (C ₂₆ H ₄₇ ClN ₅ O ₈ S)	624.2831	9.47	C	4
TD 72	CAA-S-PEG ₃ -Biotin	610.2677 (C ₂₅ H ₄₅ ClN ₅ O ₈ S)	610.2681	9.16	C	7
TD 73	FAA-LR-PEG ₃ -Biotin	776.4504 (C ₃₄ H ₆₃ FN ₉ O ₈ S)	776.4523	9.59	C	7
TD 74	FAA-A-PEG ₃ -Biotin	578.3024 (C ₂₅ H ₄₅ FN ₅ O ₇ S)	578.3052	9.18	C	15
TD 75	FAA-Y-PEG ₃ -Biotin	670.3286 (C ₃₁ H ₄₉ FN ₅ O ₈ S)	670.3282	9.67	C	6
TD 76	FAA-R-PEG ₃ -Biotin	663.3664 (C ₂₈ H ₅₂ FN ₈ O ₇ S)	663.3661	8.46	C	9
TD 77	FAA-S-PEG ₃ -Biotin	594.2973 (C ₂₅ H ₄₅ FN ₅ O ₈ S)	594.2974	8.91	C	1
TD 78	MA-LR-PEG ₃ -Biotin	842.4810 (C ₃₈ H ₆₈ N ₉ O ₁₀ S)	842.4819	10.39	C	10
TD 79	MA-A-PEG ₃ -Biotin	644.3329 (C ₂₉ H ₅₀ N ₅ O ₉ S)	644.3331	10.18	C	17
TD 80	MA-Y-PEG ₃ -Biotin	736.3591 (C ₃₅ H ₅₄ N ₅ O ₁₀ S)	736.3600	10.52	C	21
TD 81	MA-R-PEG ₃ -Biotin	729.3969 (C ₃₂ H ₅₇ N ₈ O ₉ S)	729.3961	9.33	C	39
TD 82	MA-T-PEG ₃ -Biotin	674.3435 (C ₃₀ H ₅₂ N ₅ O ₁₀ S)	674.3438	9.74	C	26
TD 83	MA-S-PEG ₃ -Biotin	660.3278 (C ₂₉ H ₅₀ N ₅ O ₁₀ S)	660.3292	9.92	C	25
TD 84	AOAA-LR-PEG ₃ - Biotin	908.4897 (C ₄₂ H ₇₀ N ₉ O ₁₁ S)	908.4933	11.61	C	99
TD 85	AOAA-A-PEG ₃ -Biotin	710.3435 (C ₃₃ H ₅₂ N ₅ O ₁₀ S)	710.3442	11.60	C	17
TD 86	AOAA-Y-PEG ₃ -Biotin	802.3737 (C ₃₉ H ₅₆ N ₅ O ₁₁ S)	802.3736	11.99	C	51
TD 87	AOAA-R-PEG ₃ -Biotin	795.4075 (C ₃₆ H ₅₉ N ₈ O ₁₀ S)	795.4080	10.65	C	38

TD 88	AOAA-T-PEG ₃ -Biotin	740.3541 (C ₃₄ H ₅₄ N ₅ O ₁₁ S)	740.3546	11.02	C	88
TD 89	AOAA-S-PEG ₃ -Biotin	726.3384 (C ₃₃ H ₅₂ N ₅ O ₁₁ S)	726.3384	11.20	C	31
TD 90	Ac-LTEKS-MA	673.3772 (C ₃₀ H ₅₃ N ₆ O ₁₁)	673.3775	5.26	A	1.5
TD 91	Ac-K(Biotin)Y-MA	832.4279 (C ₄₀ H ₆₂ N ₇ O ₁₀ S)	832.4247	10.35	B	11
TD 92	Ac-K(Biotin)L-MA	782.4486 (C ₃₇ H ₆₄ N ₇ O ₉ S)	782.4479	10.55	B	4
TD 93	Ac-K(Biotin)E-MA	798.4072 (C ₃₆ H ₆₀ N ₇ O ₁₁ S)	798.4069	10.52	B	4
TD 94	Ac-K(Biotin)M-MA	800.4050 (C ₃₆ H ₆₂ N ₇ O ₉ S ₂)	800.4009	10.48	B	7

Table 9. 5: Table of inhibitors synthesised and used in this study. The retention time (R_t) for LC-MS used a gradient of Methanol in H₂O with 0.1% of FA from 2 to 98% over 15 min. PM = Purification method.

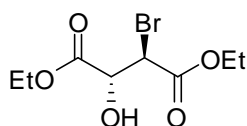
(2R, 3R)-Diethyl-2-acetoxy-3-bromosuccinate [461-462]



To neat (-)-diethyl-D-tartrate (10 mL, 58.44 mmol) cooled in ice was added hydrobromic acid 33% in solution in acetic acid (5.7 M) (40 mL) drop wise over 30 minutes. The reaction mixture was allowed to warm to room temperature and stirred for 16 h. The reaction mixture was poured into 50 g crushed ice and extracted four times with diethyl ether. The combined organic layers were washed three times with water and once with brine. After drying over magnesium sulphate all volatiles were removed under reduced pressure to afford the product (2R,3R)-diethyl-2-acetoxy-3-bromosuccinate (18.61 g, 59.88 mmol, 88% yield) as a colourless oil. The product contained acetic acid (ca. 50%) and was used in the next synthesis step without further purification.

δ_H /ppm (400 MHz; CDCl₃): 5.56 (1H, d, J = 5.4, CHOAc), 4.76 (1H, d, J = 5.4, CHBr), 4.28–4.16 (4H, m, 2 × CH₂), 2.07 (3H, s, COCH₃), 1.33–1.24 (6H, m, 2 × CH₃). δ_C /ppm (100 MHz; CDCl₃): 169.44, 166.25, 165.63 (CO), 72.66 (COAc), 62.81, 62.53 (CH₂), 43.86 (CBr), 20.70 (COCH₃), 13.94, 13.86 (CH₃).

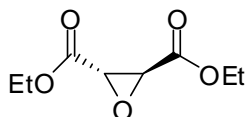
(2R,3R)-Diethyl 2-bromo-3-hydroxysuccinate (3)



A solution of acetyl chloride (2.08 mL, 29.22 mmol) was added to a solution of (2R, 3R)-diethyl-2-acetoxy-3-bromosuccinate (18.61 g, 59.88 mmol) in ethanol (60 mL). The reaction mixture was heated to a gentle reflux for 6 hours, 1 hour longer than literature procedure, due to difficulty visualising on TLC. After 6 hours all volatiles were removed by rotary evaporation. The product (2R,3R)-diethyl-2-bromo-3-hydroxysuccinate was recovered as pale yellow oil (15.30 g, 56.86 mmol, 97% yield).

δ_{H} /ppm (400 MHz; CDCl_3): 7.48 (1H, s, OH), 4.68 (1H, d, $J = 4.5$, HOCH), 4.65 (1H, d, $J = 4.5$, BrCH), 4.36–4.16 (4H, m, $2 \times \text{CH}_2$), 1.29 (6H, t, $J = 7.2$, $2 \times \text{CH}_3$). δ_{C} /ppm (100 MHz; CDCl_3): 170.32 (COCO), 166.72 (COCBr), 72.52 (COH), 62.84, 62.53 (CH_2), 47.55 (CBr), 14.00, 13.88 (CH_3).

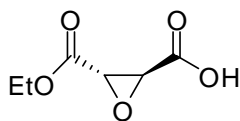
(2S,3S)-Diethyl-2,3-epoxysuccinate(4)



To an ice cooled solution of (2R,3R)-diethyl-2-bromo-3-hydroxysuccinate (15.03 g, 56.86 mmol) in diethyl ether (80 mL) stirring, a solution of 1,8-diazabicyclo-undec-7-ene (DBU) (11 mL, 73.9 mmol) in diethyl ether (40 mL) dropwise over 2 hours. Once the addition was completed the reaction mixture was stirred for a further 30 minutes. Water (10 mL) was added to the mixture. The separated organic layer was washed three times with 1 M potassium hydrogen sulphate and twice with water. After drying over magnesium sulphate all volatiles were removed under reduced pressure to afford a product (8.31 g, 44.15 mmol, 77% yield) as yellow oil.

δ_{H} /ppm (400 MHz; CDCl_3): 4.35–4.13 (4H, m, CH_2), 3.65 (2H, s, CHOCH), 1.31 (6H, t, $J = 7.2$, CH_3). ν_{max} (neat)/ cm^{-1} 2983, 2940, 2875 (CH), 1748 (CO), 1447, 1371, 1329, 1197, 1029, 901.

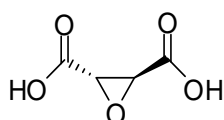
(2S,3S)-3-(Ethoxycarbonyl)oxirane-2-carboxylic acid (EtEP-OH) (5 or TD 2)



To an ice cooled solution of (2S,3S)-diethyloxirane-2,3-dicarboxylate (1.00 g, 5.31 mmol) in ethanol (13 mL) was added drop wise a solution of potassium hydroxide (0.35 g, 6.26 mmol) in ethanol (7 mL) over 15 minutes. After this addition the reaction mixture was allowed to stir in ice for 3 hours. The cooling bath was then removed and the reaction mixture was stirred for further 2 hours at room temperature. Afterwards the ethanol was removed under reduced pressure leaving an off-white solid. This salt was then taken up with 15 mL of water and washed three times with dichloromethane. The aqueous phase was acidified by addition of concentrated hydrochloric acid to pH 3 and extracted four times with ethyl acetate. The combined organic phases were dried over magnesium sulphate and concentrated under reduced pressure. After freeze drying the pure product was obtained as a pale yellow oil (0.67 g, 4.19 mmol, 79% yield).

$\delta_{\text{H}}/\text{ppm}$ (400 MHz; CDCl_3): 4.27 (2H, q, $J = 7.3$, CH_2), 3.64 (1H, d, $J = 1.6$, CH), 3.61 (1H, d, $J = 1.7$, CH), 1.32 (3H, t, $J = 7.1$, CH_3). $\delta_{\text{C}}/\text{ppm}$ (100 MHz; CDCl_3): 168.87, 167.05 (CO), 62.12 (CH_2), 51.98, 51.90 (CH), 13.95 (CH_3). $[\alpha]_{\text{D}}^{20} +269$ (c 1. in DCM). ν_{max} (neat)/ cm^{-1} : 2988 (CH), 1744 (CO), 1448, 1374, 1206, 1095, 1025, 899.

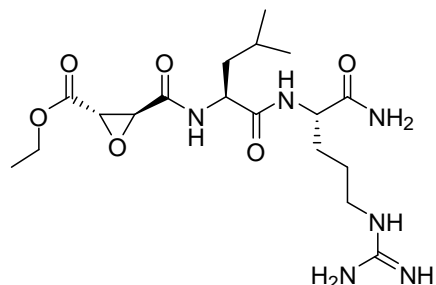
(2S,3S)-Oxirane-2,3-dicarboxylic acid[461-462]



A solution of potassium hydroxide (2.76 g, 42.4 mmol) in ethanol (50 mL) was added to a solution of (2S,3S)-diethyl-2-bromo-3-hydroxysuccinate (4.00 g, 21.2 mmol) in ethanol (150 mL) and stirred for 3 hours at room temperature. After evaporation to dryness, the residue was dissolved in 10 mL of water and the pH was adjusted to 3 with concentrated hydrochloric acid. When the mixture was cooled, the precipitated product was filtrated and freeze dried to afford a white solid. The product contained potassium chloride, and thus it

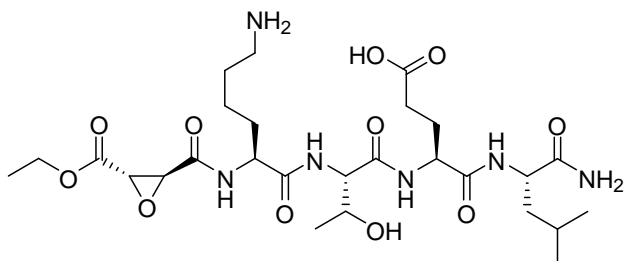
was not possible to calculate an accurate yield. δ_{H} /ppm (400 MHz; CDCl_3): 3.46 (2H, s, CHOCH). δ_{C} /ppm (100 MHz; CDCl_3): 172.71 (CO), 52.78 (CH).

(2S,3S)-Ethyl-2,3-epoxysuccinyl leucylargininamide (EtEP-LR-NH₂) (TD 5)



δ_{H} /ppm (400 MHz; D_2O): 4.334.30 (1H, m, NHCHArgCO), 4.274.22 (1H, m, NHCHLeuCO), 4.21 (2H, q, $J = 7.15$, CH_3CH_2), 3.71 (1H, d, $J = 1.86$, CHOCHCOOEt), 3.64 (1H, d, $J = 1.86$, CHOCHCOOEt), 3.13 (2H, t, $J = 6.92$, $\text{CH}_2\text{NHCNHNH}_2$), 1.84–1.51 (7H, m, $\text{CH}_2\text{CH}(\text{CH}_3)_2$, CH_2), 1.21 (3H, t, $J = 7.17$, CH_3CH_2), 0.86 (3H, d, $J = 5.80$, $\text{CH}(\text{CH}_3)_2$), 0.81 (3H, d, $J = 5.75$, $\text{CH}(\text{CH}_3)_2$). R_{f} : 10.88.

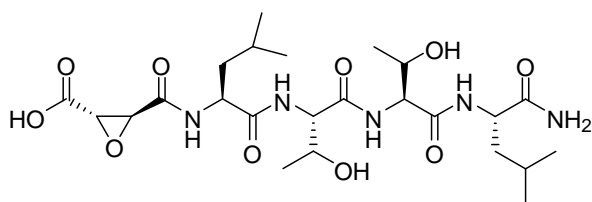
(2S,3S)-Ethyl-2,3-epoxysuccinyl lysylthreonylglutamylleucinamide (EtEP-KTEL-NH₂) (TD 9b)



Standard SPPS. Product was obtained as white solid.

δ_{H} /ppm (400 MHz; D_2O): 4.34–4.32 (2H, m, NHCHCO), 4.24–4.17 (4H, m, NHCHCO , CH_3CH_2), 4.08 (1H, p, $J = 6.29$, CHOH), 3.71 (1H, d, $J = 1.71$, CHOCHCOOEt), 3.67 (1H, d, $J = 1.73$, CHOCHCOOEt), 2.90 (2H, t, $J = 7.58$, CH_2NH_2), 2.38 (2H, t, $J = 7.31$, CH_2COOH), 2.09–2.00 (1H, m, $\text{CH}_2\text{CH}_2\text{COOH}$), 1.94–1.84, (1H, m, $\text{CH}_2\text{CH}_2\text{COOH}$), 1.82–1.65 (2H, m, $\text{CH}_2(\text{CH}_2)_3\text{NH}_2$), 1.62–1.48 (5H, m, $\text{CH}(\text{CH}_3)_2$, $\text{CH}_2\text{CH}_2\text{NH}_2$, $\text{CH}_2\text{CH}(\text{CH}_3)_2$), 1.40–1.30 (2H, m, CH_2CH_2), 1.20 (3H, t, $J = 7.14$, CH_3CH_2), 1.10 (3H, d, CH_3CHOH), 0.83 (3H, d, $J = 5.88$, $\text{CH}(\text{CH}_3)_2$), 0.78 (3H, d, $J = 5.90$, $\text{CH}(\text{CH}_3)_2$). R_{f} : 10.11.

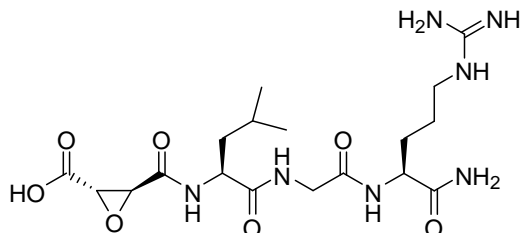
(2S,3S)-2,3-Epoxy succinyl leucylthreonylthreonylleucinamide (EP-LTTL-NH₂) (TD 14)



Standard SPPS. Product was obtained as white solid.

δ_{H} /ppm (400 MHz; D₂O): 4.40–4.35 (2H, m, NHCHCO), 4.27–4.23 (2H, m, NHCHCO), 4.15–4.09 (1H, m, CHOH), 3.65 (1H, d, $J = 1.86$, CHOCHCOOH), 3.55 (1H, d, $J = 1.86$, CHOCHCOOH), 2.94 (2H, t, $J = 7.42$, CH₂NH₂), 2.42 (2H, t, $J = 7.15$, CH₂COOH), 2.12–2.03, 1.97–1.90 (2H, m, CH₂CH₂COOH), 1.85–1.71 (2H, m, CH₂(CH₂)₃NH₂), 1.67–1.51 (5H, m, CH(CH₃)₂, CH₂CH₂NH₂, CH₂CH(CH₃)₂), 1.43–1.33 (2H, m, CH₂CH₂), 1.14 (3H, t, $J = 6.19$ CH₃CHOH), 0.87 (3H, d, $J = 5.19$, CH(CH₃)₂), 0.81 (3H, d, $J = 5.17$, CH(CH₃)₂). **ESI-MS:** $m/z = 603.2997$ ([M + H]⁺), 652.9226 ([M + Na]⁺). **R_t:** 10.78.

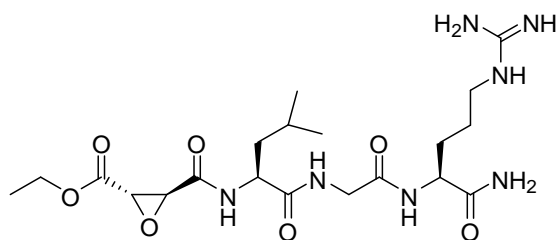
(2S,3S)-2,3-Epoxy succinyl leucylglycylargininamide (EP-LGR-NH₂) (TD 16)



Standard SPPS. Product was obtained as white solid.

δ_{H} /ppm (400 MHz; D₂O): 4.30 (1H, t, $J = 7.02$, NHCHArgCO), 4.24 (1H, dd, $J = 5.04$ and 9.04, NHCHLeuCO), 3.85 (2H, s, NH₂CH₂CO), 3.61 (1H, d, $J = 1.35$, CHOCHCOOH), 3.50 (1H, d, $J = 1.77$, CHOCHCOOH), 3.12 (2H, t, $J = 6.90$, CH₂NHCNHNH₂), 1.86–1.78 (1H, m, CH₂CH(CH₃)₂), 1.71–1.50 (6H, m, 3 × CH₂), 0.85 (3H, d, $J = 5.56$, CH(CH₃)₂), 0.81 (3H, d, $J = 5.52$, CH(CH₃)₂). **R_t:** 9.08.

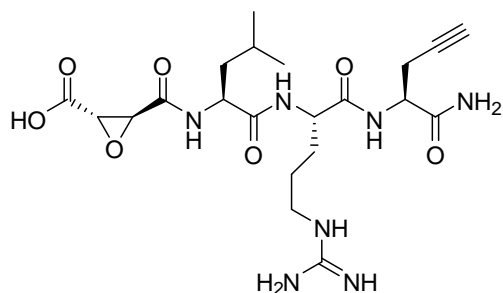
(2S,3S)-ethyl-2,3-epoxysuccinyl leucylglycylargininamide (EtEP-LRG-NH₂)



Standard SPPS. Product was obtained as white solid.

δ_{H} /ppm (400 MHz; D₂O): 4.33–4.29 (1H, m, $J = 7.02$, NHCHArgCO), 4.26–4.22 (1H, m, NHCHLeuCO), 4.20 (2H, q, $J = 7.30$, CH₃CH₂), 3.85 (2H, s, NH₂CH₂CO), 3.70 (1H, s, CHOCHCOOEt), 3.65 (1H, d, $J = 0.89$, CHOCHCOOEt), 3.12 (2H, t, $J = 6.89$, CH₂NHCNHNH₂), 1.84–1.77 (1H, m, CH₂CH(CH₃)₂), 1.71–1.49 (6H, m, CH₂), 1.20 (3H, t, $J = 7.04$, CH₃CH₂), 0.85 (3H, d, $J = 5.52$, CH(CH₃)₂), 0.80 (3H, d, $J = 5.42$, CH(CH₃)₂). **R_t:** 10.74.

(2S,3S)-2,3-Epoxy succinyl leucylarginyl propargylglycinamide (EP-LR-Pra-NH₂) (TD 21)

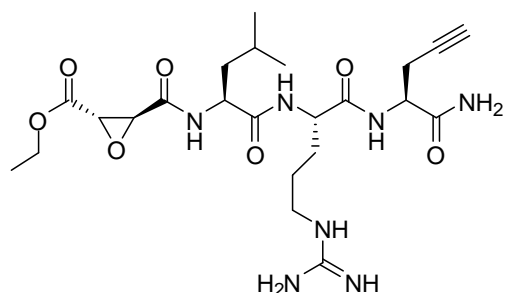


Standard SPPS. Product was obtained as white solid.

δ_{H} /ppm (400 MHz; D₂O): 4.41 (1H, t, $J = 6.53$, NHCHCO), 4.31 (2H, dt, $J = 7.23$, 10.19, NHCHCO), 3.49 (1H, d, $J = 2.00$, CHOCHCOOEt), 3.35 (1H, d, $J = 1.98$, CHOCHCOOEt), 3.13 (2H, t, $J = 6.92$, CH₂NHCNHNH₂), 2.66–2.64 (2H, m, CH₂CCH), 2.37 (1H, t, $J = 2.57$, CCH), 1.83–1.51 (7H, m, CH₂CH(CH₃)₂, CH₂), 0.85 (3H, d, $J = 5.81$, CH(CH₃)₂), 0.81 (3H, d, $J = 5.79$, CH(CH₃)₂). **R_t:** 8.02.

(2S,3S)-Ethyl-2,3-epoxysuccinyl leucylarginyl propargylglycinamide (EtEP-LR-Pra-NH₂)

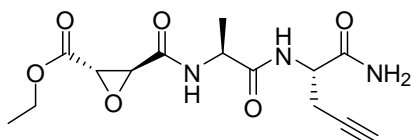
(TD 22)



Standard SPPS. Product was obtained as white solid.

δ_{H} /ppm (400 MHz; D₂O): 4.41 (1H, t, $J = 6.54$, NHCHCO), 4.31 (2H, m, NHCHCO), 4.21 (2H, q, $J = 7.15, 7.18$, CH₂OCO), 3.71 (1H, d, $J = 1.78$, CHOCHCOOEt), 3.65 (1H, d, $J = 1.81$, CHOCHCOOEt), 3.13 (2H, t, $J = 6.91$, CH₂NHCNHNH₂), 2.66–2.64 (2H, m, CH₂CCH), 2.37 (1H, t, $J = 2.55$, CCH), 1.83–1.50 (7H, m, CH₂CH(CH₃)₂, CH₂), 1.21 (3H, t, $J = 7.17$, CH₃CH₂), 0.85 (3H, d, $J = 5.75$, CH(CH₃)₂), 0.81 (3H, d, $J = 5.72$, CH(CH₃)₂). **R_t:** 9.08.

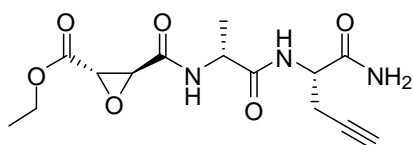
(2S,3S)-Ethyl-2,3-epoxysuccinyl alanyl propargylglycinamide (EtEP-A-Pra-NH₂) (TD 23)



Standard SPPS. Product was obtained as white solid.

δ_{H} /ppm (400 MHz; DMSO): 8.70 (1H, d, $J = 7.71$, NHCHCONH), 8.21 (1H, d, $J = 8.07$, NHCHCONH₂), 7.35 (1H, s, NH₂), 7.21 (1H, s, NH₂), 4.40–4.32 (1H, m, NHCHCO), 4.29 (1H, td, $J = 5.89, 7.97$, NHCHCO), 4.24–4.14 (2H, m, CH₂CH₃), 3.74 (1H, d, $J = 1.68$, CHOCHCOOEt), 3.62 (1H, d, $J = 1.84$, CHOCHCOOEt), 2.85 (1H, t, $J = 2.56$, CCH), 2.60 (1H, dd, $J = 2.66, 5.72$, CH₂CCH), 2.46 (1H, dd, $J = 2.57, 7.88$, CH₂CCH), 1.24 (6H, t, $J = 7.20$, CH₂CH₃ and CHCH₃). **R_t:** 7.68.

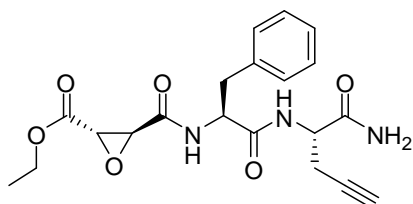
(2S,3S)-Ethyl-2,3-epoxysuccinyl-D-alanyl propargylglycinamide (EtEP-(D)A-Pra-NH₂) (TD 39)



Standard SPPS. Product was obtained as white solid.

δ_{H} /ppm (400 MHz; DMSO): 8.72 (1H, d, $J = 7.13$, NHCHCONH), 8.34 (1H, d, $J = 8.43$, NHCHCONH₂), 7.36 (1H, s, NH₂), 7.26 (1H, s, NH₂), 4.46–4.26 (2H, m, 2 × NHCHCO), 4.26–4.08 (2H, m, OCH₂CH₃), 3.73 (1H, d, $J = 1.82$, CHOCHCOOEt), 3.62 (1H, d, $J = 1.85$, CHOCHCOOEt), 2.84 (1H, t, $J = 2.56$, CCH), 2.64 (1H, dd, $J = 2.62, 5.02$, CH₂CCH), 2.45 (1H, dd, $J = 2.56, 8.81$, CH₂CCH), 1.25 (3H, d, $J = 7.00$, CHCH₃), 1.24 (3H, t, $J = 7.12$, CH₂CH₃). **R_t:** 7.70.

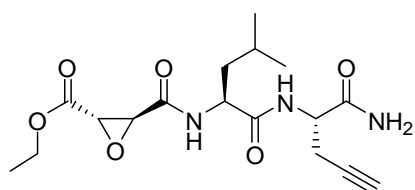
(2S,3S)-Ethyl-2,3-epoxysuccinyl phenylalanyl propargylglycinamide (EtEP-F-Pra-NH₂) (TD 24)



Standard SPPS. Product was obtained as white solid.

δ_{H} /ppm (400 MHz; DMSO): 8.64 (1H, d, $J = 8.47$, NHCHCONH), 8.42 (1H, d, $J = 7.97$, NHCHCONH₂), 7.39 (1H, s, NH₂), 7.28–7.27 (5H, m, Bn), 7.23 (1H, s, NH₂), 4.65 (1H, td, $J = 4.22, 9.78$, NHCHCONH), 4.33 (1H, td, $J = 6.21, 7.70$, NHCHCONH₂), 4.21–4.13 (2H, m, OCH₂CH₃), 3.64 (1H, d, $J = 1.66$, CHOCHCOOEt), 3.40 (1H, d, $J = 1.81$, CHOCHCOOEt), 3.07 (1H, dd, $J = 4.25, 13.67$, CHCH₂Bn), 2.88 (1H, t, $J = 2.55$, CCH), 2.79 (1H, dd, $J = 10.12, 13.68$, CHCH₂Bn), 2.55 (2H, s, CH₂CCH), 1.22 (3H, t, $J = 7.10$, CH₃). **ESI-MS:** $m/z = 402.1679$ ([M + H]⁺). **R_t:** 10.56.

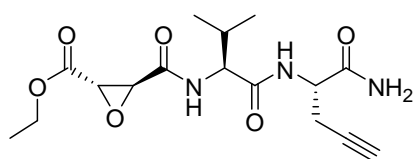
(2S,3S)-Ethyl-2,3-epoxysuccinyl leucyl propargylglycinamide (EtEP-L-Pra-NH₂) (TD 26)



Standard SPPS. Product was obtained as white solid.

δ_{H} /ppm (400 MHz; DMSO): 8.65 (1H, d, $J = 7.90$, NHCHCONH), 8.24 (1H, d, $J = 7.93$, NHCHCONH₂), 7.33 (1H, s, NH₂), 7.17 (1H, s, NH₂), 4.45–4.32 (1H, m, NHCHCO), 4.28 (1H, td, $J = 6.91$, 13.79, NHCHCO), 4.24–4.10 (2H, m, CH₂CH₃), 3.73 (1H, d, $J = 1.59$, CHOCHCOOEt), 3.59 (1H, d, $J = 1.81$, CHOCHCOOEt), 2.83 (1H, t, $J = 2.52$, CCH), 2.59 (1H, dd, $J = 2.68$, 5.77 CH₂CCH), 2.45 (1H, dd, $J = 2.49$, 7.80, CH₂CCH), 1.68–1.54 (1H, m, CH(CH₃)₂), 1.54–1.40 (2H, m, CHCH₂CH), 1.23 (3H, t, $J = 7.09$, CH₂CH₃), 0.87 (6H, dd, $J = 6.42$, 13.73, CH(CH₃)₂). **R_t:** 10.00.

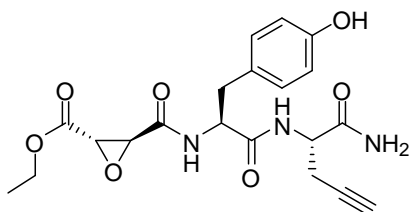
(2S,3S)-Ethyl-2,3-epoxysuccinyl valyl propargylglycinamide (EtEP-V-Pra-NH₂) (TD 27)



Standard SPPS. Product was obtained as white solid.

δ_{H} /ppm (400 MHz; DMSO): 8.58 (1H, d, $J = 8.40$, NHCHCONH), 8.28 (1H, d, $J = 8.00$, NHCHCONH₂), 7.34 (1H, s, NH₂), 7.18 (1H, s, NH₂), 4.36–4.14 (4H, m, 2 × NHCHCO, OCH₂CH₃), 3.88 (1H, d, $J = 1.65$, CHOCHCOOEt), 3.60 (1H, d, $J = 1.74$, CHOCHCOOEt), 2.86 (1H, t, $J = 2.53$, CCH), 2.51–2.44 (2H, s, CH₂CCH), 2.03 (1H, dq, $J = 6.72$, 13.41, CH(CH₃)₂), 1.24 (3H, t, $J = 7.12$, CH₂CH₃), 0.87 (6H, dd, $J = 6.82$, 10.49, CH(CH₃)₂). **R_t:** 9.01.

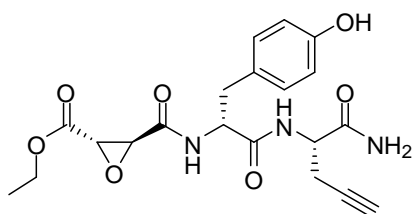
(2S,3S)-Ethyl-2,3-epoxysuccinyl tyrosyl propargylglycinamide (EtEP-Y-Pra-NH₂) (TD 28)



Standard SPPS. Product was obtained as white solid.

δ_{H} /ppm (400 MHz; DMSO): 9.21 (1H, s, OH), 8.56 (1H, d, $J = 8.47$, NHCHCONH), 8.37 (1H, d, $J = 8.02$, NHCHCONH₂), 7.39 (1H, s, NH₂), 7.23 (1H, s, NH₂), 7.05 (2H, d, $J = 8.45$, Bn), 6.65 (2H, d, $J = 8.48$, Bn), 4.56 (1H, td, $J = 4.27$, 9.56, NHCHCONH), 4.32 (1H, td, $J = 6.17$, 7.67, NHCHCONH₂), 4.24–4.10 (2H, m, OCH₂CH₃), 3.65 (1H, d, $J = 1.69$, CHOCHCOOEt), 3.43 (1H, d, $J = 1.64$, CHOCHCOOEt), 2.94 (1H, dd, $J = 4.12$, 13.84, CHCH₂Bn), 2.88 (1H, t, $J = 2.56$, CCH), 2.67 (1H, dd, $J = 9.95$, 13.79, CHCH₂Bn), 2.55 (2H, s, CH₂CCH), 1.23 (3H, t, $J = 7.10$, CH₃). **R_t:** 8.98.

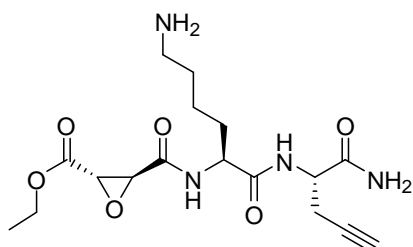
(2S,3S)-Ethyl-2,3-epoxysuccinyl -D-tyrosyl propargylglycinamide (EtEP-(D)Y-Pra-NH₂) (TD 42)



Standard SPPS. Product was obtained as white solid.

δ_{H} /ppm (400 MHz; DMSO): 9.21 (1H, s, OH), 8.63 (1H, d, $J = 8.14$, NHCHCONH), 8.46 (1H, d, $J = 8.33$, NHCHCONH₂), 7.41 (1H, s, NH₂), 7.26 (1H, s, NH₂), 7.04 (2H, d, $J = 8.44$, Bn), 6.65 (2H, d, $J = 8.45$, Bn), 4.54 (1H, td, $J = 4.69$, 9.40, NHCHCONH), 4.37 (1H, td, $J = 5.54$, 8.22, NHCHCONH₂), 4.18 (2H, q, $J = 7.12$, OCH₂CH₃), 3.67 (1H, d, $J = 1.65$, CHOCHCOOEt), 3.47 (1H, d, $J = 1.69$, CHOCHCOOEt), 2.94 (1H, dd, $J = 4.55$, 13.86, CHCH₂Bn), 2.85 (1H, t, $J = 2.56$, CCH), 2.69 (1H, dd, $J = 10.06$, 11.79, CHCH₂Bn), 2.40 (2H, ddd, $J = 2.48$, 8.18, 16.79, CH₂CCH), 1.23 (3H, t, $J = 7.10$, CH₃). **R_t:** 10.55.

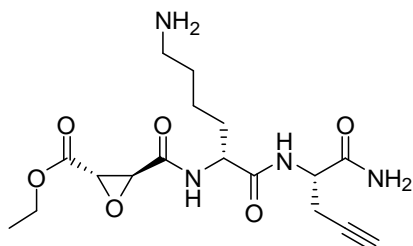
(2S,3S)-Ethyl-2,3-epoxysuccinyl lysyl propargylglycinamide (EtEP-K-Pra-NH₂) (TD 31)



Standard SPPS. Product was obtained as white solid.

δ_{H} /ppm (400 MHz; D_2O): 4.38 (1H, t, $J = 6.67$, NHCHCONH_2), 4.29 (1H, dd, $J = 6.02$, 8.22, NHCHCO), 4.17 (2H, q, $J = 7.15$, CH_2CH_3), 3.69 (1H, d, $J = 1.60$, CHOCHCOOEt), 3.65 (1H, d, $J = 1.58$, CHOCHCOOEt), 2.88 (2H, t, $J = 7.47$, CH_2NH_2), 2.67–2.57 (2H, m, CH_2CCH), 2.34 (1H, m, CCH), 1.84–1.64 (2H, m, CH_2CHCO), 1.65–1.52 (2H, m, $\text{CH}_2\text{CH}_2\text{NH}_2$), 1.41–1.27, (2H, m, $\text{CH}_2\text{CH}_2\text{CH}_2\text{NH}_2$), 1.18 (3H, t, $J = 7.16$, CH_3). R_{t} : 6.48.

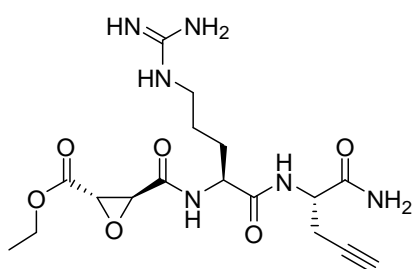
(2S,3S)-Ethyl-2,3-epoxysuccinyl -D-lysyl propargylglycinamide (EtEP-(D)K-Pra-NH₂) (TD 40)



Standard SPPS. Product was obtained as white solid.

δ_{H} /ppm (400 MHz; D_2O): 4.42 (1H, dd, $J = 5.32$, 7.83, NHCHCONH_2), 4.25 (1H, dd, $J = 6.72$, 7.79, NHCHCO), 4.22–4.13 (2H, m, CH_2CH_3), 3.70 (2H, s, CHOCHCOOEt), 2.87 (2H, t, $J = 7.66$, CH_2NH_2), 2.69–2.57 (2H, m, CH_2CCH), 2.34 (1H, t, $J = 2.59$, CCH), 1.82–1.64 (2H, m, CH_2CHCO), 1.63–1.51 (2H, m, $\text{CH}_2\text{CH}_2\text{NH}_2$), 1.44–1.26, (2H, m, $\text{CH}_2\text{CH}_2\text{CH}_2\text{NH}_2$), 1.18 (3H, t, $J = 7.20$, CH_3). **ESI-MS:** $m/z = 383.1915$ ($[\text{M} + \text{H}]^+$). R_{t} : 6.18.

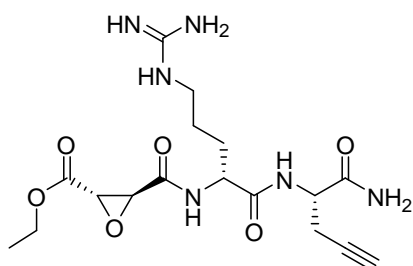
(2S,3S)-Ethyl-2,3-epoxysuccinyl arginyl propargylglycinamide (EtEP-R-Pra-NH₂) (TD 32)



Standard SPPS. Product was obtained as white solid.

δ_{H} /ppm (400 MHz; D_2O): 4.39 (1H, dd, $J = 6.33$, 7.22, NHCHCONH_2), 4.30 (1H, dd, $J = 5.96$, 8.23, NHCHCO), 4.17 (2H, q, $J = 7.15$, CH_2CH_3), 3.69 (1H, d, $J = 1.89$, CHOCHCOOEt), 3.64 (1H, d, $J = 1.89$, CHOCHCOOEt), 3.10 (2H, t, $J = 6.81$, CH_2NHCNH), 2.65–2.56 (2H, m, CH_2CCH), 2.33 (1H, t, $J = 2.63$, CCH), 1.86 – 1.64 (2H, m, CH_2CHCO), 1.57–1.50 (2H, m, $\text{CH}_2\text{CH}_2\text{CH}_2$), 1.17 (3H, t, $J = 7.16$, CH_3). R_{t} : 6.92.

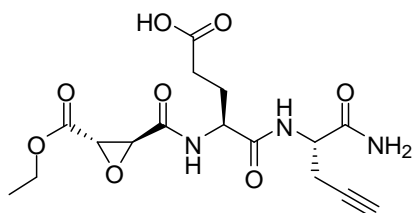
(2S,3S)-Ethyl-2,3-epoxysuccinyl -D-arginyl propargylglycinamide (EtEP-(D)R-Pra-NH₂) (TD 41)



Standard SPPS. Product was obtained as white solid.

δ_{H} /ppm (400 MHz; D₂O): 4.42 (1H, dd, $J = 5.20, 7.97$, NHCHCONH₂), 4.26 (1H, t, $J = 7.23$, NHCHCO), 4.21–4.13 (2H, m, CH₂CH₃), 3.69 (2H, dd, $J = 1.89, 4.80$, CHOCHCOOEt), 3.10 (2H, t, $J = 6.88$, CH₂NHCNH), 2.69 (1H, dd, $J = 2.66, 5.20$, CH₂CCH), 2.61 (1H, dd, $J = 2.69, 7.68$, CH₂CCH), 2.32 (1H, t, $J = 2.59$, CCH), 1.81–1.66 (2H, m, CH₂CHCO), 1.61–1.51 (2H, m, CH₂CH₂CH₂), 1.18 (3H, t, $J = 7.18$, CH₃). **R_t:** 7.08.

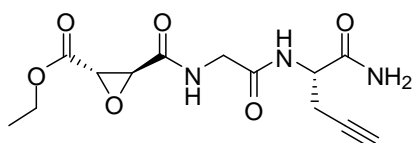
(2S,3S)-Ethyl-2,3-epoxysuccinyl glutamyl propargylglycinamide (EtEP-E-Pra-NH₂) (TD33)



Standard SPPS. Product was obtained as white solid.

δ_{H} /ppm (400 MHz; D₂O): 4.45–4.30 (2H, m, 2xNHCHCO), 4.20–4.15 (2H, m, CH₂CH₃), 3.67 (2H, d, $J = 10.15$, CHOCHCOOEt), 2.72–2.51 (2H, m, CH₂CCH), 2.39 (2H, t, $J = 6.89$, CH₂COOH), 2.33 (1H, s, CCH), 2.06 (1H, dt, $J = 6.46, 20.18$, CH₂CH₂COOH), 1.93 (1H, dt, $J = 7.07, 20.85$, CH₂CH₂COOH), 1.17 (3H, t, $J = 7.00$, CH₃). **ESI-MS:** $m/z = 384.1394$ ([M + H]⁺). **R_t:** 5.44.

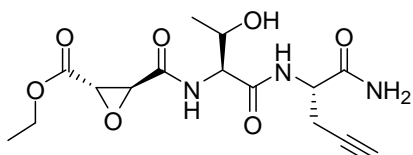
(2S,3S)-Ethyl-2,3-epoxysuccinyl glycyl propargylglycinamide (EtEP-G-Pra-NH₂) (TD 34)



Standard SPPS. Product was obtained as white solid.

δ_{H} /ppm (400 MHz; DMSO): 8.64 (1H, t, $J = 5.71$, NHCH₂CO), 8.25 (1H, d, $J = 8.16$, NHCHCO), 7.44, 7.24 (2H, s, NH₂), 4.34 (1H, td, $J = 5.59$, 7.96, NHCHCH₂), 4.25–4.12 (2H, m, OCH₂CH₃), 3.83 (2H, qd, $J = 5.77$, 16.66, NHCH₂CO), 3.72 (1H, d, $J = 1.83$, CHOCHCOOEt), 3.64 (1H, d, $J = 1.80$, CHOCHCOOEt), 2.86 (1H, t, $J = 2.56$, CCH), 2.61–2.40 (2H, m, CH₂CCH), 1.24 (3H, t, $J = 7.10$, CH₃). ESI-MS: $m/z = 312.12.1206$ ($[M + H]^+$). R_t : 7.09.

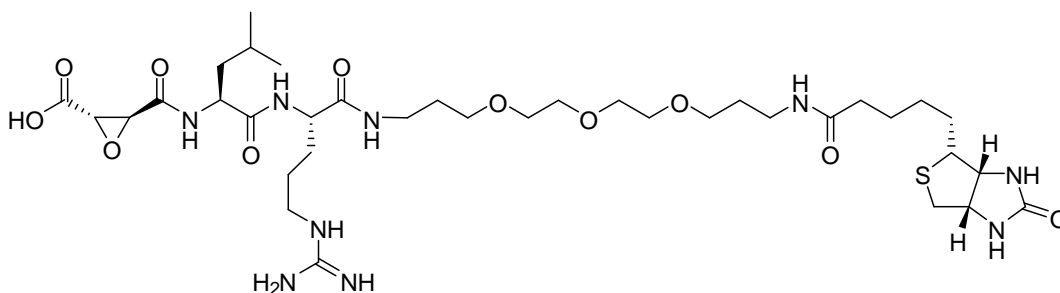
(2S,3S)-Ethyl-2,3-epoxysuccinyl threonyl propargylglycinamide (EtEP-T-Pra-NH₂) (TD 36)



Standard SPPS. Product was obtained as white solid.

δ_{H} /ppm (400 MHz; D₂O): 4.44 (1H, td, $J = 2.77$, 6.61, NHCHCONH₂), 4.35 (1H, dd, $J = 1.46$, 4.82, NHCHCO), 4.26–4.12 (3H, m, CH₂CH₃ and CHCHOH), 3.80 (1H, d, $J = 1.88$, CHOCHCOOEt), 3.72 (1H, d, $J = 1.90$, CHOCHCOOEt), 2.67–2.64 (2H, m, CH₂CCH), 2.38–2.35 (1H, m, CCH), 1.20 (3H, t, $J = 7.14$, CH₂CH₃), 1.14 (3H, dd, $J = 4.05$, 6.37, CHCH₃). R_t : 7.82.

(2S,3S)-2,3-Epoxysuccinyl leucylarginyl triethylene glycol biotin (EP-LR-PEG₃-Biotin) (TD 43)



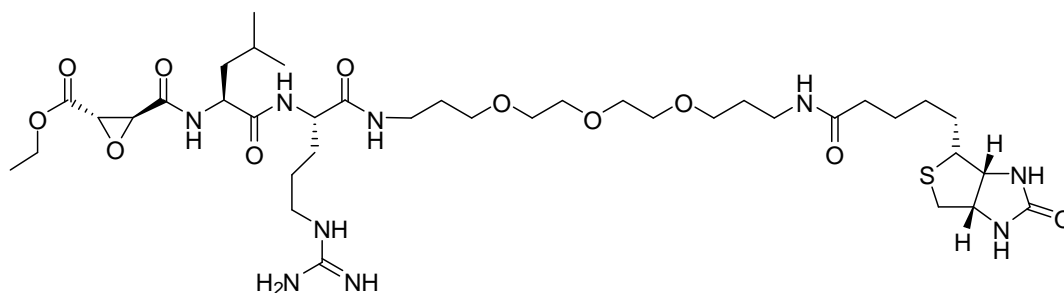
Standard SPPS. Product was obtained as white solid.

δ_{H} /ppm (400 MHz; D₂O): 4.51 (1H, dd, $J = 4.89$, 7.88, SCH₂CH), 4.32 (1H, dd, $J = 4.49$, 7.92, SCHCH), 4.28 (1H, t, $J = 7.17$, NHCHArgCO), 4.20–4.16 (1H, m, NHCHLeuCO), 3.59–3.55 (9H, m, CHOCHCOOH, O(CH₂)₂O(CH₂)₂O), 3.49–3.43 (5H, m, CHOCHCOOH, NH(CH₂)₂CH₂O), 3.26–3.09 (7H, m, 3 × CH₂NH, SCHCH), 2.89 (1H, dd, $J = 4.96$, 13.07, SCH₂CH equatorial, same site with CHCH), 2.68 (1H, d, $J = 13.04$, SCH₂CH axial), 2.15 (2H, t, $J = 7.21$ NHCOCH₂), 1.76–1.43 (15H, m, (CH₂)₃CHS, CH₂CH(CH₃)₂, PEG-CH₂CH₂NHCO, CH(CH₃)₂, (CH₂)₂CH₂NHCHNH₂), 1.34–

1.26 (2H, m, CONHCH₂CH₂CH₂O), 0.84 (3H, d, *J* = 5.58, CH(CH₃)₂), 0.80 (3H, d, *J* = 5.57, CH(CH₃)₂). *R*_t: 10.68.

(2S,3S)-Ethyl-2,3-epoxysuccinyl leucylarginyl triethylene glycol biotin (EtEP-LR-PEG₃-Biotin)

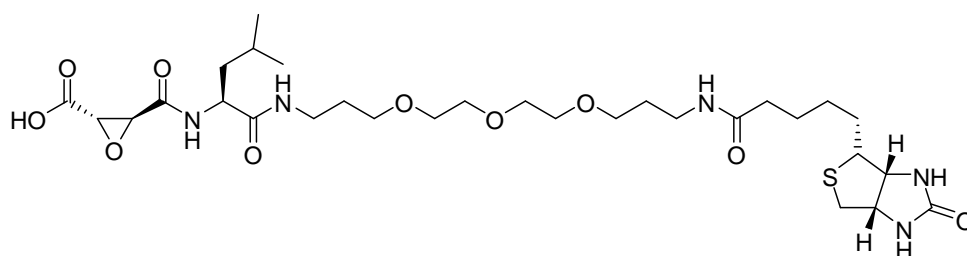
(TD 44)



Standard SPPS. Product was obtained as white solid.

δ_{H} /ppm (400 MHz; D₂O): 4.51 (1H, dd, *J* = 5.00, 7.88, SCH₂CH), 4.32 (1H, dd, *J* = 4.35, 7.74, SCHCH), 4.29 (1H, t, *J* = 8.76, NHCHArgCO), 4.22–4.15 (3H, m, NHCHLeuCO, OCH₂CH₃), 3.69 (1H, s, CHOCHCOOEt), 3.63 (1H, s, CHOCHCOOEt), 3.59–3.55 (8H, m, O(CH₂)₂O(CH₂)₂O), 3.48 (2H, d, *J* = 6.36, NH(CH₂)₂CH₂O), 3.45 (2H, d, *J* = 6.36, NH(CH₂)₂CH₂O), 3.26–3.09 (7H, m, 3 × CH₂NH, SCHCH), 2.89 (1H, dd, *J* = 4.88, 13.13, SCH₂CH equatorial, same site as CHCH), 2.68 (1H, d, *J* = 13.04, SCH₂CH axial), 2.15 (2H, t, *J* = 7.19, NHCOCH₂), 1.76–1.43 (15H, m, (CH₂)₃CHS, CH₂CH(CH₃)₂, PEG-CH₂CH₂NHCO, CH(CH₃)₂, (CH₂)₂CH₂NHCHNH₂), 1.35–1.26 (2H, m, CONHCH₂CH₂CH₂O), 1.20 (3H, t, *J* = 7.08, OCH₂CH₃), 0.84 (3H, d, *J* = 5.38, CH(CH₃)₂), 0.80 (3H, d, *J* = 5.34, CH(CH₃)₂). *R*_t: 11.09.

(2S,3S)-2,3-Epoxysuccinyl leucyl triethylene glycol biotin (EP-L-PEG₃-Biotin) (TD 49)

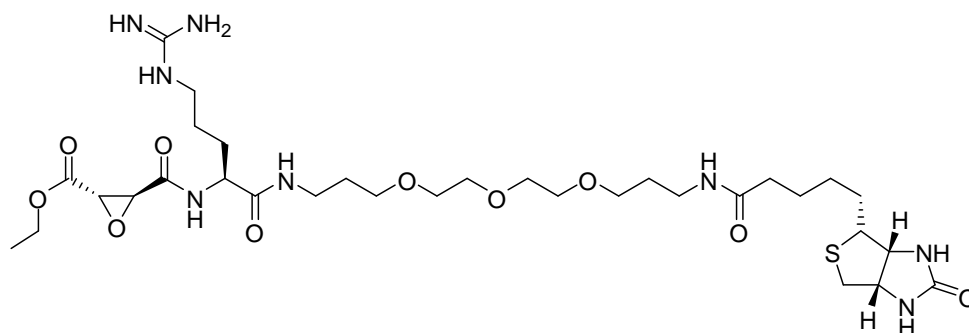


Standard SPPS. Product was obtained as white solid.

δ_{H} /ppm (400 MHz; D₂O): 4.50 (1H, dd, *J* = 4.94, 7.88, SCH₂CH), 4.32 (1H, dd, *J* = 4.50, 7.92, SCHCH), 4.22 (1H, t, *J* = 7.05, NHCHCO), 3.59–3.54 (9H, m, CHOCHCOOH, O(CH₂)₂O(CH₂)₂O), 3.49–3.43 (5H, m, CHOCHCOOH, NH(CH₂)₂CH₂O), 3.25–3.10 (5H, m, 2 × CH₂NH, SCHCH), 2.89

(1H, dd, $J = 4.96, 13.06$, SCH₂CH axial ,same site with CHCH), 2.67 (1H, d, $J = 13.03$, SCH₂CH equatorial), 2.15 (2H, t, $J = 7.21$ NHCOCH₂), 1.70–1.43 (11H, m, (CH₂)₃CHS, CH₂CH(CH₃)₂, PEG-CH₂CH₂NHCO, CH(CH₃)₂), 1.35–1.25 (2H, m, CONHCH₂CH₂CH₂), 0.83 (3H, d, $J = 5.63$, CH(CH₃)₂), 0.78 (3H, d, $J = 5.61$, CH(CH₃)₂). R_t : 11.60.

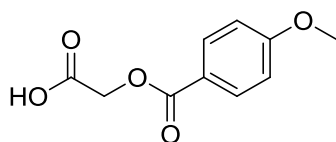
(2S,3S)-Ethyl-2,3-epoxysuccinyl arginyl triethylene glycol biotin (EtEP-R-PEG₃-Biotin) (TD 55)



Standard SPPS. Product was obtained as white solid.

δ_H /ppm (400 MHz; D₂O): 4.51 (1H, dd, $J = 4.84, 7.91$, SCH₂CH), 4.32 (1H, dd, $J = 4.49, 7.93$, SCHCH), 4.20 (1H, q, $J = 7.10$, OCH₂CH₃), 3.71 (1H, d, $J = 1.80$, CHOCHCOOEt), 3.66 (1H, d, $J = 1.86$, CHOCHCOOEt), 3.62–3.53 (8H, m, 2 × O(CH₂)₂O), 3.47 (4H, dt, $J = 6.37, 10.46$ CH₂O(CH₂)₂OCH₂), 3.26–3.10 (7H, m, 3 × CH₂NH, SCHCH), 2.90 (1H, dd, $J = 4.95, 13.09$, SCH₂CH axial ,same site with CHCH), 2.68 (1H, d, $J = 13.03$, SCH₂CH equatorial), 2.16 (2H, t, $J = 7.21$ NHCOCH₂), 1.82–1.27 (14H, m, (CH₂)₃CHS, CH₂CH₂-PEG-CH₂CH₂, CONHCH₂CH₂CH₂), 1.20 (3H, t, $J = 7.16$). R_t : 9.78.

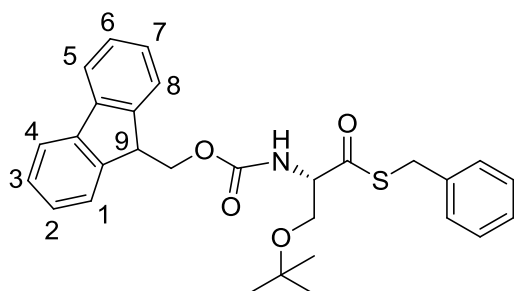
2-(4-methoxybenzoyloxy)acetic acid



Tert-butyl bromoacetate (0.886 mL, 6 mmol), methoxybenzoic acid (1.825 g, 12 mmol) and potassium fluoride (1.394 g, 24 mmol) were dissolved in DMF. After overnight stirring, Et₂O was added after which the solution was extracted with NaHCO₃ (sat. aq.).The organic layer was dried, concentrated *in vacuo* and subsequently dissolved in TFA/H₂O (95/5). The solution was concentrated after 1h, coevaporated with toluene to remove residual traces of TFA and applied to silica gel chromatography (Hexane–EtOAc) giving the white solid product

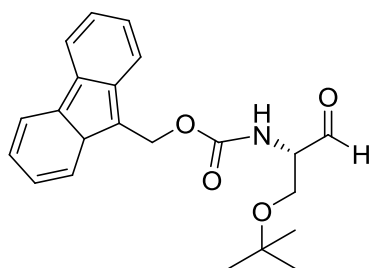
(53%, 0.664 g). $R_f = 0.59$ (Hexane–EtOAc, 2:1 + 0.1%FA). δ_H /ppm (400 MHz; $CDCl_3$): 8.07 (2H, d, $J = 5.7$, Bn), 6.97 (2H, dd, $J = 6.6, 4.9$, Bn), 4.91 (2H, s, CH_2), 3.90 (3H, s, OCH_3). δ_C /ppm (100 MHz; $CDCl_3$): 132.10 (Bn), 121.24 (Bn), 113.80 (Bn), 60.34 (CH_2), 55.50 (OCH_3).

Fmoc-L-(OtBu)Ser-SBn (15)



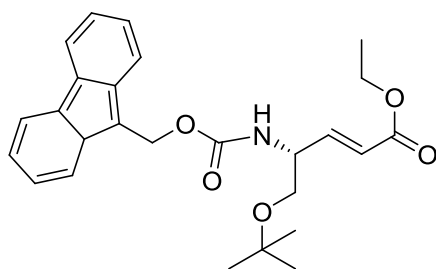
DMAP (0.159 g, 1.30 mmol, 0.1 eq.) and DCC (2.820 g, 13.69 mmol, 1.05 eq.) were added sequentially to a solution of Fmoc-L-Ser(tBu)-OH (5.000 g, 13.04 mmol, 1eq.) and benzyl mercaptan (3.06 mL, 26.08 mmol, 2 eq.) in THF (65 mL) at RT. The cloudy reaction mixture was stirred at RT for 18 h and then filtered. After concentrating the filtrate it was partitioned between EtOAc (70 mL) and 1.0 M HCl (40 mL). The organic layer was dried over Na_2SO_4 and concentrated. Subsequent addition of Et_2O (15 mL) resulted in a white precipitate that was collected by vacuum filtration and washed with petroleum ether. The product was obtained as a yellow solid and used without further purification (5.826 g, 8.18 mmol, 91% yield). $R_f = 0.66$ (Hexane–EtOAc, 2:1); mp 163–165 °C; δ_H /ppm (400 MHz; $CDCl_3$): 7.79 (2H, d, $J = 7.50$, Fmoc-4 and 5), 7.66–7.63 (2H, m, Fmoc-1 and 8), 7.41 (2H, app td, $J = 2.72, 7.38, 7.35$, Fmoc-3 and 6), 7.36–7.30 (7H, m, Bn, Fmoc-2 and 7), 5.78 (1H, d, $J = 8.94$, NH), 4.55 (2H, dd, $J = 6.94, 10.40$, Fmoc- CH_2), 4.42–4.35 (1H, m, $S\alpha$), 4.29 (1H, t, $J = 6.99$, Fmoc-9), 4.16 (2H, dd, $J = 13.76, 35.67$, $S\beta$), 3.96 (1H, dd, $J = 2.50, 9.02$, Bn- CH_2), 3.57 (1H, dd, $J = 3.44, 9.03$, Bn- CH_2), 1.15 (9H, s, tBu). ν_{max} (neat)/ cm^{-1} : 3322, 2954, 2936, 2908, 2853, 1724 (CO), 1684, 1625, 1554, 1509, 1392, 1331, 1234, 1174, 1071, 965, 698. HRMS (ESI): $m/z =$ found 490.2039 ($[M + H]^+$), 512.1859 ($[M + Na]^+$), (required 490.2052; $C_{29}H_{31}NO_4S [M + H]^+$).

Fmoc-L-(OtBu)Ser-H (16)



Triethylsilane (6.50 mL, 40.90 mmol, 5eq.) was slowly added to a suspension of **Fmoc-L-(OtBu)Ser-SBn** (4 g, 8.18 mmol, 1 eq.) and Pd/C (10%, 2.20 g) in acetone (200 mL) at RT previously degassed with N₂. The reaction mixture was stirred for 15 min and then filtered through Celite. The filtrate was concentrated to afford 6.21 g of brown oily crude, which was used without further purification. $R_f = 0.57$ (Hexane–EtOAc, 3:2); δ_H /ppm (400 MHz; CDCl₃): 9.66 (1H, s, CHO), 7.81 (2H, d, $J = 7.34$, Fmoc-4 and 5), 7.65 (2H, dt, $J = 5.74, 11.49$, Fmoc-1 and 8), 7.44 (2H, t, $J = 7.29$, Fmoc-3 and 6), 7.36 (7H, dd, $J = 4.98, 9.52$, Fmoc-2 and 7), 5.70 (1H, d, $J = 7.30$, NH), 4.46 (2H, dd, $J = 1.41, 6.38$, Fmoc-CH₂), 4.41 (2H, dd, $J = 5.84, 10.17$, S α), 4.29 (1H, t, $J = 6.86, 13.98$, Fmoc-9), 4.00 (1H, dd, $J = 3.10, 9.30$, S β), 3.67 (1H, dd, $J = 4.20, 9.38$, S β), 1.20 (9H, s, tBu). **ESI-MS:** $m/z =$ found 390.1681 [M + Na]⁺, (required 390.1681; C₂₂H₂₅NO₄Na).

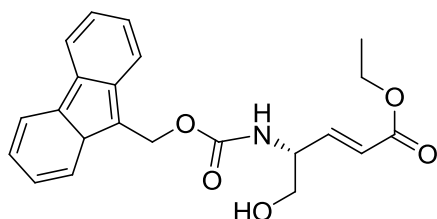
Ethyl-3-[Fmoc-L-Ser(tBu)]-(E)-propenoate (17)



(Carbethoxymethylene)-triphenyl phosphane (2.137 g, 6.135 mmol, 1.5 eq.) was added to a solution of crude **Fmoc-L-(OtBu)Ser-H** (6.210 g) in THF (60 mL) at RT. The resulting mixture was stirred for 24 h and then concentrated under reduced pressure. The residue was purified by silica gel flash column chromatography (Hexane–EtOAc, 9:1) to afford a white solid (0.662 g, 37% yield from **Fmoc-L-(OtBu)Ser-SBn**). $R_f = 0.51$ (Hexane–EtOAc, 2:1); mp 50–53 °C; δ_H /ppm (400 MHz; CDCl₃): 7.80 (2H, d, $J = 7.49$, Fmoc-4 and 5), 7.63 (2H, d, $J = 7.12$, Fmoc-1 and 8), 7.43 (2H, t, $J = 7.40$, Fmoc-3 and 6), 7.35 (2H, t, $J = 7.11$, Fmoc-2 and

7), 6.96 (1H, dd, $J = 4.64, 15.63$, $S\alpha$ -CH), 6.00 (1H, d, $J = 15.66$, CHCOOEt), 5.31 (1H, d, $J = 7.85$, NH), 4.49 (1H, s, $S\alpha$), 4.45 (2H, d, $J = 6.88$, Fmoc- CH_2), 4.28 (1H, t, $J = 6.92$, Fmoc-9), 4.23 (2H, q, $J = 7.08$, COOCH₂CH₃), 3.54–3.47 (2H, m, $S\beta$), 1.33 (3H, t, $J = 7.15$, COOCH₂CH₃), 1.21 (9H, s, tBu). ν_{\max} (neat)/cm⁻¹: 3342, 1725 (CO), 1697, 1541, 1182, 1032, 881, 738. **ESI-MS**: m/z = found 438.2276 ([M + H]⁺), 460.2100 ([M + Na]⁺), (required 460.2100, C₂₆H₃₁NO₅Na).

Ethyl-3-Fmoc-L-Ser-(E)-propenoate (18)



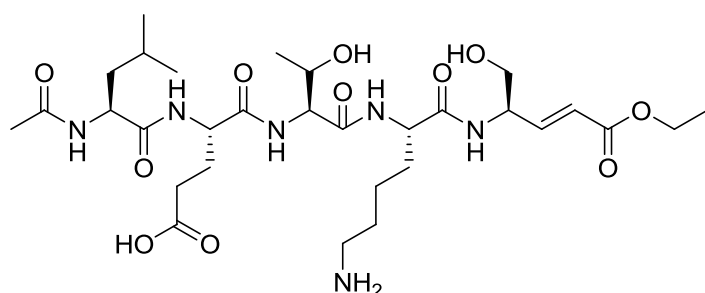
TFA (2 mL) was added to a solution of **ethyl-3-[Fmoc-L-Ser(tBu)]-(E)-propenoate** (0.318 g, 0.833 mmol) in DCM (8 mL) at RT. The reaction mixture was stirred for 2 h, and then the volatiles were removed under reduced pressure. The residue was triturated with a 1:1 mixture of Et₂O and hexanes (3 mL), and the resulting white solid was collected by filtration and washed thoroughly with cold Et₂O–Hexane (1:1) to afford the product (0.197 g, 0.518 mmol, 62% yield). $R_f = 0.51$ (Hexane–EtOAc, 2:1); mp 110–113 °C; δ_H /ppm (400 MHz; CDCl₃): 7.78 (2H, d, $J = 7.50$, Fmoc-4 and 5), 7.60 (2H, d, $J = 7.05$, Fmoc-1 and 8), 7.41 (2H, t, $J = 7.40$, Fmoc-3 and 6), 7.32 (2H, t, $J = 7.26$, Fmoc-2 and 7), 6.92 (1H, dd, $J = 4.19, 15.67$, $S\alpha$ -CH), 6.00 (1H, d, $J = 15.70$, CHCOOEt), 5.50 (1H, d, $J = 7.80$, NH), 4.50–4.36 (3H, m, Fmoc- CH_2 , $S\alpha$), 4.29–4.09 (3H, m, CH₂CH₃, Fmoc-9), 3.75 (2H, s, $S\beta$), 1.31 (3H, t, $J = 7.14$, COOCH₂CH₃). δ_C /ppm (100 MHz; CDCl₃): 166.26, 156.23 (CO), 145.10 (CH-Fmoc), 143.72, 141.33 (C-Fmoc), 127.78, 127.12, 125.00, 122.67, 120.03 (CH-Fmoc, CHCH), 66.91, 64.03, 60.76 (CH₂), 53.80 (CH-Fmoc), 47.19 (C α), 14.21 (CH₃). ν_{\max} (neat)/cm⁻¹: 3320, 2974, 2878, 2839, 1716 (CO), 1692, 1535, 1461, 1377, 1287, 1212, 1016, 736. **ESI-MS**: m/z = 382.1650 ([M + H]⁺), 405.1526 ([M + Na]⁺), (required 382.1576, C₂₂H₂₄NO₅).

Ethyl-3-[Fmoc-L-Ser(ClTrt resin)]-(E)-propenoate

2-Chlorotrityl chloride resin (0.199 g, 1 eq.) was added to a solution of **ethyl-3-Fmoc-L-Ser-(E)-propenoate** (0.197 g, 0.518 mmol, 2 eq.) and pyridine (83.73 μ L, 1.035 mmol, 4 eq.) in

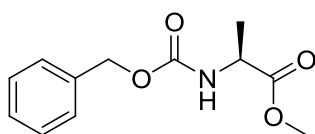
dry THF (2 mL). The reaction mixture was heated at 60 °C with stirring for 6 h. The resin was filtered and washed with DCM–MeOH–DiPEA (17:2:1) and 3 × DCM, 2 × DMF, 2 × DCM. Finally, the resin was dried *in vacuo* over KOH. The yield of the loading was 72% as determined by the Fmoc-test.

Ethyl-3-[-L-Ser-Lys-Thr-Glu-Leu-Ac]-(*E*)-propenoate (Ac-LETKS-MA) (TD 90)



δ_{H} /ppm (400 MHz; D_2O): 6.84 (1H, ddd, $J = 3.41, 4.81, 15.79$, CHCHCOOEt), 5.90 (1H, ddd, $J = 1.72, 11.57, 15.86$, CHCHCOOEt), 4.88 (2H, d, $J = 3.76$, OH), 4.55 (1H, dt, $J = 3.12, 6.34$, NHCHCO), 4.49 (1H, dd, $J = 5.03, 9.04$, NHCHCO), 4.36 (2H, dd, $J = 5.39, 8.99$, NHCHCO), 4.30 (1H, dd, $J = 5.55, 8.79$, NHCHCO), 4.25–3.98 (6H, m, $\text{CH}_2(\text{OEt})$, 4 × NHCHCO), 3.70–3.56 (2H, m, CH(Thr), NHCHCH), 2.91 (2H, t, $J = 7.33$, CH_2), 2.78 (1H, t, $J = 6.26$, $\text{CH}_2(\text{Ser})$), 2.60–2.55 (1H, m, $\text{CH}_2(\text{Ser})$), 2.44–2.25 (2H, m, CH_2), 2.13–2.00 (1H, m, CH(Leu)), 1.94 (3H, s, $\text{CH}_3(\text{Ac})$), 1.81–1.29 (10H, m, 5 × CH_2), 1.19 (3H, t, $J = 7.18$, $\text{CH}_3(\text{Thr})$), 1.16–1.09 (3H, m, $\text{CH}_3(\text{OEt})$), 0.84 (3H, d, $J = 6.24$, $\text{CH}_3(\text{Leu})$), 0.80 (3H, d, $J = 6.24$, $\text{CH}_3(\text{Leu})$). R_{f} : 5.26.

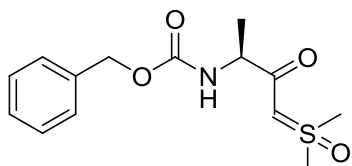
Z-L-Ala-OMe (20)



Z-L-Ala-OH (2.000 g, 8.961 mmol, 1.0 eq.) was dissolved in DMF (10 mL) and cooled to 0 °C. K_2CO_3 (1.240 g, 8.961 mmol, 1.0 eq.) was added and the resulting suspension was stirred for 10 min at 0 °C before methyl iodide (1.120 mL, 17.921 mmol, 2.0 eq.) was added. The reaction was monitored by TLC (Et_2O –hexane, 4:1), and no starting material was detected after 30 min at 0 °C and a further 1 h at room temperature with continued stirring. The reaction mixture was filtered and the filtrate diluted with EtOAc (40 mL), washed with H_2O (15 mL) and brine (3 × 10 mL), and finally dried over Na_2SO_4 , filtered and concentrated *in*

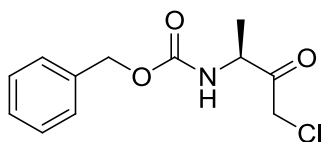
vacuo to afford a yellow oily crude product (1.92 g, 90.4%); $R_f = 0.35$ (Hexane–EtOAc, 4:1); δ_H/ppm (400 MHz; CDCl_3): 7.39–7.30 (5H, m, Bn), 5.42 (1H, s, NH), 5.15–5.09 (2H, m, CH_2), 4.43–4.36 (1H, m, $\text{A}\alpha$), 3.75 (3H, s, OMe), 1.42 (3H, d, $J = 7.20$, $\text{A}\beta$). **ESI-MS:** $m/z =$ found 260.0890 ($[\text{M} + \text{Na}]^+$), (required 260.0899, $\text{C}_{12}\text{H}_{15}\text{NO}_4\text{Na}$).

Z-L-Ala-dimethylsulfoxonium (21)



A solution of 1 M potassium *tert*-butoxide in THF (12.650 mL, 12.653 mmol, 3.0 eq.) was added at RT to a suspension of trimethylsulfoxonium iodide (2.780 g, 12.653 mmol, 3.0 eq.) in THF (20 mL). The mixture was heated at reflux for 2h and then was cooled to 0 °C. A solution of **Z-L-Ala-OMe** (1.000 g, 4.218 mmol, 1.0 eq.) in THF (10 mL) was added dropwise at 0 °C and the resultant solution was stirred for 4 h at 0 °C. The reaction was quenched with water (20 mL). The organic layer was separated and washed with brine (2 × 20 mL). The solvent was removed at reduced pressure to afford the crude sulphur ylide (0.447 g) as a white off solid. $R_f = 0.24$ (5% MeOH in DCM); mp 129–132 °C; δ_H/ppm (400 MHz; CDCl_3): 7.37–7.31 (5H, m, Bn), 5.88 (1H, d, $J = 7.60$, NH), 5.12–5.03 (2H, m, CH_2), 4.58 (1H, s, CHS), 4.23–4.11 (1H, m, $\text{A}\alpha$), 3.35 (3H, s, SMe), 3.33 (3H, s, SMe), 1.32 (3H, d, $J = 6.90$, $\text{A}\beta$). ν_{max} (neat)/ cm^{-1} : 3314, 2930, 2906, 2853, 1687 (CO), 1552, 1391, 1320, 1257, 1174, 1002, 1037, 855, 744. **ESI-MS:** $m/z =$ found 298.1122 ($[\text{M} + \text{H}]^+$), (required 298.1113, $\text{C}_{14}\text{H}_{20}\text{NO}_4\text{S}$).

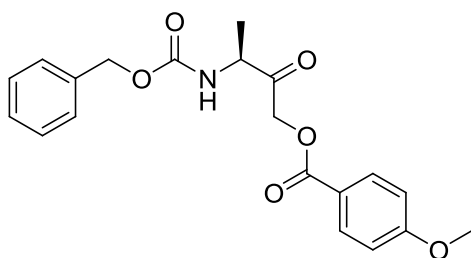
Z-L-Ala-CMK (22)



The crude **Z-L-Ala-dimethylsulfoxonium** (0.410 g, 1.381 mmol, 1.0 eq.) was dissolved in THF (5 mL) and the solution was cooled to 0 °C. Lithium chloride (0.066 g, 1.574 mmol, 1.1 eq.) and methanesulfonic acid (0.099 mL, 1.519 mmol, 1.1 eq.) were added. The temperature was slowly raised to 70 °C and stirring was continued at that temperature for 2 h. After cooling, the reaction was quenched by addition of water (4 mL). The reaction mixture was

extracted using 2:1 hexane–EtOAc (20 mL). To remove DMSO, the organic layer was washed with saturated NaHCO₃ (20 mL), water (2 × 20 mL), and brine (20 mL) and dried over Na₂SO₄. Removal of solvent afforded the crude chloromethyl ketone. The crude product was purified by silica gel flash column chromatography (Hexane–EtOAc, 9:1) to afford white solid Z-Ala-methyl chloromethyl ketone (0.221 g, 62.7%). R_f = 0.46 (Hexane–EtOAc, 2:1); m.p 50–52 °C; **δ_H/ppm (400 MHz; CDCl₃):** 7.38–7.33 (5H, m, Bn), 5.71 (1H, d, *J* = 6.40, NH), 5.17–5.04 (2H, m, BnCH₂), 4.62 (1H, p, *J* = 7.20, Aα), 4.33–4.21 (2H, m, CH₂Cl), 1.38 (3H, d, *J* = 7.20, Aβ). **δ_C/ppm (100 MHz; CDCl₃):** 201.71 (OCO), 155.87 (OCCH₂), 136.28 (C-Bn), 128.61, 128.33, 128.14 (CH-Bn), 67.16 (CH₂), 53.62 (CH), 49.61 (CH₂), 17.33 (CH₃). **ν_{max} (neat)/cm⁻¹:** 3324, 1741 (CO), 1680, 1531, 1396, 1273, 1048, 1007, 739. **C.I.-MS:** *m/z* = found 273.1008 ([M + NH₄⁺]), (required 273.1006, C₁₂H₁₈N₂O₃Cl).

Z-L-Ala-AOMK (23)



The **Z-L-Ala-CMK** (0.150 mg, 0.588 mmol, 1.0 eq.) and 4-methoxybenzoic acid (0.179 g, 1.176 mmol, 2.0 eq.) were dissolved in DMF (2.35 mL). Potassium fluoride (0.136 g, 2.352 mmol) was added to the reaction mixture and the reaction stirred overnight at RT. The reaction mixture was diluted with DCM (10 mL), washed with water, brine, dried over Na₂SO₄, filtered and concentrated under vacuum. Purification by SiO₂ column chromatography using Hexane–EtOAc, 9:1 yielded the desired white solid product (0.988 g, 45 %). R_f = 0.52 (Hexane–EtOAc, 2:1); m.p 92–95°C; **δ_H/ppm (400 MHz; CDCl₃):** 8.06 (2H, d, *J* = 8.80, CH-BnOMe), 7.41–7.34 (5H, m, Bn), 6.99–6.93 (2H, m, CH-BnOMe), 5.48 (1H, d, *J* = 6.20, NH), 5.18–5.10 (2H, m, CH₂), 5.04 (2H, q, *J* = 17.00, CH₂), 4.62 (1H, p, *J* = 7.20, Aα), 3.90 (3H, s, OMe), 1.48 (3H, d, *J* = 7.10, Aβ). **δ_C/ppm (100 MHz; CDCl₃):** 203.16 (OCO), 165.50 (CO), 163.88, 155.70 (C), 132.33, 132.06, 128.60, 128.28, 128.15, 113.83, 113.75 (CH-Bn), 67.11, 66.19 (CH₂), 55.50 (OCH₃), 53.62 (CH), 17.70 (CH₃). **ν_{max} (neat)/cm⁻¹:** 2924, 2854,

1748 (CO), 1723, 1688, 1608, 1552, 1530, 1395, 1252, 1171, 1099, 1026, 846, 766. **ESI.-MS:**
 m/z = found 372.1434 ($[M + H]^+$), 394.1254 ($[M + Na]^+$), (required 394.1267, $C_{20}H_{21}NO_6Na$).

Chapter 10 References

1. Brown, P.O. and D. Botstein, *Exploring the new world of the genome with DNA microarrays*. Nature Genetics, 1999. **21**: p. 33-37.
2. Golub, T.R., et al., *Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring*. Science, 1999. **286**(5439): p. 531-537.
3. Patterson, S.D. and R.H. Aebersold, *Proteomics: the first decade and beyond*. Nature Genetics 2003. **33**: p. 311 - 323.
4. Yates, J.R., *Mass spectral analysis in proteomics* Annual Review of Biophysics and Biomolecular Structure, 2004. **33**(1): p. 297-316.
5. Domon, B., *Review - Mass spectrometry and protein analysis*. Science, 2006. **312**(5771): p. 212-217.
6. Cravatt, B.F., G.M. Simon, and J.R. Yates Iii, *The biological impact of mass-spectrometry-based proteomics*. Nature, 2007. **450**(7172): p. 991-1000.
7. Kobe, B. and B.E. Kemp, *Active site-directed protein regulation*. Nature, 1999. **402**(6760): p. 373-376.
8. Cravatt, B.F., A.T. Wright, and J.W. Kozarich, *Activity-Based Protein Profiling: From Enzyme Chemistry to Proteomic Chemistry*. Annual Review of Biochemistry, 2008. **77**(1): p. 383-414.
9. Fonovic, M.B., Matthew, *Activity Based Probes for Proteases: Applications to Biomarker Discovery, Molecular Imaging and Drug Screening*. Current Pharmaceutical Design, 2007. **13**: p. 253-261.
10. Liu, Y., M.P. Patricelli, and B.F. Cravatt, *Activity-based protein profiling: The serine hydrolases*. Proceedings of the National Academy of Sciences of the United States of America, 1999. **96**(26): p. 14694-14699.
11. Modaber, I., *Clostridium difficile* Acta Medica Iranica, 1975. **18**(3-4): p. 111-128.
12. Katayama, H.Y. and Y. Oda, *Chemical proteomics for drug discovery based on compound-immobilized affinity chromatography*. Journal of Chromatography B- Analytical Technologies in the Biomedical and Life Sciences, 2007. **855**(1): p. 21-27.
13. Rix, U. and G. Superti-Furga, *Target profiling of small molecules by chemical proteomics*. Nature Chemical Biology, 2009. **5**(9): p. 616-624.
14. Bantscheff, M., A. Scholten, and A.J.R. Heck, *Revealing promiscuous drug-target interactions by chemical proteomics*. Drug Discovery Today, 2009. **14**(21-22): p. 1021-1029.
15. Cohen, A.A., et al., *Dynamic Proteomics of Individual Cancer Cells in Response to a Drug*. Science, 2008. **322**(5907): p. 1511-1516.
16. Lee, A.Y.H., et al., *Quantitative Analysis of Histone Deacetylase-1 Selective Histone Modifications by Differential Mass Spectrometry*. Journal of Proteome Research, 2008. **7**(12): p. 5177-5186.
17. Liang, X.Q., et al., *Quantification of change in phosphorylation of BCR-ABL kinase and its substrates in response to imatinib treatment in human chronic myelogenous leukemia cells*. Proteomics, 2006. **6**(16): p. 4554-4564.
18. Song, D., et al., *Antitumor activity and molecular effects of the novel heat shock protein 90 inhibitor, IPI-504, in pancreatic cancer*. Molecular Cancer Therapeutics, 2008. **7**(10): p. 3275-3284.

19. Zhang, J.H., et al., *Protein profile in neuroblastoma cells incubated with S- and R-enantiomers of ibuprofen by iTRAQ-coupled 2-D LC-MS/MS analysis: Possible action of induced proteins on Alzheimer's disease*. *Proteomics*, 2008. **8**(8): p. 1595-1607.
20. Sui, J.J., et al., *Comparative proteomics analysis of vascular smooth muscle cells incubated with S- and R-enantiomers of Atenolol using iTRAQ-coupled two-dimensional LC-MS/MS*. *Molecular & Cellular Proteomics*, 2008. **7**(6): p. 1007-1018.
21. Yamanaka, H., et al., *Quantitative proteomic analysis of rat liver for carcinogenicity prediction in a 28-day repeated dose study*. *Proteomics*, 2007. **7**(5): p. 781-795.
22. Brehmer, D., et al., *Cellular targets of gefitinib*. *Cancer Research*, 2005. **65**(2): p. 379-382.
23. Wissing, J., et al., *Chemical proteomic analysis reveals alternative modes of action for pyrido 2,3-d pyrimidine kinase inhibitors*. *Molecular & Cellular Proteomics*, 2004. **3**(12): p. 1181-1193.
24. Godl, K., et al., *Proteomic characterization of the angiogenesis inhibitor SU6668 reveals multiple impacts on cellular kinase signaling*. *Cancer Research*, 2005. **65**(15): p. 6919-6926.
25. Graves, P.R., et al., *Discovery of novel targets of quinoline drugs in the human purine binding proteome*. *Molecular Pharmacology*, 2002. **62**(6): p. 1364-1372.
26. Scholten, A., et al., *Analysis of the cGMP/cAMP interactome using a chemical proteomics approach in mammalian heart tissue validates sphingosine kinase type 1-interacting protein as a genuine and highly abundant AKAP*. *Journal of Proteome Research*, 2006. **5**(6): p. 1435-1447.
27. Schulze, W.X., L. Deng, and M. Mann, *Phosphotyrosine interactome of the ErbB-receptor kinase family*. *Molecular Systems Biology*, 2005. **1**.
28. Hanke, S. and M. Mann, *The Phosphotyrosine Interactome of the Insulin Receptor Family and Its Substrates IRS-1 and IRS-2*. *Molecular & Cellular Proteomics*, 2009. **8**(3): p. 519-534.
29. Hagenstein, M.C. and N. Sewald, *Chemical tools for activity-based proteomics*. *Journal of Biotechnology*, 2006. **124**(1): p. 56-73.
30. Lottspeich, F., *Proteome analysis: A pathway to the functional analysis of proteins*. *Angewandte Chemie-International Edition*, 1999. **38**(17): p. 2477-2492.
31. Rix, U., et al., *Chemical proteomic profiles of the BCR-ABL inhibitors imatinib, nilotinib, and dasatinib, reveal novel kinase and nonkinase targets*. *Blood*, 2007. **110**: p. 4055-4063.
32. Bantscheff, M., et al., *Quantitative chemical proteomics reveals mechanisms of action of clinical ABL kinase inhibitors*. *Nature Biotechnology*, 2007. **25**(9): p. 1035-1044.
33. Evans, M.J. and B.F. Cravatt, *Mechanism-based profiling of enzyme families*. *Chemical Reviews*, 2006. **106**(8): p. 3279-3301.
34. Shi, H., et al., *Small molecule microarray-facilitated screening of affinity-based probes (AfBPs) for gamma-secretase*. *Chemical Communication* 2009(33): p. 5030-2.
35. Tanaka, Y., M.R. Bond, and J.J. Kohler, *Photocrosslinkers illuminate interactions in living cells*. *Molecular Biosystems*, 2008. **4**(6): p. 473-480.
36. Kalesh, K.A., et al., *The use of click chemistry in the emerging field of catalomics*. *Organic & Biomolecular Chemistry*, 2010. **8**(8): p. 1749-62.

37. Schmidinger, H., A. Hermetter, and R. Birner-Gruenberger, *Activity-based proteomics: enzymatic activity profiling in complex proteomes*. *Amino acids*, 2006. **30**(4): p. 333-350.
38. Uttamchandani, M., C.H.S. Lu, and S.Q. Yao, *Next Generation Chemical Proteomic Tools for Rapid Enzyme Profiling*. *Accounts of Chemical Research*, 2009. **42**(8): p. 1183-1192.
39. Cohen, M.S., H. Hadjivassiliou, and J. Taunton, *A clickable inhibitor reveals context-dependent autoactivation of p90 RSK*. *Nature Chemical Biology*, 2007. **3**(3): p. 156-160.
40. Yee, M.-c., et al., *A cell-permeable, Activity-based Probe for Protein and Lipid Kinases*. *Journal of Biological Chemistry*, 2005. **280**(32): p. 29053-29059.
41. Patricelli, M.P., et al., *Functional Interrogation of the Kinome Using Nucleotide Acyl Phosphates*. *Biochemistry*, 2006. **46**(2): p. 350-358.
42. Kumar, S., et al., *Activity-based probes for protein tyrosine phosphatases*. *Proceedings of the National Academy of Sciences of the United States of America*, 2004. **101**(21): p. 7943-7948.
43. Kalesh, K.A., et al., *Peptide-based activity-based probes (ABPs) for target-specific profiling of protein tyrosine phosphatases (PTPs)*. *Chemical Communications*, 2010. **46**(4): p. 589-591.
44. Deigner, H.P., et al., *Novel reversible, irreversible and fluorescent inhibitors of platelet-activating factor acetylhydrolase as mechanistic probes*. *Atherosclerosis*, 1999. **144**(1): p. 79-90.
45. Schmidinger, H., et al., *Novel fluorescent phosphonic acid esters for discrimination of lipases and esterases*. *Chembiochem*, 2005. **6**(10): p. 1776-1781.
46. Vocadlo, D.J. and C.R. Bertozzi, *A strategy for functional proteomic analysis of glycosidase activity from cell lysates*. *Angewandte Chemie-International Edition*, 2004. **43**(40): p. 5338-5342.
47. Hekmat, O., et al., *Active-site peptide "fingerprinting" of glycosidases in complex mixtures by mass spectrometry - Discovery of a novel retaining beta-1,4-glycanase in *Cellulomonas fimi**. *Journal of Biological Chemistry*, 2005. **280**(42): p. 35126-35135.
48. Adam, G.C., B.F. Cravatt, and E.J. Sorensen, *Profiling the specific reactivity of the proteome with non-directed activity-based probes*. *Chemistry & Biology*, 2001. **8**(1): p. 81-95.
49. Weerapana, E., G.M. Simon, and B.F. Cravatt, *Disparate proteome reactivity profiles of carbon electrophiles*. *Nature Chemical Biology*, 2008. **4**(7): p. 405-407.
50. Sadaghiani, A.M., et al., *Design, synthesis, and evaluation of in vivo potency and selectivity of epoxysuccinyl-based inhibitors of papain-family cysteine proteases*. *Chemistry & Biology*, 2007. **14**(5): p. 499-511.
51. Lee, J. and M. Bogyo, *Development of Near-Infrared Fluorophore (NIRF)-Labeled Activity-Based Probes for in Vivo Imaging of Legumain*. *Acs Chemical Biology*, 2009. **5**(2): p. 233-243.
52. Deu, E., et al., *Use of Activity-Based Probes to Develop High Throughput Screening Assays That Can Be Performed in Complex Cell Extracts*. *PLoS ONE*, 2010. **5**(8): p. e11985.
53. Kato, D., et al., *Activity-based probes that target diverse cysteine protease families*. *Nature Chemical Biology*, 2005. **1**(1): p. 33-38.

54. Simon, G.M. and B.F. Cravatt, *Activity-based Proteomics of Enzyme Superfamilies: Serine Hydrolases as a Case Study*. Journal of Biological Chemistry, 2010. **285**(15): p. 11051-11055.
55. Medzihradszky, K.F., et al., *O-sulfonation of serine and threonine - Mass spectrometric detection and characterization of a new posttranslational modification in diverse proteins throughout the eukaryotes*. Molecular & Cellular Proteomics, 2004. **3**(5): p. 429-440.
56. Bogyo, M., et al., *Covalent modification of the active site threonine of proteasomal beta subunits and the Escherichia coli homolog HslV by a new class of inhibitors*. Proceedings of the National Academy of Sciences of the United States of America, 1997. **94**(13): p. 6629-6634.
57. Bottcher, T., M. Pitscheider, and S.A. Sieber, *Natural Products and Their Biological Targets: Proteomic and Metabolomic Labeling Strategies*. Angewandte Chemie-International Edition, 2010. **49**(15): p. 2680-2698.
58. Wang, C.C., et al., *Biochemical analysis of the 20 S proteasome of Trypanosoma brucei*. Journal of Biological Chemistry, 2003. **278**(18): p. 15800-15808.
59. Chehade, K.A.H., et al., *An improved preparation of the activity-based probe JPM-OEt and in situ applications*. Synthesis-Stuttgart, 2005(2): p. 240-244.
60. Greenbaum, D.C., et al., *Small Molecule Affinity Fingerprinting: a Tool for Enzyme Family Subclassification, Target Identification, and Inhibitor Design*. Chemistry & Biology, 2002. **9**(10): p. 1085-1094.
61. Ekici, O.D., et al., *Aza-peptide Michael acceptors: A new class of inhibitors specific for caspases and other clan CD cysteine proteases*. Journal of medicinal chemistry, 2004. **47**(8): p. 1889-1892.
62. Dragovich, P.S., et al., *Structure-based design, synthesis, and biological evaluation of irreversible human rhinovirus 3C protease inhibitors. 2. Peptide structure-activity studies*. Journal of medicinal chemistry, 1998. **41**(15): p. 2819-2834.
63. Heal, W.P., S.R. Wickramasinghe, and E.W. Tate, *Activity based chemical proteomics: profiling proteases as drug targets*. Current Drug Discovery Technologies, 2008. **5**(3): p. 200-12.
64. Berger, A.B., K.B. Sexton, and M. Bogyo, *Commonly used caspase inhibitors designed based on substrate specificity profiles lack selectivity*. Cell Research, 2006. **16**(12): p. 961-963.
65. Borodovsky, A., et al., *Small-Molecule Inhibitors and Probes for Ubiquitin- and Ubiquitin-Like-Specific Proteases*. Chembiochem, 2005. **6**(2): p. 287-291.
66. Hemelaar, J., et al., *Chemistry-Based Functional Proteomics: Mechanism-Based Activity-Profiling Tools for Ubiquitin and Ubiquitin-like Specific Proteases*. Journal of Proteome Research, 2004. **3**(2): p. 268-276.
67. Kurpiers, T. and H.D. Mootz, *Bioorthogonal Ligation in the Spotlight*. Angewandte Chemie-International Edition, 2009. **48**(10): p. 1729-1731.
68. Heal, W.P., et al., *Site-specific N-terminal labelling of proteins in vitro and in vivo using N-myristoyl transferase and bioorthogonal ligation chemistry*. Chemical Communications, 2008(4): p. 480-482.
69. Agard, N.J., J.A. Prescher, and C.R. Bertozzi, *A new bioorthogonal ligation for labelling of azide-bearing glycoconjugates*. Glycobiology, 2004. **14**(11): p. 513.

70. Grabarek, J., et al., *Activation of chymotrypsin-like serine protease(s) during apoptosis detected by affinity-labeling of the enzymatic center with fluoresceinated inhibitor*. International Journal of Oncology, 2002. **20**(2): p. 225-233.
71. Waggoner, A., *Fluorescent labels for proteomics and genomics*. Current opinion in chemical biology, 2006. **10**(1): p. 62-66.
72. Verhelst, S.H.L. and M. Bogyo, *Solid-phase synthesis of double-headed epoxysuccinyl activity-based probes for selective targeting of papain family cysteine proteases*. ChemBiochem, 2005. **6**(5): p. 824-+.
73. Greenbaum, D., et al., *Chemical approaches for functionally probing the proteome*. Molecular & Cellular Proteomics, 2002. **1**(1): p. 60-68.
74. Patricelli, M.P., et al., *Direct visualization of serine hydrolase activities in complex proteomes using fluorescent active site-directed probes*. Proteomics, 2001. **1**(9): p. 1067-1071.
75. Kessler, B., et al., *Extended peptide-based inhibitors efficiently target the proteasome and reveal overlapping specificities of the catalytic beta-subunits*. Chemistry & Biology, 2001. **8**(9): p. 913-929.
76. Jewett, J.C. and C.R. Bertozzi, *Cu-free click cycloaddition reactions in chemical biology*. Chemical Society Reviews, 2010. **39**(4): p. 1272-1279.
77. Verhelst, S.H.L., M. Fonovic, and M. Bogyo, *A mild chemically cleavable linker system for functional proteomic applications*. Angewandte Chemie-International Edition, 2007. **46**(8): p. 1284-1286.
78. Leriche, G., et al., *Optimization of the Azobenzene Scaffold for Reductive Cleavage by Dithionite; Development of an Azobenzene Cleavable Linker for Proteomic Applications*. European Journal of Organic Chemistry, 2010(23): p. 4360-4364.
79. Landi, F., et al., *Synthesis and application of a new cleavable linker for "click"-based affinity chromatography*. Organic & Biomolecular Chemistry, 2010. **8**(1): p. 56-59.
80. Geurink, P.P., et al., *A Cleavable Linker Based on the Levulinoyl Ester for Activity-Based Protein Profiling*. Angewandte Chemie-International Edition, 2010. **49**(38): p. 6802-6805.
81. Speers, A.E. and B.F. Cravatt, *A tandem orthogonal proteolysis strategy for high-content chemical proteomics*. Journal of the American Chemical Society, 2005. **127**(28): p. 10018-10019.
82. Weerapana, E., A.E. Speers, and B.F. Cravatt, *Tandem orthogonal proteolysis-activity-based protein profiling (TOP-ABPP) - a general method for mapping sites of probe modification in proteomes*. Nature Protocols, 2007. **2**(6): p. 1414-1425.
83. Kolb, H.C., M.G. Finn, and K.B. Sharpless, *Click chemistry: Diverse chemical function from a few good reactions*. Angewandte Chemie-International Edition, 2001. **40**(11): p. 2004-+.
84. Salisbury, C.M. and B.F. Cravatt, *Click chemistry-led advances in high content functional proteomics*. Qsar & Combinatorial Science, 2007. **26**(11-12): p. 1229-1238.
85. Srinivasan, R., et al., *Methods of using click chemistry in the discovery of enzyme inhibitors*. Nature Protocols, 2007. **2**(11): p. 2655-2664.
86. Kalesh, K.A., et al., *Click chemistry as a high-throughput amenable platform in catalomics*. Qsar & Combinatorial Science, 2007. **26**(11-12): p. 1135-1144.
87. Berry, A.F.H., et al., *Rapid Multilabel Detection of Geranylgeranylated Proteins by Using Bioorthogonal Ligation Chemistry*. ChemBiochem, 2010. **11**(6): p. 771-773.

88. Gubbens, J. and A. de Kroon, *Proteome-wide detection of phospholipid-protein interactions in mitochondria by photocrosslinking and click chemistry*. *Molecular Biosystems*, 2010. **6**(10): p. 1751-1759.
89. Johnson, D.S., E. Weerapana, and B.F. Cravatt, *Strategies for discovering and derisking covalent, irreversible enzyme inhibitors*. *Future Medicinal Chemistry*, 2010. **2**(6): p. 949-964.
90. Adam, G.C., et al., *(-)-FR182877 is a potent and selective inhibitor of carboxylesterase-1*. *Angewandte Chemie-International Edition*, 2003. **42**(44): p. 5480-5484.
91. Luo, Y., et al., *Activity-based protein profiling reagents for protein arginine deiminase 4 (PAD4): Synthesis and in vitro evaluation of a fluorescently labeled probe*. *Journal of the American Chemical Society*, 2006. **128**(45): p. 14468-14469.
92. Fuwa, H., et al., *Divergent synthesis of multifunctional molecular probes to elucidate the enzyme specificity of dipeptidic gamma-secretase inhibitors*. *ACS Chemical Biology*, 2007. **2**(6): p. 408-418.
93. Wang, J., et al., *Rapid assembly of matrix metalloprotease inhibitors using click chemistry*. *Organic Letters*, 2006. **8**(17): p. 3821-3824.
94. Agard, N.J., et al., *A comparative study of bioorthogonal reactions with azides*. *ACS Chemical Biology*, 2006. **1**(10): p. 644-648.
95. Rostovtsev, V.V., et al., *A stepwise Huisgen cycloaddition process: Copper(I)-catalyzed regioselective "ligation" of azides and terminal alkynes*. *Angewandte Chemie-International Edition*, 2002. **41**(14): p. 2596-+.
96. Speers, A.E., G.C. Adam, and B.F. Cravatt, *Activity-based protein profiling in vivo using a copper(I)-catalyzed azide-alkyne 3+2 cycloaddition*. *Journal of the American Chemical Society*, 2003. **125**(16): p. 4686-4687.
97. Ovaa, H., et al., *Chemistry in living cells: Detection of active proteasomes by a two-step labeling strategy*. *Angewandte Chemie-International Edition*, 2003. **42**(31): p. 3626-3629.
98. Alexander, J.P. and B.F. Cravatt, *Mechanism of carbamate inactivation of FAAH: Implications for the design of covalent inhibitors and in vivo functional probes for enzymes*. *Chemistry & Biology*, 2005. **12**(11): p. 1179-1187.
99. Evans, M.J., et al., *Target discovery in small-molecule cell-based screens by in situ proteome reactivity profiling*. *Nature Biotechnology*, 2005. **23**(10): p. 1303-1307.
100. Wright, A.T., J.D. Song, and B.F. Cravatt, *A Suite of Activity-Based Probes for Human Cytochrome P450 Enzymes*. *Journal of the American Chemical Society*, 2009. **131**(30): p. 10692-10700.
101. Ning, X., et al., *Visualizing Metabolically Labeled Glycoconjugates of Living Cells by Copper-Free and Fast Huisgen Cycloadditions*. *Angewandte Chemie International Edition*, 2008. **47**(12): p. 2253-2255.
102. Fernandez-Suarez, M., et al., *Redirecting lipoyl acid ligase for cell surface protein labeling with small-molecule probes*. *Nature Biotechnology* 2007. **25**(12): p. 1483-1487.
103. Strassberger, V., et al., *Chemical proteomic and bioinformatic strategies for the identification and quantification of vascular antigens in cancer*. *Journal of Proteomics*, 2010. **73**(10): p. 1954-1973.
104. Elschenbroich, S., et al., *Isolation of cell surface proteins for mass spectrometry-based proteomics*. *Expert Review of Proteomics*, 2010. **7**(1): p. 141-154.

105. Mondal, K. and M.N. Gupta, *The affinity concept in bioseparation: Evolving paradigms and expanding range of applications*. Biomolecular Engineering, 2006. **23**(2-3): p. 59-76.
106. Aye, T.T., et al., *Selectivity in Enrichment of cAMP-dependent Protein Kinase Regulatory Subunits Type I and Type II and Their Interactors Using Modified cAMP Affinity Resins*. Molecular & Cellular Proteomics, 2009. **8**(5): p. 1016-1028.
107. Daub, H., *Characterisation of kinase-selective inhibitors by chemical proteomics*. Biochimica Et Biophysica Acta-Proteins and Proteomics, 2005. **1754**(1-2): p. 183-190.
108. Patricelli, M.P., et al., *Functional interrogation of the kinome using nucleotide acyl phosphates*. Biochemistry, 2007. **46**(2): p. 350-358.
109. Oda, Y., et al., *Quantitative chemical proteomics for identifying candidate drug targets*. Analytical Chemistry, 2003. **75**(9): p. 2159-2165.
110. Yamamoto, K., et al., *A versatile method of identifying specific binding proteins on affinity resins*. Analytical Biochemistry, 2006. **352**(1): p. 15-23.
111. Jenkins, R.E., et al., *Glutathione-S-transferase pi as a model protein for the characterisation of chemically reactive metabolites*. Proteomics, 2008. **8**(2): p. 301-315.
112. Piffeteau, A. and M. Gaudry, *Biotin uptake-influx, efflux and countertransport in Escherichia coli K12*. Biochimica Et Biophysica Acta, 1985. **816**(1): p. 77-82.
113. Cicmanec, J.F. and H.C. Lichstein, *Uptake of extracellular biotin by Escherichia coli biotin prototrophs*. J. Bacteriol., 1978. **133**(1): p. 270-278.
114. Detmers, F.J.M., et al., *Kinetics and specificity of peptide uptake by the oligopeptide transport system of Lactococcus lactis*. Biochem., 1998. **37**(47): p. 16671-16679.
115. Walker, J.R. and E. Altman, *Biotinylation facilitates the uptake of large peptides by Escherichia coli and other gram-negative bacteria*. Appl. Environ. Microbiol., 2005. **71**(4): p. 1850-1855.
116. Wilchek, M. and E.A. Bayer, *Introduction to avidin-biotin technology* Methods in Enzymology, 1990. **184**: p. 5-13.
117. Fraenkel-Conrat, H., et al., *Avidin. I. Isolation and characterization of the protein and nucleic acid*. Archives of Biochemistry and Biophysics, 1952. **39**(1): p. 80-96.
118. Tausig, F. and F.J. Wolf, *Streptavidin-Substance with avidin-like properties produced by microorganisms* Biochemical and Biophysical Research Communications, 1964. **14**(3): p. 205-&.
119. Chaiet, L. and F.J. Wolf, *Properties of streptavidin biotin-binding protein produced by Streptomyces* Archives of Biochemistry and Biophysics, 1964. **106**(1-3): p. 1-&.
120. Vermette, P., et al., *Immobilization and surface characterization of NeutrAvidin biotin-binding protein on different hydrogel interlayers*. Journal of Colloid and Interface Science, 2003. **259**(1): p. 13-26.
121. Scientific, T. *NeutrAvidin Protein and Conjugates* 2010; Available from: <http://www.piercenet.com/browse.cfm?fldID=01030702>.
122. Alan J. Barrett, N.D.R.J.F.W., *Handbook of Proteolytic Enzymes*. 1998, London: Academic Press.
123. Alan J. Barrett, G.S., *Proteinase inhibitors*, ed. G.S. Alan J. Barrett. 1986, the University of Michigan: Elsevier.
124. Rawlings, N.D. and A.J. Barrett, *Evolutionary families of peptidases* Biochemical Journal, 1993. **290**: p. 205-218.

125. Barrett, A.J., *Classification of peptidases*. Proteolytic Enzymes: Serine and Cysteine Peptidases, ed. M.I. Enzymology. Vol. 244. 1994: Academic Press Inc., 525 B Street, Suite 1900, San Diego, CA 92101-4495. 1-15.
126. Rawlings, N.D., F.R. Morton, and A.J. Barrett, *MEROPS: the peptidase database*. Nucleic Acids Research, 2006. **34**: p. D270-D272.
127. Schechter I, B.A., *On the size of the active site in proteases. I. Papain*. Biochemical and biophysical research communications, 1967. **27**(2): p. 157.
128. Website, T.E.
129. Powers, J.C., et al., *Irreversible inhibitors of serine, cysteine, and threonine proteases*. Chemical Reviews, 2002. **102**(12): p. 4639-4750.
130. Gorrell, M.D., *Dipeptidyl peptidase IV and related enzymes in cell biology and liver disorders*. Clinical Science, 2005. **108**(4): p. 277-292.
131. DeClerck, Y.A., et al., *Proteases and protease inhibitors in tumor progression*, in *Chemistry and Biology of Serpins*, F.C. Church, et al., Editors. 1997. p. 89-97.
132. Kidd, D., Y.S. Liu, and B.F. Cravatt, *Profiling serine hydrolase activities in complex proteomes*. Biochemistry, 2001. **40**(13): p. 4005-4015.
133. Pan, Z.Y., et al., *Development of activity-based probes for trypsin-family serine proteases*. Bioorganic & Medicinal Chemistry Letters, 2006. **16**(11): p. 2882-2885.
134. Clark, J.D., et al., *Cytosolic phospholipase A2*. Journal of Lipid Mediators and Cell Signalling, 1995. **12**(2-3): p. 83-117.
135. Mignatti, P. and D.B. Rifkin, *Plasminogen activators and angiogenesis*, in *Attempts to Understand Metastasis Formation I*. 1996. p. 33-50.
136. Piñeiro-Sánchez, M.L., et al., *Identification of the 170-kDa Melanoma Membrane-bound Gelatinase (Seprase) as a Serine Integral Membrane Protease*. Journal of Biological Chemistry, 1997. **272**(12): p. 7595-7601.
137. Bouma, B.N., et al., *Human plasma prekallikrein. Studies of its activation by activated factor XII and of its inactivation by diisopropyl phosphorfluoridate*. Biochemistry, 1980. **19**(6): p. 1151-1160.
138. Nomura, D.K., et al., *Monoacylglycerol Lipase Regulates a Fatty Acid Network that Promotes Cancer Pathogenesis*. Cell, 2010. **140**(1): p. 49-61.
139. Gilmore, B.F., et al., *Expedited Solid-Phase Synthesis of Fluorescently Labeled and Biotinylated Aminoalkane Diphenyl Phosphonate Affinity Probes for Chymotrypsin- and Elastase-Like Serine Proteases*. Bioconjugate Chemistry, 2009. **20**(11): p. 2098-2105.
140. Mahrus, S. and C.S. Craik, *Selective Chemical Functional Probes of Granzymes A and B Reveal Granzyme B Is a Major Effector of Natural Killer Cell-Mediated Lysis of Target Cells*. Chemistry & Biology, 2005. **12**(5): p. 567-577.
141. Bochtler, M., et al., *The proteasome* Annual Review of Biophysics and Biomolecular Structure, 1999. **28**(1): p. 295-317.
142. Verdoes, M., et al., *A panel of subunit-selective activity-based proteasome probes*. Organic & Biomolecular Chemistry, 2010. **8**(12): p. 2719-2727.
143. Iwata, Y., et al., *Macrophage cathepsin L, a factor in the erosion of subchondral bone in rheumatoid arthritis*. Arthritis & Rheumatism, 1997. **40**(3): p. 499-509.
144. Lavrik, I.N., A. Golks, and P.H. Krammer, *Caspases: pharmacological manipulation of cell death*. The Journal of Clinical Investigation, 2005. **115**(10): p. 2665-2672.
145. Berdowska, I., *Cysteine proteases as disease markers*. Clinica Chimica Acta, 2004. **342**(1-2): p. 41-69.

146. Denault, J.B. and G.S. Salvesen, *Caspases: Keys in the ignition of cell death*. Chemical Reviews, 2002. **102**(12): p. 4489-4499.
147. Yan, S.Q., M. Sameni, and B.F. Sloane, *Cathepsin B and human tumor progression*. Biological Chemistry, 1998. **379**(2): p. 113-123.
148. Shaw, E., *Peptidyl diazomethanes as inhibitors of cysteine and serine proteinases*, in *Methods in Enzymology*, J.B. Alan, Editor. 1994, Academic Press. p. 649-656.
149. Palmer, J., et al., *Vinyl sulfones as mechanism-based cysteine protease inhibitors* Journal of medicinal chemistry, 1995. **38**(17): p. 3193-3196.
150. James, K.E., et al., *Design, synthesis, and evaluation of aza-peptide epoxides as selective and potent inhibitors of caspases-1,-3,-6, and-8*. Journal of medicinal chemistry, 2004. **47**(6): p. 1553-1574.
151. Goursalin, B.J., et al., *Epoxy succinyl dipeptides as selective inhibitors of Cathepsin-B* Journal of medicinal chemistry, 1993. **36**(6): p. 720-725.
152. Bromme, D., et al., *Potent and selective inactivation of cysteine proteinases with N-peptidyl-O-acyl-hydroxylamines* Biochemical Journal, 1989. **263**(3): p. 861-866.
153. Smith, R., et al., *New inhibitors of cysteine proteinases - Peptidyl acyloxymethyl ketones and the quiescent nucleofuge strategy* Journal of the American Chemical Society, 1988. **110**(13): p. 4429-4431.
154. Pliura, D., et al., *Comparative behavior of calpain and cathepsin-B toward peptidyl acyloxymethyl ketones, sulfonium methyl ketones and other potential inhibitors of cysteine proteases* Biochemical Journal, 1992. **288**: p. 759-762.
155. Hanada, K., et al., *Isolation and characterization of E-64, a new thiol protease inhibitor* Agricultural and biological chemistry, 1978. **42**(3): p. 523-528.
156. Hang, H.C., et al., *Mechanism-based probe for the analysis of cathepsin cysteine proteases in living cells*. Acs Chemical Biology, 2006. **1**(11): p. 713-723.
157. Greenbaum, D., et al., *Epoxide electrophiles as activity-dependent cysteine protease profiling and discovery tools*. Chemistry & Biology, 2000. **7**(8): p. 569-581.
158. van Swieten, P.F., et al., *Development of an isotope-coded activity-based probe for the quantitative profiling of cysteine proteases*. Bioorganic & Medicinal Chemistry Letters, 2004. **14**(12): p. 3131-3134.
159. Beckham, S.A., et al., *A major cathepsin B protease from the liver fluke Fasciola hepatica has atypical active site features and a potential role in the digestive tract of newly excysted juvenile parasites*. The International Journal of Biochemistry & Cell Biology, 2009. **41**(7): p. 1601-1612.
160. Dai, Y., L. Hedstrom, and R.H. Abeles, *Inactivation of Cysteine Proteases by (Acyloxy)methyl Ketones Using S'-P' Interactions†*. Biochemistry, 2000. **39**(21): p. 6498-6502.
161. Blum, G., et al., *Comparative Assessment of Substrates and Activity Based Probes as Tools for Non-Invasive Optical Imaging of Cysteine Protease Activity*. PLoS ONE, 2009. **4**(7): p. e6374.
162. Yuan, F., et al., *A selective activity-based probe for the papain family cysteine protease dipeptidyl peptidase I cathepsin C*. Journal of the American Chemical Society, 2006. **128**(17): p. 5616-5617.
163. Coleman, J.E., *Zinc enzymes*. Current opinion in chemical biology, 1998. **2**(2): p. 222-234.

164. Bursavich, M.G. and D.H. Rich, *Designing Non-Peptide Peptidomimetics in the 21st Century: Inhibitors Targeting Conformational Ensembles*. Journal of medicinal chemistry, 2002. **45**(3): p. 541-558.
165. Page-McCaw, A., A.J. Ewald, and Z. Werb, *Matrix metalloproteinases and the regulation of tissue remodelling*. Nature Reviews Molecular Cell Biology, 2007. **8**(3): p. 221-233.
166. Overall, C.M. and C. Lopez-Otin, *Strategies for MMP inhibition in cancer: innovations for the post-trial era*. Nature Reviews Cancer 2002. **2**(9): p. 657-672.
167. Chang, C. and Z. Werb, *The many faces of metalloproteases: cell growth, invasion, angiogenesis and metastasis*. Trends in Cell Biology, 2001. **11**(11): p. S37-S43.
168. Saghatelian, A., et al., *Activity-based probes for the proteomic profiling of metalloproteases*. Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(27): p. 10000-10005.
169. Chan, E.W.S., et al., *Developing Photoactive Affinity Probes for Proteomic Profiling: Hydroxamate-based Probes for Metalloproteases*. Journal of the American Chemical Society, 2004. **126**(44): p. 14435-14446.
170. Wang, J., et al., *"Click" synthesis of small molecule probes for activity-based fingerprinting of matrix metalloproteases*. Chemical Communications, 2006(36): p. 3783-3785.
171. Li, Y.-M., et al., *Photoactivated [gamma]-secretase inhibitors directed to the active site covalently label presenilin 1*. Nature, 2000. **405**(6787): p. 689-694.
172. Shearman, M.S., et al., *L-685,458, an Aspartyl Protease Transition State Mimic, Is a Potent Inhibitor of Amyloid β -Protein Precursor γ -Secretase Activity*. Biochemistry, 2000. **39**(30): p. 8698-8704.
173. Aebersold, R. and M. Mann, *Mass spectrometry-based proteomics*. Nature, 2003. **422**(6928): p. 198-207.
174. Murray, P.R., *Matrix-assisted laser desorption ionization time-of-flight mass spectrometry: usefulness for taxonomy and epidemiology*. Clinical Microbiology and Infection, 2010. **16**(11): p. 1626-1630.
175. Walther, T. and M. Mann, *Mass spectrometry-based proteomics in cell biology*. The journal of cell biology, 2010. **190**(4): p. 491-500.
176. Pandey, A. and M. Mann, *Proteomics to study genes and genomes*. Nature, 2000. **405**(6788): p. 837-846.
177. Spengler, B., et al., *Postionization of laser-desorbed organic and inorganic compounds in a time of flight mass-spectrometer* Analytical Instrumentation, 1988. **17**(1-2): p. 173-193.
178. Fenn, J.B., et al., *Electrospray ionization for mass-spectrometry of large biomolecules*. Science, 1989. **246**(4926): p. 64-71.
179. Smith, D. and P.R. Cromey, *An inexpensive bakeable quadrupole mass spectrometer* Journal of Physics E-Scientific Instruments, 1968. **1**(5): p. 523-&.
180. Yinon, J. and M. Ganz, *Trap current regulated ion-source power-supply for a mass spectrometer* Review of Scientific Instruments, 1975. **46**(12): p. 1707-1708.
181. Wiley, W.C., *Bendix Time-of-Flight mass spectrometer* Science, 1956. **124**(3226): p. 817-820.
182. Marshall, A.G., M.B. Comisarow, and G. Parisod, *Theory of Fourier- transform ion-cyclotron resonance mass spectrometry-III. 1. Relaxation and spectral-line shape in*

- Fourier-transform ion resonance spectroscopy* Journal of Chemical Physics, 1979. **71**(11): p. 4434-4444.
183. Comisarow, M.B., *Theory of Fourier-transform ion-cyclotron resonance mass-spectrometry. 2. Signal modeling for ion-cyclotron resonance* Journal of Chemical Physics, 1978. **69**(9): p. 4097-4104.
 184. Bairoch, A., et al., *The Universal Protein Resource (UniProt) 2009*. Nucleic Acids Research, 2009. **37**: p. D169-D174.
 185. Cooks, R.G., et al., *Ion trap mass spectrometry*. Chemical & Engineering News, 1991. **69**(12): p. 26-&.
 186. Olsen, J.V. and M. Mann, *Improved peptide identification in proteomics by two consecutive stages of mass spectrometric fragmentation*. Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(37): p. 13417-13422.
 187. Makarov, A., *Electrostatic Axially Harmonic Orbital Trapping: A High-Performance Technique of Mass Analysis*. Analytical Chemistry, 2000. **72**(6): p. 1156-1162.
 188. Hu, Q., et al., *The Orbitrap: a new mass spectrometer*. Journal of Mass Spectrometry, 2005. **40**(4): p. 430-443.
 189. Hardman, M. and A.A. Makarov, *Interfacing the orbitrap mass analyzer to an electrospray ion source*. Analytical Chemistry, 2003. **75**(7): p. 1699-1705.
 190. Han, X.M., A. Aslanian, and J.R. Yates, *Mass spectrometry for proteomics*. Current opinion in chemical biology, 2008. **12**(5): p. 483-490.
 191. Makarov, A., et al., *Performance Evaluation of a Hybrid Linear Ion Trap/Orbitrap Mass Spectrometer*. Analytical Chemistry, 2006. **78**(7): p. 2113-2120.
 192. Meier, K. and J. Seibl, *Channel electron multipliers as detectors in organic mass-spectrometry* Journal of Physics E-Scientific Instruments, 1973. **6**(2): p. 133-135.
 193. Henzel, W.J., et al., *Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases*. Proceedings of the National Academy of Sciences of the United States of America, 1993. **90**(11): p. 5011-5015.
 194. Zhu, W.H., J.W. Smith, and C.M. Huang, *Mass Spectrometry-Based Label-Free Quantitative Proteomics*. Journal of Biomedicine and Biotechnology, 2010.
 195. Shukla, A.K. and J.H. Futrell, *Tandem mass spectrometry: dissociation of ions by collisional activation*. Journal of Mass Spectrometry, 2000. **35**(9): p. 1069-1090.
 196. Cordero, M.M., et al., *Sequencing peptides without scanning the reflectron-post-source decay with a curved-field reflectron time-of-flight mass-spectrometer* Rapid Communications in Mass Spectrometry, 1995. **9**(14): p. 1356-1361.
 197. Zubarev, R.A., N.L. Kelleher, and F.W. McLafferty, *Electron Capture Dissociation of Multiply Charged Protein Cations. A Nonergodic Process*. Journal of the American Chemical Society, 1998. **120**(13): p. 3265-3266.
 198. Zubarev, R., *Protein primary structure using orthogonal fragmentation techniques in Fourier transform mass spectrometry*. Expert Review of Proteomics, 2006. **3**(2): p. 251-261.
 199. Syka, J.E.P., et al., *Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry*. Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(26): p. 9528-9533.

200. Perkins, D.N., et al., *Probability-based protein identification by searching sequence databases using mass spectrometry data*. Electrophoresis 1999. **20**(18): p. 3551-3567.
201. Eng, J.K., A.L. McCormack, and J.R. Yates, *An approach to correlate tandem mass-spectral data of peptides with amino-acid-sequences in a protein database* Journal of the American Society for Mass Spectrometry, 1994. **5**(11): p. 976-989.
202. Craig, R., J.P. Cortens, and R.C. Beavis, *Open source system for analyzing, validating, and storing protein identification data*. Journal of Proteome Research, 2004. **3**(6): p. 1234-1242.
203. Geer, L.Y., et al., *Open mass spectrometry search algorithm*. Journal of Proteome Research, 2004. **3**(5): p. 958-964.
204. Liu, H., R.G. Sadygov, and J.R. Yates, *A Model for Random Sampling and Estimation of Relative Protein Abundance in Shotgun Proteomics*. Analytical Chemistry, 2004. **76**(14): p. 4193-4201.
205. Mueller, L.N., et al., *An assessment of software solutions for the analysis of mass spectrometry based quantitative proteomics data*. Journal of Proteome Research, 2008. **7**(1): p. 51-61.
206. Elias, J.E., et al., *Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations*. Nature Methods, 2005. **2**(9): p. 667-675.
207. Goujon, M., et al., *A new bioinformatics analysis tools framework at EMBL-EBI*. Nucleic Acids Research, 2010. **38** Suppl: p. W695-9.
208. Tateno, Y., et al., *DNA Data Bank of Japan (DDBJ) for genome scale research in life science*. Nucleic Acids Research, 2002. **30**(1): p. 27-30.
209. Benson, D.A., et al., *GenBank*. Nucleic Acids Research, 2010. **38**: p. D46-D51.
210. Pruitt, K.D., et al., *NCBI Reference Sequences: current status, policy and new initiatives*. Nucleic Acids Research, 2009. **37**: p. D32-D36.
211. Geer, L.Y., et al., *The NCBI BioSystems database*. Nucleic Acids Research, 2010. **38**: p. D492-D496.
212. Kersey, P.J., et al., *Ensembl Genomes: Extending Ensembl across the taxonomic space*. Nucleic Acids Research, 2010. **38**: p. D563-D569.
213. David, F.P.A. and Y.L. Yip, *SSMap: A new UniProt-PDB mapping resource for the curation of structural-related information in the UniProt/Swiss-Prot Knowledgebase*. BMC Bioinformatics, 2008. **9**.
214. Martens, L., et al., *PRIDE: The proteomics identifications database*. Proteomics, 2005. **5**(13): p. 3537-3545.
215. Ogorzalek Loo, R.R., et al., *Top-down, bottom-up, and side-to-side proteomics with virtual 2-D gels*. International Journal of Mass Spectrometry, 2005. **240**(3): p. 317-325.
216. Garbis, S., G. Lubec, and M. Fountoulakis, *Limitations of current proteomics technologies*. Journal of Chromatography A, 2005. **1077**(1): p. 1-18.
217. Cravatt, B.F., G.M. Simon, and J.R. Yates, *The biological impact of mass-spectrometry-based proteomics*. Nature, 2007. **450**(7172): p. 991-1000.
218. Schulze, W.X. and B. Usadel, *Quantitation in Mass-Spectrometry-Based Proteomics*, in *Annual Review of Plant Biology, Vol 61*. 2010, Annual Reviews: Palo Alto. p. 491-516.

219. Oda, Y., et al., *Accurate quantitation of protein expression and site-specific phosphorylation*. Proceedings of the National Academy of Sciences of the United States of America, 1999. **96**(12): p. 6591-6596.
220. Ong, S.-E., et al., *Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics*. Molecular & Cellular Proteomics, 2002. **1**(5): p. 376-386.
221. Gygi, S.P., et al., *Quantitative analysis of complex protein mixtures using isotope-coded affinity tags*. Nature Biotechnology 1999. **17**(10): p. 994-999.
222. Ross, P.L., et al., *Multiplexed Protein Quantitation in*. Molecular & Cellular Proteomics, 2004. **3**(12): p. 1154-1169.
223. Desiderio, D.M. and M. Kai, *Preparation of stable isotope-incorporated peptide internal standards for field desorption mass spectrometry quantification of peptides in biologic tissue*. Biological Mass Spectrometry, 1983. **10**(8): p. 471-479.
224. Hattan, S.J. and K.C. Parker, *Methodology utilizing MS signal intensity and LC retention time for quantitative analysis and precursor ion selection in proteomic LC-MALDI analyses*. Analytical Chemistry, 2006. **78**(23): p. 7986-7996.
225. Lu, P., et al., *Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation*. Nature Biotechnology 2007. **25**(1): p. 117-124.
226. Hu, J., et al., *The importance of experimental design in proteomic mass spectrometry experiments: Some cautionary tales*. Briefings in Functional Genomics & Proteomics, 2005. **3**(4): p. 322-331.
227. Old, W.M., et al., *Comparison of Label-free Methods for Quantifying Human Proteins by Shotgun Proteomics*. Molecular & Cellular Proteomics, 2005. **4**(10): p. 1487-1502.
228. Hall IC, O.T.E., *Intestinal flora in bew-born infants: with a description of a new pathogenic anaerobe Bacillus difficilis*. American journal of Diseases of children, 1935. **49**: p. 390-402.
229. Bartlett, J.G., *Clostridium difficile -History of its role as an enteric pathogen and the current state of knowlege about the organism*. Clinical Infectious Diseases, 1994. **18**: p. S265-S272.
230. Kelly, C.P. and J.T. LaMont, *Clostridium difficile — More Difficult Than Ever*. New England Journal of Medicine, 2008. **359**(18): p. 1932-1940.
231. Rupnik, M., M.H. Wilcox, and D.N. Gerding, *Clostridium difficile infection: new developments in epidemiology and pathogenesis*. Nature Reviews Microbiology, 2009. **7**(7): p. 526-536.
232. Bartlett, J.G., *Clostridium difficile - Old and new observations*. Journal of Clinical Gastroenterology, 2007. **41**(5): p. S24-S29.
233. Sullivan, N.M., S. Pellett, and T.D. Wilkins, *Purification and characterisation of toxin A and toxin B of Clostridium difficile*. Infection and Immunity, 1982. **35**(3): p. 1032-1040.
234. Rifkin, G.D., et al., *Antibiotic induced colitis implication of toxin neutralized by Clostridium sordellii antitoxin*. Lancet, 1977. **2**(8048): p. 1103-1106.
235. Bartlett, J.G., et al., *Clindamycin associated colitis due to a toxin producing species of Clostridium in hamsters*. Journal of Infectious Diseases, 1977. **136**(5): p. 701-705.
236. Voth, D.E. and J.D. Ballard, *Clostridium difficile toxins: Mechanism of action and role in disease*. Clinical Microbiology Reviews, 2005. **18**(2): p. 247-+.

237. Just, I. and R. Gerhard, *Large clostridial cytotoxins*. Rev Physiol Biochem Pharmacol, 2004. **152**: p. 23-47.
238. Bartlett, J.G., *New drugs for Clostridium difficile infection*. Clinical Infectious Diseases, 2006. **43**(4): p. 428-431.
239. Wilcox, M.H., et al., *Financial burden of hospital-acquired Clostridium difficile infection*. Journal of Hospital Infection, 1996. **34**(1): p. 23-30.
240. Kyne, L., et al., *Health care costs and mortality associated with nosocomial diarrhea due to Clostridium difficile*. Clinical Infectious Diseases, 2002. **34**(3): p. 346-353.
241. Hafiz, S. and C.L. Oakley, *Clostridium difficile - Isolation and characteristics*. Journal of Medical Microbiology, 1976. **9**(2): p. 129-8.
242. Saif, N.A. and J.S. Brazier, *The distribution of Clostridium difficile in the environment of South Wales*. Journal of Medical Microbiology, 1996. **45**(2): p. 133-137.
243. Matsuki, S., et al., *Colonization by Clostridium difficile of neonates in a hospital, and infants and children in three day-care facilities of Kanazawa, Japan*. International Microbiology, 2005. **8**(1): p. 43-48.
244. Heap, J.T., et al., *The CloStron: A universal gene knock-out system for the genus Clostridium*. Journal of Microbiological Methods, 2007. **70**(3): p. 452-464.
245. Mohammed, S., et al., *The multidrug-resistant human pathogen Clostridium difficile has a highly mobile, mosaic genome*. Nature Genetics, 2006. **38**(7): p. 779-786.
246. Kuehne, S.A., et al., *The role of toxin A and toxin B in Clostridium difficile infection*. Nature, 2010. **467**(7316): p. 711-713.
247. Thelestam, M. and E. Chaves-Olarte, *Cytotoxic effects of the Clostridium difficile toxins*. Current topics in Microbiology and Immunology ed. C. difficile. Vol. 250. 2000. 85-96.
248. Jank, T., T. Giesemann, and K. Aktories, *Rho-glucosylating Clostridium difficile toxins A and B: new insights into structure and function*. Glycobiology, 2007. **17**(4): p. 15R-22R.
249. Pothoulakis, C., *Effects of Clostridium difficile Toxins on Epithelial Cell Barrier*. Annals of the New York Academy of Sciences, 2000. **915**(1): p. 347-356.
250. Borriello, S.P., et al., *Virulence factors of Clostridium difficile*. Reviews of Infectious Diseases, 1990. **12**: p. S185-S191.
251. Karjalainen, T., et al., *Cloning of a genetic determinant from Clostridium difficile involved in adherence to tissue-culture cells and mucus*. Infection and Immunity, 1994. **62**(10): p. 4347-4355.
252. Cerquetti, M., et al., *Characterization of surface layer proteins from different Clostridium difficile clinical isolates*. Microbial Pathogenesis, 2000. **28**(6): p. 363-372.
253. Kawata, T., et al., *Demonstration and preliminary characterisation of a regular array in the cell wall of Clostridium difficile*. Fems Microbiology Letters, 1984. **24**(2-3): p. 323-328.
254. Takeoka, A., et al., *Purification and characterisation of S-layer proteins from Clostridium difficile GAI-0714*. Journal of General Microbiology, 1991. **137**: p. 261-267.
255. Calabi, E., et al., *Molecular characterization of the surface layer proteins from Clostridium difficile*. Molecular Microbiology, 2001. **40**(5): p. 1187-1199.
256. Waligora, A.-J., et al., *Characterization of a Cell Surface Protein of Clostridium difficile with Adhesive Properties*. Infection and Immunity, 2001. **69**(4): p. 2144-2153.

257. Savariau-Lacomme, M.P., et al., *Transcription and analysis of polymorphism in a cluster of genes encoding surface-associated proteins of Clostridium difficile*. Journal of Bacteriology, 2003. **185**(15): p. 4461-4470.
258. Calabi, E., et al., *Binding of Clostridium difficile surface layer proteins to gastrointestinal tissues*. Infection and Immunity, 2002. **70**(10): p. 5770-5778.
259. Wright, A., et al., *Immunoreactive cell wall proteins of Clostridium difficile identified by human sera*. Journal of Medical Microbiology, 2008. **57**(6): p. 750-756.
260. Wright, A., et al., *Proteomic analysis of cell surface proteins from Clostridium difficile*. Proteomics, 2005. **5**(9): p. 2443-2452.
261. Fagan, R.P., et al., *Structural insights into the molecular organization of the S-layer from Clostridium difficile*. Molecular Microbiology, 2009. **71**(5): p. 1308-1322.
262. Cerquetti, M., et al., *Purification and characterisation of an immunodominant 36 kDa antigen present on the cell surface of Clostridium difficile*. Microbial Pathogenesis, 1992. **13**(4): p. 271-279.
263. Karjalainen, T., et al., *Molecular and genomic analysis of genes encoding surface-anchored proteins from Clostridium difficile*. Infection and Immunity, 2001. **69**(5): p. 3442-3446.
264. Calabi, E. and N. Fairweather, *Patterns of sequence conservation in the S-layer proteins and related sequences in Clostridium difficile*. Journal of Bacteriology, 2002. **184**(14): p. 3886-3897.
265. Eidhin, D.N., et al., *Sequence and phylogenetic analysis of the gene for surface layer protein, slpA, from 14 PCR ribotypes of Clostridium difficile*. J Med Microbiol, 2006. **55**(Pt 1): p. 69-83.
266. Sharp, J. and I.R. Poxton, *The cell-wall proteins of Clostridium difficile*. Fems Microbiology Letters, 1988. **55**(1): p. 99-103.
267. Pantosti, A., et al., *Immunoblot analysis of serum immunoglobulin-G response to surface proteins of Clostridium difficile in patients with antibiotic associate diarrhea*. Journal of Clinical Microbiology, 1989. **27**(11): p. 2594-2597.
268. Poxton, I.R., et al., *Variation in the cell surface proteins of Clostridium difficile*. Anaerobe, 1999. **5**(3-4): p. 213-215.
269. Karjalainen, T., et al., *Clostridium difficile genotyping based on slpA variable region in S-layer gene sequence: an alternative to serotyping*. Journal of Clinical Microbiology, 2002. **40**(7): p. 2452-2458.
270. Cerquetti, M., et al., *Purification and characterization of an immunodominant -36kDa antigen present on the cell-surface of Clostridium difficile*. Microbial Pathogenesis, 1992. **13**(4): p. 271-279.
271. Kawata, T. and K. Masuda, *Electron-microscopy of a regular array in the cell-wall of Clostridium difficile* Journal of Electron Microscopy, 1984. **33**(3): p. 297-297.
272. McCoubrey, J. and I.R. Poxton, *Variation in the surface layer proteins of Clostridium difficile*. Fems Immunology and Medical Microbiology, 2001. **31**(2): p. 131-135.
273. Sharp, J. and I.R. Poxton, *The cell wall proteins of Clostridium difficile* Fems Microbiology Letters, 1988. **55**(1): p. 99-103.
274. Takeoka, A., et al., *Purification and characterisation of S-layer proteins from Clostridium difficile GAI-0714* Journal of General Microbiology, 1991. **137**: p. 261-267.

275. Wright, A.E., *Proteomic and Bioinformatic Analysis of Surface Proteins in Clostridium difficile*, in *Department of Biological Sciences*. 2006, Imperial College London: London. p. 132.
276. Savariau-Lacomme, M.-P., et al., *Transcription and Analysis of Polymorphism in a Cluster of Genes Encoding Surface-Associated Proteins of Clostridium difficile*. *Journal of Bacteriology*, 2003. **185**(15): p. 4461-4470.
277. Sebahia, M., et al., *The multidrug-resistant human pathogen Clostridium difficile has a highly mobile, mosaic genome*. *Nature Genetics*, 2006. **38**(7): p. 779-786.
278. Waligora, A.J., et al., *Characterization of a cell surface protein of Clostridium difficile with adhesive properties*. *Infection and Immunity*, 2001. **69**(4): p. 2144-2153.
279. Emerson, J.E., et al., *A novel genetic switch controls phase variable expression of CwpV, a Clostridium difficile cell wall protein*. *Molecular Microbiology*, 2009. **74**(3): p. 541-556.
280. Janoir, C., et al., *Cwp84, a surface-associated protein of Clostridium difficile, is a cysteine protease with degrading activity on extracellular matrix proteins*. *Journal of Bacteriology*, 2007. **189**(20): p. 7174-7180.
281. Poilane, I., et al., *Protease activity of Clostridium difficile strains*. *Canadian Journal of Microbiology*, 1998. **44**(2): p. 157-161.
282. Steffen, E.K. and D.J. Hentges, *Hydrolytic enzymes of anaerobic bacteria isolated from human infections*. *Journal of Clinical Microbiology*, 1981. **14**(2): p. 153-156.
283. Egerer, M., et al., *Auto-catalytic cleavage of Clostridium difficile toxins a and B depends on cysteine protease activity*. *Journal of Biological Chemistry*, 2007. **282**(35): p. 25314-25321.
284. Reineke, J., et al., *Autocatalytic cleavage of Clostridium difficile toxin*. *Nature*, 2007. **446**(7134): p. 415-419.
285. Emerson, J., *Transcriptional studies on Clostridium difficile and its surface proteins*. 2007.
286. Emerson, J., and N. F. Fairweather, *Surface structures of C. difficile and other clostridia*. In H. Bruggemann and G. Gottschalk (ed.) ed. *Clostridia - Molecular Biology in the Post-genomic Era*. 2009, Norfolk, UK: Caister Academic Press.
287. Coulombe, R., et al., *Structure of human procathepsin L reveals the molecular basis of inhibition by the prosegment*. *Embo Journal*, 1996. **15**(20): p. 5492-5503.
288. Olsen, J.G., et al., *Tetrameric dipeptidyl peptidase I directs substrate specificity by use of the residual pro-part domain*. *Febs Letters*, 2001. **506**(3): p. 201-206.
289. Cygler, M., et al., *Structure of rat procathepsin B: Model for inhibition of cysteine protease activity by the proregion*. *Structure*, 1996. **4**(4): p. 405-416.
290. Groves, M.R., et al., *The prosequence of procaricain forms an alpha-helical domain that prevents access to the substrate-binding cleft*. *Structure*, 1996. **4**(10): p. 1193-1203.
291. Meno, K., et al., *The crystal structure of recombinant proDer p 1, a major house dust mite proteolytic allergen*. *Journal of Immunology*, 2005. **175**(6): p. 3835-3845.
292. Wang, S.X., et al., *Structural basis for unique mechanisms of folding and hemoglobin binding by a malarial protease*. *Proceedings of the National Academy of Sciences of the United States of America*, 2006. **103**(31): p. 11503-11508.
293. Kaulmann, G., et al., *The crystal structure of a Cys25 -> Ala mutant of human procathepsin S elucidates enzyme-prosequence interactions*. *Protein Science*, 2006. **15**(11): p. 2619-2629.

294. Stack, C.M., et al., *Structural and functional relationships in the virulence-associated cathepsin L proteases of the parasitic liver fluke, Fasciola hepatica*. Journal of Biological Chemistry, 2008. **283**(15): p. 9896-9908.
295. Podobnik, M., et al., *Crystal structure of the wild-type human procathepsin B at 2.5 angstrom resolution reveals the native active site of a papain-like cysteine protease zymogen*. Journal of Molecular Biology, 1997. **271**(5): p. 774-788.
296. Sivaraman, J., et al., *Crystal structure of wild-type human procathepsin K*. Protein Science, 1999. **8**(2): p. 283-290.
297. Laskowski, M. and I. Kato, *Protein inhibitors of proteases* Annual Review of Biochemistry, 1980. **49**: p. 593-626.
298. Laurie, A.T.R. and R.M. Jackson, *Methods for the prediction of protein-ligand binding sites for Structure-Based Drug Design and virtual ligand screening*. Current Protein & Peptide Science, 2006. **7**(5): p. 395-406.
299. Janciauskiene, S., *Conformational properties of serine proteinase inhibitors (serpins) confer multiple pathophysiological roles*. Biochimica Et Biophysica Acta-Molecular Basis of Disease, 2001. **1535**(3): p. 221-235.
300. Huang, L. and C.H. Chen, *Proteasome Regulators: Activators and Inhibitors*. Current Medicinal Chemistry, 2009. **16**(8): p. 931-939.
301. Njoroge, F.G., et al., *Challenges in modern drug discovery: A case study of boceprevir, an HCV protease inhibitor for the treatment of hepatitis C virus infection*. Accounts of Chemical Research, 2008. **41**(1): p. 50-59.
302. Fear, G., S. Komarnytsky, and I. Raskin, *Protease inhibitors and their peptidomimetic derivatives as potential drugs*. Pharmacology & Therapeutics, 2007. **113**(2): p. 354-368.
303. Katunuma, N., et al., *Structure based development of novel specific inhibitors for cathepsin L and cathepsin S in vitro and in vivo*. Febs Letters, 1999. **458**(1): p. 6-10.
304. Tanrikulu, Y. and G. Schneider, *Pseudoreceptor models in drug design: bridging ligand- and receptor-based virtual screening*. Nature Reviews Drug Discovery, 2008. **7**(8): p. 667-677.
305. Stafford-Smith, M., et al., *The Association of ϵ -Aminocaproic Acid with Postoperative Decrease in Creatinine Clearance in 1502 Coronary Bypass Patients*. Anesthesia & Analgesia, 2000. **91**(5): p. 1085-1090.
306. listed, N.a., *Amicar and hyperfibrinolysis*. The medical letter on drugs and therapeutics, 1967. **9**(3): p. 10-1.
307. Tanizawa, K., S. Ishii, and Y. Kanaoka, *Proteolytic enzymes. 6. Aromatic amidines as competitive inhibitors of trypsin* The Journal of biochemistry, 1971. **69**(5): p. 893.
308. Beynon, R.J.a.B., J.S., *Proteolytic Enzymes: A Practical Approach*. 1990.
309. Hial, V. and C. Diniz, *Purification and properties of a human urinary kallikrein (kininogenase)* Biochemistry, 1974. **13**(21): p. 4311-4318.
310. ENSINCK, J.W., et al., *Use of Benzamidine as a Proteolytic Inhibitor in the Radioimmunoassay of Glucagon in Plasma*. The Journal of Clinical Endocrinology & Metabolism 1972. **35**(3): p. 463-467.
311. Markwardt, F., J. Hoffmann, and E. Körbs, *The influence of synthetic thrombin inhibitors on the thrombin-antithrombin reaction*. Thrombosis Research, 1973. **2**(4): p. 343-348.

312. Markwardt, F., J. Drawert, and P. Walsmann, *Synthetic low molecular weight inhibitors of serum kallikrein*. *Biochemical Pharmacology*, 1974. **23**(16): p. 2247-2256.
313. Lawson, W., et al., *Studies on the inhibition of human thrombin effects of plasma and plasma constituents* *Folia Haematologica Inter. Magazin fuer klinische und experimentelle* 1982. **109**(1): p. 52-60.
314. Tyner, J., et al., *High-throughput sequencing screen reveals novel, transforming RAS mutations in myeloid leukemia patients*. *Blood*, 2009. **113**(8): p. 1749-1755.
315. Simon, S., et al., *Rapid typing of transmissible spongiform encephalopathy strains with differential ELISA*. *Emerging infectious diseases*, 2008. **14**(4): p. 608-616.
316. Shull, E.R.P., et al., *Differential DNA damage signaling accounts for distinct neural apoptotic responses in ATLD and NBS*. *Genes & Development*, 2009. **23**(2): p. 171-180.
317. Tomkinson, N.P., I.J. Galpin, and R.J. Beynon, *Synthetic analogs of chymostatin-Inhibition of chymotrypsin and Streptomyces griseus proteinase-A* *S. Biochemical Journal*, 1992. **286**: p. 475-480.
318. Feinstein, G., C. Malemud, and A. Janoff, *Inhibition of human leukocyte elastase and chymotrypsin-like protease by elastatinal and chymostatin* *Biochimica et biophysica acta*, 1976. **429**(3): p. 925-932.
319. Hopgood, M.F., S.E. Knowles, and F.J. Ballard, *Proteolysis of N-ethylmaleimide-modified aldolase loaded into erythrocyte-ghosts-prevention by inhibitors of calpain* *Biochemical Journal*, 1989. **259**(1): p. 237-242.
320. Mason, R.W., *Characterization of the active-site of human multicatalytic proteinase* *Biochemical Journal*, 1990. **265**(2): p. 479-484.
321. Ortiznavarrete, V., et al., *Subunit of the 20S proteasome (multicatalytic proteinase) encoded by the major histocompatibility complex* *Nature*, 1991. **353**(6345): p. 662-664.
322. Rivett, A.J., *The multicatalytic proteinase of mammalian cells* *Archives of Biochemistry and Biophysics*, 1989. **268**(1): p. 1-8.
323. Ascenzi, P., et al., *The bovine basic pancreatic trypsin inhibitor (Kunitz inhibitor: A milestone protein*. *Current Protein & Peptide Science*, 2003. **4**(3): p. 231-251.
324. Kunitz, M. and J.H. Northrop, *Isolation from beef pancreas of crystalline trypsinogen, trypsin, a trypsin inhibitor, and an inhibitor-trypsin compound* *The Journal of General Physiology*, 1936. **19**(6): p. 991-1007.
325. Blow, D.M., J. Janin, and R.M. Sweet, *Mode of action of soybean trypsin inhibitor (Kunitz) as a model for specific protein-protein interactions*. *Nature*, 1974. **249**(5452): p. 54-57.
326. Silverman, G.A., et al., *The serpins are an expanding superfamily of structurally similar but functionally diverse proteins - Evolution, mechanism of inhibition, novel functions, and a revised nomenclature*. *Journal of Biological Chemistry*, 2001. **276**(36): p. 33293-33296.
327. Rawlings, N.D., D.P. Tolle, and A.J. Barrett, *Evolutionary families of peptidase inhibitors*. *Biochemical Journal*, 2004. **378**(3): p. 705-716.
328. Kim, S.H., et al., *Comparative-study on amino acid sequences of Kunitz-type soybean trypsin-inhibitors, TIA, TIB, and TIC*. *Journal of Biochemistry*, 1985. **98**(2): p. 435-448.
329. Umezawa, H., *Structures and activities of protease inhibitors of microbial origin*. *Methods in Enzymology*, 1976. **45**: p. 678-95.

330. Kondo, S.I., et al., *Isolation and characterization of leupeptins produced by Actinomycetes*. Chemical & pharmaceutical bulletin, 1969. **17**(9): p. 1896-901.
331. Kuramochi, H., H. Nakata, and S. Ishii, *Mechanism of association of a specific aldehyde inhibitors, leupeptin, with bovine trypsin* The Journal of biochemistry, 1979. **86**(5): p. 1403-1410.
332. Aoyagi, T., et al., *Biological activities of leupeptins* Journal of Antibiotics, 1969. **22**(11): p. 558-&.
333. Aoyagi, T., et al., *Leupeptins new protease inhibitors from actinomycetes* Journal of Antibiotics, 1969. **22**(6): p. 283-&.
334. Suda, H., et al., *Structure of Bestatin* The Journal of Antibiotics, 1976. **29**(1): p. 100-101.
335. Umezawa, H., et al., *Bestatin, an inhibitor of aminopeptidase-B, produced by actinomycetes* The Journal of Antibiotics, 1976. **29**(1): p. 97-99.
336. Hirayama, Y., et al., *Chemotherapy with ubenimex corresponding to patient age and organ disorder for 18 cases of acute myelogeneous leukemia in elderly patients--effects, complications and long-term survival*. Gan To Kagaku Ryoho, 2003. **30**(8): p. 1113-8.
337. Umezawa, H., *Low-molecular-weight enzyme inhibitors of microbial origin* Annual Review of Microbiology, 1982. **36**: p. 75-99.
338. Vaccari, M., et al., *Effects of the protease inhibitor antipain on cell malignant transformation*. Anticancer Research, 1999. **19**(1A): p. 589-596.
339. Lee, B.R., et al., *Aorsin, a novel serine proteinase with trypsin-like specificity at acidic pH*. Biochemical Journal, 2003. **371**: p. 541-548.
340. Aoki, H., et al., *Purification and characterization of collagenolytic proteases from the hepatopancreas of Northern shrimp (Pandalus eous)*. Journal of Agricultural and Food Chemistry, 2003. **51**(3): p. 777-783.
341. Sreedharan, S.K., et al., *Demonstration that 1-trans-epoxysuccinyl-L-leucylamido-(4-guanidino)butane (E-64) is one of the most effective low M(r) inhibitors of trypsin-catalysed hydrolysis. Characterization by kinetic analysis and by energy minimization and molecular dynamics simulation of the E-64-beta-trypsin complex*. Biochemical Journal, 1996. **316**: p. 777-786.
342. Katunuma, N. and E. Kominami, *Structure, properties, mechanisms, and assays of cysteine protease inhibitors - cystatins and E-64 derivatives* Biothiols, Pt A, 1995. **251**: p. 382-397.
343. Marciniszyn, J., J. Hartsuck, and J. Tang, *Pepstatin inhibition mechanism*. Advances in Experimental Medicine and Biology, 1977. **95**: p. 199-210.
344. Umezawa, H., et al., *Pepstatin, a new pepsin inhibitor produced by actinomycetes* Journal of Antibiotics, 1970. **23**(5): p. 259-&.
345. Li, Y., et al., *Presenilin 1 is linked with gamma-secretase activity in the detergent solubilized state*. Proceedings of the National Academy of Sciences of the United States of America, 2000. **97**(11): p. 6138-6143.
346. Katoh, I., et al., *Inhibition of retroviral protease activity by an aspartyl proteinase-inhibitor* Nature, 1987. **329**(6140): p. 654-656.
347. Arima, K., et al., *Isolation and characterization of a serine protease from the sprouts of Pleioblastus hindsi Nakai*. Phytochemistry, 2000. **54**(6): p. 559-565.
348. Suda, H., et al., *Thermolysin inhibitor produced by actinomycetes- Phosphoramidon* The Journal of Antibiotics, 1973. **26**(10): p. 621-623.

349. Komiyama, T., et al., *Studies on inhibitory effect of phosphoramidon and its analogs on thermolysin*. Archives of Biochemistry and Biophysics, 1975. **171**(2): p. 727-731.
350. H. Zollner, E., *Handbook of Enzyme Inhibitors*, in *Handbook of Enzyme Inhibitors*. 1993, VCH Verlagsgesellschaft, Weinheim. p. 551.
351. Bithell, T.C., *Study of inhibitory effect of ethylenediamine tetra-acetic acid on thrombin-fibrinogen reaction* Biochemical Journal, 1964. **93**(2): p. 431-&.
352. Cremona, T. and T.P. Singer, *Mechanism of reversible inhibition of Zn-flavoproteins by chelators- Chelation or resolution* Biochimica et biophysica acta, 1962. **57**(2): p. 412-&.
353. Banfi, G., G.L. Salvagno, and G. Lippi, *The role of ethylenedimine tetraacetic acid (EDTA) as in vitro anticoagulant for diagnostic purposes*. Clinical Chemistry and Laboratory Medicine, 2007. **45**(5): p. 565-576.
354. Shrihari, J.S., et al., *Role of EDTA chelation therapy in cardiovascular diseases*. National Medical Journal of India, 2006. **19**(1): p. 24-26.
355. Knight, C., *Human Cathepsin-B -Application of the substrate N-benzyloxycarbonyl-L-arginyl-L-arginine 2-naphthylamide to a study of the inhibition by leupeptin* The Biochemical journal, 1980. **189**(3): p. 447-453.
356. Matsuishi, M., et al., *Purification and some properties of cathepsin H from rabbit skeletal muscle*. International Journal of Biochemistry & Cell Biology, 2003. **35**(4): p. 474-485.
357. Nishimura, Y., et al., *Inhibitory effect of leupeptin on the intracellular maturation of lysosomal cathepsin-L in primary cultures of rat hepatocytes* Biological & pharmaceutical bulletin, 1995. **18**(7): p. 945-950.
358. Ning, M.C.Y.J. and R.J. Beynon, *Preparation and purification of the proteinase inhibitor, leupeptin, from culture filtrates of Streptomyces lavendulae*. International Journal of Biochemistry, 1986. **18**(9): p. 813-820.
359. Barrett, A.J., et al., *L-trans-epoxysuccinyl-leucylamido (4-guanidino)butane (E-64) and its analogs as inhibitors of cysteine proteinases including cathepsins B, H and L* Biochemical Journal, 1982. **201**(1): p. 189-198.
360. Goffin, C. and J.M. Ghuysen, *Multimodular penicillin binding proteins: An enigmatic family of orthologs and paralogs*. Microbiology and Molecular Biology Reviews, 1998. **62**(4): p. 1079-+.
361. Bar-On, P., et al., *Kinetic and structural studies on the interaction of cholinesterases with the anti-Alzheimer drug rivastigmine*. Biochemistry, 2002. **41**(11): p. 3555-3564.
362. Zhou, S.F., et al., *Mechanism-based inhibition of cytochrome P450 3A4 by therapeutic drugs*. Clinical Pharmacokinetics, 2005. **44**(3): p. 279-304.
363. Potashman, M.H. and M.E. Duggan, *Covalent Modifiers: An Orthogonal Approach to Drug Design*. Journal of medicinal chemistry, 2009. **52**(5): p. 1231-1246.
364. Shimada, K., et al., *Poly- (L) -malic acid - A new protease inhibitor from Penicillium cyclopium* Biochemical and Biophysical Research Communications, 1969. **35**(5): p. 619-&.
365. Hanada, K., et al., *Structure and synthesis of E-64, a new thiol protease inhibitor* Agricultural and biological chemistry, 1978. **42**(3): p. 529-536.
366. Hanada, K., et al., *Inhibitory activities of E-64 derivatives on papain*. Agricultural and biological chemistry, 1978. **42**(3): p. 537-541.
367. Matsumoto, K., et al., *Structural basis of inhibition of cysteine proteases by E-64 and its derivatives*. Biopolymers, 1999. **51**(1): p. 99-107.

368. Varughese, K.I., et al., *Crystal structure of a papain E-64 complex* Biochemistry, 1989. **28**(3): p. 1330-1332.
369. Zhao, B.G., et al., *Crystal structure of human osteoclast cathepsin K complex with E-64*. Nature Structural Biology, 1997. **4**(2): p. 109-111.
370. Turk, D., et al., *Crystal structure of Cathepsin-B inhibited with CA030 at 2.0-angstrom resolution- a basis for the design of specific epoxysuccinyl inhibitors* Biochemistry, 1995. **34**(14): p. 4791-4797.
371. Gavira, J.A., et al., *Structure of the mexicain-E-64 complex and comparison with other cysteine proteases of the papain family*. Acta Crystallographica Section D-Biological Crystallography, 2007. **63**: p. 555-563.
372. Ghosh, R., et al., *Structural insights into the substrate specificity and activity of ervatamins, the papain-like cysteine proteases from a tropical plant, Ervatamia coronaria*. FEBS Journal, 2008. **275**(3): p. 421-434.
373. Varughese, K.I., et al., *Crystal structure of an actinidin E-64 complex* Biochemistry, 1992. **31**(22): p. 5172-5176.
374. Schaschke, N., et al., *E-64 analogues as inhibitors of cathepsin B. On the role of the absolute configuration of the epoxysuccinyl group*. Bioorganic & Medicinal Chemistry, 1997. **5**(9): p. 1789-1797.
375. Schaschke, N., et al., *Epoxysuccinyl peptide-derived affinity labels for cathepsin B*. Febs Letters, 2000. **482**(1-2): p. 91-96.
376. Turk, D. and G. Guncar, *Lysosomal cysteine proteases (cathepsins): promising drug targets*. Acta Crystallographica Section D-Biological Crystallography, 2003. **59**: p. 203-213.
377. Turk, D., B. Turk, and V. Turk, *Papain-like lysosomal cysteine proteases and their inhibitors: drug discovery targets?* Biochemical Society Symposia, 2003(70): p. 15-30.
378. Summer, R., *Characterisation of a protease activity associated with maturation of the S-layer of Clostridium difficile* 2007.
379. van Roosmalen, M.L., et al., *Type I signal peptidases of gram-positive bacteria*. Biochimica Et Biophysica Acta-Molecular Cell Research, 2004. **1694**(1-3): p. 279-297.
380. Scott, J.R. and D. Zahner, *Pili with strong attachments: Gram-positive bacteria do it differently*. Molecular Microbiology, 2006. **62**(2): p. 320-330.
381. Bruser, T., *The twin-arginine translocation system and its capability for protein secretion in biotechnological protein production*. Applied Microbiology and Biotechnology, 2007. **76**(1): p. 35-45.
382. Czaplewski, C., et al., *Binding modes of a new epoxysuccinyl-peptide inhibitor of cysteine proteases. Where and how do cysteine proteases express their selectivity?* Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology, 1999. **1431**(2): p. 290-305.
383. Mori, K. and H. Iwasawa, *Pheromone synthesis. 35. Stereoselective synthesis of optically-active forms of delta-multistriatin, the attractant for european populations of the smaller european elm bark beetle* Tetrahedron, 1980. **36**(1): p. 87-90.
384. Meara, J.P. and D.H. Rich, *Mechanistic studies on the inactivation of papain by epoxysuccinyl inhibitors*. Journal of medicinal chemistry, 1996. **39**(17): p. 3357-3366.
385. Shi, G.P., et al., *Molecular-cloning and expression of human alveolar macrophage Cathepsin-S, an elastinolytic cysteine protease* Journal of Biological Chemistry, 1992. **267**(11): p. 7258-7262.

386. Murata, M., et al., *Novel epoxysuccinyl peptides-selective inhibitors of cathepsin-B, in vitro* Febs Letters, 1991. **280**(2): p. 307-310.
387. Towatari, T., et al., *Novel epoxysuccinyl peptides- A selective inhibitor of Cathepsin-B in vivo* Febs Letters, 1991. **280**(2): p. 311-315.
388. Tamai, M., et al., *Study on thiol protease inhibitors. 4. Relationship between structure and papain inhibitory activity of epoxysuccinyl amino-acid derivates* Agricultural and biological chemistry, 1981. **45**(3): p. 675-679.
389. Korn, A., S. Rudolphbohner, and L. Moroder, *A convenient synthesis of convenient synthesis of optically pure (2R, 3R)-2,3-Epoxysuccinyl-dipeptides* Tetrahedron, 1994. **50**(28): p. 8381-8392.
390. Sadaghiani, A.M., S.H.L. Verhelst, and M. Bogyo, *Solid-phase methods for the preparation of epoxysuccinate-based inhibitors of cysteine proteases*. Journal of Combinatorial Chemistry, 2006. **8**(6): p. 802-804.
391. McCoy, E.C., et al., *Superficial antigens of campylobacter (vibrio) fetus- Characterization of an anti-phagocytic component* Infection and Immunity, 1975. **11**(3): p. 517-525.
392. Jonquieres, R., et al., *Interaction between the protein InlB of Listeria monocytogenes and lipoteichoic acid: a novel mechanism of protein association at the surface of Gram-positive bacteria*. Molecular Microbiology, 1999. **34**(5): p. 902-914.
393. Deneve, C., et al., *Effects of Subinhibitory Concentrations of Antibiotics on Colonization Factor Expression by Moxifloxacin-Susceptible and Moxifloxacin-Resistant Clostridium difficile Strains*. Antimicrobial Agents and Chemotherapy, 2009. **53**(12).
394. Deneve, C., et al., *Antibiotics involved in Clostridium difficile-associated disease increase colonization factor gene expression*. Journal of Medical Microbiology, 2008. **57**(6): p. 732-738.
395. Lawley, T.D., et al., *Proteomic and Genomic Characterization of Highly Infectious Clostridium difficile 630 Spores*. Journal of Bacteriology, 2009. **191**(17): p. 5377-5386.
396. Wright, M., *N-Myristoyltransferase in Disease: Probing Myristoylation by Chemical Proteomics*. 2009.
397. Chan, T.R., et al., *Polytriazoles as copper(I)-stabilizing ligands in catalysis*. Organic Letters, 2004. **6**(17): p. 2853-2855.
398. Donnelly, P.S., et al., *'Click' cycloaddition catalysts: copper(I) and copper(II) tris(triazolylmethyl)amine complexes*. Chemical Communications, 2008(21): p. 2459-2461.
399. Albert, B.I. and K. Koide, *How rapidly do Epoxides nonspecifically form covalent bonds with Thiols in water?* Chembiochem, 2007. **8**: p. 1912-+.
400. MacKinnon, A.L. and J. Taunton, *Target Identification by Diazirine Photo-Cross-Linking and Click Chemistry*. Current Protocols in Chemical Biology. 2009: John Wiley & Sons, Inc.
401. Dubowchik, G.M. and S. Radia, *Monomethoxytrityl (MMT) as a versatile amino protecting group for complex prodrugs of anticancer compounds sensitive to strong acids, bases and nucleophiles*. Tetrahedron Letters, 1997. **38**(30): p. 5257-5260.
402. Invitrogen, *Dynabeads MyOne Streptavidin C1*. 2010.
403. Perkins, D.N., et al., *Probability-based protein identification by searching sequence databases using mass spectrometry data*. Electrophoresis, 1999. **20**(18): p. 3551-3567.

404. Lahm, H.W. and H. Langen, *Mass spectrometry: A tool for the identification of proteins separated by gels*. Electrophoresis 2000. **21**(11): p. 2105-2114.
405. Graves, P.R. and T.A.J. Haystead, *A functional proteomics approach to signal transduction*. Recent Progress in Hormone Research, 2003. **58**(1): p. 1-24.
406. Han, X., A. Aslanian, and J. Yates, *Mass spectrometry for proteomics*. Current opinion in chemical biology, 2008. **12**(5): p. 483-490.
407. Lim, H., et al., *Identification of 2D-gel proteins: A comparison of MALDI/TOF peptide mass mapping to [mu] LC-ESI tandem mass spectrometry*. Journal of the American Society for Mass Spectrometry, 2003. **14**(9): p. 957-970.
408. Keller, A., et al., *Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search*. Analytical Chemistry, 2002. **74**(20): p. 5383-5392.
409. Scheler, C., et al., *Peptide mass fingerprint sequence coverage from differently stained proteins on two-dimensional electrophoresis patterns by matrix assisted laser desorption/ionization mass spectrometry (MALDI-MS)*. Electrophoresis 1998. **19**(6): p. 918-927.
410. Janoir, C., et al., *Cwp84, a surface-associated protein of Clostridium difficile, is a cysteine protease with degrading activity on extracellular matrix proteins*. Journal of Bacteriology, 2007. **189**: p. 7174-7180.
411. Kirby, J.M., et al., *Cwp84, a Surface-associated Cysteine Protease, Plays a Role in the Maturation of the Surface Layer of Clostridium difficile*. Journal of Biological Chemistry, 2009. **284**(50): p. 34666-34673.
412. Elias, J.E., et al., *Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations*. Nat Meth, 2005. **2**(9): p. 667-675.
413. Kirby, J.M., et al., *Cwp84, a surface-associated cysteine protease, plays a role in the maturation of the surface layer of Clostridium difficile*. Vol. 284. 2009. 34666-73.
414. Cravatt, B., A. Wright, and J. Kozarich, *Activity-based protein profiling: From enzyme chemistry*. Annual Review of Biochemistry, 2008. **77**: p. 383-414.
415. Leung, D., G. Abbenante, and D.P. Fairlie, *Protease inhibitors: Current status and future prospects*. Journal of medicinal chemistry, 2000. **43**(3): p. 305-341.
416. Schoellmann, G. and E. Shaw, *Direct evidence for presence of histidine in active center of chymotrypsin* Biochemistry, 1963. **2**(2): p. 252-&.
417. Song, Q.Z., et al., *Interleukin-1 beta-converting enzyme-like protease cleaves DNA-dependent protein kinase in cytotoxic T cell killing*. Journal of Experimental Medicine, 1996. **184**(2): p. 619-626.
418. Piguet, P.F., et al., *TNF-induced enterocyte apoptosis and detachment in mice: Induction of caspases and prevention by a caspase inhibitor, ZVAD-fmk*. Laboratory Investigation, 1999. **79**(4): p. 495-500.
419. Rauber, P., et al., *The synthesis of peptidylfluoromethanes and their properties as inhibitors of serine proteases and cysteine proteases* Biochemical Journal, 1986. **239**(3): p. 633-640.
420. Poulos, T.L., et al., *Polypeptide halomethyl ketones bind to serine proteases as analogs of tetrahedral intermediate- X-Ray crystallographic comparison of lysine-polypeptide and phenylalanine- Polypeptide chloromethyl ketone- inhibited subtilisin* Journal of Biological Chemistry, 1976. **251**(4): p. 1097-1103.
421. Towatari, T., et al., *Purification and properties of a new cathepsin from rat-liver* Journal of Biochemistry, 1978. **84**(3): p. 659-671.

422. Gamcsik, M.P., et al., *Structure and stereochemistry of tetrahedral inhibitor complexes of papain by direct NMR observation* Journal of the American Chemical Society, 1983. **105**(20): p. 6324-6325.
423. Drenth, J., K.H. Kalk, and H.M. Swen, *Binding of chloromethyl ketone substrate analogs to crystalline papain* Biochemistry, 1976. **15**(17): p. 3731-3738.
424. Shaw, E., *CysteinyI proteinases and their selective inactivation* Advances in Enzymology and Related Areas of Molecular Biology, 1990. **63**: p. 271-347.
425. Abrahamson, M., et al., *Human cystatin-C role of the N-terminal segment in the inhibition of human cysteine proteases and in its inactivation by leukocyte elastase* Biochemical Journal, 1991. **273**: p. 621-626.
426. Green, G.D.J. and E. Shaw, *Peptidyl diazomethyl ketones are specific inactivators of thiol proteinases* Journal of Biological Chemistry, 1981. **256**: p. 1923-1928.
427. Buttle, D.J. and J. Saklatvala, *Lysosomal cysteine endopeptidases mediate interleukin-1-stimulated cartilage proteoglycan degradation* Biochemical Journal, 1992. **287**: p. 657-661.
428. Kirschke, H., P. Wikstrom, and E. Shaw, *Active center differences between cathepsins L and B: The S1 binding region*. Febs Letters, 1988. **228**(1): p. 128-130.
429. Crawford, C., et al., *The design of peptidyl diazomethane inhibitors to distinguish between the cysteine proteinases calpain-II, cathepsin-L and cathepsin-B* Biochemical Journal, 1988. **253**(3): p. 751-758.
430. Krantz, A., et al., *Peptidyl (acyloxy) methyl ketones and the quiescent affinity label concept- The departing group as a variable structural element in the design of inactivators of cysteine proteases* Biochemistry, 1991. **30**(19): p. 4678-4687.
431. Wagner, B.M., et al., *In-vivo inhibition of Cathepsin-B by peptidyl (acyloxy)methyl ketones*. Journal of medicinal chemistry, 1994. **37**(12): p. 1833-1840.
432. Bromme, D., et al., *Potent inactivation of Cathepsin-S and Cathepsin-L by peptidyl (acyloxy)methyl ketones* Biological Chemistry Hoppe-Seyler, 1994. **375**(5): p. 343-347.
433. Brady, K., *Bimodal inhibition of caspase-1 by aryloxymethyl and acyloxymethyl ketones*. Biochemistry, 1998. **37**(23): p. 8508-8515.
434. Krantz, A., *Peptidyl (acyloxy)methanes as quiescent affinity labels for cysteine proteases in Proteolytic Enzymes: Serine and Cysteine Peptidases*. 1994. p. 656-671.
435. Somoza, J., et al., *Crystal structure of human cathepsin V*. Biochemistry, 2000. **39**(41): p. 12543-12551.
436. Govardhan, C.P. and R.H. Abeles, *Inactivation of cysteine proteases*. Archives of Biochemistry and Biophysics, 1996. **330**(1): p. 110-114.
437. Hanzlik, R. and S. Thompson, *Vinylogous amino-acid ester- a new class of inactivators for thiol proteases* Journal of medicinal chemistry, 1984. **27**(6): p. 711-712.
438. Thompson, S., P. Andrews, and R. Hanzlik, *Carboxyl-modified amino-acids and peptides as protease inhibitors* Journal of medicinal chemistry, 1986. **29**(1): p. 104-111.
439. Roush, W., et al., *Vinyl sulfonate esters and vinyl sulfonamides: Potent, irreversible inhibitors of cysteine proteases*. Journal of the American Chemical Society, 1998. **120**(42): p. 10994-10995.
440. Dragovich, P.S., et al., *Structure-based design, synthesis, and biological evaluation of irreversible human rhinovirus 3C protease inhibitors. 1. Michael acceptor structure-activity studies*. Journal of medicinal chemistry, 1998. **41**(15): p. 2806-2818.

441. Horn, M.A., P.F. Heinstejn, and P.S. Low, *Biotin-mediated delivery of exogenous macromolecules into soybean cells* Plant Physiology, 1990. **93**(4): p. 1492-1496.
442. Getman, D., et al., *Discovery of a novel class of potent HIV-1 protease inhibitors containing the (R)-(hydroxymethyl) urea isostere* Journal of medicinal chemistry, 1993. **36**(2): p. 288-291.
443. Wood, W.J.L., L. Huang, and J.A. Ellman, *Synthesis of a diverse library of mechanism-based cysteine protease inhibitors*. Journal of Combinatorial Chemistry, 2003. **5**(6): p. 869-880.
444. Wang, D.J., et al., *One-carbon chain extension of esters to alpha-chloroketones: A safer route without diazomethane*. Journal of Organic Chemistry, 2004. **69**(5): p. 1629-1633.
445. Hoffman, R.V., W.S. Weiner, and N. Maslouh, *Highly stereoselective synthesis of anti-N-protected-alpha-amino epoxides*. Journal of Organic Chemistry, 2001. **66**(17): p. 5790-5795.
446. Honda, Y., et al., *Synthesis of optically active alpha-aminoalkyl (alpha'-halomethyl ketone: A cross-Claisen condensation approach*. Organic Letters, 2002. **4**(3): p. 447-449.
447. Kiefer, F., et al., *The SWISS-MODEL Repository and associated resources*. Nucleic Acids Research, 2009. **37**: p. D387-D392.
448. Staub, I. and S.A. Sieber, *beta-Lactam Probes As Selective Chemical-Proteomic Tools for the Identification and Functional Characterization of Resistance Associated Enzymes in MRSA*. Journal of the American Chemical Society, 2009. **131**(17): p. 6271-6276.
449. Staub, I. and S.A. Sieber, *beta-lactams as selective chemical probes for the in vivo labeling of bacterial enzymes involved in cell wall biosynthesis, antibiotic resistance, and virulence*. Journal of the American Chemical Society, 2008. **130**(40): p. 13400-13409.
450. Puri, A.W., et al., *Rational Design of Inhibitors and Activity-Based Probes Targeting Clostridium difficile Virulence Factor TcdB*. Chemical Biology, 2010. **17**(11): p. 1201-11.
451. Dang, T.H.T., et al., *Chemical Probes of Surface Layer Biogenesis in Clostridium difficile*. Acs Chemical Biology, 2010. **5**(3): p. 279-285.
452. Emerson, J., *Transcriptional studies on Clostridium difficile and its surface proteins*. 2007.
453. Wust, J. and U. Hardegger, *Transferable resistance to clindamycin, erythromycin, and tetracycline in Clostridium difficile*. Antimicrobial Agents and Chemotherapy, 1983. **23**(5): p. 784-6.
454. Hussain, H.A., A.P. Roberts, and P. Mullany, *Generation of an erythromycin-sensitive derivative of Clostridium difficile strain 630 (630 Delta erm) and demonstration that the conjugative transposon Tn916 Delta E enters the genome of this strain at multiple sites*. Journal of Medical Microbiology, 2005. **54**(2): p. 137-141.
455. Stubbs, S.L.J., et al., *PCR targeted to the 16S-23S rRNA gene intergenic spacer region of Clostridium difficile and construction of a library consisting of 116 different PCR ribotypes*. Journal of Clinical Microbiology, 1999. **37**(2): p. 461-463.
456. Heard, S.R., et al., *The epidemiology of Clostridium difficile with use of a typing scheme: nosocomial acquisition and cross-infection among immunocompromised patients*. The Journal of Infectious Diseases 1986. **153**(1): p. 159-62.

457. Delmee, M., G. Bulliard, and G. Simon, *Application of a technique for serogrouping Clostridium difficile in an outbreak of antibiotic-associated diarrhoea*. The Journal of Infectious Diseases 1986. **13**(1): p. 5-9.
458. Delmee, M., G. Bulliard, and G. Simon, *Application of a technique for serogrouping Clostridium difficile in an outbreak of antibiotic associated diarrhea*. J. Infect. Dis., 1986. **13**(1): p. 5-9.
459. Dubreuil, J.D., et al., *Antigenic differences among Campylobacter fetus S-layer proteins*. Journal of Bacteriology, 1990. **172**(9): p. 5035-5043.
460. Nesvizhskii, A.I., et al., *A statistical model for identifying proteins by tandem mass spectrometry*. Analytical Chemistry, 2003. **75**(17): p. 4646-4658.
461. Sadaghiani, A.M., S.H.L. Verhelst, and M. Bogyo, *Solid-phase methods for the preparation of epoxysuccinate-based inhibitors of cysteine proteases*. J. Comb. Chem., 2006. **8**(6): p. 802-804.
462. Bogyo, M., et al., *Selective targeting of lysosomal cysteine proteases with radiolabeled electrophilic substrate analogs*. Chem. Biol., 2000. **7**(1): p. 27-38.

Chapter 11 Appendix

Appendix 1: Strain dependency of inhibitors/ABPs.

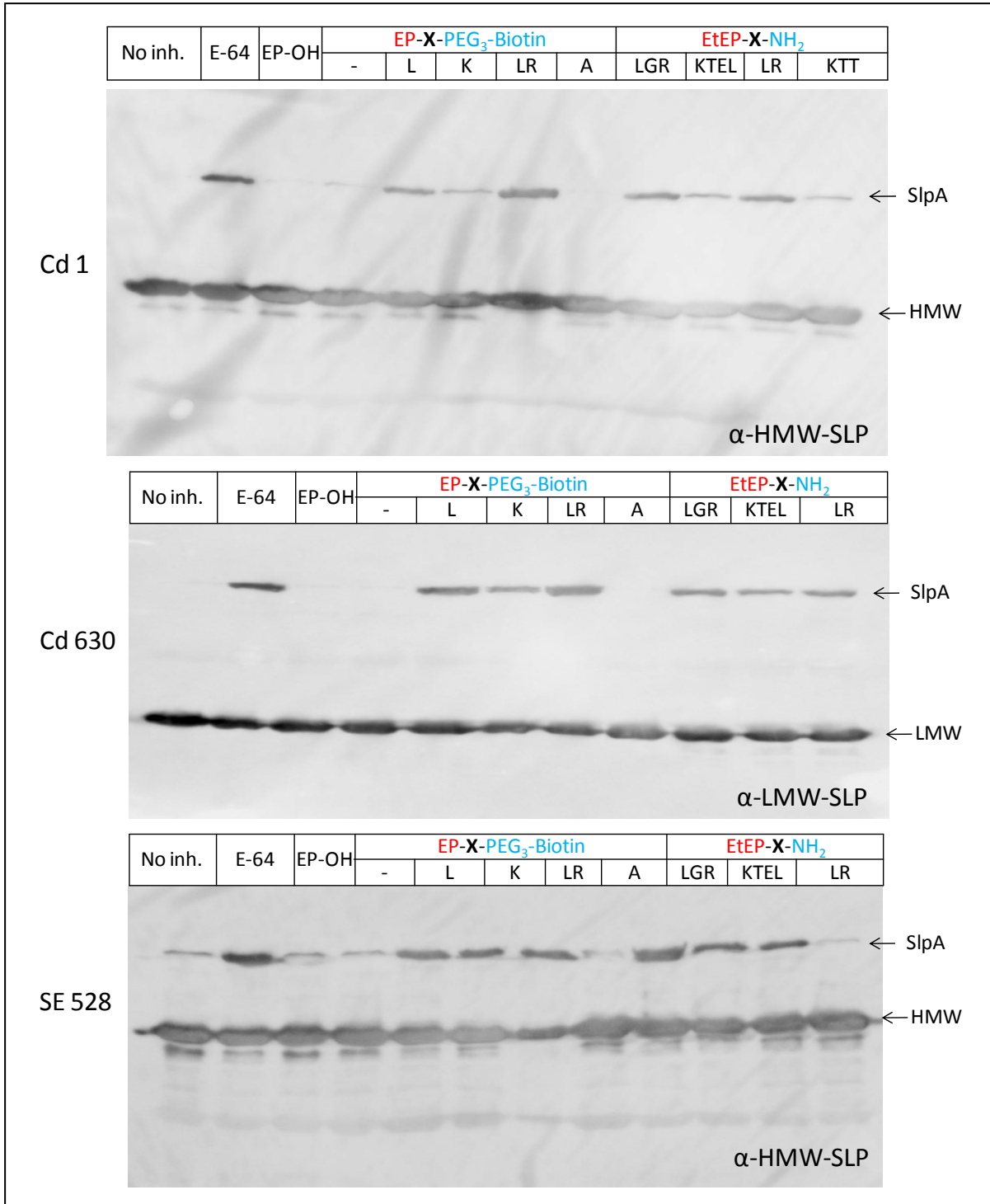


Figure 11. 1: Activity of inhibitors/ABPs, as measured by Western blot. Indicated strains were treated with inhibitors/ABPs at 250 μ M overnight. Whole cell lysis were analysed by SDS-PAGE.

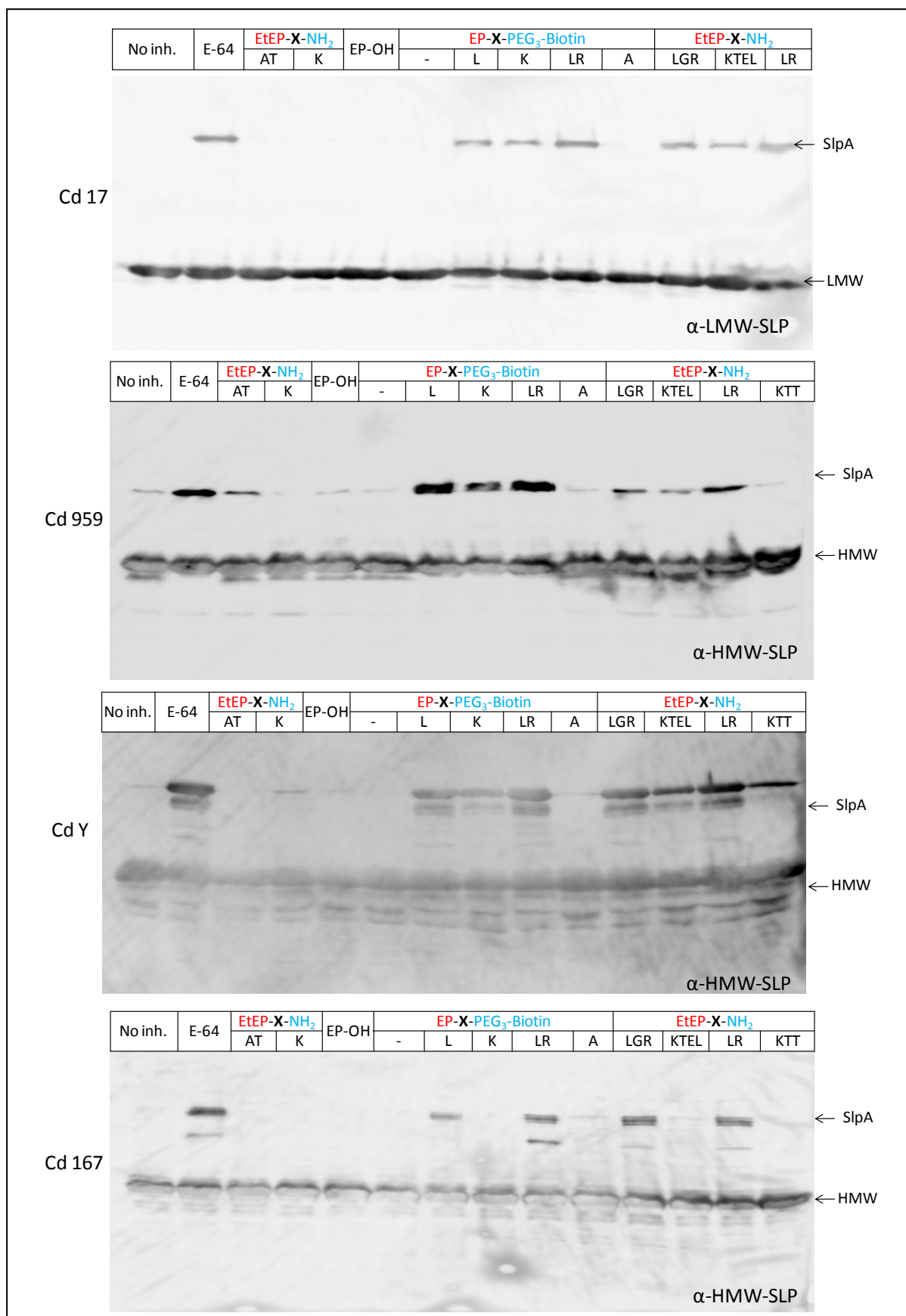


Figure 11. 2: Activity of inhibitors/ABPs, as measured by Western blot. Indicated strains were treated with inhibitors/ABPs at 250 μ M overnight. Whole cell lysis were analysed by SDS-PAGE.

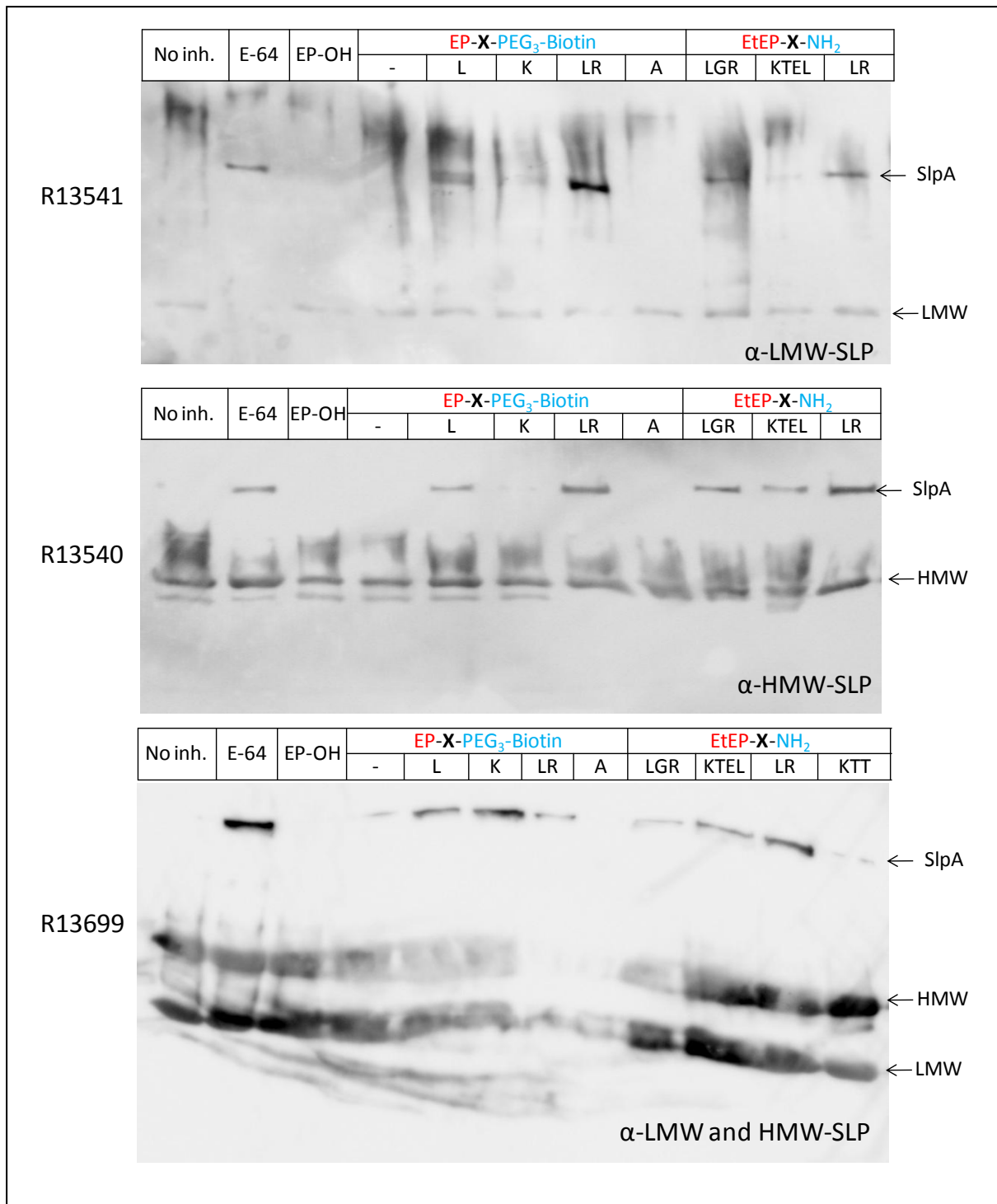


Figure 11. 3: Activity of inhibitors/ABPs, as measured by Western blot. Indicated strains were treated with inhibitors/ABPs at 250 μ M overnight. Whole cell lysis were analysed by SDS-PAGE.

Appendix 2: Proteomics data using Orbitrap LC-ESI MS/MS

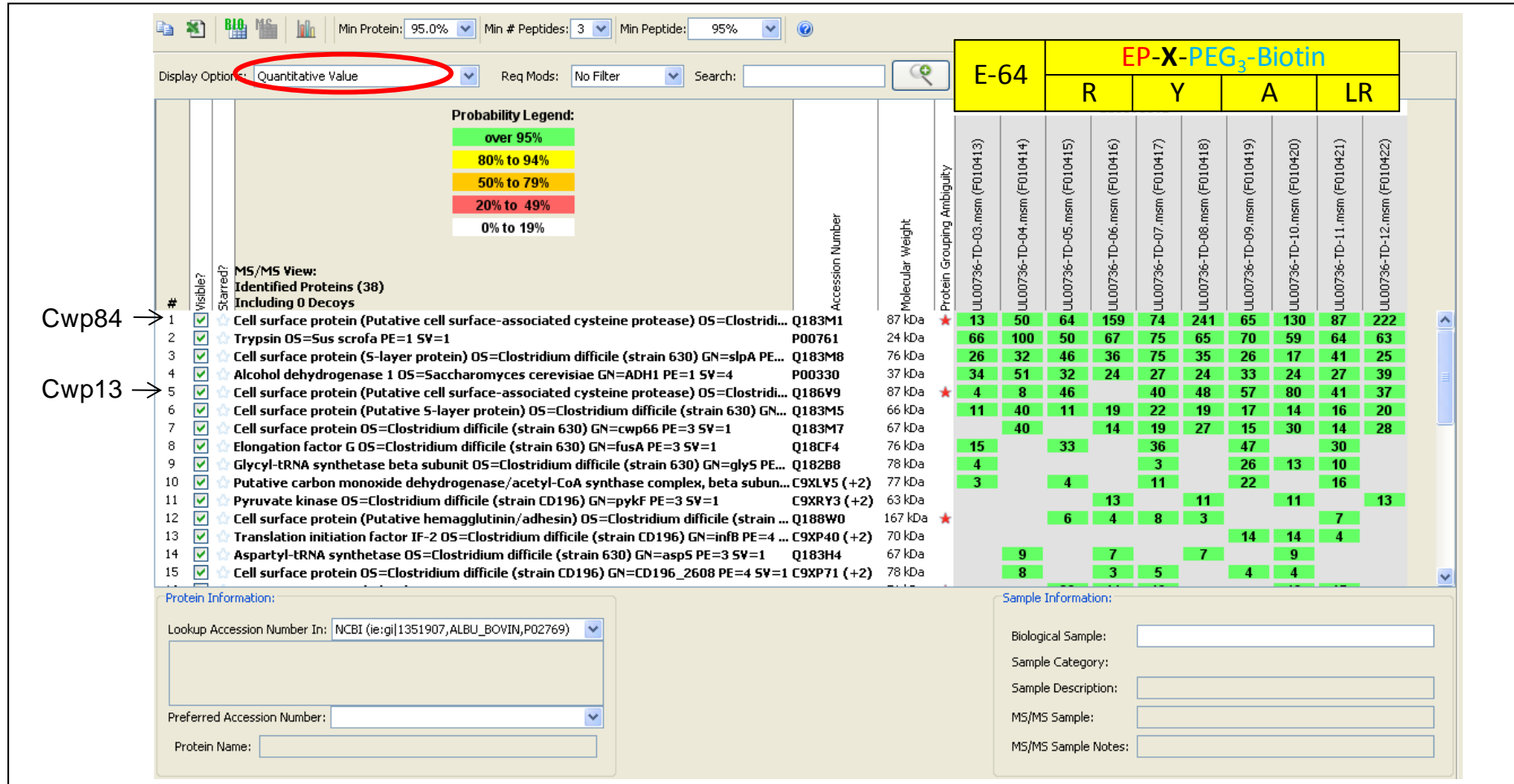


Figure 11. 4: List of identified protein of wild typ Cd 630 under the view of Scaffold software. Quantitative values are displayed.

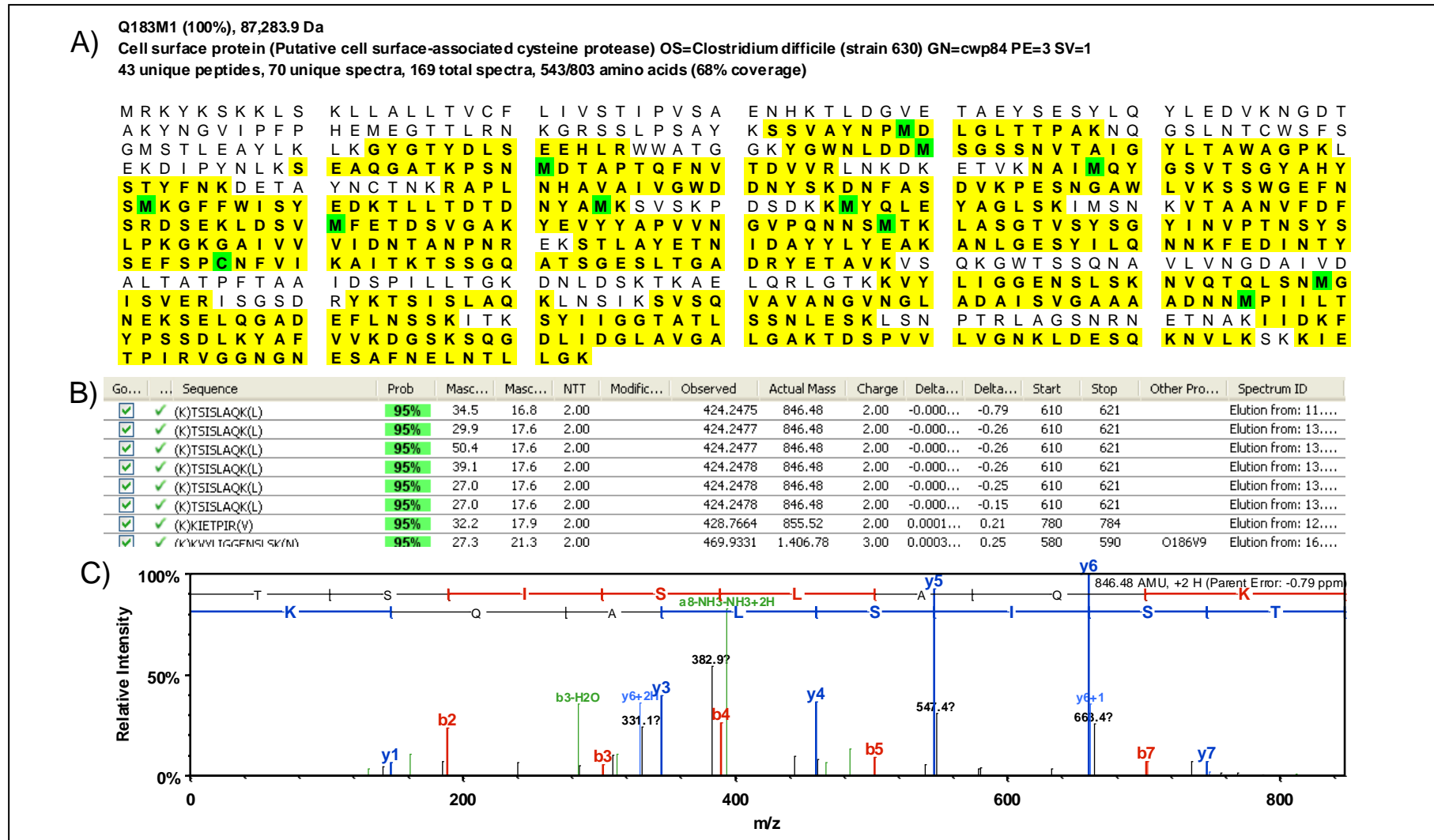


Figure 11. 5: Data for Cwp84, generated using Scaffold software. A) Cwp84 sequence, matched peptides are indicated in yellow. B) Short cut view of identified peptide list. C) Sample of mass spectra, illustrated for peptide TSISLAQK.

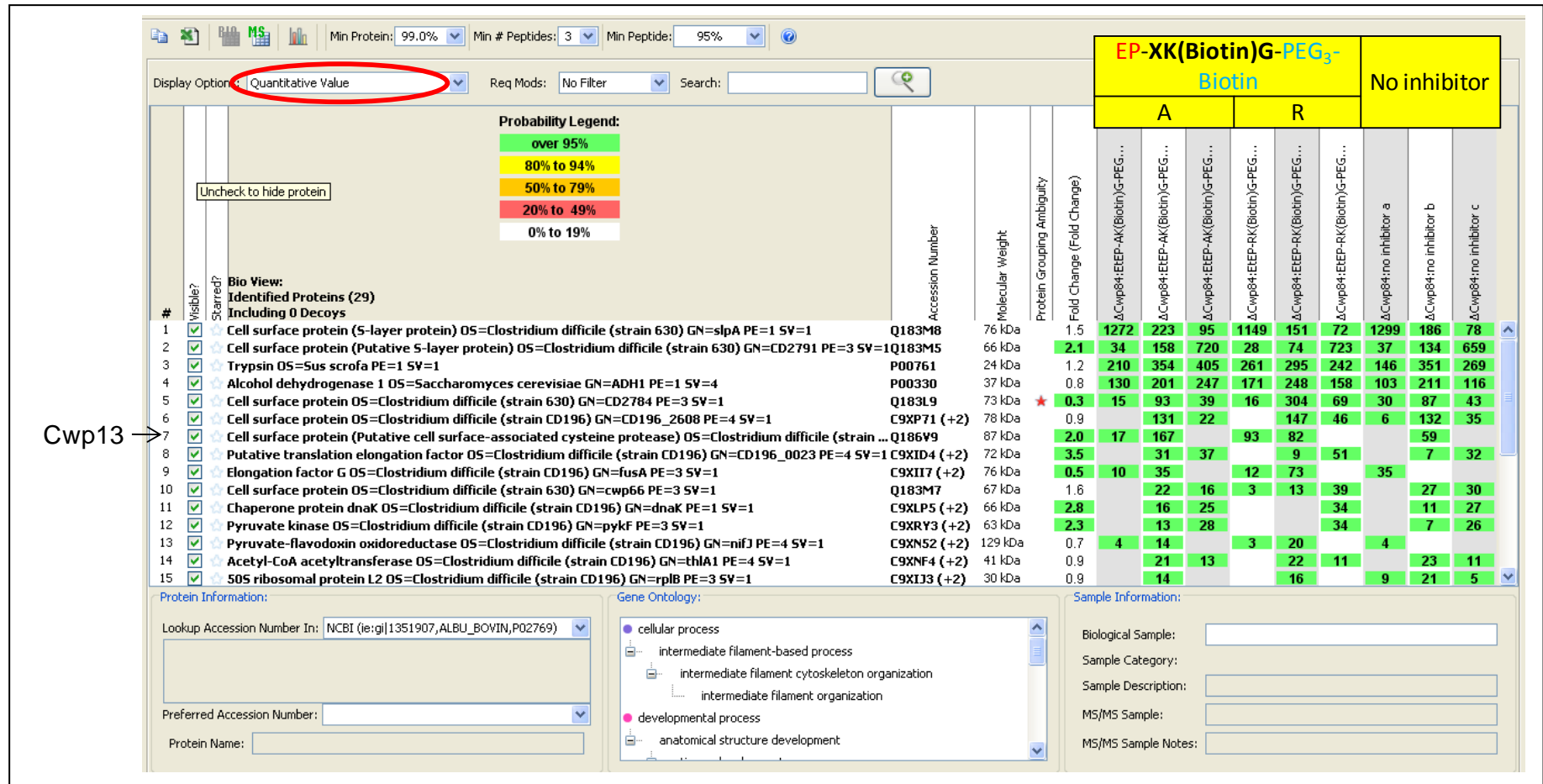


Figure 11. 7: List of identified protein of 630Δcwp84 strain under the view of Scaffold software. Quantitative values are displayed. Each samples (probes) was divided in three small samples and treated separately to enhance coverage. For evaluation the results, the values of the three small samples should be combined additively.