Structural and Mechanistic Studies of Porcine Pancreatic Elastase



James D. S. Linyard

St John's College
University of Oxford
Hilary term, 2023

A Thesis Submitted to the Board of the Faculty of Mathematical, Physical and Life Sciences of the University of Oxford in Partial Fulfilment of the Requirements for the Degree of Doctor of Philosophy

Dedicated to my wife.

Abstract

Serine peptidases (SPs) were among the earliest discovered enzymes and have since become some of the best characterised of all enzyme classes. SPs play important roles in many physiological processes including the immune response, blood clotting, and digestion. Due to the inherent destructive power of SPs, many have been implicated in disease. Despite a generally accepted SP mechanism being proposed more than 50 years ago, many of its details remain a matter of debate. One aspect that remains elusive is the presence of (a) tetrahedral intermediate(s). Such intermediates have not been observed by spectroscopic methods in a catalytic reaction between a SP and a substrate, but are inferred from studies using inhibitors that form analogous tetrahedral complexes. There are, however, several proposed tetrahedral intermediates that have been observed by crystallographic methods. The work presented herein aimed to further characterise one such intermediate and to prepare for time-resolved studies with PPE.

In **Chapter 2**, a proposed tetrahedral intermediate complex, formed between PPE and the heptapeptide inhibitor BCM7 *in crystallo* following 'pH-jump' studies, was further investigated. The pH-jumps were reproduced under novel conditions. The results from these studies suggest that the proposed tetrahedral intermediate is not, at least solely, a tetrahedral species. Furthermore, preparations were made to probe the reaction mechanism of PPE using time-resolved serial femtosecond crystallographic studies.

In **Chapter 3**, alternative inhibitors for time-resolved crystallography with PPE were investigated. Two variants of BCM7 were explored, revealing a novel crystallisation condition and space group for PPE. However, both variants were found to be unsuitable for time-resolved studies. The reaction of PPE with Sivelestat was subsequently explored. A reaction intermediate was captured *in crystallo*, supporting the proposed mechanism of reaction for Sivelestat with PPE. However, the use of Sivelestat for further time-resolved studies was ruled out.

Chapter 4 describes the optimisation of expression conditions and the purification method for recombinant ProPPE from *E. coli*. Initial attempts at expression manifested near complete insoluble expression, but reconfiguration of the gene, the use of alternate expression strains, and optimisation of post-induction incubation time and temperature resulted in improved soluble expression of ProPPE. Initial purification attempts also produced poor results. Optimisation of the protocol enabled pure and catalytically active protein, although these were not obtained in a single purification. However, a method to achieve pure and active PPE is proposed, although further optimisation is required to improve the final yield.

In **Chapter 5**, three inhibitor scaffolds were screened for inhibitory potential against PPE as a model for HNE inhibition. Five novel thiophene γ-lactam scaffolds were identified with single-digit micromolar IC_{50} values. The oxindole and benzoxaborole scaffolds were both found to be poor inhibitors of PPE with less than 50% inhibition at 200 μM. Crystallographic analyses with the benzoxaboroles rationalised the poor inhibitory potential of the scaffold, but guided suggestions of how to improve the potency of the scaffold in the future.

Overall, the body of work presented herein provides insight into a proposed intermediate of the PPE deacylation reaction and prepares the ground work for time-resolved studies. Additionally, the work lays the foundations to enable site-directed mutagenesis studies with PPE to complement time-resolved studies.

Acknowledgements

Firstly, I would like to thank my supervisor, Professor Chris Schofield for allowing me to work within his group. It has been an absolute pleasure working with Chris; his passion and scientific knowledge is unmatched. Thank you for the endless stream of fantastic ideas, but also for allowing me to pursue my own scientific interests, and for supporting the XFEL studies which allowed me, and others, to travel around the world to perform them.

A special thanks goes to Dr Patrick Rabe, my day-to-day supervisor, for teaching me nearly everything I know about research. Patrick has been a constant support throughout my DPhil, putting up with endless questions, preparing countless numbers of shipments, and sharing a lab space (and chemicals) with me, although I am sure he has already discovered that I was never the messy one. Special thanks for your support on our late nights preparing grams of enzyme, during data collections at Diamond, PAL-XFEL, and LCLS, and at the Rabe group drinks.

Thank you to Dr Stephen Marshall for ideas and analysis during the difficult protein optimisation work, to Dr Pauline Lang for help with the preparations for time-resolved crystallography, and to Dr Adam Hardy for ensuring that the lab runs smoothly.

Thank you to my second supervisor Dr Allen Orville, particularly for his enthusiasm during the aforementioned data collections, at LCLS and PAL-XFEL. Thank you also goes to everyone who contributed towards data collection at Diamond (Dr Pierre Aller, Johan Glerup, Dr Anastasiia Shilova, Dr Nicholas Devenish, Dr Tiankun Zhou, Dr Sam Horrell, Dr Robin Owen, Dr Danny Axford, and Dr Graeme Winter) and PAL-XFEL (Dr Jaehyun Park, Sehan Park, Gisu Park, and Dr Hyojung Hyun) and other researchers for making the trips enjoyable.

Thank you to the BBSRC and St John's College for funding this work.

Extra thanks to the proof-readers of this work: Chris, Patrick, Dr Klemensas Simelis, and Stephen. Their detailed feedback was well received.

A shout out goes to the friends I have made from the CJS group that have made the last four years enjoyable. Particularly 'the boys' (Dr Klem 'Dr Klum' Simelis, Dr Raphael 'Ralph' Reinbold, Dr Freddie 'The Tross' Foster, Dr William Figg, Dr Tim Suits, Dr Jos Kamps, Dr Roman Belle, Siegfried 'The Eagle' Thun-Hohenstein, and Dr Tobias John), the tea time crew (Emily 'Classical Vibes' Freeman, Giorgia 'Knuckles' Fiorini, and Amelia 'Meals' Brasnett), the OG bay squad (Rebecca Southwart, and Thomas Corner), with whom I have enjoyed many 'interesting' conversations, often supported by a few rounds of Guinness golf. Also, to the other members of the CJS group, past and present, who have made the lab a pleasant place to work. Special shout out to the Formula 1 enthusiasts (Will, Roman, and Tim) for never divulging who put the number 7 mug on my desk, and to Knuckles, Meals, and Patrick for our time spent together at LCLS and the redwood park.

Thank you to my parents for the continued support and encouragement I have received throughout my academic endeavours. Thanks also goes to Toni, Gary, and Shaun for similar support.

Most importantly I would like to thank Chloe, who has tolerated life with a professional student for all these years. Thank you for keeping me sane particularly at the end, for all the late-night taxis, and for being a constant joy. I genuinely could not have done it without your unwavering love, support, and encouragement.

Declaration and Contribution List

I declare that the work discussed and figures presented in this thesis are entirely my own, unless a reference to a person or a published source has been made, with the exception of the following work:

Chapter 2 and 3:

Prior to shipment to PAL-XFEL, chipless chip frames were prepared with the help of Dr Patrick Rabe and Dr Pauline Lang. SFX fixed-target data collection at PAL-XFEL was performed by Dr Jaehyun Park, Sehan Park, Gisu Park, and Dr Hyojung Hyun (all PAL-XFEL). SFX data processing was performed by Dr Pierre Aller, Dr Tiankun Zhou, Dr Agata Butryn, Dr Jos Kamps, Dr Robert Bosman, and Dr Aaron Brewster.

SSX fixed-target drop-on-chip data was collected and processed by Dr Pierre Aller, Johan Glerup, Dr Anastasiia Shilova, Dr Nicholas Devenish, Dr Tiankun Zhou, Dr Sam Horrell, Dr Robin Owen, Dr Danny Axford, and Dr Graeme Winter (all Diamond Light Source). Piezoelectric ejection parameters were optimised by Johan Glerup and Dr Pierre Aller (both Diamond Light Source).

Chapter 4:

pNIC28-Bsa4 vector was kindly provided by Tika Malla (University of Oxford). A modified pCOLD-I vector was kindly provided by Dr Patrick Rabe (University of Oxford). Vectors containing the genes for MBP, SUMO, Trx, and GST were kindly provided by Eidarus Salah (University of Oxford).

Chapter 5:

All 5’5-fused thiophene γ-lactams and all benzoxaborole compounds tested were synthesised by Gayatri (University of Oxford). All oxindole compounds tested were synthesised by Dr Lewis Ibbotson (University of Oxford).

Contents

Abstract ..... **i**
Acknowledgements ..... **ii**
Declaration and Contribution List ..... **iv**
Contents ..... **v**
Abbreviations ..... **ix**
Chapter 1: Introduction **1**
1.1. Peptidases 1
1.2. Serine Peptidases 3
1.2.1. General Classification of Serine Peptidases 3
1.2.2. Catalytic Residues 4
1.2.3. Tertiary Structure of the Peptidase Unit 5
1.2.4. Specificity 7
1.2.5. Regulation 11
1.3. Elastases 12
1.3.1. Activation and General Properties 13
1.3.2. Structure 14
1.3.3. Specificity 16
1.4. The Serine Protease Mechanism 18
1.4.1. Discovery of the Catalytic Residues 18
1.4.2. The Classical Serine Peptidase Mechanism 19
1.4.3. Site-Directed Mutagenesis 21
1.4.4. Protons Involved in Catalysis 24
1.4.5. Structural Evidence for the Acyl-Enzyme and Catalytic Water 27
1.4.6. Structural Evidence for the Tetrahedral Intermediate 31
1.5. X-Ray Free Electron Lasers (XFELs) and Time-Resolved Crystallography 34
1.5.1. Background 34
1.5.2. Examples of Time-Resolved SFX Experiments 36
1.6. Thesis Objectives 37
Chapter 2: Time-Resolved Crystallography with PPE **39**
2.1. Introduction 39
2.1.1. β-Casomorphins (BCMs) 39
2.1.2. PPE and BCM7 40
2.1.3. “Time-Resolved” Crystallography with PPE:BCM7 42
2.2. Results 44
2.2.1. Screening for Novel PPE Crystallisation Conditions 44
2.2.2. Reproducing ‘pH-Jumps’ in Macrocrystals 51

2.2.3. Ambient Temperature Serial Crystallography with BCM7 64
2.3. Discussion and Perspectives 76
Chapter 3: Alternate Inhibitors for Serial Crystallography with PPE 82
3.1. Introduction 82
3.1.1. Derivatives of BCM7 82
3.1.2. Sivelestat 84
3.1.3. Aims 87
3.2. Results 88
3.2.1. Alternative Peptide Inhibitors 88
3.2.2. Non-Peptidic Elastase Inhibitor: Sivelestat 92
3.3. Discussion and Perspectives 95
Chapter 4: Recombinant Expression and Purification of PPE 97
4.1. Introduction 97
4.2. Results 99
4.2.1. His_6-ProPPE Expression Trials 99
4.2.2. ProPPE-His_6 and *N*–Terminal Solubility Tags Expression Trials 102
4.2.3. ProPPE-His_6 Purification Trials 107
4.2.4. Further ProPPE-His_6 Expression Optimisation 121
4.2.5. Purification of ProPPE-His_6 Examining $CaCl_2$ Stabilisation (Purification 3) 128
4.3. Discussion 136
Chapter 5: PPE Inhibition Studies 142
5.1. Part 1: Bicyclic γ-Lactams 142
5.1.1. Introduction 142
5.1.2. Results 148
5.1.3. Discussion and Perspectives 165
5.2. Part 2: Benzoxaboroles 167
5.2.1. Introduction 167
5.2.2. Results 170
5.2.3. Discussion and Perspectives 183
Chapter 6: Summary and Conclusions 186
Chapter 7: Materials and Methods 195
7.1. General Information 195
7.2. General Protein Crystallography Methods 195
7.2.1. Materials 195
7.2.2. Sitting Drop Vapour Diffusion Crystal Screening 195
7.2.3. Hanging-Drop Vapour Diffusion 196
7.2.4. Batch Crystallisation 197
7.2.5. pH-Jump Methodology 198
7.2.6. Soaking of Sivelestat, Bicyclic γ-Lactams, and Benzoxaborole Compounds 198

7.2.7. Data Collection for Single Crystals 198
7.2.8. Pal-XFEL SFX Fixed-Target Data Collection 199
7.2.9. Diamond Light Source I24 SSX Fixed-Target Data Collection 199
7.2.10. Diamond Light Source I24 SSX Fixed-Target DoC Data Collection 200
7.2.11. SSX and SFX Data Processing 200
7.2.12. Model Building and Refinement Strategy 201
7.3. General Molecular Biology and Microbiology Methods 201
7.3.1. Materials 201
7.3.2. Preparation of Media and Agar Plates 202
7.3.3. Agarose Gel Electrophoresis 203
7.3.4. SDS-PAGE 203
7.3.5. Western Blotting 204
7.3.6. Bacterial Transformation and Plasmid DNA Isolation 205
7.3.7. Preparation Transformed Cell Glycerol Stocks 206
7.4. Cloning 206
7.4.1. Materials 206
7.4.2. Ligation Independent Cloning (LIC) 206
7.4.3. Gibson Cloning General Method 209
7.4.4. Insertion and Deletion PCRs 211
7.5. Recombinant Protein Expression and Purification 213
7.5.1. Materials 213
7.5.2. Cell Culture and Protein Expression 214
7.5.3. Cell Lysis 215
7.5.4. HisPur purification for Expression Trials 216
7.5.5. Ammonium Sulphate Precipitation 216
7.5.6. HisTrap Immobilised Metal Affinity Chromatography (IMAC) 216
7.5.7. Cation Exchange Chromatography (CEX) 218
7.5.8. Size Exchange Chromatography (SEC) 218
7.5.9. Enzyme Storage 219
7.6. Enzyme Absorbance Assays 219
7.6.1. Materials 219
7.6.2. pNA Standard Curve 219
7.6.3. Kinetic Measurements 219
7.6.4. Enzyme Inhibition Assay 220
7.7. Protein LC-MS 221
7.7.1. Materials 221
7.7.2. Method 221
Chapter 8: References 223
Appendix A: Zymogen Sequences for PPE, HNE and CTR A ..... a

Appendix B: Comparison of PPE numbering systems c
Appendix C: Crystal and Refinement Statistics ..... d
Appendix D: Protein Sequences ..... q

Abbreviations

A peptidase	Aspartic peptidase
A260/A280	UV light absorbance ratio (260/280 nm)
Ac-	*N*–Acetyl group
Ac-NPI	*N*–Acetyl-aparagine-proline-isoleucine
Ac-OPI	*N*–Acetyl-ornithine-proline-isoleucine
ADE	Acoustic droplet ejector
ADE-DoT	Acoustic droplet ejection combined with droplet on tape
AFC	7-Amino-4-trifluormethylcoumarin
ALI	Acute lung injury
AMC	7-Amino-4-methylcoumarin
AmpC	Class C β-lactamase
B. Taurus	*Bos Taurus*
BCM7	β-casomorphin 7
B-factors	Atomic displacement parameter or Temperature factor
Bis-tris	Bis(2-hydroxyethyl)amino-tris(hydroxymethyl)methane
BL21	BL21(DE3) Competent Cells
BlaC	*Mycobacterium tuberculosis* β-lactamase
BOC	*t*-Butyloxycarbonyl
BPTI	Bovine pancreatic trypsin inhibitor
BSA	Bovine serum albumin
C peptidase	Cysteine peptidase
CcO	Cytochrome *c* oxidase
CELA	Chymotrypsin-like elastase family member
CEX	Cation-exchange chromatography
CG	Cathepsin G
CHES	2-(Cyclohexylamino)ethanesulfonic acid
CKD	Chronic kidney disease
COPD	Chronic obstructive pulmonary disease
CRP	*E. coli* cAMP-regulatory protein
CTAB	Cetyltrimethyl ammonium bromide
CTR	Chymotrypsin
dCTP	Deoxycytosine triphosphate
dGTP	Deoxyguanosine triphosphate
DIFP	Diisopropyl fluorophosphate
DM-1	Diabetes mellitus type-I
DMSO	Dimethyl sulphoxide
dNTPs	Deoxynucleoside triphosphates
DoC-Tr	Drop-on-chip time-resolved
DPPI	Dipeptidyl peptidase I
DsbA	Thiol:disulfide interchange protein
DsbC	Disulphide bond isomerase C
DTT	Dithiothreitol

E. coli	*Escherichia coli*
E:P	Enzyme product complex
E:S	Enzyme-substrate complex
EAVA	Excite and visit again
ECM	Extra cellular matrix
EDTA	Ethylenediaminetetraacetic
EFSA	European Food Safety Authority
ESI-MS	Electrospray ionisation-mass spectrometry
FAB-MS/MS	Fragmentation-mass spectrometry
G peptidase	Glutamic peptidase
GFP	Green fluorescent protein
GI	Gastrointestinal
GST	Glutathione *S*–transferase
GTEx	Genotype-Tissue Expression project
H. sapiens	*Homo sapiens*
HAT	Human African trypanosomiasis
H-bond	Hydrogen bond
HCV	Hepatitis C virus
HEPES	*N*–2-Hydroxyethylpiperazine-*N'*–2-ethanesulfonic acid
His_{10}-tag	Decahistidine tag
His_6-tag	Hexahistidine tag
His_8-tag	Octahistidine tag
HNE	Human neutrophil elastase
HPE	Human pancreatic elastase
HSE	Human sputum elastase
IC_{50}	Half-maximal inhibitory concentration
IEX	Ion-exchange chromatography
IMAC	Immobilised metal affinity chromatography
IPNS	Isopenicillin *N* synthetase
IPTG	Isopropylthio-β-galactoside
K. pneumoniae	*Klebsiella pneumoniae*
k_{cat}	Turnover number
KLD	Kinase, ligase, DpnI
K_m	Michaelis constant
KR2	*Krokinobacter eikastus* Rhodopsin 2
LAMA	Liquid application method for time-resolved analyses
LBHB	Low barrier hydrogen bond
LC-MS	Liquid chromatography-mass spectroscopy
Lemo	Lemo21(DE3) Competent *E. coli*
LIC	Ligation independent cloning
M peptidase	Metallopeptidase
MBP	Maltose binding protein
MD	Molecular dynamics
MeOSuc-	Methyl-*O*-succinyl functional group
MES	2-(*N*–Morpholino)ethanesulfonic acid

MIB	Malonate, imidazole, boric acid buffer
MIP	Monoisopropyl phosphoryl
MISC	Mix and inject serial crystallography
MMT	DL-Malic acid, MES, Tris buffer
MPD	2-Methyl-2,4-pentanediol
MQ water	Water from a Milli-Q® Reference Water Purification System with a Quantum® TEX Polishing Cartridge and a Millipak® Express 40 Filter
MS	Mass spectroscopy
MW	Molecular weight
MWCO	Molecular weight cut-off
N lyase	Asparagine lyase
$N_2(l)$	Liquid nitrogen
NDSB-201	Non-detergent sulfobetaine-201
NEB® 5α	NEB® 5-α Competent *E. coli*
NETs	Neutrophil extracellular traps
Ni-NTA resin	Nickel-charged nitrilotriacetic acid resin
NMR	Nuclear magnetic resonance spectroscopy
NSP4	Neutrophil serine proteinase 4
NSPs	Neutrophil serine peptidases
Ntn	*N*–Terminal nucleophile
Nu	Nucleophile
OD_{600}	Optical density measured at 600 nm wavelength
OmpA	Outer membrane protein A
OMTKY3	Turkey ovomucoid inhibitor
OXA-10	Class D β-lactamase
OXA-48	Class D β-lactamase
P. aeruginosa	*Pseudomonas aeruginosa*
P1/P2	Product 1/ Product 2
PAH	Pulmonary arterial hypertension
PAL-XFEL	Pohang Accelerator Laboratory X-ray Free Electron Laser
PARs	Protease activated receptors
PBPs	Penicillin binding proteins
PBS	Phosphate buffered saline
PBS-T	Phosphate buffered saline with 0.1% Tween® 20
PDB	Protein data bank
PE	Pancreatic elastase
PEG 1500/3350	Polyethylene glycol average molecular weight 1500/3350 g mol^{-1}
PhoA	Acid phosphatase
pI	Protein isoelectric point
pIC_{50}	The negative logarithm of the IC_{50}
pK_a	The negative logarithm of the acid dissociation constant (K_a)
PMSF	Phenylmethylsulfonyl fluoride
P_n/P_n'	Peptide substrate nomenclature
pNA	*p*-Nitroanilide
PPE	Porcine pancreatic elastase

PR3	Proteinase 3
ProPPE	Pro-porcine pancreatic elastase
PSII	Photosystem II
PTFE	Polytetrafluoroethylene
RCL	Reactive Centre Loop
RMSD	Root mean square deviation
ROCK2	Rho-activated kinase
ROS	Reactive oxygen species
RSCC	Real space correlation coefficient
S peptidase or SP	Serine peptidase
SAR	Structure activity relationship
SARS-CoV-2 M^{Pro}	Severe acute respiratory syndrome coronavirus 2 main protease
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
SCP-B	Scytalidocarboxyl peptidase B
SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SEC	Size-exclusion chromatography
Serpins	Serine protease inhibitors
SFX	Serial femtosecond crystallography
SHuffle	SHuffle® T7 Competent *E. coli*
SIDS	Sudden infant death syndrome
SIHB	Short ionic hydrogen bond
SIRS	Systemic inflammatory response syndrome
S_n/S_n'	Peptidase subsite nomenclature
SOC	Super Optimal broth with Catabolite repression outgrowth medium
SP	Serine peptiadse
SPGA	Streptogrisin A
SPGB	Streptogrisin B
SSX	Serial synchrotron crystallography
Subtilisin BPN'	Subtilisin from *Bacillus amyloliquefaciens*
Suc-	*N*–Succinyl group
SUMO	Small ubiquitin modifier protein
T peptidase	Threonine peptidase
T. brucei	*Tranpanosoma brucei*
TAE	Tris:acetic acid:EDTA buffer
TB	Terrific broth medium
*t*Bu-ester	*t*-Butyl ester
TES buffer	Tris:EDTA:sucrose buffer
TPCK	L–1–Tosyl-amido–2–phenylethyl chloromethyl ketone
TR	Trypsin
Tris	Tris(hydroxymethyl)aminomethane
tRNA	Transfer ribonucleic acid
TRX	Thioredoxin A
TS	Transition state
TTON	Total turnover number
v/v	Volume by volume

w/v	Weight by volume
XFEL	X-ray free electron laser
A_1-PI	α_1-Antitrypsin or α_1-proteinase inhibitor
αLP	α-Lytic protease

Chapter 1: Introduction

1.1. Peptidases

Peptide bonds are chemically stable under normal physiological conditions with half-lives of more than 500 years[1]. Peptidases are enzymes that catalyse the breakage of peptide bonds, enhancing the rate of reaction by 10^{10} or more[1]. The name peptidase is often used interchangeably with protease and proteinase. Based on combined analyses from the Human Genome Project[2] and Genotype-Tissue Expression (GTEx) project[3], it is estimated that the human genome consists of between 20,000–21,000 protein encoding genes. According to the Degradome database, there are 588 human peptidase encoding genes equating to approximately 3% of the total number of human genes[4].

Peptidases can be classified based on the key active residues involved in catalysis[5]. Five classes of peptidases have been identified in mammals: aspartic (A), cysteine (C), serine (S), threonine (T), and metallo- (M) peptidases (Figure 1.1)[5]. There are two additional peptidase classes of which there are currently no known mammalian homologues: glutamic (G) proteases and asparagine (N) peptide lyases. Glutamic proteases were first classified as aspartic proteases[6]. However, crystallisation of scytalidocarboxyl peptidase B (SCP-B or scytalidoglutamic peptidase) from the fungus *Scytalidium lignicola,* revealed the presence of a catalytic glutamic acid residue rather than an aspartic acid residue[7-9].

Each of the six classes, except N, incorporate a water molecule into their substrate(s) and are thus hydrolases (*vide infra*). Asparagine peptide lyases are distinct as their mechanism does not employ water to break the peptide bond[10]. They are mentioned here briefly because they were originally identified as aspartic peptidases before subsequent reclassification[11]. Peptidases can also be distinguished by the substrate cleavage location. Endopeptidases cleave within the substrate peptide chain. Exopeptidases cleave up to three residues from the *N*–terminus (aminopeptidase) or *C*–terminus (carboxypeptidase) of their substrates[12].

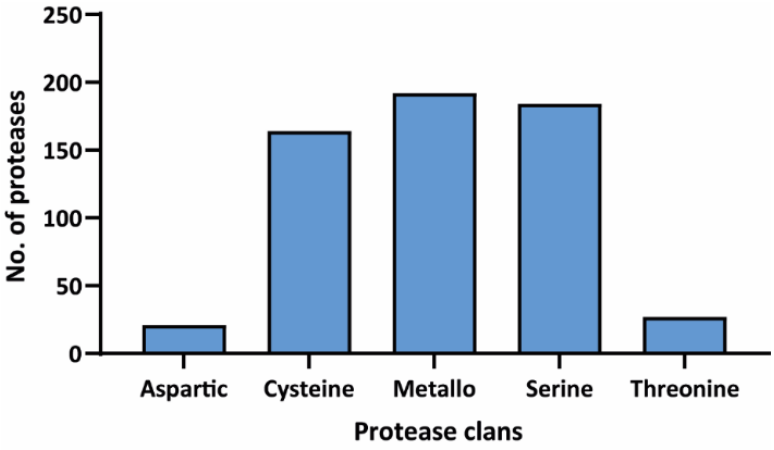


Figure 1.1 | The number of functional human proteases divided by clan. The data derived from the Degradome database[4].

Peptidases have historically been attributed to degradative roles in humans. The earliest discovered peptidases, i.e., pepsin, trypsin (TR), and chymotrypsin (CTR), function within the digestive system[13-15], and the most well-known nucleophilic threonine peptidases are the proteasomes[16]. However, the view of peptidase action has changed. Research has demonstrated that peptidases are important post-translation modifiers that regulate a wide array of cellular functions[17]. Often these reactions require a single precise proteolytic cleavage rather than degradation to completion, something known as limited proteolysis. For example, enzymes and hormones may be expressed as inactive precursors (zymogens) which are often activated by limited proteolysis[18-20]. Limited proteolytic cleavage cascades are involved in both immunity (complement cascade)[21] and blood coagulation[22]. Unsurprisingly, peptidases represent a growing class of drug targets. More than 10% of human peptidases have been linked to 135 hereditary disorders according to the Degradome database[4]. Peptidases have also been implicated in inflammatory diseases[23], and cancer[24]. Many exogenous peptidases are virulence factors for pathogenic fungi (e.g., secreted aspartic proteases (SAPs) from *Paracoccidioides brasillensis*[25]), bacteria (e.g., caseinolytic proteases (Clps) from various species[26]), and viruses (e.g., non-structural protein 3/4a (NS3/4a) from hepatitis C virus (HCV)[27]). Special attention has been paid to the inhibition of viral peptidases in recent years due to the onset of the 2019 severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pandemic[28].

1.2. Serine Peptidases

1.2.1. General Classification of Serine Peptidases

The MEROPS database classifies peptidases into clans and families[5]. A clan contains families of peptidases that appear to derive from similar evolutionary origins. Therefore, peptidases from a clan generally have similar tertiary structures[5]. Clans are labelled with two letters: the first derived from the catalytic residue type (A, C, S, etc.) and the second assigned alphabetically depending on when the clan was formed. Certain clans have been shown to include more than one catalytic type, likely arising from mutation of the nucleophilic residue and are thus labelled as P clans[12]. Families of peptidases are classified on similarity of the amino acid sequence within the peptidase unit, i.e., the part of the protein that exhibits peptidase activity[5]. A full explanation of the MEROPS classification system can be found in references 5 and 12. An overview of clans that include serine peptidases (SPs) are described in Table 1.1.

The largest clan, as determined by both number of families and peptidases, is PA. Originally, PA contained only SPs, however, the discovery of cysteine peptidases with similar tertiary structure to TR led to its reclassification. The following sections will highlight the main features of SPs. Special attention will be paid to the S1 family from clan PA, subclan S. These chymotrypsin-like (CTR-like) peptidases have traditionally been numbered using the chymotrypsinogen numbering system. As most of the CTR-like peptidase structures in the protein data bank (PDB) are deposited with this numbering system, and for ease of comparison, this numbering system will be used herein, unless stated otherwise. See Appendix A for a sequence alignment of CTR, porcine pancreatic elastase (PPE), and human neutrophil elastase (HNE). See Appendix B for a comparison of the PPE numbering vs CTR numbering.

Table 1.1 | SP clans as classified by the MEROPS database. For mixed type catalytic clans (PA, PB, PC, and PE) the subclan is added in brackets. The number of families reported in mixed type clans only includes those with SPs. If the clan type peptidase from the mixed clan is not a SP, a family type peptidase is reported instead. The catalytic residues are written in the order of amino acid sequence. † The family P1 from the PE clan contains *N*–terminal nucleophile (Ntn) peptidases where members of the family can have either a serine or threonine nucleophile and are thus a mixed family[29,30]. ‡ Some families of the SE and SH clans only require a dyad whereas others contain an additional residue. The data derives from the MEROPS database (accessed 10th March 2023)[5].

Clan (subclan)	No. of S families	Catalytic residues	Clan type protease
PA(S)	14	H,D,S or H,E,S	Chymotrypsin A (*Bos taurus*)
PB(S)	1	S (Ntn)	**S45 Family type protease:** penicillin G acylase precursor (*Escherichia coli*)
PC(S)	1	S,H,E	Dipeptidase E (*Salmonella enterica*)
PE†	1	S or T	No clan type. **Family type (P1):** DmpA aminopeptidase (*Ochrobactrum anthropi*)
SB	2	D,H,S or E,D,D,S	Subtilisin Carlsberg (*Bacillus licheniformis*)
SC	7	S,D,H	Serine carboxypeptidase D (*Triticum aestivum*)
SE	3	S,K,(S/Y)‡	D-Ala-D-Ala carboxypeptidase B (*Streptomyces lividans*)
SF	2	S,K/H	UmuD protein (*E. coli*)
SH	5	H,S,(H)‡ or D,H,S	Cytomegalovirus assemblin (human herpesvirus 5)
SJ	3	S,K	Lon-A peptidase (*E. coli*)
SK	3	S,H,D or S,K or S,H,S,E or K,S,S or S,S,K	Peptidase Clp (type 1) (*E. coli*)
SO	1	S,K	Escherichia coli phage K1F endosialidase CIMCD self-cleaving protein (Enterobacteria phage K1F)
SP	1	H,S	Nucleoporin 145 (*Homo sapiens*)
SR	1	K,S	Lactoferrin (*H. sapiens*)
SS	1	S,E,H	Murein tetrapeptidase LD-carboxypeptidase (*Pseudomonas aeruginosa*)
ST	1	S,H	Rhomboid-1 (*Drosophila melanogaster*)
Unclassified	8	multiple	multiple

1.2.2. Catalytic Residues

Members of the serine subclan (PA(S)) ubiquitously contain the 'classical' His, Asp, Ser catalytic triad[12]. Only members of the family S65 differ in this regard; the S65 family has putatively been annotated with a glutamate residue instead of an aspartate. However, this has not been confirmed *in crystallo*. Site-directed mutagenesis of the putative catalytic Glu residue to Asn or Ala did not affect auto-catalytic processing but did reduce hydrolysis of an artificial substrate[31]. SPs from other clans often employ the Ser, His, Asp catalytic triad, although the order of amino acids may vary. For example,

subtilisin Carlsberg (clan SB) has a Asp, His, Ser triad[32], while the prolyl oligopeptidase (clan SC) utilises Ser, Asp, His[33]. Alternatively, SPs have been annotated with monads, dyads, triads, and tetrads with alternate residues, including, lysine and tyrosine[34].

1.2.3. Tertiary Structure of the Peptidase Unit

The peptidases from the S1 family have similar tertiary structure in the peptidase unit. The fold, which is exemplified by CTR, is termed a double β-barrel. CTR can be divided into two structural domains, both of which fold into an open-ended β-barrel (Figure 1.2, A). The active site bridges a cleft between the two domains. The CTR catalytic triad residues H57 and D102 are located on the *N*–terminal domain and S195 in the *C*–terminal domain. Peptidases from each of the six S1 subfamilies (S1A-F) exhibit similar tertiary structure (Figure 1.2, A-C).

Peptidases from the PA(C) subclan utilise a cysteine nucleophile and are exemplified by the poliovirus 3C proteinase. These peptidases also exhibit a similar double β-barrel fold (Figure 1.2, C)[35]. It is hypothesised that PA(C) peptidases are evolutionarily related to the PA(S) peptidases, but the active site serine was subsequently mutated[36,37].

Other SP clans exhibit a range of folds. The clan SB is exemplified by subtilisin Carlsberg (Figure 1.2, D). The subtilisin family (S8) possesses a catalytic triad (Asp, His, Ser) which superimposes closely with the positions of the catalytic triad residues from CTR (Figure 1.2, E), despite little primary nor overall tertiary structure similarity[38,39]. Hence, subtilisin and CTR are often used as examples of convergent evolution[40]. Subtilisin has nine α-helices, two of which contain the catalytic S221 and H64[39]. Within the α-helices, there is a 7 stranded parallel β-sheet which contains the final catalytic residue D32[12]. A smaller two-stranded anti-parallel β-sheet is located at the *C*–terminus (Figure 1.2, D). The remaining SP clans exhibit various folds that usually include both α and β structural elements such as α/β/β/α-, and α/β/α- sandwiches[12]. However, the peptidase units of clan ST are all α. The ST clan consists of transmembrane rhomboid SPs, such as the *E. coli* GlpG rhomboid peptidase (Figure 1.2, F)[41].

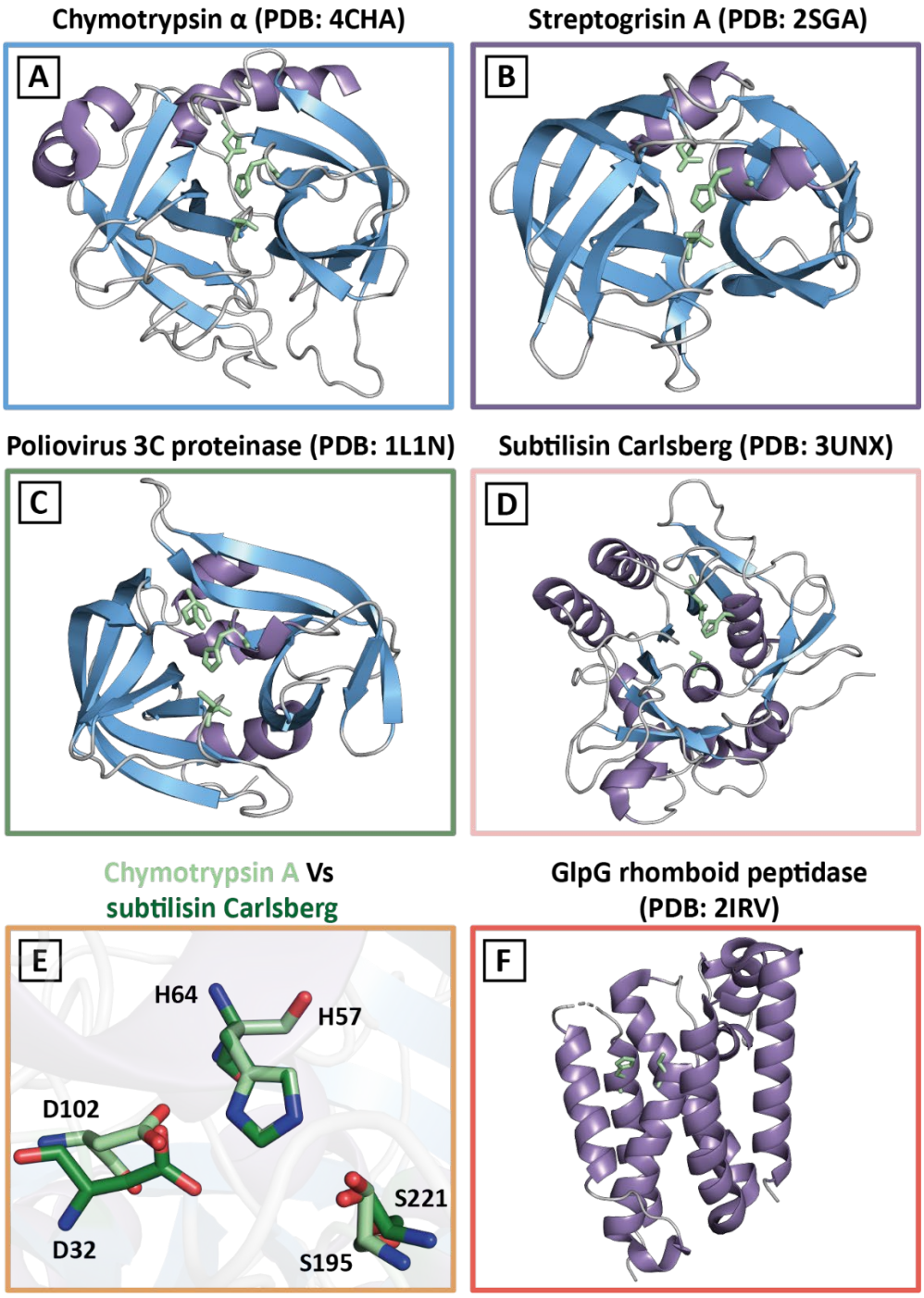


Figure 1.2 | Overviews from crystal structures of proteases showing the tertiary structures. A) A structure of CTR-α from *B. Taurus,* (classification: PA(S) S1A S01.001)[42]; **B)** a structure of streptogrisin A from *Streptomyces griseus* (classification: PA(S) S1F S01.261)[43]; **C)** a structure of poliovirus 3C proteinase from Human poliovirus 1 (classification: PA(C) C3 C03.001)[44]; **D)** a structure of subtilisin Carlsberg from *Bacillus licheniformis* (classification: SB S8A S08.001)[45]; **E)** a structure alignment of the sidechain atoms of the three catalytic residues of CTR-α (pale green carbons) and subtilisin Carlsberg (dark green carbons); **F)** a structure of GlpG rhomboid peptidase from *E. coli K-12* (classification: ST S54 S54.016)[41]. **A-D** and **F)** α-Helices are purple, β-strands are blue, catalytic residues are shown in green.

1.2.4. Specificity

Peptidase specificity can broadly be divided into endo- and exo-peptidases. Clan PA(S) is almost exclusively formed of endopeptidases, which cleave within the peptide chain (Figure 1.3, A). However, family S46 contains several bacterial dipeptidyl-peptidases (Dpp) which cleave dipeptides from the *N*–terminal[12]. Dpp7 from *Porphyromonas gingivalis* is essential for the production of dipeptides which are transported into the cell to be used as an amino acid source[46]. Outside of PA(S) clan, there is much more variability in specificity and amino-[47], tripeptidyl-[48], carboxy-[49], and di-peptidases[50] have been identified (Figure 1.3, A).

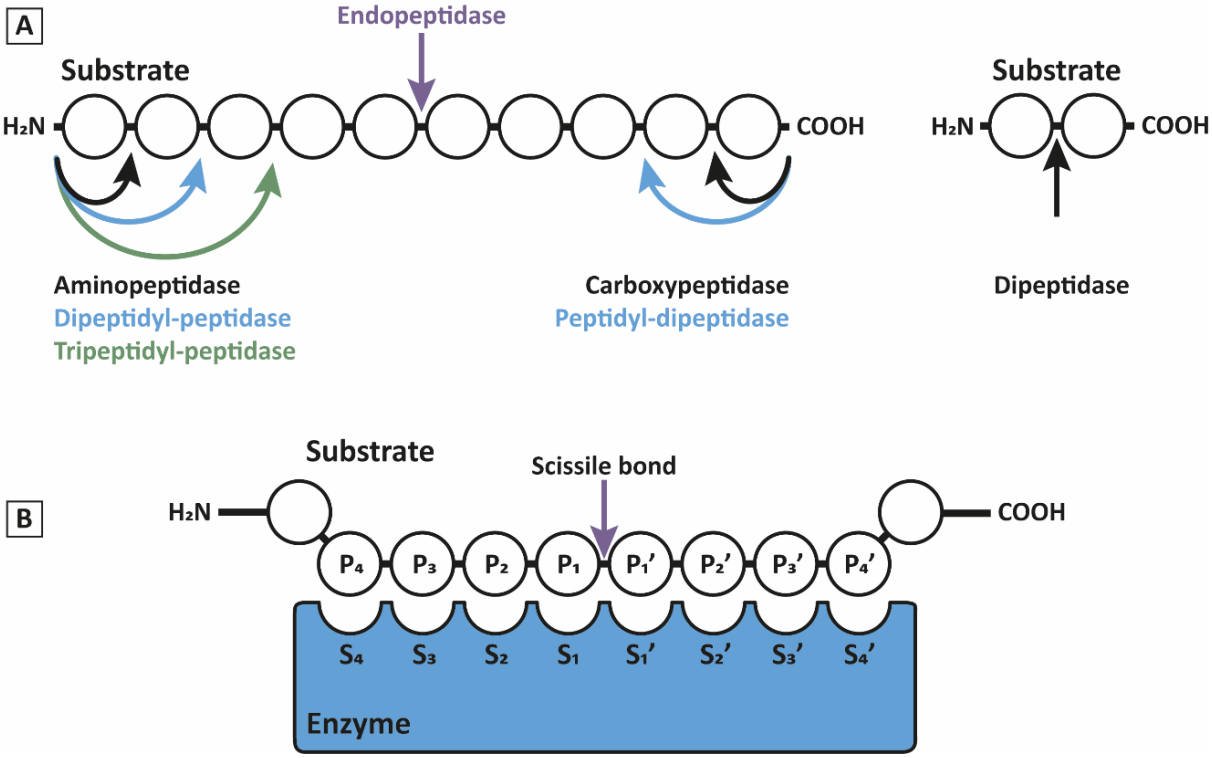


Figure 1.3 | Peptidase subsite and specificity nomenclature. A) Nomenclature for peptidase classification by specificity. Each circle represents one amino acid. Endopeptidases are shown above the substrate chain, exopeptidases are shown below. The colour of the exopeptidase group name corresponds to the arrow indicating the location of the substrate chain that is cleaved by that group of peptidases; **B)** nomenclature of peptidase substrate binding sites and the substrate residues as proposed by Schechter and Berger[51] using an endopeptidase as an example. The figure was adapted from Rawlings, 2020[52].

The specificity of a peptidase is determined by interactions between the peptide substrate and substrate binding sites. Schechter and Berger described the conventionally used nomenclature for the peptide substrate and substrate binding sites in 1968[51]. The peptide substrate residues are labelled P while the substrate binding sites are labelled S. Each peptide residue and binding pocket is given a number starting from 1 on each side of the scissile bond. The numbers increase each residue or interacting binding site away from the scissile bond (e.g., P_1, P_2, P_3, or S_1, S_2, S_3, etc.). The residues and the binding sites *C*–terminal to the scissile bond are termed primed sites (e.g., P_1', P_2', P_3', or S_1', S_2', S_3' etc.) to distinguish them from the sites *N*–terminal to the bond (Figure 1.3, B)[51].

Family S1 SPs form an antiparallel β-sheet with peptide substrates, which is stabilised by several hydrogen bonds[53]. The S_1 specificity pocket is the primary determinant of substrate specificity for most peptidases from the S1 family[12]. The peptidases from this family are generally divided into three specificity types: CTR-like peptidases preferentially cleave bonds after large hydrophobic residues (e.g., Phe, Tyr, Trp, etc.), trypsin-like (TR-like) peptidases cleave after Arg or Lys, while elastase-like peptidases prefer small hydrophobic residues in the P_1 position (Figure 1.4). The narrow and predominantly hydrophobic S_1 pocket is formed by the backbone atoms of residues 190-192 and 214-216, and the side chain of residue 213 (Figure 1.5)[53]. In both CTR and TR, residues 216 and 226 are both glycine residues which enable the S_1 pocket to accommodate large P_1 sidechains. However, CTR and TR differ at residue 189[54]. The sidechain of S189 in CTR is sufficiently small to enable large and predominantly hydrophobic amino acids to preferentially bind in the S_1 pocket (Figure 1.5, A). In contrast, D189 of TR interacts directly or indirectly with the sidechains of Arg or Lys, respectively. Therefore, TR preferentially cleaves the bond after basic residues (Figure 1.5, B)[54]. Elastase preferentially binds small, hydrophobic residues at P_1. This preference can be attributed to the replacement of G216 and G226 of CTR with V216 and T226 in elastase. The sidechains of these residues protrude into the S_1 specificity pocket of elastase, creating a smaller hydrophobic pocket (Figure 1.5, C)[53].

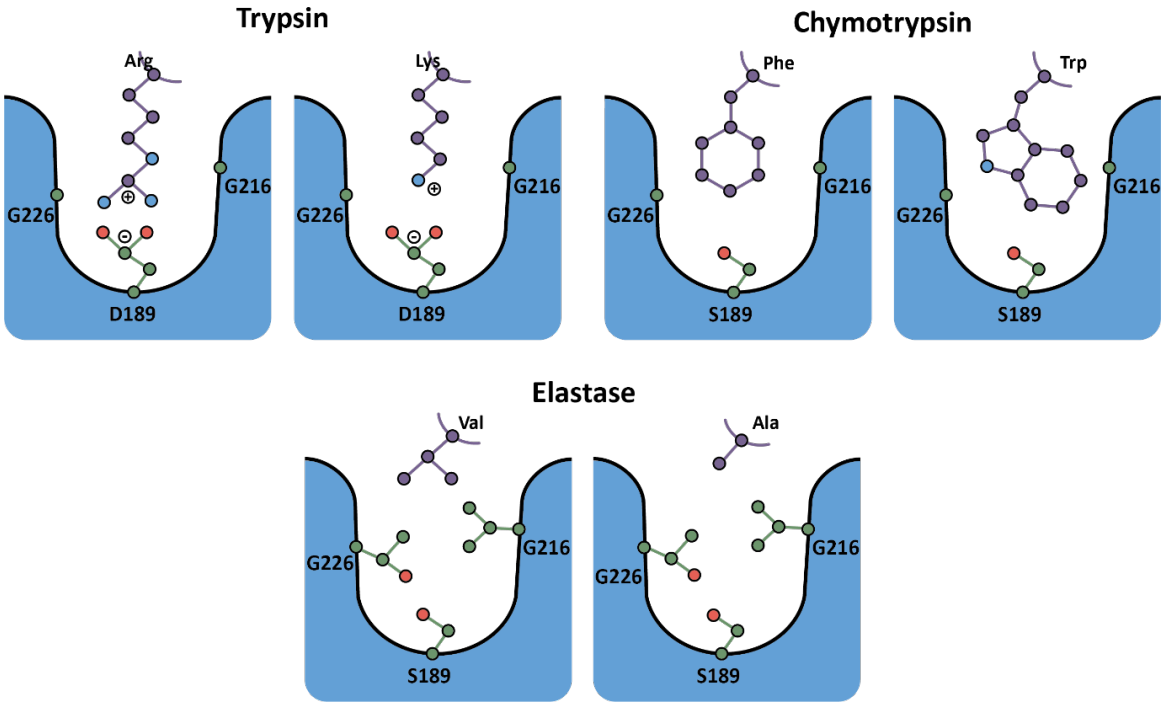


Figure 1.4 | Schematic of the S_1 specificity pocket and preferred substrate residues of CTR, TR, and elastase. The figure was adapted from Brändén and Tooze, 2009[54].

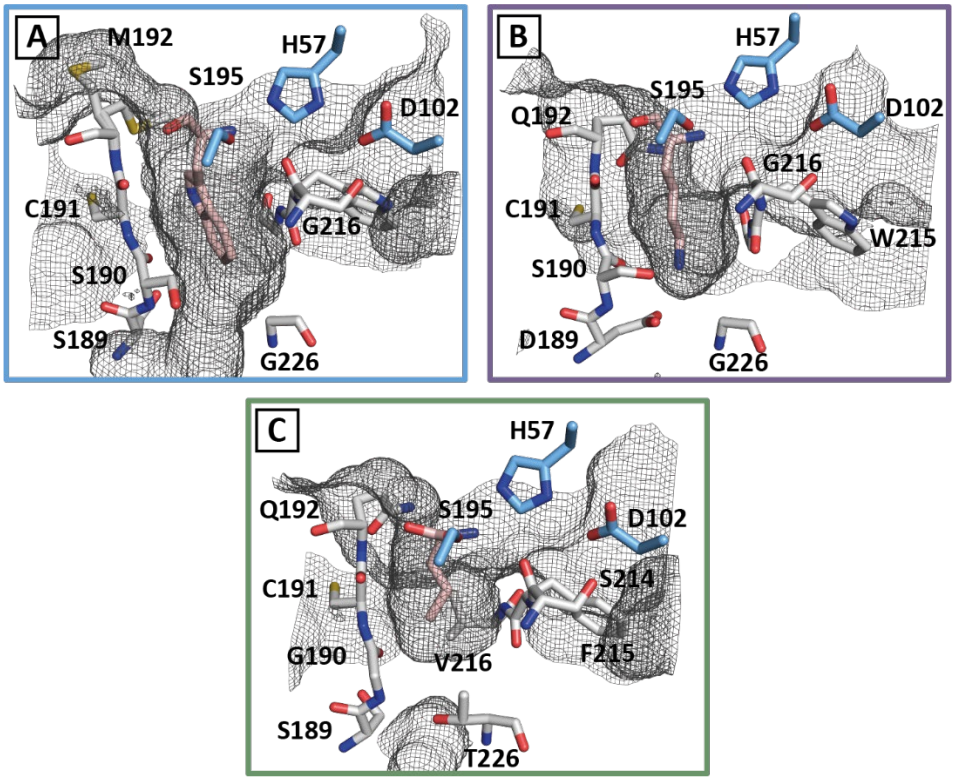


Figure 1.5 | Active site views from reported crystal structures of SPs showing P_1 residues of inhibitors bound within the S_1 subsite of CTR α, TR, and porcine pancreatic elastase (PPE). A) An active site view of the S_1 pocket from α-CTR in complex with a mutant bovine pancreatic trypsin inhibitor (BPTI) with a Trp in the P_1 position; **B)** an active site view of the S_1 pocket from β-TR in complex with a mutant cyclic nonapeptide inhibitor with a Lys in the P_1 position; **B)** an active site view of the S_1 pocket from PPE in acyl-enzyme complex with a heptapeptide inhibitor with an Ile in the P_1 position. The catalytic triad carbons are light blue, other enzyme carbons involved in forming the S_1 pocket are white, the P_1 residue carbons are pink, and the surface of the enzyme rendered around the S_1 residues is represented as a grey mesh.

While the S_2-S_n and S_1'-S_n' sites are generally less important, they can still contribute to the specificity of peptidases. For example, enteropeptidase is responsible for the hydrolytic activation of TR from the inactive zymogen, trypsinogen[12]. Enteropeptidase has TR-like activity, cleaving after Lys15 of the trypsinogen propeptide, but demonstrates strict specificity for the sequence DDDDK[55]. This was rationalised by the presence of Lys99 which is surrounded by the Asp sidechains of the P_2-P_4 residues of the substrate[56]. Lys99 likely makes salt-bridge interactions with the P_2 and P_4 sidechain carboxylates, while the P_3 side chain carboxyl group hydrogen bonds with the hydroxy group of Tyr174[56]. The specificity of SPs partially arises from insertions or mutations in eight loops surrounding the active site (Figure 1.6)[57]. For example, thrombin and several kallikreins have large insertions in the 60- and 99-loops, respectively (Figure 1.6). The 60-loop shields the thrombin active site. The large 99-loop insertion in kallikrein enzymes, known as the kallikrein loop, contains post-translational modifications which may contribute to substrate specificity[58].

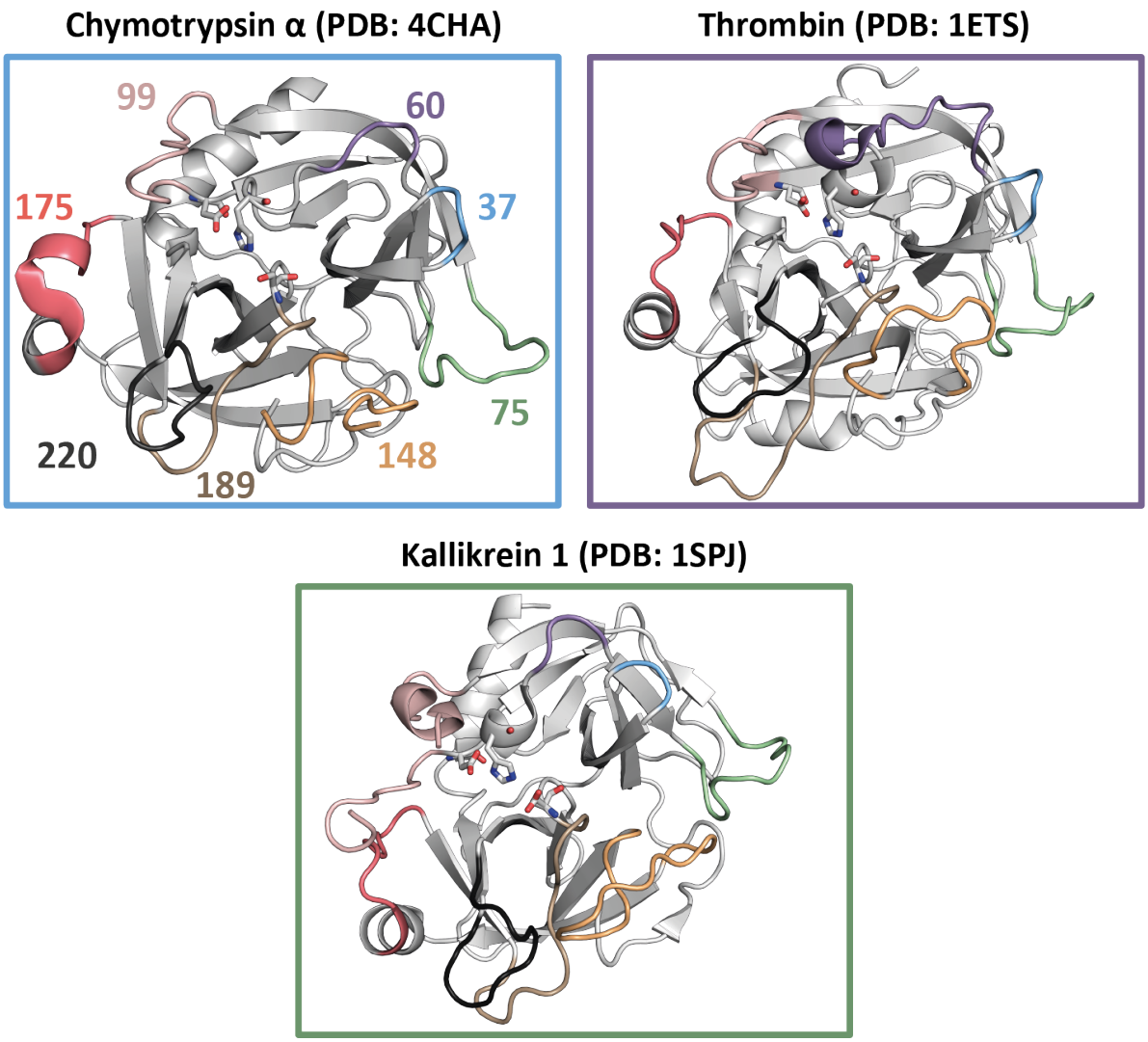


Figure 1.6 | Views from reported crystal structures of peptidases from the S1 family. Structural overview of CTR-α, thrombin, and kallikrein 1, highlighting the eight surface loops surrounding the active site. The loops are named according to the approximate centre residue in CTR. The catalytic triad residues are shown as sticks. The figure was adapted from Goettig *et al.* 2019[57].

1.2.5. Regulation

The unregulated or imbalanced activity of SPs has been implicated in various disorders. Peptidases are regulated at both the gene and protein levels. At the gene level, peptidase expression may be tissue specific, such as the neutrophil SPs (NSPs) which are expressed by only a few cell types, such as neutrophils from which their name derives[59,60]. At the protein level, peptidases may be compartmentalised. The NSPs are localised to azurophilic granules and are only released in response to signals such as infection[61]. A common method of regulation of SPs involves the biosynthesis of peptidases as zymogens, which are subsequently activated by limited proteolysis or through a

conformational change brought about by the environment or substrate/allosteric interactions[62]. Finally, an array of endogenous serine protease inhibitors (serpins) are produced to regulate the activity of SPs. These proteins inhibit peptidases via a suicide mechanism[63,64]. A motif called the "reactive centre loop" (RCL) is cleaved by the target peptidase. Following cleavage of the scissile bond, the serpin undergoes a major conformational change which transports the peptidase to opposite pole of the serpin and irreversibly traps the SP[63,64].

1.3. Elastases

Elastic fibres are a widespread component of the extra cellular matrix (ECM) and are particularly important in tissues that require repeated physical deformation such as the major blood vessels and lungs[65]. The fibres are made predominantly from cross-linked elastin proteins which have a high percentage of hydrophobic amino acids[66]. The hydrophobic nature and the primary amino acid sequence of tropoelastin (the monomeric and uncrosslinked form of elastin) provides the elastic ability in which the protein can stretch to more than eight times its original length[65]. In the polymeric form, the hydrophilic regions of elastin are highly modified by cross-linking, making elastin resistant to the proteolytic activity of many peptidases[66]. This resistance is demonstrated by the exceedingly long half-life of approximately 70 years[67].

Elastases are a group of peptidases that are capable of solubilising insoluble elastin fibres. Peptidases from several classes (A, C, M, and S) have been identified as elastases, although many will only hydrolyse elastin under certain conditions, such as in an acidic environment[68]. However, pancreatic and neutrophil elastases are potent elastinolytic enzymes. There are three enzymes named "pancreatic elastase" (PE; or CTR-like elastase family members (CELAs))[12]. The peptidases are encoded by the CELA genes. PE I (EC: 3.4.21.36; gene: CELA1) was first identified from porcine pancreatic preparations and distinguished from the activity of TR and CTR by Balo and Banga in 1949[69,70]. Confusingly, this enzyme is not expressed in the pancreas of humans, but is instead present in the skin[71]. There are two isoforms of human PE II (EC: 3.4.21.71; genes: CELA2A and CELA2B) which share

high sequence identity. PE IIA is expressed in the human pancreas, while PE IIB is inactive and has not been identified at the protein level[72-74]. There are also two isoforms of PE III (EC:3.4.21.70; genes: CELA3A and CELA3B). Both isoforms are expressed in the pancreas, but the enzyme products exhibit low elastinolytic activity and are often referred to as protease E rather than PE III[75,76]. Both the PE I and II genes are expressed in the pancreas of pigs[12]. As the work presented in this thesis is related to porcine PE I (PPE), the following sections will predominantly focus on PPE with some reference to HNE (EC: 3.4.21.37; gene: ELANE).

1.3.1. Activation and General Properties

PPE is a digestive enzyme that is stored as a zymogen (proelastase (ProPPE)) within acinar cells of the pig pancreas[12]. The PPE gene encodes a 16-residue signal peptide and a 10-residue propeptide, which terminates at Arg15 (Uniprot: P00772). The Arg15-Val16 bond of ProPPE is cleaved by TR in the duodenum[77], following activation of TR by enteropeptidase[56]. The active form of PPE has 240 residues and the molecular weight is 25.9 kDa[78]. Neither PPE nor human PE I are glycosylated, and the primary amino acid sequences of the active enzymes exhibit 89% identity. HNE is produced as a pre-pro-protein with a 27-residue signal peptide and a short pro-dipeptide at the *N*–terminus, and a 20-residue extension at the *C*–terminus (Uniprot: P08246)[12]. Only the *N*–terminal dipeptide is required to keep HNE in the zymogen state and is cleaved by dipeptidyl peptidase I (DPPI or cathepsin C)[79]. The *C*–terminal extension is not present in the mature enzyme and is likely removed during post-translational processing[12]. Mature HNE has 218 amino acids and two glycosylation sites, both of which are asparagine residues[80]. Variation in glycosylation is responsible for the observation of several isoforms which exhibit similar specificity but different molecular weights (29–33 kDa)[53]. The amino acid sequences of active HNE and PPE are 39% identical.

PPE and HNE have similar pH optima, i.e., 8.8[81] and 8.0–8.5[82], respectively. The isoelectric point (pI) of PPE is 9.5[83], while HNE is 10.5[84]. PPE demonstrates proteolytic activity between pH 6.0 and 10.5. Below pH 6.0, PPE demonstrates reduced activity due to protonation of H57, but remains properly folded.

Above pH 10.5 the *N*–terminal may become solvent exposed causing loss of enzymatic activity[81,83]. Similar effects are observed below pH 4.2, however, short exposure to pH 2.6 is reversible[85].

1.3.2. Structure

Shotton *et al*. first crystallised PPE in 1968[86] and solved the X-ray crystal structure of *p*-toluenesulfonylated PPE in 1970[87]. The first unmodified structure was solved to high resolution (1.65 Å) by Meyer *et al*. 1988 in 70% methanol[88]. Subsequently, a structure of native PPE was solved to 1.1 Å[89] and an enzyme:inhibitor complex to sub-angstrom resolution (0.94 Å)[90]. The X-ray crystal structure of HNE was solved in 1986 by Bode *et al*.[91] while the structure of native unmodified HNE was solved in 2011[92].

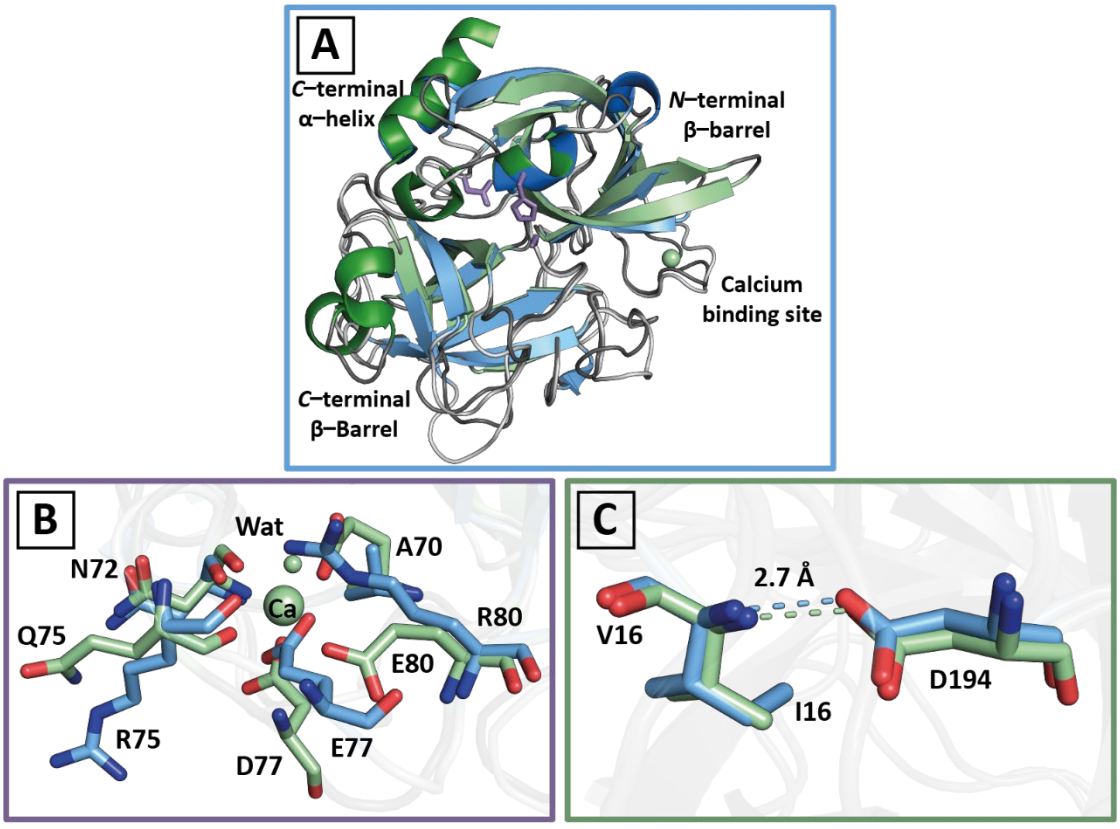


Figure 1.7 | Superimposition of apo-PPE (PDB: 1QNJ) and apo-HNE (PDB: 3Q76). A) Structural overview; **B)** a view of the PPE metal-binding site showing the residues that coordinate the Ca^{2+} ion; **C)** a view of the buried *N*–terminus. PPE is shown in green and black, HNE is shown in blue and white, and the catalytic triad residues are shown in purple. The calcium ion and water in the PPE structure are represented by large and small green spheres, respectively.

As with other S1 family SPs, both PPE and HNE exhibit a double β-barrel structure (Figure 1.7, A). Each β-barrel is formed of six anti-parallel β-strands and each barrel is orientated perpendicularly to one another. Both PPE and HNE have a common, large *C*–terminal α-helix and a smaller α-helix that holds H57 at the active site. The position of several additional small α-helical sections differs between the structures. PPE has a calcium binding site that is conserved in CTR and TR-like SPs, and which is important for stability, but exhibits no effect on activity (Figure 1.7, B)[93]. Several loop residues between E70-E80 coordinate the calcium ion in PPE, with one additional water molecule forming octahedral calcium coordination[93]. HNE does not have a metal binding site. E80 is replaced by R80 in HNE and the guanidinium cation replaces the calcium ion and plays a stabilising role (Figure 1.7, B)[95].

There are currently no reported crystal structures of proelastase, therefore, the exact conformational changes that occur after activation by TR are unknown, but can be inferred from structural studies of CTR and TR[94]. Chymotrypsinogen and ProPPE are activated by the proteolytic action of TR, which cleaves an Arg-Ile or Arg-Val bond, respectively[12]. Upon cleavage the *N*–terminal residue of mature PPE (V16) is buried (Figure 1.7, C), and the amino group of V16 forms a salt-bridge with the sidechain of D194, the residue adjacent to the catalytic serine (S195)[95]. Residues 16 of CTR, TR, and HNE form identical interactions (Figure 1.7, C). Whilst there is little movement in the catalytic triad, residues that form the oxyanion hole (G193) and the S_1 specificity pocket (190–192 and 214–216) are reorientated to enable substrate binding. The substrate binding site is therefore likely to be inaccessible in chymotrypsinogen[94]. Conformational changes that occur following cleavage of the chymotrypsinogen propeptide are shown in (Figure 1.8).

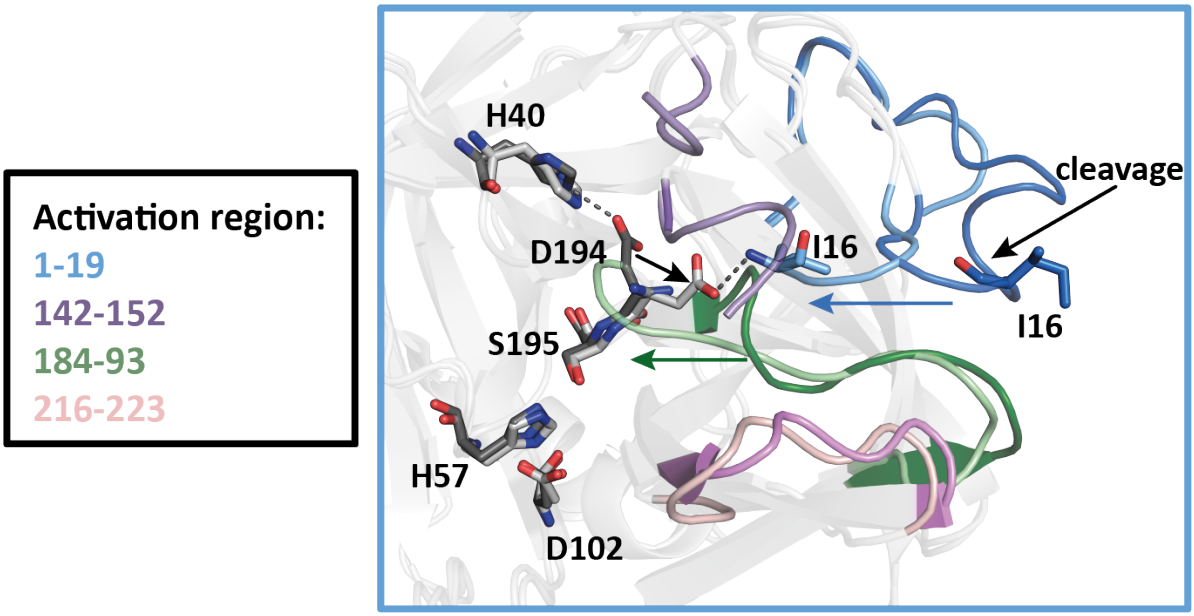


Figure 1.8 | Comparison of reported crystal structures of chymotrypsinogen and CTR α. The conformational changes involved in activation are highlighted. The amino acid numbering for each region is shown on the left, with the colours corresponding to each region. The darker shades correspond to inactive chymotrypsinogen (PDB: 1CHG) and the lighter shades correspond to active CTR-α (PDB: 4CHA). Note, the loop residues 142–152 are not fixed in the chymotrypsinogen structure and thus not modelled. Upon activation, part of the loop region becomes more rigid and is modelled, but much of the loop remains unfixed.

1.3.3. Specificity

The substrate specificity of PPE and HNE has been extensively studied by both kinetic and crystallographic means. Following the report of the structures of PPE and HNE in complex with inhibitors that extended from the primed to non-primed subsites, Bode *et al.* proposed the subsite residues, as detailed in Figure 1.9[53,91,96]. The substrate forms an antiparallel β-sheet with S214 and V216. The backbone of the P_3 residue forms two backbone H-bonds with V216 while a third H-bond is formed between the P_1 amide and the carbonyl of S214[53]. S214 is conserved in many CTR-like SPs and the Oγ atom is also positioned to interact with the sidechain of D102 of the catalytic triad. Thus, S214 has been proposed to link hydrolysis to substrate binding[53].

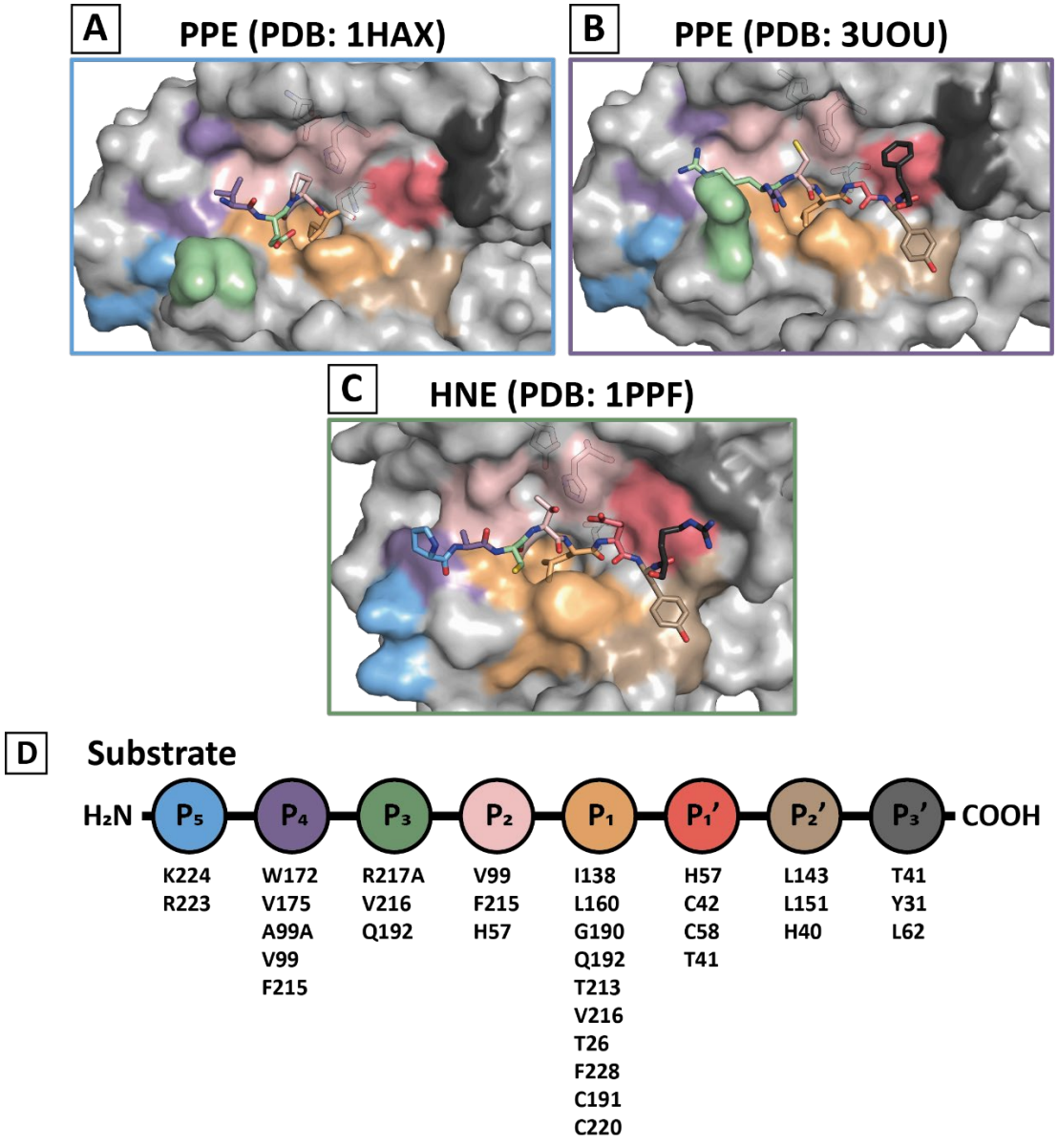


Figure 1.9 | Substrate binding sites of PPE and HNE. A) A view from a crystal structure of PPE in complex with a heptapeptide inhibitor (YPFVEPI); **B)** a view from a crystal structure of PPE in complex with a Kunitz-type protease inhibitor (ShPI-1 (K13L mutant)); **C)** a view from a crystal structure of HNE in complex with a domain of the turkey ovomucoid inhibitor; **D)** schematic of the substrate residues. The enzyme residues identified below the peptide chain correspond the binding sites for PPE (according to Bode *et al*. 1989[53]). **A-C)** The surface of the enzyme and the carbons of the inhibitor (sticks) are coloured according to their subsite shown in **D**. Some enzyme residues contribute to more than one binding site and are coloured based on proximity to the scissile bond (S_1 to S_5/S_3').

The S_1 subsites of PPE and HNE are most important in determining substrate specificity. This is the only true binding 'pocket' whereas the other subsites form depressions in the enzyme surface[53]. The S_1 specificity pocket of PPE is slightly smaller than HNE[53]. As a result, PPE preferentially hydrolyses peptides with Ala, Val, and Ile P_1 residues, whereas HNE can accommodate larger hydrophobic residues such as Leu and Met[97-99]. The active site of PPE and HNE is relatively large, extending from S_5 to S_3'[100,101].

A P_2 proline was found to productively orientate the substrate within the active site and appears to inhibit larger peptides binding in reverse or in multiple copies within the active site[102]. Generally, the $S_{3–5}$ residues do not contribute to specificity of PPE nor HNE, although increasing the chain length of substrates *N*–terminal to the scissile bond can increase the catalytic efficiency (k_{cat}/K_m) by orders of magnitude[103-105]. Many of the peptide substrates used to analyse PPE specificity employ a hydrophobic P_3 residue, which is predominantly Ala. However, peptide inhibitors with polar P_3 residues have been reported and will be discussed in detail in Chapter 2 and 3[106,107].

The specificity of each of the prime subsites has also been examined. Atlas *et al*. demonstrated that there are three primed subsites in PPE and that productive binding in these sites can increase the catalytic activity towards peptide substrates[101]. S_1' has hydrophobic character and accepts Ala residues. However, P_1' Lys residues can also result in higher affinity for PPE[108]. A similar observation was also made for HNE which is proposed to have a larger and less specific S_1' subsite. Both small hydrophobic and aromatic P_3' residues have been demonstrated to improve the affinity for PPE[101,108].

1.4. The Serine Protease Mechanism

The most well studied family of SPs is the S1 family, all of which possess the classical His, Asp, Ser catalytic triad[5]. Many other families of SPs have a similar catalytic triad, including the subtilisin peptidases. Although the order of residues may differ, the superimposition of the three-dimensional orientation of the catalytic triad residues is almost identical (Figure 1.2, E). Thus, these peptidases likely hydrolyse peptides via a similar mechanism. The following discussion will focus on the proposed mechanism for enzymes with the classical triad. For reviews of other catalytic triads see references 12 and 109.

1.4.1. Discovery of the Catalytic Residues

The catalytic Ser was originally identified by inhibition experiments with CTR using the nerve gas diisopropyl fluorophosphate (DIFP)[83,110,111]. The role of the catalytic His residue was subsequently

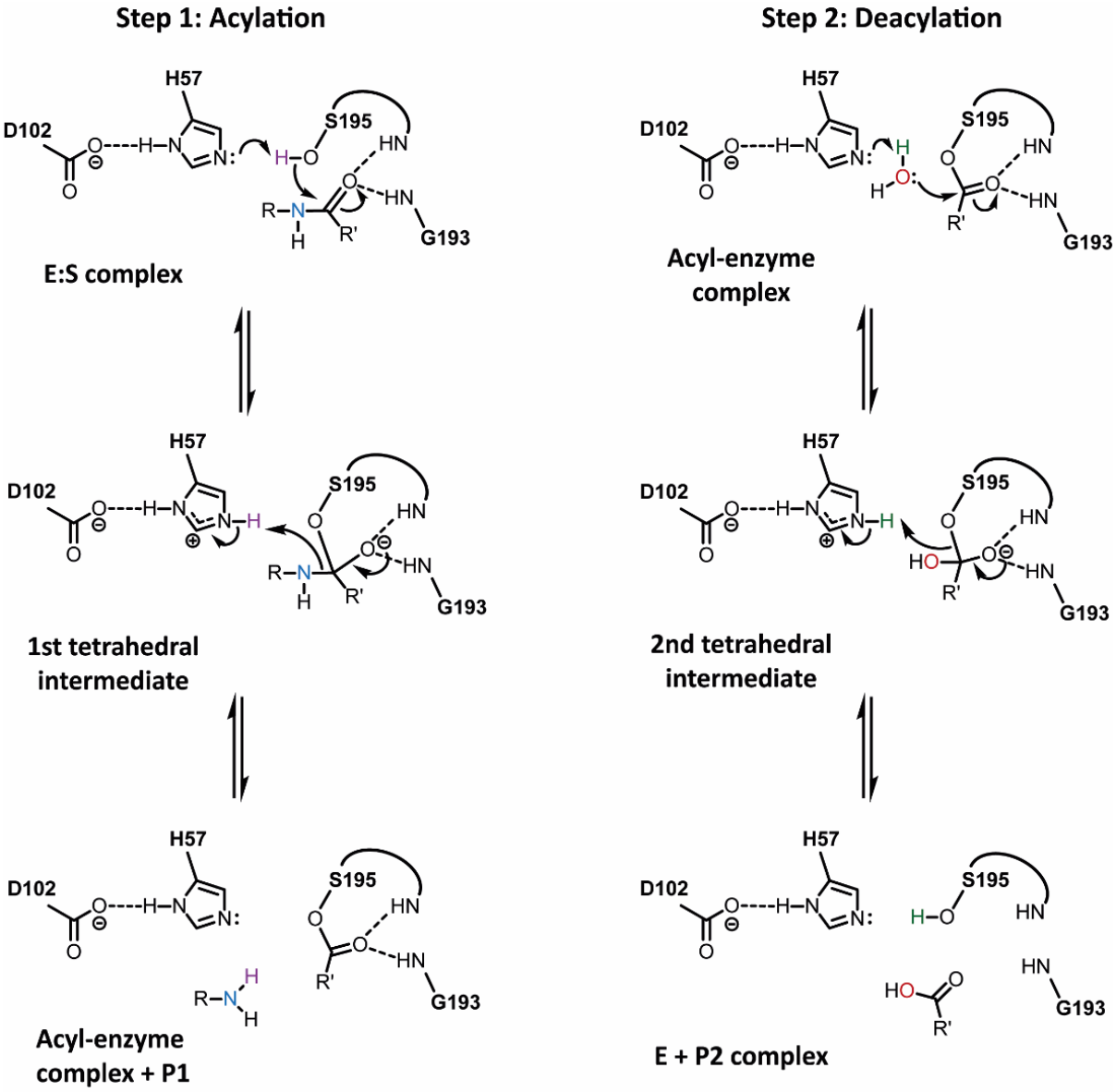
inferred due to the pH dependence of rate constants with CTR and TR, which demonstrated that the pK_a of the reaction was between 6–7[112,113]. Thus, it was proposed that the imidazole ring of a histidine residue was involved in catalysis. Subsequently, Schoellmann *et al*. demonstrated that L–1–tosyl-amido–2–phenylethyl chloromethyl ketone (TPCK) irreversibly inhibited CTR by reaction with a His residue[114]. The involvement of the 'buried' catalytic Asp residue was proposed by Blow *et al.* following the report of a high-resolution structure of CTR-α[115-117]. These residues were identified as H57, D102, and S195.

1.4.2. The Classical Serine Peptidase Mechanism

The generally accepted mechanism of SP peptide catalysis can broadly be divided into two stages[117,118]. In the first stage, often known as the "acylation reaction", the peptidase breaks the peptide bond, liberates the *C*–terminal fragment of the peptide substate, and forms a covalent linkage between the *N*–terminal fragment and S195 (acyl-enzyme complex). In the second stage, termed the "deacylation reaction", the acyl-enzyme intermediate is cleaved to liberate the *N*–terminal fragment of the substrate. The two stages follow closely related mechanisms which differ in the nucleophile utilised to attack the peptide/ester bond and the leaving group (Scheme 1).

In the acylation reaction, following formation of the Michaelis complex, S195 attacks the scissile amide carbon. S195 is activated by H57 which acts as a general base, accepting the Oγ proton from S195. While the pK_a of H57 is ~7 in the ground state[119], evidence has accumulated demonstrating that the pK_a of H57 is raised to approximately 10–11 in the presence of inhibitors that mimic the tetrahedral intermediate[120-123]. This treatment moves the pK_a much closer to that of the S195 which is estimated to be ~15, perhaps explaining how H57 is able to act as a general base to increase the nucleophilicity of S195[124]. The attack by S195 on the substrate carbonyl leads to formation of an anionic tetrahedral intermediate and a cationic imidazolium ion. The positive charge on the latter is stabilised by an electrostatic interaction with the negatively charged sidechain of D102. The anionic tetrahedral intermediate is stabilised by H-bond interactions with a region of the active site known as the

"oxyanion hole"[125]. In CTR-like SPs, the oxyanion hole is formed by the backbone amides of S195 and G193. The unstable tetrahedral intermediate subsequently collapses and H57 assists again, this time acting as a general acid, to protonate the amino leaving group of the *C*–terminal fragment of the substrate (product 1 (P1)). The *N*–terminal fragment remains covalently bound to S195. In the second stage, H57 plays an identical role but instead activates a water. The activated hydroxyl attacks the ester carbonyl carbon leading to formation of a second tetrahedral intermediate. Collapse of this intermediate liberates product 2 (P2), the *N*–terminal fragment. H57 again acts as a general acid, donating a proton to S195 to regenerate the enzyme to the ground state[117,118]. Thus, during peptide catalysis peptidases perform both amidase and esterase reactions. Unsurprisingly, peptidases often demonstrate potent esterase activity.



Scheme 1 | The proposed mechanism of serine protease peptide hydrolysis. The schematic shows the oxyanion hole formed from the amide nitrogen atoms of S195 and G193, as would be observed in CTR-like SPs. The composition of the oxyanion hole from other SP families is variable and may include sidechain interactions.

1.4.3. Site-Directed Mutagenesis

1.4.3.1. The Catalytic Triad Residues

The residues of the catalytic triad of several SP enzymes have been probed by site-directed mutagenesis. Early work employed chemical modification of these residues of CTR. A decrease in CTR activity of 10^4 was achieved by converting S195 to dehydroalanine[126] or by methylating H57 Nε2[127,128]. Subsequent site-directed mutagenesis, which was also extended to D102, produced similar effects. At pH 8.0, both S195A or H57A of rat anionic TR led to a ~10^5 reduction in k_{cat}[129]. The D102N mutant also

exhibited a large decrease in k_{cat} (10^4). However, the effects were not additive as the activity of the triple catalytic triad mutant did not decrease below that of the S195A nor H57A single mutants (10^5). The K_m of all mutants was less affected by the mutations (< 12-fold difference) suggesting a limited role in substrate binding for the catalytic triad[129].

Similar results were observed for the subtilisin from *Bacillus amyloliquefaciens* (subtilisin BPN')[32]. The single mutants, H64A and S221A (subtilisin numbering), both reduced the k_{cat} of subtilisin BPN' by 10^5, while the D32A mutation led to a decrease of only 10^3 [32]. All of the double mutants and the triple mutant exhibited approximately the same drop in k_{cat} as either the single H64A or S221A mutants. Thus, loss of either the catalytic Ser or His residue is sufficient to effectively disable the catalytic triad[32]. However, it should be noted that the rate of hydrolysis in both enzymes remained orders of magnitude above the uncatalyzed reaction (≥ 10^3). The background catalysis has been attributed to the TS stabilisation deriving from the oxyanion hole and desolvation of the substrate[118]. Importantly, the activity of subtilisin BPN' could be completely inhibited in the presence of SP inhibitors such as the turkey ovomucoid inhibitor (OMTKY3), indicating that the residual activity is not an artefact[130].

Interestingly, the catalytic activity of certain mutants can be partially rescued by changes in pH, or additional mutations. Furthermore, subtilisin BPN' and TR mutants have demonstrated the possibility of substrate assisted catalysis[131-133]. The H57A mutants hydrolyse peptide substrates with a histidine residue in either P_2 or P_1' position up to 300 times faster than those without[131-133]. Alternatively, raising the pH to ~10 for certain mutants may partially restore activity[32,134,135]. The k_{cat} of the D102N mutant of TR was reduced by 10^4 at neutral pH, but was restored to approximately 6% of wt-TR at pH 10[134]. This observation has been described as hydroxide dependence. Finally, the activity of the D102 mutant can be partially rescued by mutation of close contacts of D102[136]. S214 H-bonds with the sidechain of D102 in the wild-type enzyme and has been proposed to be a fourth member of the catalytic tetrad. The TR D102S mutant has a 10^4 lower k_{cat} compared to the wt. However, the k_{cat} of the D102S, S214E double mutant was 600-fold higher than the single mutant[136]. The double mutant with a S214D

mutation was also higher, but only by 10-fold. Thus, it was proposed that the sidechain carboxylates of the mutant 214 residues were able to act in place of D102, stabilising H57, to partially restore the catalytic activity[136].

1.4.3.2. The Oxyanion Hole

The oxyanion hole is a crucial feature of SPs, which contributes to stabilisation of the tetrahedral oxyanionic intermediate by donating H-bonds[125]. In CTR-like SPs, the oxyanion hole is formed from the backbone amides of S195 and G193[125], which are therefore not amenable to site-directed mutagenesis (Figure 1.10, A). However, many other SP families contain an oxyanion hole, including the subtilisin peptidases. The oxyanion hole of subtilisin BPN' appears to be predominantly formed by the sidechain amide of N155 and backbone amide of S221 (Figure 1.10, B)[118]. However, the hydroxy group of T220 also appears to play a role, although the distance between the group and the tetrahedral oxyanion is approximately 4 Å and is thus too distant to form a H-bond[118]. Modelling has suggested that the sidechain may rotate upon formation of the tetrahedral intermediate[137]. T220 mutants lacking a hydroxy group were calculated to exhibit approximately 2 kcal mol^{-1} less stabilisation energy on catalysis[138]. In comparison, subtilisin Asn155 was calculated to provide between approximately 2–5 kcal mol^{-1} of stabilisation energy, based on the decrease in k_{cat} upon mutation to Leu[139], Thr, Gln, Asp or His[140]. Interestingly, both residues appear to function cooperatively, potentially suggesting different mechanisms[118]. The double mutant (N155A and T220A) was calculated to reduce stabilisation by 6 kcal mol^{-1}, i.e., the approximate sum calculated for the single mutants[138]. By contrast, for the double mutant of N155A and the catalytic S221A the reduction in k_{cat} was similar to that of the S221A single mutant[130]. Analogously to the catalytic triad residues, the oxyanion hole mutations predominantly affected the turnover number while binding efficiency was only marginally altered[130,138-140].

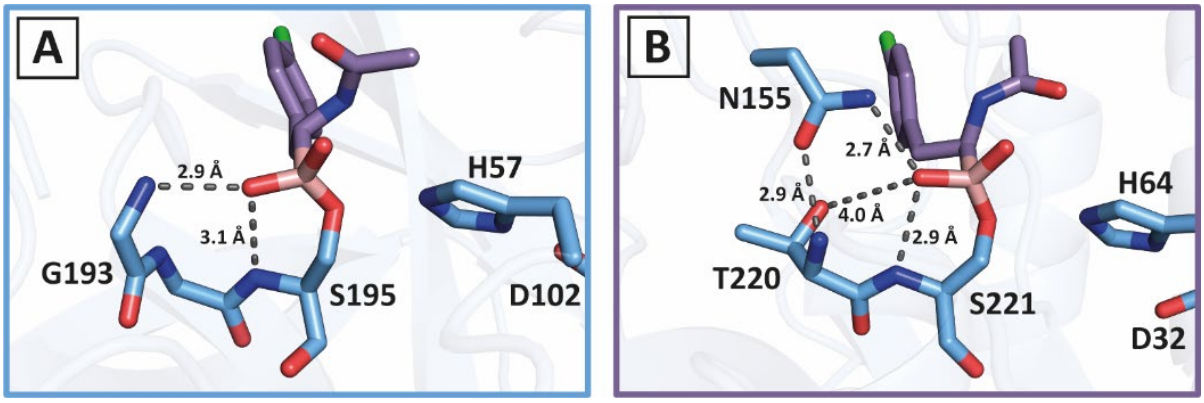


Figure 1.10 | Oxyanion holes of CTR and subtilisin complexed with boronic acid inhibitors. A) An active site view of the tetrahedral complex between CTR and L-para-chloro-1-acetamino boronic acid highlighting the oxyanion hole interactions (PDB: 1VGC); **B)** an active site view of the tetrahedral complex between subtilisin and D-para-chlorophenyl-1-acetamino boronic acid inhibitor highlighting the oxyanion hole interactions (PDB: 1AVT). The enzyme carbons are light blue, the inhibitor carbons are purple, and the tetrahedral boron is coloured pink.

1.4.4. Protons Involved in Catalysis

While the SP mechanism was proposed more than 50 years ago, many details of the reaction remain a matter of debate[118]. The location and movement of the protons during catalysis has been particularly problematic.

1.4.4.1. The H57-D102 H-Bond and the One- Versus Two-Proton Mechanism

A proton located between H57 (Nδ1) and D102 was first assigned in ^{1}H nuclear magnetic resonance (NMR) spectra by Robillard and Shulman in 1972[141]. The bond was assigned to a single low-field resonance (-15 to -18 ppm) due to the pH dependence, which demonstrated a pK_a of 7.2, and pH independence when the Nε2 of H57 was alkylated. As the signal remained when Nε2 of H57 was alkylated, this also indicated that the H-bond was on the Nδ1 of H57, as predicted by Blow *et al.*[117]. α-Lytic protease (αLP), a SP with homology to elastase, has a single His residue (H57). Due to this, αLP is amenable to enrichment with ^{13}C and ^{15}N histidine. In 1985, H57 ^{15}Nδ1 enriched αLP demonstrated the presence of the H57-D102 H-bond due to spin-spin coupling with the low field peak[142].

The two-proton transfer theory suggested that there was concerted transfer of protons from the S195 Oγ to the Nε2 of H57 and of the H57 Nδ1 proton to D102. Instead of a cationic H57 and anionic D102, both would be neutral. Based on NMR studies with H57 ^{13}Cε1 labelled αLP, Hunkapillar *et al.* proposed

the two-proton transfer mechanism[143,144]. Two pH-dependent chemical shifts were observed, with pK_a values of 3.3 and 6.7, and a C-H spin coupling that was interpreted to be indicative that a H57 imidazolium ion was only present below pH 3.8. From these results it was proposed that the pK_a values of H57 and D102 were essentially reversed (to 4.5 and 6.7, respectively) due to the microenvironment of the residues. However, much evidence has accumulated against the two proton transfer proposal. The previously mentioned ^{1}H NMR experiments[141] and ^{15}Nδ1 enriched αLP studies demonstrated that the pK_a of H57 is usually *ca*. 7 in the ground state[119]. Furthermore, using transition state (TS) analogues, the pK_a of H57 can increase to more than 10[120-123]. A neutron diffraction structure of TR with a tetrahedral intermediate analogue (monoisopropyl phosphoryl; MIP) also demonstrated that the proton was associated with H57 rather than D102[145]. While these studies do not explicitly preclude proton transfer during peptide hydrolysis, the one proton mechanism is generally accepted.

1.4.4.2. Low Barrier Hydrogen Bonds

The most common point of contention is the nature of the hydrogen bond formed between H57 and D102. Low barrier hydrogen bonds (LBHBs) are unusually strong (12–24 kcal mol^{-1}) and short (O–H–O = < 2.55 Å; O–H–N = < 2.65 Å; i.e., less than the sum of the van der Waals radii) H-bonds, where the hydrogen atom is delocalised and shared between the heteroatoms[146]. The importance of the LBHB in a SP was first proposed by Frey *et al.* in 1994[146]. It was proposed that formation of a LBHB would increase the reactivity of H57 as a general base and would therefore better facilitate nucleophilic attack of S195. The original definition of a LBHB required several features, including the heteroatom distance, equal pK_a values of the heteroatoms, equal sharing of the proton, and that the LHBH was sequestered from solvent[146]. The work built on the ^{1}H NMR studies of Robillard and Shulman[141], whereby Frey *et al.* suggested that the unusually low field chemical shifts (δ > 16 ppm) of the proton between H57 and D102 was indicative of a LBHB, when considered with other factors such as deuterium isotope effects[118]. However, much of this was disputed. For example, the pK_a values of H57 and D102 are not matched[119,147], and ^{15}N NMR demonstrated that the proton is 85% localised to Nδ1

when H57 is protonated (100% when it is neutral)[148]. Furthermore, site-directed mutagenesis of D102 argued against the presence, or at least the importance, of a LBHB[136,149]. Computational analyses of the energetics also favoured an electrostatic, stabilising role for D102[150]. LBHBs have nonetheless been assigned based on the H-bond heteroatom distance in X-ray crystal structures, but this can be misleading[151]. High-resolution X-ray and neutron diffraction crystallography has demonstrated that the hydrogens in short bonds may be localised. In a 0.78 Å structure of subtilisin, the hydrogen of the D32-H64 Nδ1 bond was located 1.5 Å and 1.2 Å from the heteroatoms, respectively[152]. Combined neutron and X-ray diffraction structures of PPE in complex with a tetrahedral intermediate analogue, demonstrated that while the H-bond length is indicative of a LBHB (2.6 Å), the hydrogen is localised to the H57 Nδ1 (X-ray = 0.8 Å; neutron = 0.96 Å)[90]. Thus, the authors argued in favour of a strong but conventional hydrogen bond (short ionic hydrogen bond (SIHB)) rather than a LBHB[90]. NMR evidence continues to accumulate for a strong H-bond between D102 and H57[124], but the actual nature of the bond remains a matter of debate.

1.4.4.3. The S195-H57 H-Bond

It is unlikely that S195 is deprotonated prior to attack on the amide bond, as the resultant alkoxide ion would likely be rapidly reprotonated by reaction with water[153]. The presence of a S195 Oγ-H57 Nε2 bond has, however, been questioned. In the early ^{1}H NMR work, Robillard and Shulman were unable to directly detect the proton, potentially due to solvent exchange[154]. However, an extra 0.6 ppm shift upfield at high pH after chemical modification of H57 was attributed to loss of the H57-S195 (N:---H–O) bond[154]. Blow *et al.* had also proposed the presence of this H-bond due to the distance between H57 and S195 (3.0 Å) in the original CTR crystal structure, despite performing crystallisation at pH 4.0 which would mean that H57 was protonated and likely a H-bond donor[117]. Later, the presence of the S195-H57 bond was confirmed, again utilising ^{15}N labelled αLP. This work demonstrated that the ^{15}Nε2 was affected by serine modification, but not the ^{15}Nδ1 of H57. The chemical shifts were observed only at high pH (9.0), when H57 is deprotonated, suggesting that Nε2 is the H-bond acceptor and that when

H57 is protonated it does not form H-bond interaction with S195, at least in the resting state[155]. However, there was discrepancy between the crystallographic and spectroscopic evidence for the H-bond in αLP. X-ray crystal structures of αLP suggested that the S195-H57 distance was too long (>3Å) for formation of a H-bond[156,157]. Solid-state NMR with ^{15}N labelled αLP crystals revealed that the discrepancy was caused by an increase in the pK_a of H57 (7.9) *in crystallo*, likely due to the high concentration of lithium sulphate (Li_2SO_4) required for crystallisation[158]. When crystals were examined at a higher pH, where H57 was neutral, the presence of the S195-H57 bond was confirmed[158].

1.4.4.4. The Moving Histidine Mechanism

The moving His or His-flip mechanism was proposed to explain how H57, after deprotonating S195, is able to protonate the leaving group of the tetrahedral intermediate. It was suggested that if the protonated H57 Nε2 remained close to S195, that protonation of S195 would be favoured over the leaving group[153]. Based on the observation that the αLP H57-D102 H-bond NMR signal disappeared at low pH in the presence of tetrahedral analogue inhibitors. It was proposed that the imidazolium ring of H57 flipped placing the Nδ1 closer to the leaving group[155]. However, there are several problems with the His-flip mechanism as highlighted by Hedstrom[118]. Firstly, H57 is surrounded by the P_2 and P_1' residues of the substrate and other enzyme residues (55 and 58) producing steric hinderance. A His-flip may also place the proton further from the leaving group based on analysis of tetrahedral intermediate analogue structures[159,160]. Finally, while the exact nature of the H57-D102 bond is unknown, evidence suggests a strong H-bond exists between H57 and D102 which would need to be broken to enable a ring flip[124]. Instead, the transfer may simply be facilitated by small movements of H57 that do not necessitate that the H57 Nδ1-D102 H-bond is broken[118].

1.4.5. Structural Evidence for the Acyl-Enzyme and Catalytic Water

The acyl-enzyme intermediate is one of the more accepted aspects of the SP mechanism. The formation of an acyl-enzyme was inferred prior to the mechanistic proposal by Blow *et al*.[117]. The reaction of CTR with *para*-nitrophenol esters produced an initial rapid 'burst' of hydrolysis which was

approximately equivalent to the enzyme concentration[161]. Subsequent hydrolysis was slower and linear. The observation was justified by rapid acylation of the enzyme, which liberated the nitrophenol group, followed by a rate-limiting deacylation step. Shortly after Blow's proposal, the presence of a common acyl-enzyme intermediate during, peptide, ester, and anilide hydrolysis was demonstrated[162]. Acetyl-phenyl (AcPhe) derivatives (-*O*-methyl, -*p*-trimethylammonium anilide, or -AlaNH_2) were hydrolysed in the presence of nucleophiles (Nu) which produces mixed acyltransfer products (AcPhe-COOH or AcPhe-Nu). Should a common intermediate exist, the partition ratio between the products would be the same irrespective of the initial substrate bond, as was observed[162].

The first acyl-enzyme X-ray crystal structures were formed between peptidases and ester substrates. These were often 'poor' substrates with low catalytic activity that were crystallised at low pH to further reduce the rate of hydrolysis. A structure of a CTR-α indoleacryloyl complex was reported by Henderson in 1970[125]. The use of the poor substrate 'trapped' the acyl-enzyme complex, however, it also bound in an essentially non-productive manner, explaining the poor activity. Both the Oγ of S195 and H57 had shifted out of position relative to the apo-structure (2.5 Å and 0.3 Å, respectively). A potential hydrolytic water molecule that could effect deacylation was within H-bonding distance to H57. However, due to the distorted conformation of the complex, the water was not in position to attack the carbonyl carbon[125]. Despite this, the existence of the oxyanion hole was proposed. It was noted that a 45˚ rotation of the complex would place the carbonyl oxygen within bonding distance of the backbone amides of S195 and G193[125]. Additionally, Henderson proposed that these interactions would be strengthened in the tetrahedral intermediates and would therefore stabilise the intermediate[125].

In 1989, Dixon and Matthews reported the first peptidyl acyl-enzyme complex formed with γ-CTR[163]. The tetrapeptide was identified as a potential γ-CTR autolysis product, with density consistent with Pro-Gly-Ala-Tyr. It was proposed that this corresponded to residues 225-228 (Pro-Gly-Val-Tyr) with the isopropyl group of V227 in free rotation[163]. The P_1 Tyr sidechain was present within the S_1 specificity

pocket and there was evidence of H-bonds between the backbones of the tetrapeptide and G216. The carbonyl oxygen was not 'fully' located within the oxyanion hole, and only in position to form a H-bond with the amide of G193 (2.79 Å), although high B-factors (46.5 $Å^2$) suggested some movement in the position[163]. A potential hydrolytic water was observed H-bonded to H57 (3.02 Å), although the angle of attack on the carbonyl was not ideal (71˚ rather than 107˚)[164]. The authors noted that were the complex rotated to orientate the carbonyl oxygen in the oxyanion hole, the water would be approaching from a more ideal angle[163]. Shortly after, the same group demonstrated that the S195 Oγ-carbonyl carbon distance reduced from 1.82 Å at pH 10.5 to 1.54 Å at pH 2.0[165]. This was interpreted as an acyl-enzyme complex at low pH and enzyme-product complex at high pH, demonstrating that the complex was able to turn over[165].

The three-dimensional structure of the CTR-like bacterial SP, *Streptomyces griseus* Proteinase A (SGPA), was first reported in 1978[166] and to high resolution (1.8 Å) in 1979[167]. During molecular dynamics (MD) simulations on the high-resolution structure, several phosphate ions (deriving from the precipitant) migrated into a solvent channel[168]. To investigate further, the group solved the 3D-structure of SGPA after soaking in the crystals in potassium arsenate (KH_2AsO_4)[169]. Unexpectedly, in the KH_2AsO_4 soaked structure additional active site density was observed and was interpreted as a tetrapeptide acyl-enzyme complex with S195. The stability of the SGPA acyl-enzyme complex was justified by the low pH (4.1) of the crystallisation conditions which would reduce the catalytic activity of SGPA[169]. An unusual tetrapeptide was modelled into the density, Gly-Ala-Ser-(β-hydroxy)Asp (P_4-P_1), although in the discussion the density was assigned to a mixture of oligopeptides, likely with P_1 Leu, Val, and Phe residues[169]. The oligopeptides likely derived from the solution that SGPA is prepared from (Pronase). Nonetheless, unlike the peptidyl acyl-enzyme complexes reported by Dixon *et al.*[163,165], the SGPA complex was modelled with the carbonyl oxygen orientated within the oxyanion hole (2.6 Å and 3.0 Å to Gly193 and S195 amides, respectively). A hydrolytic water was assigned close to the H57 Nε2

(3.1 Å) and the carbonyl carbon (2.7 Å), and with a better angle of attack (89˚). A second water molecule was located 1.7 Å from this water, each being assigned to 50% occupancy[169].

In 1997, the unusually stable acyl-enzyme complex of PPE and a heptapeptide human β-casomorphin 7 (BCM7) was reported at high resolution[106,170]. This complex was demonstrated to be catalytically active, and unlike previous structures the composition was known. The carbonyl oxygen of the peptide was orientated in the oxyanion hole, and a candidate hydrolytic water was in a good position for attack on the carbonyl carbon (Figure 1.11, A). This complex will be discussed in more detail in Chapter 2.

Atomic resolution acyl-enzyme complexes of bovine TR and peptide-paranitroanilide (pNA) substrates have also been reported (Figure 1.11, B & C)[171]. The structure of the acyl-enzyme complexes were obtained under pseudosteady-state conditions with two substrates that were apparently rate limited by the deacylation reaction: succinyl-Ala-Ala-Pro-Arg-pNA (AAPR) and succinyl-Ala-Ala-Pro-Lys-pNA (AAPK)[171]. The carbonyl oxygens of both complexes were H-bonded to the oxyanion amides. However, the refined water-carbonyl carbon distance in the TR:AAPK structure was substantially shorter than the other TR:AAPR complex or the PPE:BCM7 complex (2.5 Å Vs 3.0 Å and 3.2 Å, respectively).

In theory, stabilisation of the tetrahedral intermediate by enzyme interactions, such as those with the oxyanion hole, should place strain on the ground state[118]. Note that the 'ground state' here could refer to either the E:S or acyl-enzyme complexes. Resonance Raman spectroscopy with acyl-enzyme complexes of CTR and subtilisin revealed multiple carbonyl features that were interpreted as being indicative of multiple conformations, therefore supporting the carbonyl strain theory[172,173]. While the ester links in high resolution crystal structures of the PPE:BCM7 and TR:AAPR acyl-enzyme complexes were both refined as being planar (Figure 1.11, A & B)[170,171], the acyl-enzyme complex of TR:AAPK was refined in a distorted manner (Figure 1.11, C). The carbonyl carbon was 0.2 Å out of plane which added an additional 13˚ to the wat–C–$O_{carbonyl}$ angle[171]. The distortion and the short water-carbonyl carbon distance (2.5 Å) was attributed to the structure being further along the deacylation pathway, where

the water had approached the carbonyl carbon leading to the minor pyramidal distortion of the carbonyl[171]. However, other factors must be considered. The TR crystals were soaked under different conditions (AAPR: Bis Tris pH 6.0 Vs AAPK: Na citrate pH 4.0), the occupancy of the peptide was 30% lower than AAPR (~70% Vs ~100%, respectively), and there were alternate and unusual conformations of H57 and S195 in the AAPK structure (Figure 1.11, C). Whether the TR:AAPK structure truly represents a midpoint along the trajectory requires further validation.

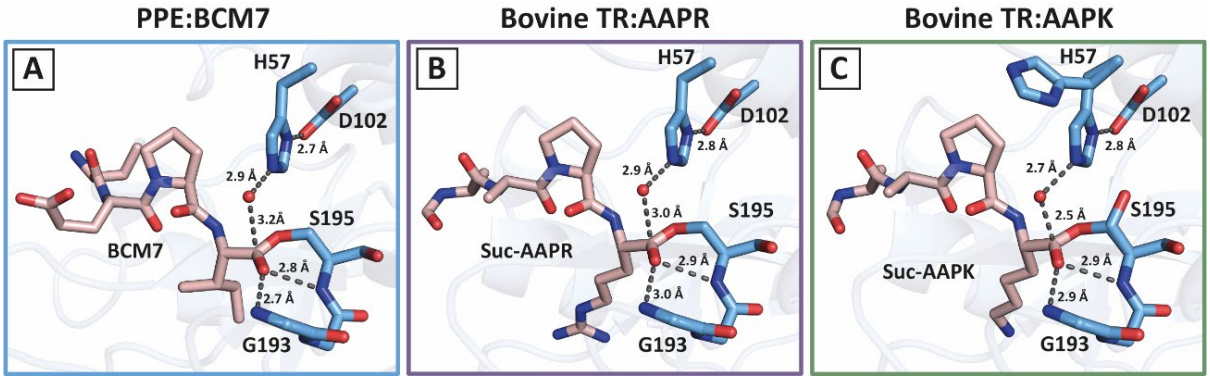


Figure 1.11 | Active site views of peptide acyl-enzyme complexes of PPE and bovine TR. A) An acyl-enzyme complex between PPE and YPFVEPI (PDB: 1HAX); **B)** an acyl-enzyme complex between bovine TR and Suc-AAPR (PDB: 2AGE); **C)** an acyl-enzyme complex between bovine TR and Suc-AAPK (PDB: 2AGG). The enzyme carbons are light blue, the peptide substrate/inhibitor carbons are pink, and waters are shown as red spheres.

1.4.6. Structural Evidence for the Tetrahedral Intermediate

Transition state stabilisation is a fundamental aspect of enzyme catalysis as suggested by Pauling[174,175]. Site-directed mutagenesis on the oxyanion residues of subtilisin and inhibition of SPs using tetrahedral analogues has indicated that the oxyanion hole interactions provide approximately 6 kcal mol^{-1} of stabilisation to the tetrahedral intermediate[118,120]. However, tetrahedral intermediates exhibit extremely short lifetimes, with formation and breakdown of the intermediate being essentially concerted with the best leaving groups[118]. Unsurprisingly, structural evidence for the accumulation of tetrahedral intermediates has been controversial. The evidence supporting the formation of a tetrahedral intermediate derives from solution chemistry[176] and from the crystal structures of SPs with TS analogues. Some of the earliest known SP inhibitors (DIFP and phenylmethylsulfonyl fluoride (PMSF)) form tetrahedral complexes, however, due to the large sulphur and phosphorus heteroatoms,

they are likely poor analogues of the catalytic tetrahedral intermediate[109]. Peptide boronic acid and aldehyde derivatives, are both better analogues of the tetrahedral intermediate (Figure 1.12, A & B)[109]. Note that the former carries a negative charge at the boron heteroatom rather than on the carbonyl oxygen in the true intermediate. Consistent with the stabilisation of the TS, these inhibitors can exhibit high inhibitory potential with K_i values in the nanomolar range[177].

There are, nonetheless, several reports of tetrahedral intermediates observed *in crystallo* with active substituents. Following the discovery of the tetrapeptide acyl-enzyme complex of γ-CTR[163], Yennawar *et al*. soaked crystals of γ-CTR in hexane[178]. This was in response to reports that nonaqueous solvents could modify the specificity of enzymes[179,180] or alter the thermodynamic favourability of reactions[181]. Soaking γ-CTR crystals in hexane produced additional density in the P_1' subsite and surrounding the carbonyl carbon[178]. The extra density was interpreted as an Asp residue, with the amino group having attacked the carbonyl carbon to form a tetrahedral intermediate, essentially in reverse of the acylation reaction (Figure 1.12, D). It was suggested that hexane had altered the equilibrium to favour binding of Asp (or P_1') with the S_1' subsite, as had been observed previously[181]. The exact origin of the Asp nucleophile is unknown but was likely derived from autolytic cleavage activity of CTR, and the Asp may be the *N*–terminal residue of an oligopeptide that only interacts with the S_1' subsite[181].

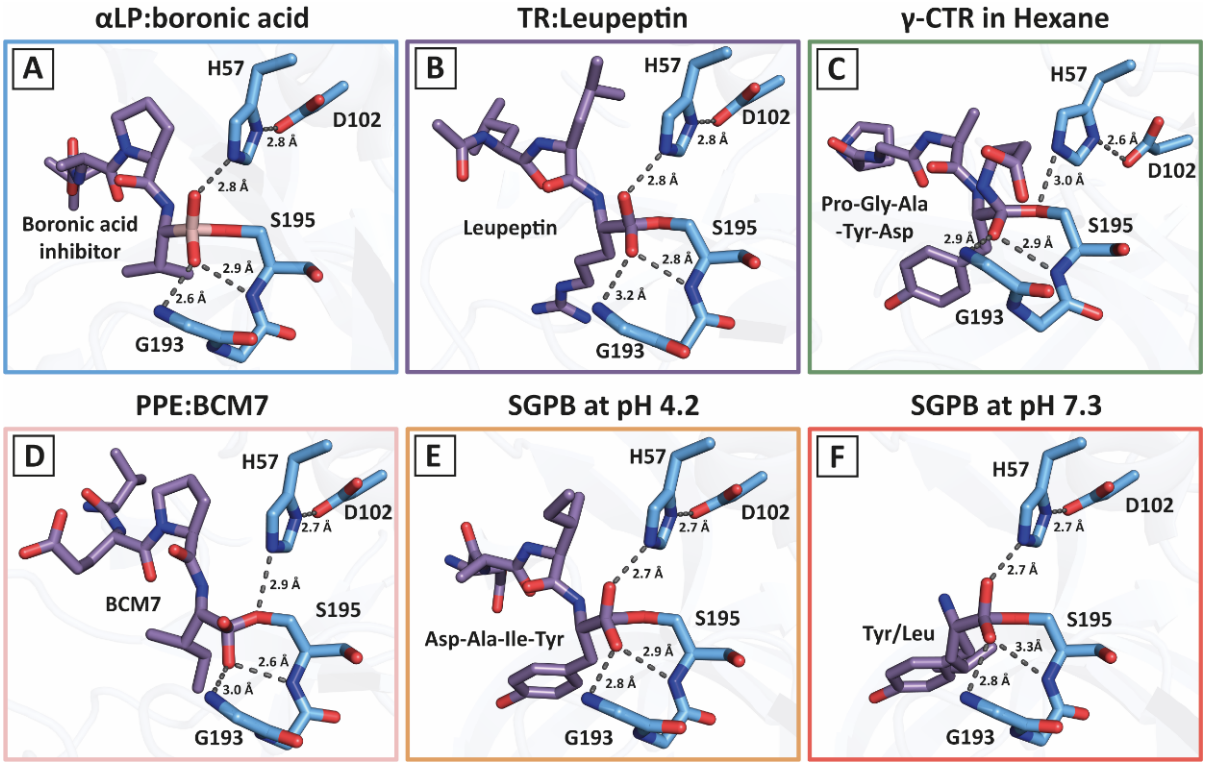


Figure 1.12 | Active site views of reported SP tetrahedral complexes. A) A covalent complex between αLP and a peptidyl boronic acid inhibitor (PDB: 1P01); **B)** a covalent complex between bovine TR and leupeptin (PDB: 2AGI); **C)** a covalent complex between γ-CTR and a pentapeptide (PGAYD) soaked in hexane (PDB:1GMD); **D)** a covalent complex between PPE and BCM7 (PDB: 1HAZ); **E)** A covalent complex between SGPB and a tetrapeptide (DAIY) at pH 4.2 (PDB: 2QA9); **F)** A covalent complex between SGPB and a peptide (Y/L) at pH 7.3 (PDB: 2QAA). The enzyme carbons are light blue, the substrate/inhibitor carbons are purple, and boron atoms are pink.

Crystals of the catalytically active PPE:BCM7 acyl-enzyme complex[106] were subject to 'pH-jumps' by Wilmouth *et al.* in 2001[170]. The resultant electron density in one of the structures was consistent with a tetrahedral intermediate (Figure 1.12, E), however, due to the presence of sulphate in the crystallisation solution, the assignment has been questioned[118]. This intermediate will be discussed in more detail in Chapter 2.

A subsequent tetrahedral intermediate was proposed in a complex of Streptogrisin B (SGPB) and peptides of unknown origins[182]. Two structures were refined as tetrahedral intermediates at different pH values (4.2 and 7.3; Figure 1.12, F & G). At pH 4.2 there was continuous density for a tetrapeptide bound in an antiparallel β-sheet, modelled as Asp-Ala-Ile-Tyr (P_4 to P_1), while at pH 7.3 only disjointed density was present[182]. A glycerol molecule was modelled into the S_2 subsite, and two amino acids (Tyr and Leu) were modelled into the S_1 subsites in an equal ratio. As with the γ-CTR tetrahedral

intermediate, the peptides likely derive from an autocatalytic origin[182]. Thus, the exact composition of the peptides is unknown. While these complexes are proposed to be tetrahedral, the interpretation of the electron density maps is not convincing. At pH 4.2, the complex is more indicative of a E:P structure as the electron density between the carbonyl carbon and S195 Oγ is broken (Figure 1.13, A). The electron density surrounding the carbonyl carbon also appears to be planar rather than tetrahedral and is therefore more consistent with a carboxyl group in a product (Figure 1.13, A). At pH 7.3, there is continuous density between the carbonyl carbon and S195 Oγ in the 2*Fo*–*Fc* map, but there is also negative density across this bond in the *Fo*–*Fc* map (Figure 1.13, B). Additionally, there is a large area of positive density above the carbonyl group as displayed in Figure 1.13, B, which is unaccounted for in the model. Thus, both of the proposed tetrahedral intermediates are tenuous.

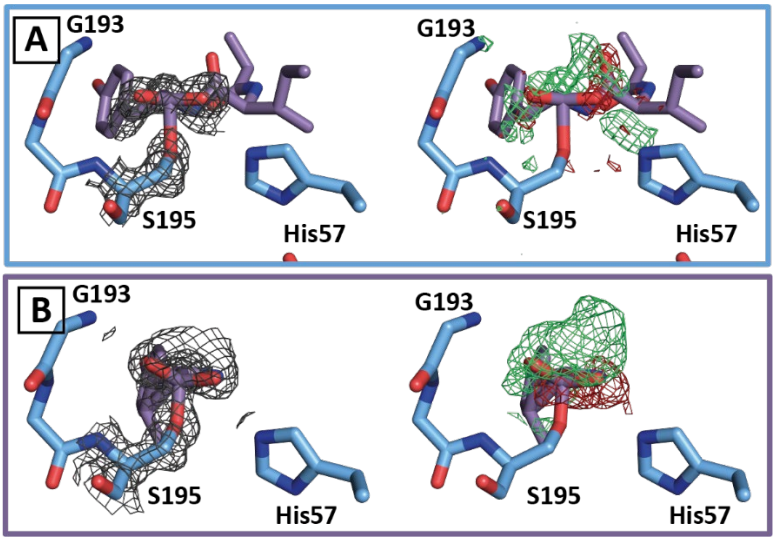


Figure 1.13 | Active site views of SGPB tetrahedral structures with 2*Fo*–*Fc* and *Fo*–*Fc* maps. A) A complex between SPGB and a tetrapeptide (DAIY) at pH 4.2 (PDB: 2QA9); **B)** a complex between SPGB and a peptide (Y/L) at pH 7.3 (PDB: 2QAA). The enzyme carbons are light blue, the substrate carbon are purple, the 2*Fo*–*Fc* maps shown as black mesh are contoured to 1 σ, and carved around the sidechain of S195 and the carbonyl atoms (C, Cα, $O_{carbonyl}$), and the *Fo*–*Fc* maps shown as green and red mesh are contoured to 3/–3 σ, respectively.

1.5. X-Ray Free Electron Lasers (XFELs) and Time-Resolved Crystallography

1.5.1. Background

Since the first report of a protein structure solved by X-ray crystal diffraction in 1958, X-ray crystallography has become one of the most important techniques for structural analysis[183]. There are

now more than 173,000 protein structures deposited in the PDB (as of March 2023) solved by X-ray diffraction alone (85.7% of the total structures). In the 1990s, the field of biomacromolecule crystallography started to look much as it does today. Third-generation synchrotrons with substantially brighter beams compared with previous sources enabled atomic-resolution diffraction. However, the high intensity beams also generate prohibitive levels of radiation damage in samples[184]. The diffusion of free radicals formed from the ionising radiation can be sufficiently slowed to enable high-resolution structure solution by performing data collection at a cryogenic temperature (~100 K)[185]. The combination of these two techniques has dominated biomacromolecule X-ray crystallography ever since. However, there are limitations. Radiation damage is not completely eradicated and cryocooling can alter the observed conformation states of proteins leading to unidentified mechanistic details[186,187]. Additionally, freezing samples limits the ability to perform time-resolved experiments.

X-ray free electron laser (XFEL) radiation sources offer the opportunity to perform high-resolution, time-resolved, crystallography studies to probe the mechanisms of enzymes at ambient temperature. XFEL beams have much higher brilliance and shorter pulse lengths compared with third generation synchrotrons, which facilitates high-resolution data collection[187]. Approximately the same number of photons as are delivered from a synchrotron beam in 1 s as can be delivered into the sample with a 1 to 10 fs XFEL pulse, although the actual dose varies between sources[188]. The extremely high dose of ionising radiation destroys the sample, but diffraction 'out-runs' most of the effects, and thus cryocooling is unnecessary[187]. This method is known as diffraction-before-destruction[189]. As the sample is destroyed after a single shot, diffraction data are collected from tens of thousands of crystals in random orientations, and subsequently merged to form a structure[187]. The methodology of serial data collection using XFEL radiation sources is known as serial femtosecond crystallography (SFX)[190,191]. Importantly, the experimental setup can additionally facilitate temporal control over mixing or activation times resulting in the production of snapshots of the enzyme reactions.

1.5.2. Examples of Time-Resolved SFX Experiments

Many sample delivery and reaction initiation approaches are now available for time-resolved crystallographic studies[192]. The early time-resolved experiments were predominantly focused on pump-probe activation methodology. Microcrystals of enzymes in the ground state (e.g., dark) are first illuminated by an optical 'pump' to initiate the reaction. Subsequently, after a specific time delay, the crystals are illuminated with the X-ray 'probe' for diffraction data collection[192]. The pump-probe technique enables ultra-fast data collection following reaction activation (fs-ms)[193,194]. The photoinduced conformational changes induced in chromophores and associated residues of rhodopsins[193-196], green fluorescent protein (GFP)[197,198], and photosystem II (PSII)[199-201] are among the most common targets investigated for pump-probe activation. The versatility of the pump-probe method is exemplified by successive publications studying rhodopsins. Nogly *et al*. and Nass Kovacs *et al.* followed the rapid photo-induced conformational changes of all-*trans* to 13-*cis* retinal of the *Halobacterium salinarum* NRC-1 proton pump from femto- to picosecond timeframes[193,195]. Skopintsev *et al*. subsequently followed the reaction of *Krokinobacter eikastus* Rhodopsin 2 (KR2) from 800 fs all the way up to the millisecond timeframe using pump-probe methodology[194]. The movements of waters (ps), and conformational changes of sidechain and backbone atoms (ns-ms) were observed[194].

Alternatively, reactions can be initiated for time-resolved studies through the addition of substrates or co-factors, referred to as chemical mixing or diffusion methodologies[192]. This technique requires microcrystal slurries prepared in the absence of a component required for reaction, i.e., a substrate or inhibitor, or a co-factor such as oxygen. To start the reaction the crystal slurry is mixed with or exposed to the initiator solution/atmosphere, and data is collected after a time-delay[192]. Olmos *et al*. utilised the mix and inject serial crystallography (MISC) technique to monitor cleavage of the cephalosporin antibiotic, ceftriaxone, by a *Mycobacterium tuberculosis* β-lactamase (BlaC)[202]. The first structure of an

oxygen intermediate of cytochrome *c* oxidase (C*c*O) was captured by MISC using oxygen saturated buffer as the reaction initiator[203].

Alternatively, samples can be mixed after injection as for the acoustic droplet ejection (ADE) on-tape method[204]. With the ADE tape technique, crystal slurries are ejected onto a conveyor belt. The conveyor belt delivers the slurry through a reaction region before arriving at the diffraction region, with the belt speed able to determine the mixing time. Isopenicillin *N* synthetase (IPNS) crystals have been probed using this technique[205]. Anaerobically prepared slurries containing all reaction components, except oxygen, were passed through an oxygen reaction chamber before diffraction data was collected. Binding of O_2 to the catalytic iron, led to correlated conformational changes within the active site and in distal regions of the enzyme[205]. These conformational changes altered the active site cavity to enable cyclisation of the substrate[205]. The ADE on-tape method has an additional benefit for metalloproteins whereby X-ray emission spectroscopy data can be collected concomitantly to diffraction data which provides data on the oxidation state of the catalytic metal atoms[205-208]. Data collection for these mixing experiments is presently limited to millisecond to second timescales but can clearly still provide useful mechanistic data for many enzyme systems.

1.6. Thesis Objectives

The work presented herein was comprised of biochemical studies aimed at investigating the mechanism of PPE and performing preliminary inhibition studies on PPE as a model for HNE and other SPs.

The aim of the work described in Chapter 2 was to further characterise the reported tetrahedral intermediate captured from the crystal complex of PPE and the catalytically active heptapeptide (BCM7). Novel crystallisation conditions were sought to remove sulphate from the precipitant and prior experiments were reproduced under these novel conditions. In addition, preparations were made for conducting time-resolved serial crystallography experiments using modern techniques.

Alternate options for time-resolved crystallography with PPE were explored as described in Chapter 3. More potent derivatives of BCM7 are first explored, followed by the only approved synthetic inhibitor of HNE. Attempts are made to crystallise each of the inhibitors with the ultimate goal of performing time-resolved crystallographic experiments with the enzyme:inhibitor complexes.

Site-directed mutagenesis is an invaluable tool for probing the mechanistic details of enzymes. Thus far, experiments of this nature have not been conducted with PPE. Therefore, the work presented in Chapter 4 focused on the recombinant expression and purification of PPE from *Escherichia coli*.

The aim of the work described in Chapter 5 was to perform preliminary inhibition studies on novel derivatives of a γ-lactam scaffold which has previously demonstrated inhibitory potential against PPE and HNE, and to assess the structure-activity relationships of these inhibitors by means of dose-response assays and crystallographic structure assessment. Subsequently, similar experiments were performed with a benzoxaborole scaffold to analyse the inhibitory potential against PPE.

Chapter 2: Time-Resolved Crystallography with PPE

2.1. Introduction

2.1.1. β-Casomorphins (BCMs)

Due to the widespread consumption of dairy products, especially bovine milk, research has been dedicated to developing our understanding the potential health benefits and adverse effects of milk protein derivatives. Up to 80% of the proteinaceous content of bovine milk is comprised of casein[209]. β-Casomorphin (BCM) peptides are derived from β-casein and have attracted much attention due to their opioid receptor agonist activity[210-212]. The consumption of β-casein and the subsequent release of BCMs has been implicated with the increased risk of various human non-transmissible diseases, including autism[213], sudden infant death syndrome (SIDS)[214], diabetes mellitus type-I (DM-1)[215], cardiovascular diseases[216]. In 2009, the European Food Safety Authority (EFSA) released a comprehensive review of the available data, looking in particular at the alleged negative effects of BCM7 (*vide infra*) on human health[217]. The report concluded that a cause-effect relationship could not be established due to a lack of evidence, the use of poor models, and often contradictory results[218-220].

A heptapeptide termed BCM 1-7 (BCM7) was the first BCM and the first exogenous opioid peptide to have its peptide sequence revealed following tryptic digests of bovine β-casein[221]. Subsequently, natural and synthetic BCMs of various lengths have been discovered with varying potency towards opioid receptors (Table 2.1)[222-225]. Analogous BCMs with opioid agonist activity have been isolated from the milk of other mammals, including sheep[221], water buffalos[226], and humans[223]. Bovine BCM7 corresponds to Tyr60–Ile66 of β-casein and is released by the proteolytic activity of pepsin (cleaves Leu58–Val59), leucine amino peptidase (LAP; cleaves Val59–Tyr60), and pancreatic elastase (cleaves Ile66–His67)[224]. The peptides exhibit a common *N*–terminal motif (YXF), in which the tyrosine is crucial for opioid agonist activity[227]. This differs from the typical endogenous opioid receptor agonist motif (YGGFM/L) exhibited by enkalphins, endorphins and dinorphins[228]. In general, BCMs preferentially target μ–opioid receptors, although the peptides can exhibit agonist activity towards δ– and κ–opioid

receptors (agonist activity: μ > δ >> κ) [210-212]. Generally, human and bovine BCM opioid agonist activity follows a similar pattern depending on peptide length (BCM5 > BCM8 > BCM4 > BCM7)[212].

Table 2.1 | Bovine and Human β-casomorphin sequences. Table adapted from Jinsmaa and Yoshikawa (1999)[224], and Nyberg *et al*. (2013)[225].

Name	Sequence
Bovine BCM4	^{60}Tyr-Pro-Phe-Pro^{63} (YPFP)
Bovine BCM5	^{60}Tyr-Pro-Phe-Pro-Gly^{64} (YPFPG)
Bovine BCM7	^{60}Tyr-Pro-Phe-Pro-Gly-Pro-Ile^{66} (YPFPGPI)
Bovine BCM8	^{60}Tyr-Pro-Phe-Pro-Gly-Pro-Ile-His/Pro^{67} (YPFPGPIH)
Bovine BCM21	^{60}Tyr-Pro-Phe-Pro-Gly-Pro-Ile-His/Pro-Asn-Ser-Leu-Pro-Gln-Asn-Ile-Pro-Pro-Leu-Thr-Gln-Thr^{80} (YPFPGPIHNSLPQNIPPLTQT)
Human BCM4	^{51}Tyr-Pro-Phe-Val-Glu^{55} (YPFVE)
Human BCM5	^{51}Tyr-Pro-Phe-Val-Glu-Pro^{56} (YPFVEP)
Human BCM7	**^{51}Tyr-Pro-Phe-Val-Glu-Pro-Ile^{57} (YPFVEPI)**
Human BCM8	^{51}Tyr-Pro-Phe-Val-Glu-Pro-Ile-Pro^{57} (YPFVEPIP)
Morphiceptin	Tyr-Pro-Phe-Pro-NH_2

2.1.2. PPE and BCM7

In 1997, human BCM7 was shown to form a stable acyl–enzyme complex with PPE at pH 3.5, initially in studies employing electrospray ionisation mass spectrometry (ESI-MS)[106]. Formation of the complex was consistent with loss of a water and peptide fragmentation mass spectrometry (FAB-MS/MS) studies indicated that the reaction was mediated through the *C*–terminal Ile7 residue of BCM7[106]. ^{18}O exchange experiments revealed that at pH 7.5 the reaction was reversible, and the use of PMSF-inhibited PPE suggested that the reaction was mediated through the active site nucleophilic serine[106]. BCM7 was also found to be a competitive inhibitor of PPE (Ki = 0.76 mM). X-ray crystallographic analysis of the PPE:BCM7 complex at pH 5.0 (PDB: 1QIX[106]; 1HAX[170]) exhibited continuous density from Ile7 of BCM7 to S195 of PPE that was consistent with formation of an ester linkage (Figure 2.1)[106]. BCM7 formed an antiparallel β-sheet with PPE (S214-V216); only electron density corresponding to the four *C*–terminal residues of BCM7 (VEPI) was observed in the active site. The three *N*–terminal residues (YPF) were likely forced out of the active site cleft due to a symmetry interaction and as a consequence were disordered[106]. Binding of BCM7 also led to apparent rotation of the sidechain of R217A out of the

active site. A potential 'hydrolytic' water was observed equidistant from the ester carbon and Nε2 of H57 (3.02 Å)[106].

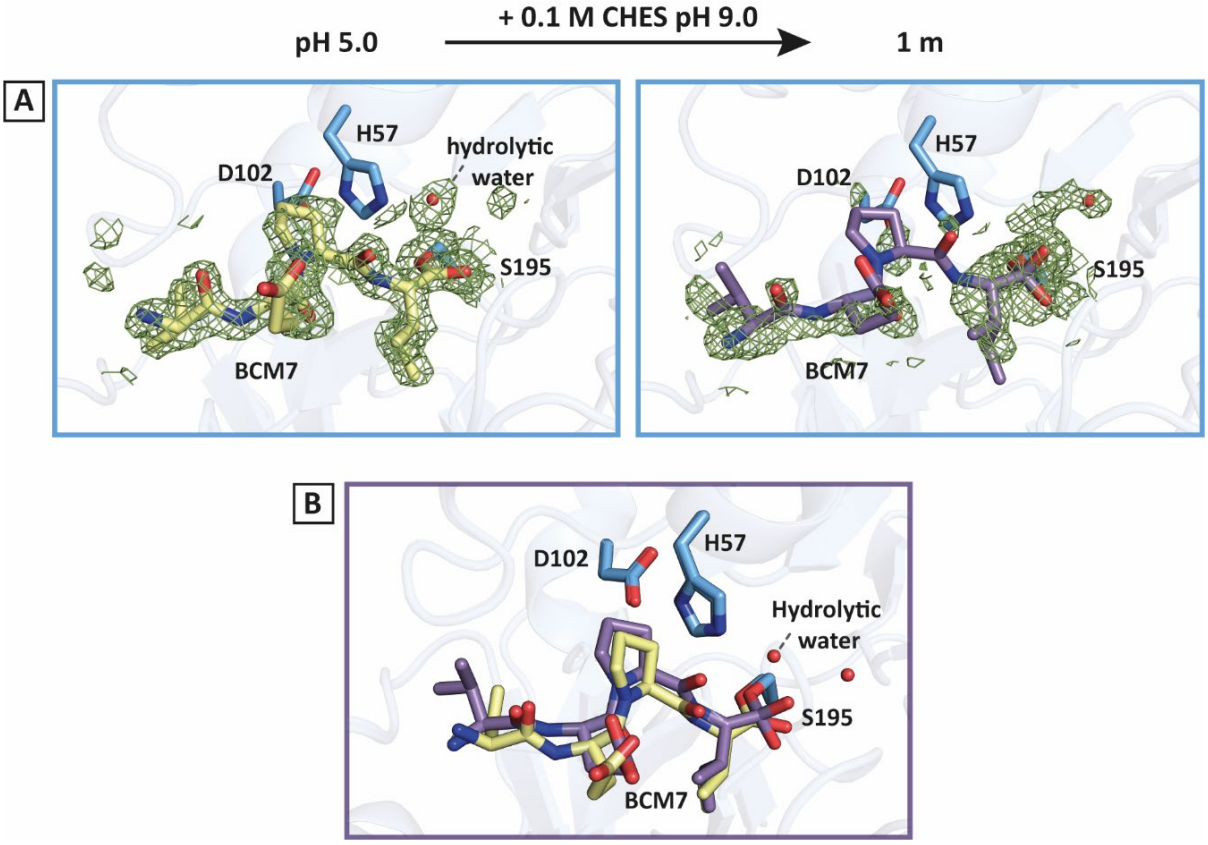


Figure 2.1 | Comparison of the reported crystal structures of the PPE:BCM7 complex before and after soaking in high pH buffer. A) An active site view of the PPE:BCM7 complex at pH 5 (PDB: 1HAX; left) and after soaking at pH 9 for 1 min (PDB: 1HAZ; right); **B)** superimposition of 1HAX and 1HAZ. The enzyme carbons are shown in light blue, BCM7 carbons at pH 5 are yellow and at pH 9 are purple. The 2*Fo*-*Fc* electron density maps shown as green mesh are contoured to a 1 σ confidence level and carved around S195, BCM7, and the 'hydrolytic' water (red spheres). The figure was adapted from Wilmouth *et al*. (2001)[170].

The unexpected stability of the PPE: BCM7 complex was originally attributed to a combination of low pH and the interaction of Q192 and Glu5 of BCM7 (P_3), an interaction that was confirmed *in crystallo*[106]. At pH 5.0, the imidazole group of H57 should be protonated, diminishing the ability of His57 to act as a general base and deprotonate the hydrolytic water required for deacylation, and thus reducing the rate of reaction. The possibility that a polar group in the P3 position may favour productive binding was first proposed by Meyer *et al*. (1988) following capture of a high resolution (1.6 Å) crystal structure of apo-PPE revealing the presence of Q192 within the S3 specificity pocket[88]. Mutation of Glu5 of BCM7 to Ala (BCM7–E5A) and Gln (BCM7–E5Q), increased the K_i for PPE (>4 mM and 1.8 mM, respectively)

supporting the proposal that this residue was important for stability (Table 2.2). Furthermore, bovine BCM7, which contains a P3 glycine, did not inhibit PPE. However, due to differences at the P4 position the lack of inhibitory activity could not definitively be attributed to the P3 glycine. Contrastingly, Q192 of PPE is replaced by F192 in human neutrophil elastase (HNE) and would therefore be unexpected that a Glu or Gln would result in favourable binding[106]. It is therefore likely that the stability of the O–acyl intermediate is only partially mediated through this interaction[106].

Table 2.2 | Kinetic analysis of BCM7 truncations and P3 mutants. The table was adapted from Wilmouth *et al*. (1997)[106]. All peptides presented had a *C*–terminal carboxylic acid.

Sequence	PPE Ki (mM)	HNE Ki (mM)
NH_2-YPFV**E**PI	0.76	0.24
NH_2-YPFV**A**PI	> 4	0.89
NH_2-YPFV**Q**PI	1.8	0.11

2.1.3. "Time-Resolved" Crystallography with PPE:BCM7

Following capture of the catalytically active PPE:BCM7 acyl–enzyme complex *in crystallo*, attempts were made to capture the proposed transient tetrahedral intermediate of the deacylation reaction by inducing a pH-jump across the crystal[170]. Increasing the pH of the crystal alters the donor: acceptor relationship between H57 and the hydrolytic water, thereby enabling H57 to activate the water, and the subsequent attack of the water on the ester carbon. To increase the pH, crystals of the PPE:BCM7 complex, prepared at pH 5.0, were soaked in a 'pH-jump solution' before freezing the crystals in liquid nitrogen (N_2(l)).

Three 'pH-jumped' structures were reported (1HAZ: pH 9.0, 1 min; 1HAY: pH 10, 10 s; and 1HB0: pH 10, 2 min). At pH 10, a 10 s soak was found to be too short with no observed changes to the structure, while a 2 min soak led to complete deacylation of the active site[170]. However, soaking in buffer at pH 9 for 1 min, led to changes that were suggestive of a tetrahedral intermediate with continuous density to S195 (Figure 2.1). Electron density for all the residues of BCM7, except Ile7, was poor. This was attributed to synchronised hydrolysis and product release, whereby the change in geometry about the

carbonyl occurs concomitant to movement of the peptide and disruption of the β-sheet H-bonds linking BCM7 and PPE[170]. A small movement in H57 (0.2–0.4 Å) was linked to the movement in the same β-sheet through a H-bond between the Oγ of S214 and one of the sidechain oxygen atoms of D102[88].

The 'oxyanion' was located within the oxyanion hole (2.6 Å and 2.9 Å from the backbone amides of S195 and G193, respectively) while the hydroxy oxygen of the intermediate was also located close to the oxyanion hole (3.9 Å and 3.1 Å from the backbone amides of S195 and G193, respectively). While this exciting result was one of the first proposed tetrahedral intermediates captured *in crystallo* with a SP, the assignment has been questioned[118]. Of particular concern was the presence of sulphate within the precipitant solution. Sulphate associates with S195 and H57 in apo-PPE structures and could be misinterpreted as a tetrahedral structure[88,89]. Wilmouth *et al*. (2001) addressed this and suggested that because sulphate is an H-bond acceptor, association with H57 at higher pH would be less favourable[170]. However, this doesn't preclude association and a 1.1 Å structure of apo-PPE at pH 7-8 revealed clear evidence of sulphate binding within the active site[89].

The aim of the work presented in this chapter was to search for alternative crystallisation conditions for PPE lacking sulphate and subsequently to reproduce pH-jumps with the PPE:BCM7 under these conditions. Furthermore, PPE:BCM7 microcrystal slurries were optimised and initial serial synchrotron crystallography (SSX) and SFX experiments were performed.

2.2. Results

2.2.1. Screening for Novel PPE Crystallisation Conditions

2.2.1.1. Initial Screening

To search for novel crystallisation conditions, PPE (SERVA Electrophoresis GmbH) was screened against commercial crystallisation kits (SaltRx, NeXtal Anions, NeXtal Cations, JCSG Plus and Pact Premier). The screens employed apo-PPE (50 mg mL^{-1} in 50 mM acetate buffer pH 5.0) with the aim of increasing the flexibility in subsequent experiments with PPE complexes. The screens produced a number of hits, although most were immediately discarded due to the presence of sulphate or pH > 6 which would likely be unfavourable for formation of the acyl-enzyme complex with BCM7. Crystals from the conditions that lacked sulphate, and which were buffered to a pH below 6, were sent for data collection at Diamond Light Source (Figure 2.2).

All the tested crystals diffracted to a high resolution (< 1.5 Å) and were isomorphous with the reported structures (e.g., PDB: 1QNJ), despite minor differences in crystal morphology (Figure 2.2)[89]. Interestingly, additional electron density was located in the vicinity of S195 in most of the structures (Figure 2.2, A-E). In all cases, the additional electron density was consistent with components of the buffer (phosphate, succinate, citrate, imidazole, and malate, from A-E, respectively). Only the condition using sodium fluoride (NaF) as the precipitant did not appear to have buffer components associated with the active site. In the NaF condition, two waters and a glycerol molecule, derived from the cryoprotectant, were located close to S195 (Figure 2.2, F). Due to their comparable size and because the apparent electron density was loosely associated within the active site, it was not possible to distinguish between fluoride and water; it is thus possible that either of these water molecules were fluoride ions. Nonetheless, the fluoride condition was further examined as glycerol was not essential for crystallisation and could be removed.

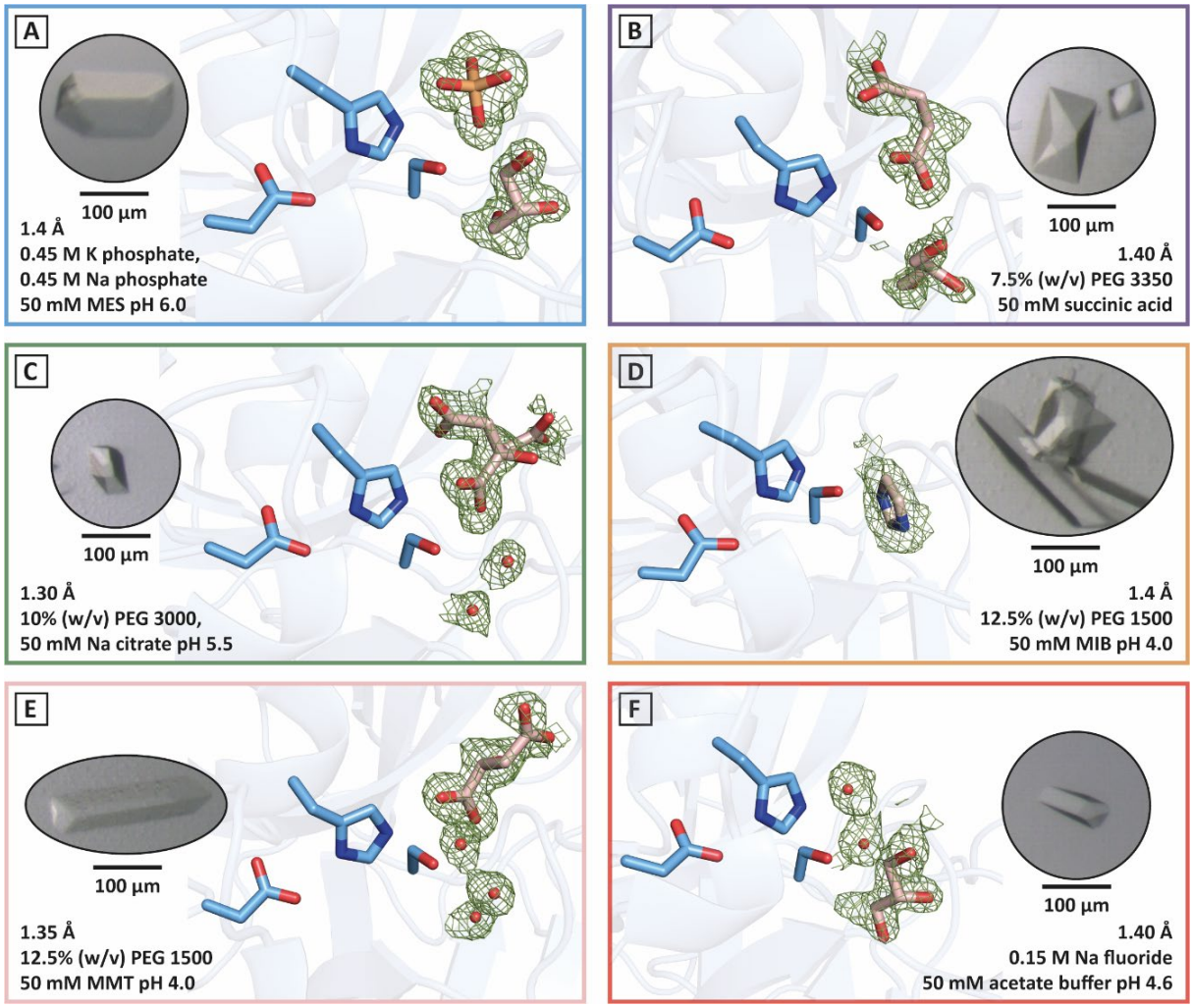


Figure 2.2 | Views from the active sites of crystal structures from the apo-PPE crystal screening trials. A) 1.40 Å resolution crystal structure of PPE precipitated with Na/K phosphate and acetate buffer; **B)** 1.40 Å resolution crystal structure of PPE precipitated with PEG 3350 and succinic acid; **C)** 1.30 Å resolution crystal structure of PPE precipitated with PEG 3000 and Na citrate; **D)** 1.40 Å resolution crystal structure of PPE precipitated with PEG 1500 and malonate, imidazole, boric acid (MIB) buffer; **E)** 1.35 Å resolution crystal structure of PPE precipitated with PEG 1500 and DL-malic acid, MES, Tris (MMT) buffer; **F)** 1.40 Å resolution crystal structure of PPE precipitated with NaF and acetate buffer. PPE carbons are light blue, ligand carbons are pink, phosphorus is orange, and waters are shown as red spheres. The Polder omit maps shown as green mesh are contoured to 3 σ.

Subsequent optimisation revealed that apo-PPE crystallised under a range of precipitant concentrations (0.075–0.2 M NaF, 0.025–0.1 M acetate buffer, pH 4.5–5.0; Figure 2.3, A). The crystals formed after one day and continued to grow for approximately a week. Increasing concentrations of fluoride promoted extension along a single axis of the crystals without variation in unit cell parameters (Figure 2.3, A). Comparison with the PPE:BCM7 complex (PDB: 1HAX) revealed that a glycerol molecule was positioned in the S_1 specificity pocket (Figure 2.3, B & C). One of two water molecules (or fluoride ions) associated with S195 was located in the position of the hydrolytic water in the PPE:BCM7

complex. The second water was positioned in place of the BCM7 acyl-enzyme carbonyl oxygen, within the oxyanion hole formed by the backbone amides of S195 and G193 (water O: backbone HN distance, 2.78 Å and 2.60 Å, respectively).

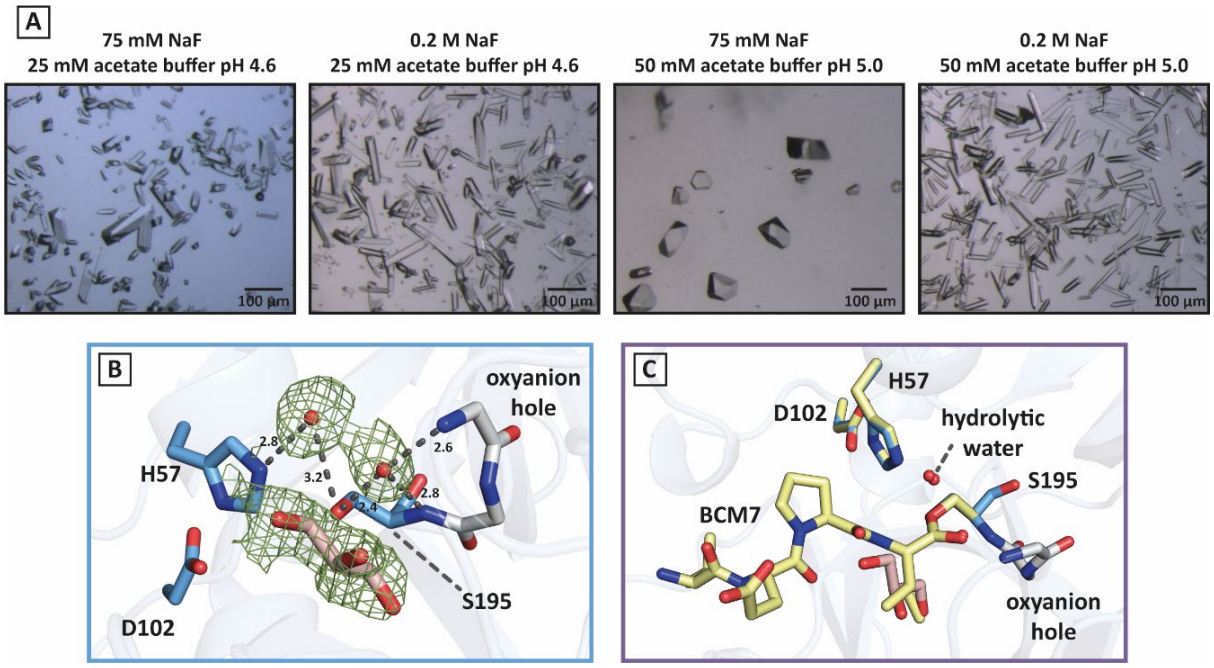


Figure 2.3 | Apo-PPE crystallised with NaF. A) Images of apo-PPE crystals in NaF precipitant; **B)** a view from the active site of a 1.40 Å resolution crystal structure of apo-PPE crystallised in NaF precipitant and soaked with glycerol cryoprotectant before flash freezing (from Figure 2.2, F); **C)** superimposition of the apo-PPE crystal structure in NaF (**B**) with the reported PPE:BCM7 complex structure crystallised in Na_2SO_4 precipitant (PDB: 1HAX). The catalytic triad carbons are light blue, the oxyanion hole carbons are white, glycerol carbons are pink, and all carbons from the 1HAX structure are yellow. The Polder omit map in **B**, shown as green mesh, is contoured to 3 σ.

Due to the inherent reactivity and the basic nature of fluoride ions which could interfere with the planned pH-jump method, the use of other halides was then investigated as precipitants. The sodium salts of chlorine, bromine, and iodine were all found to be capable of crystallising PPE (Figure 2.4, A). The larger halide ions reduced the maximum concentration of precipitant tolerated (NaCl, 10–200 mM; NaBr, 12.5–150 mM; NaI, 12.5–50 mM). Analogously to fluoride, minor differences in crystal morphologies were observed depending on the concentration of precipitant (Figure 2.4, A).

Crystal structures with each of the halides were solved after soaking in cryoprotectant (20% v/v glycerol) and NaCl additionally without any cryoprotectant (Figure 2.4, B–E). The four structures were refined with an acetate molecule associated with S195 and H57. One oxygen atom of

the acetate was positioned in the place of the assigned hydrolytic water[106] (Figure 2.4, F). The second acetate oxygen atom was directed towards the oxyanion hole but its position did not superimpose with the carbonyl oxygen of the BCM7 acyl-enzyme complex (Figure 2.4, F). Crystals from the NaCl and NaBr conditions soaked in cryoprotectant exhibited additional density around the serine that was consistent with a glycerol molecule and was refined as such (Figure 2.4, C-D). The position of the glycerol carbons were similar in both the chloride and fluoride conditions. However, the oxygen atoms at positions 1 and 3 of glycerol were orientated towards S195 in the presence of chloride compared with the fluoride structure where they were orientated away (Figure 2.3, B), although there appeared to be some flexibility in the orientation of the glycerol molecule in each condition. The NaCl crystal that was not cryoprotected had a water (or chloride ion) in place of the glycerol, while the NaI structure had an iodide ion bound in place of glycerol (Figure 2.4, B & E).

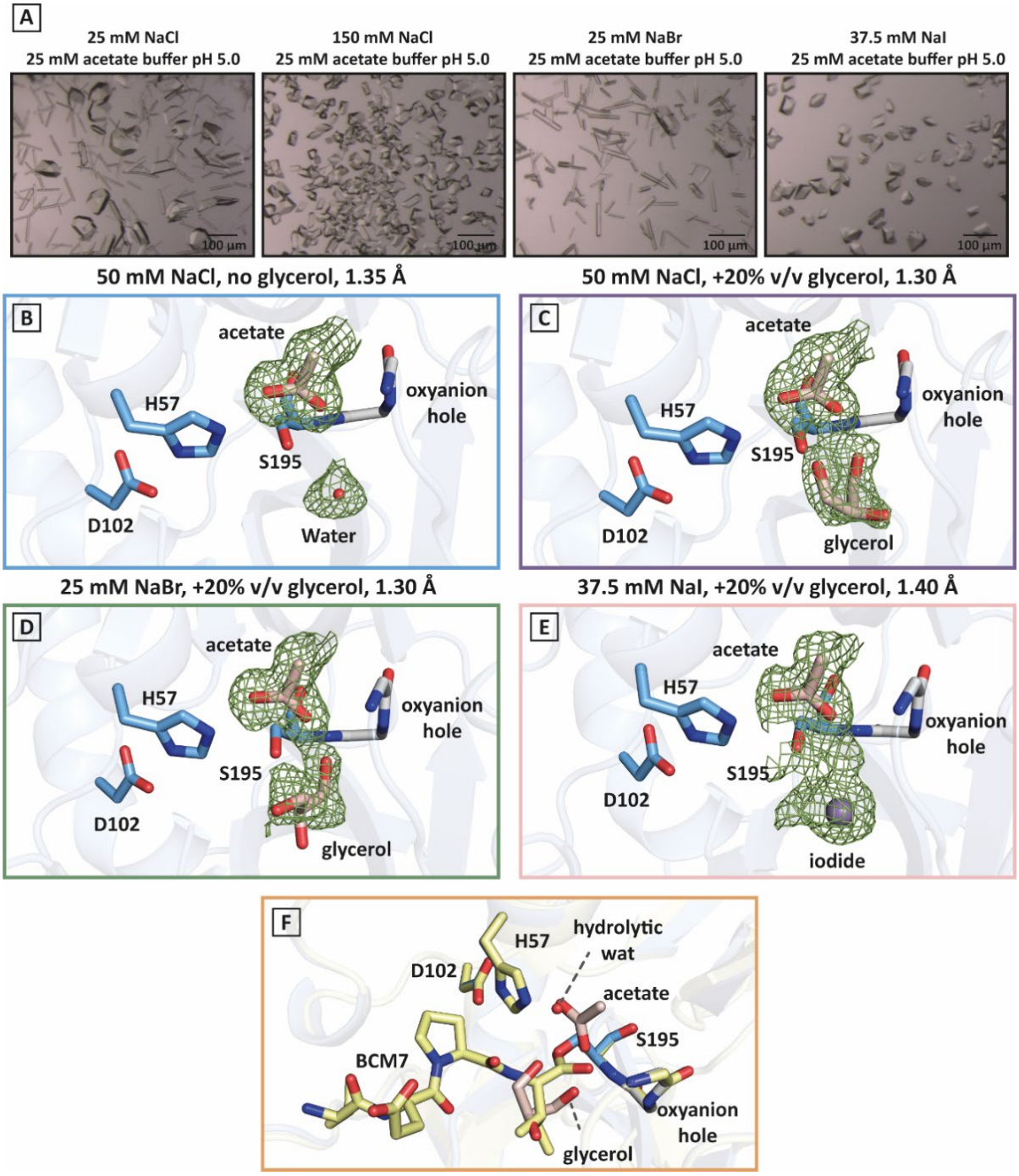


Figure 2.4 | Crystallisation of apo-PPE with halide precipitants. A) Images of apo-PPE crystals in NaCl, NaBr and NaI precipitants; **B-F)** active site views from the structures of apo-PPE crystallised with various halides; **B)** apo-PPE crystallised in NaCl precipitant and flash frozen without soaking in cryoprotectant; **C)** apo-PPE crystallised in NaCl precipitant and soaked with glycerol cryoprotectant before flash freezing; **D)** apo-PPE crystallised in NaBr precipitant and soaked with glycerol cryoprotectant before flash freezing; **E)** apo-PPE crystallised in NaI precipitant and soaked with glycerol cryoprotectant before flash freezing; **F)** superimposition of the apo-PPE crystal structure (NaCl + glycerol) and the reported PPE:BCM7 complex structure crystallised in Na_2SO_4 precipitant (PDB: 1HAX). The catalytic triad carbons are light blue, the oxyanion hole carbons are white, ligand carbons are pink, and all carbons from the 1HAX structure are yellow. The Polder omit maps in **B-E**, shown as green mesh, are contoured to 3 σ.

2.2.1.2. Conditions with BCM7

Co-crystallisation of PPE:BCM7 (30 eq of BCM7) was examined in the presence of all four of the halides investigated. Only fluoride (0.15–0.25 M) and chloride (0.1–0.25 M) yielded crystals (Figure 2.5, A & B). Both conditions produced prismatic crystals and were isomorphous with reported structures of PPE:BCM7 (PDB: 1HAX). The refined structures superimposed well with PPE:BCM7 obtained in the presence of sulphate (PDB: 1HAX), with only minor deviations in RMSD (RMSD for backbone atoms (C, Cα, N, O) = 0.15 Å and 0.14 Å for NaF and NaCl compared to 1HAX, respectively).

For clarity, BCM7 residues will henceforth be referred to by the three-letter code while PPE residues will be referred to by the single-letter code.

Only the *C*–terminal portion of BCM7 (4-Val-Glu-Pro-Ile-7) was visible in both structures forming an antiparallel β-sheet with interactions between Glu5 and V216, and Ile7 and S214 (Figure 2.5, C-F). A water was in position to hydrogen bond with the sidechain of Glu5, the carbonyl oxygen of Pro6 and the sidechain of Q192 (Figure 2.5, F). Continuous density from Ile7 to S195 of PPE was modelled as an ester bond, as previously observed[106,170]. The carbonyl oxygen of Ile7 was located within the oxyanion hole within hydrogen bond distance to the backbone amides of S195 (2.7 Å) and G193 (2.8 Å). The sidechain of Ile7 is located within the S_1 specificity pocket, surrounded by the sidechains of V216 and T213. The hydrolytic water was positioned between Nε2 of H71 (2.7 Å) and the ester carbon (2.8 Å).

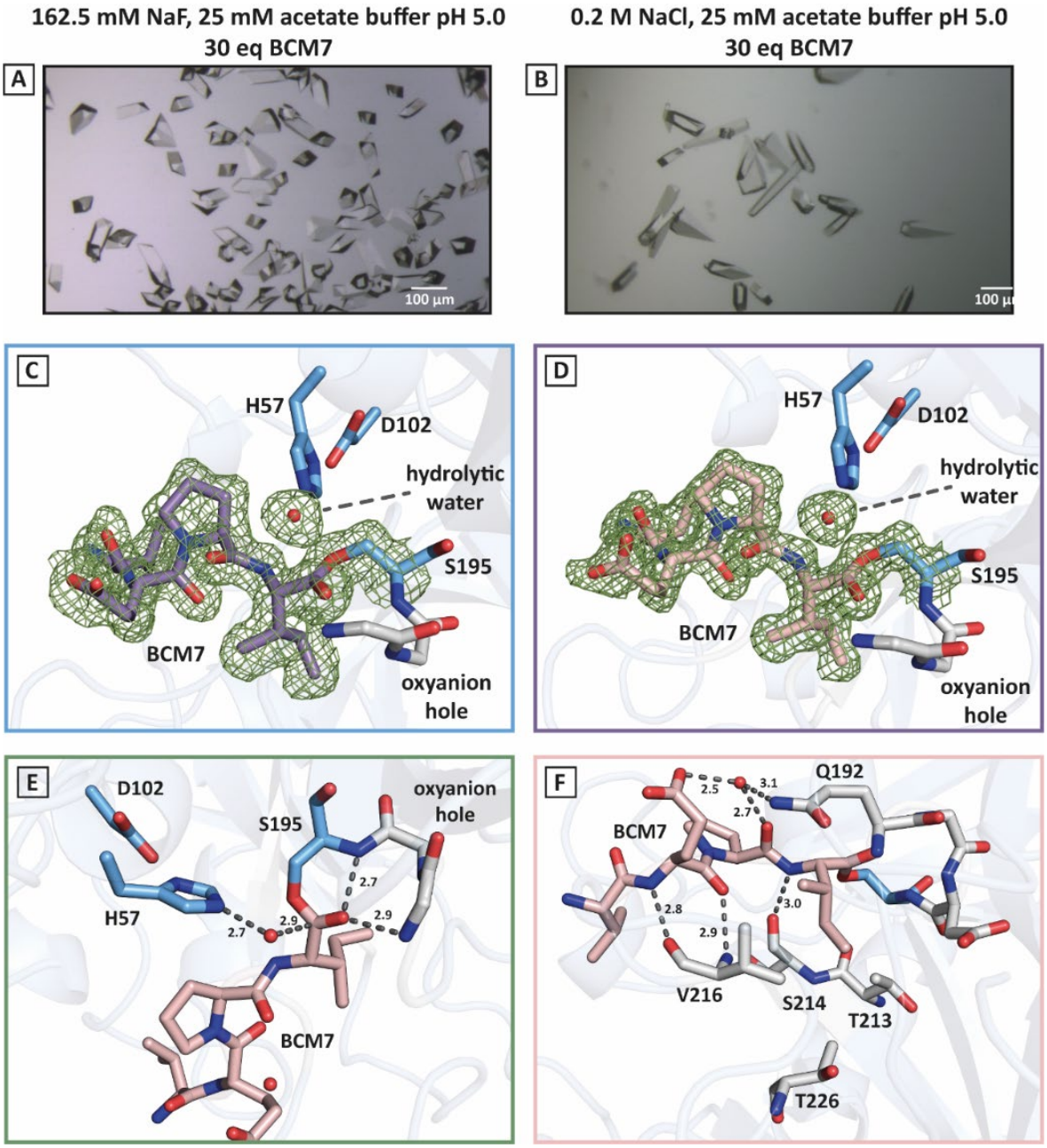


Figure 2.5 | Crystallisation of PPE:BCM7 with NaF and NaCl precipitants. A) An image of PPE co-crystallised with BCM7 (30 eq) in NaF precipitant (1.40 Å resolution); **B)** an image of PPE co-crystallised with BCM7 (30 eq) in NaCl precipitant (1.30 Å resolution); **C)** a view of the active site from a crystal structure of the PPE:BCM7 complex in NaF precipitant; **D-F)** views of the active site from a crystal structure of PPE:BCM7 complex in NaCl precipitant with a Polder omit map shown as green mesh and contoured to 3 σ; **D)** PPE:BCM7 structure with a Polder omit map shown as green mesh and contoured to 3 σ; **E)** contacts between acyl-enzyme complex, the catalytic water, and oxyanion hole; **F)** antiparallel β-sheet contacts between BCM7 and PPE, and the Glu5-Asn192 interaction. PPE catalytic triad carbons are light blue, other enzyme carbons are white, BCM7 carbons purple (NaF) or pink (NaCl).

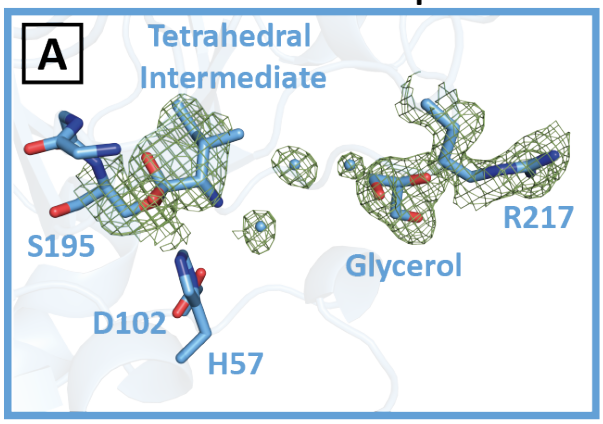
2.2.2. Reproducing 'pH-Jumps' in Macrocrystals

2.2.2.1. pH jumps with CHES and glycerol

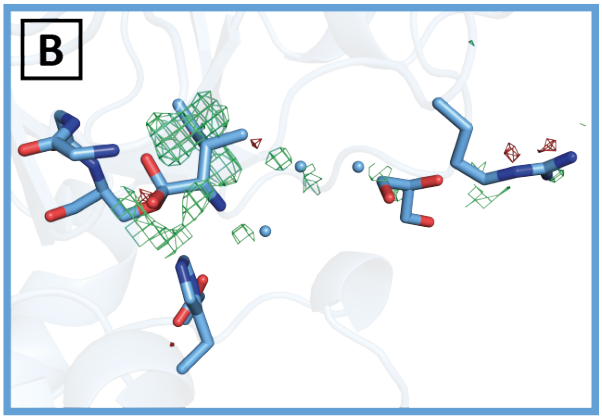
To characterise a potential tetrahedral PPE:BCM7 intermediate, Wilmouth *et al*. conducted studies in which crystals of the PPE:BCM7 complex, produced at pH 5.0 in Na_2SO_4 precipitant, were subject to 'pH-jump' experiments where the pH was raised in order to promote reaction. Here, initial pH-jump experiments were aimed at reproducing those previously performed, in the absence of sulphate[170].

Crystals of PPE:BCM7 produced at pH 5.0 were transferred to a drop containing the relevant precipitant (NaCl or NaF) supplemented with 100 mM CHES pH 9.0 and 20% v/v glycerol for up to 2 min before flash freezing. From these experiments, a single crystal (1.45 Å, 0.225 M NaCl, 1 m 30 s incubation) exhibited similar density to the previously assigned intermediate (Figure 2.6, A) (PDB: 1HAZ). A tetrahedral intermediate was modelled into the density, but the result was not convincing. Examination of the m*F*o-D*F*c difference map revealed additional positive density surrounding the sidechain of Ile7 (Figure 2.6, B), with similar positive density shown by the isomorphous difference map calculated against PPE:BCM7 (NaCl, pH 5.0) (Figure 2.6, C). This additional density was observed between the intermediate and the sidechain of Ile7, and between the hydrolytic water and the intermediate. Furthermore, only disjointed density was manifest for the BCM7 chain and the sidechain of R217A was modelled into the apo conformation (i.e., as observed without BCM7) indicating that BCM7 had left the active site (Figure 2.6, C). These observations were also observed in the reported structure and were attributed to disruption of the hydrogen bonding network between PPE and BCM7 concomitant to formation of the tetrahedral intermediate[170].

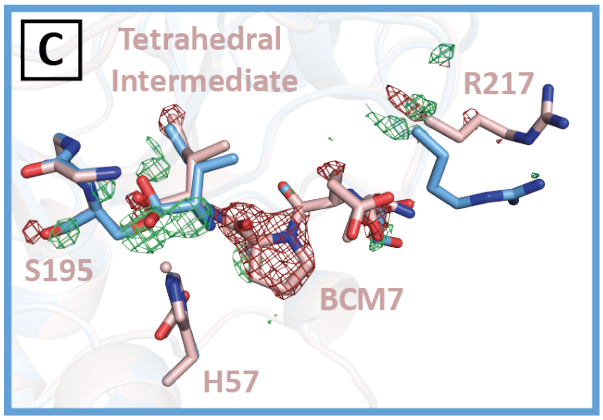
Polder omit map



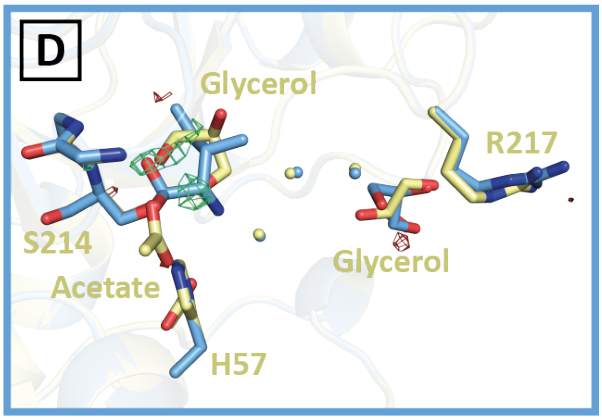
m*Fo*-D*Fc*



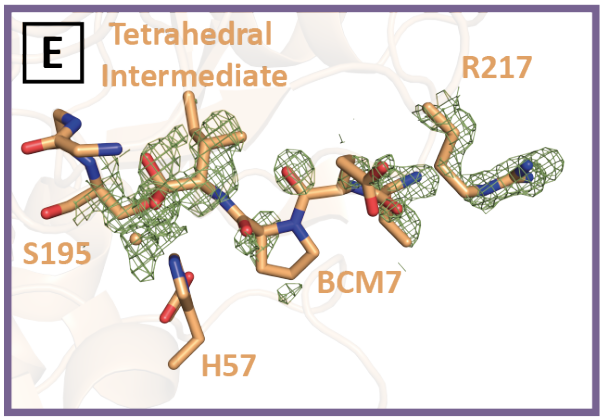
isomorphous difference map
(PPE:BCM7)



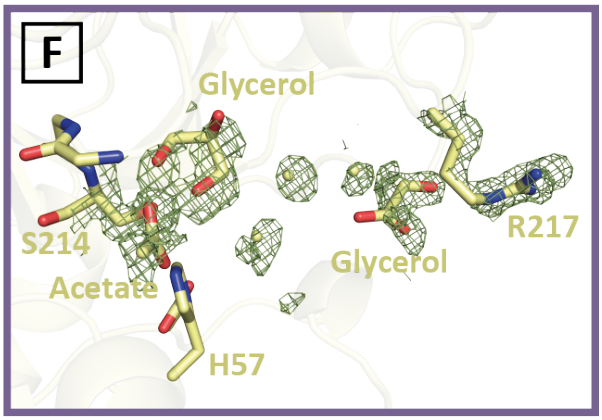
isomorphous difference map
(apo-PPE)



2*Fo*-*Fc*



2*Fo*-*Fc*



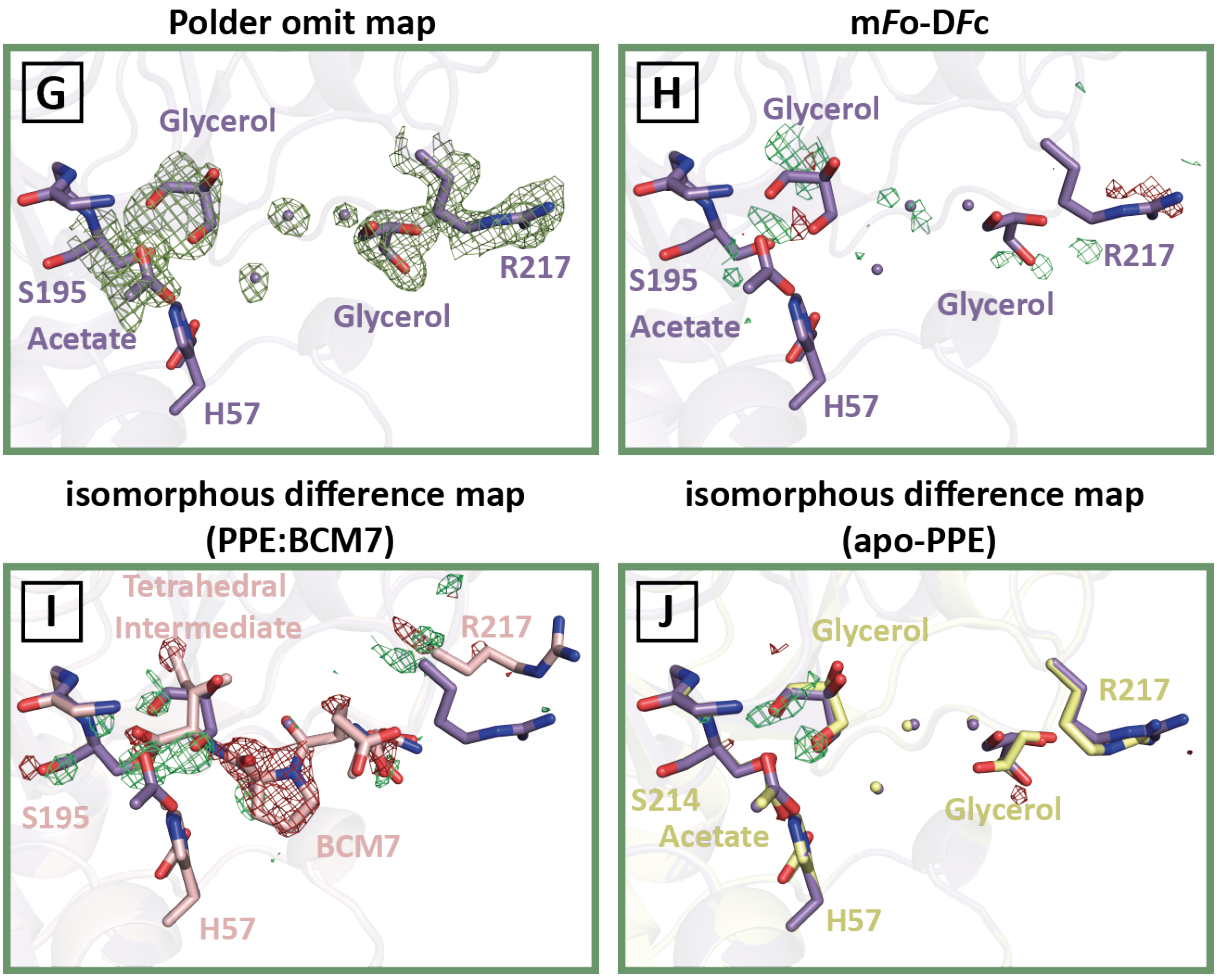


Figure 2.6 | Reproduction of pH jumps in NaCl with 0.1 M CHES pH 9.0. A-D) A view of the active site from a 1.45 Å resolution crystal structure of PPE:BCM7 after pH-jump refined with a potential tetrahedral intermediate of Ile7 (light blue carbons) with various maps; **A)** a Polder omit map (dark green) contoured to 3 σ; **B)** an m*Fo*-D*Fc* difference map, the positive density (green) and negative density (red) is contoured to 3 and -3 σ, respectively; **C)** an isomorphous difference map calculated by subtracting density of the PPE:BCM7 in NaCl at pH 5.0 (Figure 2.5, D), the positive density (green) and negative density (red) is contoured to 3 and -3 σ, respectively; **D)** an isomorphous difference map calculated by subtracting density of apo-PPE in NaCl at pH 5.0 (Figure 2.4, C). Positive density (green) and negative density (red) contoured to 3 and -3 σ, respectively; **E)** the reported structure of the PPE:BCM7 tetrahedral intermediate (PDB: 1HAZ; orange carbons) with 2*Fo*-*Fc* map (dark green) contoured to 1 σ); **F)** the crystal structure of apo-PPE in NaCl at pH 5.0 (Figure 2.4, C) overlaid with the 2*Fo*-*Fc* map of 1HAZ (dark green) contoured to 1 σ; **G-J)** a view of the active site from the same crystal structure of PPE:BCM7 potential tetrahedral complex from **A-D** but modelled with an acetate and glycerol in place of the tetrahedral Ile7 (purple carbons); **G)** a Polder omit map (dark green) contoured to 3 σ; **H)** an m*Fo*-D*Fc* difference map, the positive density (green) and negative density (red) is contoured to 3 and -3 σ, respectively; **I)** an isomorphous difference map calculated by subtracting density of PPE:BCM7 in NaCl at pH 5.0, the positive density (green) and negative density (red) contoured to 3 and -3 σ, respectively; **J)** an isomorphous difference map calculated by subtracting density of apo-PPE in NaCl at pH 5.0, the positive density (green) and negative density (red) contoured to 3 and -3 σ, respectively.

To be more confident in the assignment of the structure, precipitant and buffer components (CHES, acetate, and glycerol) were modelled into the density. Although CHES contains a sulphate group that could fit into the tetrahedral density, the carbon chain and ring of CHES could not be modelled (data not shown). The association of CHES with the active site cleft (S_3-S_4) was observed, but not within distance to interact with S195 and was therefore ruled out (data not shown).

The new apo-PPE crystal structures at pH 5.0 revealed that both acetate and glycerol can associate with S195 (Figure 2.4, C). Superimposition of an apo-PPE structure revealed that acetate and glycerol, in the apo-PPE orientation, could fit the additional electron density observed in both 1HAZ and the newly collected 'intermediate' obtained here by pH-jump methodology (Figure 2.6, E & F). Thus, acetate and glycerol were modelled into the new structure using the same orientation as observed in the apo-PPE structure (Figure 2.6, G-J). The additional density observed in the isomorphous difference map calculated against PPE:BCM7 could be explained by the presence of acetate and glycerol (Figure 2.6, D, G-J). The oxygen atoms of acetate superimposed with the hydrolytic water and one of the oxygens of the potential tetrahedral intermediate. The additional density between these positions is superimposed by the acetate carbons, with C2 in the position of the previously assigned water molecule. The positioning of the C1 and C3 hydroxy groups of glycerol explains the positive density surrounding the Ile7 sidechain. Glycerol also does not overlay completely with Ile7 sidechain, and this supports the negative density observed at the Cδ of Ile7. Along the BCM7 chain, negative density is primarily observed around Pro6. The absence of negative density for Val4 and Glu5 of BCM7 can be justified by re-entry of three waters and a glycerol into the active site, as observed in the apo-structures (Figure 2.6, I). The glycerol molecule, which overlaps with Val4, was also in a similar position in both the pH-jump and apo-structures, but orientated differently, likely due to the flexibility of glycerol and weak association with the active site. An isomorphous difference map revealed little difference between the pH-jumped structure and apo-PPE (Figure 2.6, D & J). Positive density that was observed around the C1 and C3 hydroxyls of glycerol may represent the flexibility of glycerol which was observed in multiple conformations in the apo-PPE crystal structures (Figure 2.3, B; Figure 2.4, C).

2.2.2.2. pH-Jumps under Alternate Conditions

To further examine whether the electron density assigned as a tetrahedral intermediate reflects a true intermediate, components of the pH-jump solution were altered to examine whether the reported and observed density could be reproduced. First, crystals were soaked in pH-jump buffer in the absence of

cryo-protectant. Secondly, with the aim of further ruling out the interference of CHES, Tris was used as the pH-jump buffer. As all other buffers that successfully crystallised PPE at low pH associated with the active site, the use of acetate buffer was not changed. None of the 87 crystal structures collected in the absence of glycerol exhibited density resembling the tetrahedral intermediate. In most of the structures, BCM7 was either bound or completely released. The reaction time in these experiments was much faster, in some cases exhibiting full deacylation in less than 10 s. This is likely due to the reduced viscosity of the pH-jump solution in the absence of glycerol, enabling faster diffusion throughout the crystal[205].

Based on the evidence described above, it seems unlikely that the structure obtained or the previously assigned structure (PDB: 1HAZ), at least solely, represents a tetrahedral intermediate (Figure 2.6). In contrast, however, acetate was not observed in the active site of any of the crystal structures that had been pH-jumped and where the PPE:BMC7 complex was completely deacylated. A similar trend was observed for the glycerol molecule, but to a lesser extent as one structure did contain electron density consistent with a glycerol molecule (Figure 2.7, A). At higher pH, acetate is the predominant form of acetic acid and thus it is a H-bond acceptor. As with sulphate, it is therefore unlikely that at higher pH the interaction between acetate and H57 is more favourable[170].

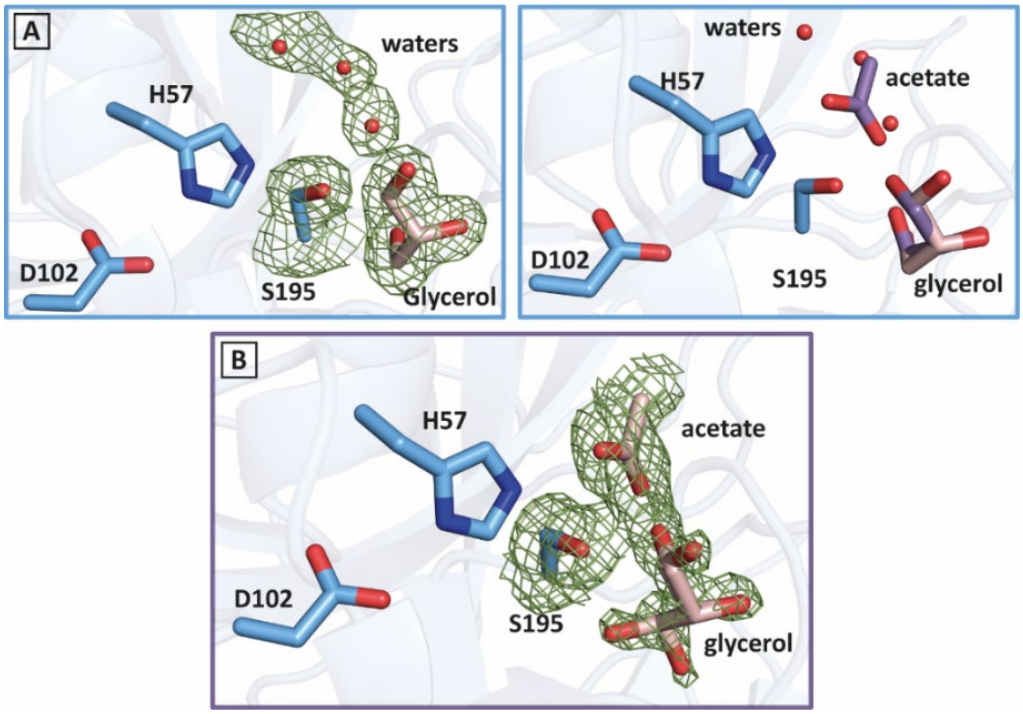


Figure 2.7 | Views from the active sites of crystals of PPE after pH-jumps. A) A structure of the PPE:BCM7 complex after soaking in Tris buffer pH 9.0 + 20% v/v glycerol for 45 s with polder omit map electron density consistent with a glycerol and several waters in place of the acetate (left), and a superimposition of the structure with the apo-PPE structure (Figure 2.4, C) (right); **B)** a structure of apo:PPE crystallised in acetate buffer at pH 5.0 and soaked in CHES pH 9.0 + 20% v/v glycerol for 30 s. Enzyme carbons are shown in light blue, ligand carbons from the soaked crystal structures are shown in pink, and ligand carbons from the apo-PPE structure (Figure 2.4, C) are shown in purple. The Polder omit map electron density shown as green mesh is contoured to 3 σ.

To further investigate the association of acetate and glycerol with the active site of PPE at high pH, crystals of apo-PPE crystallised in acetate buffer pH 5.0 were subjected to pH-jumps in various buffer (NaCl or NaF at the same concentration as the mother liquor mixed 1:1 v/v with 200 mM Tris pH 7.15, 8.0 and 8.85, 200 mM CHES pH 8.0 and 9.0, or 200 mM glycine pH 10.0, all with 20% v/v glycerol final [C]) for approximately 30 s. Most of the pH-jumped apo-PPE structures did not contain electron density for acetate within the active site but weak density for glycerol was observed (data not shown). However, one of the structures soaked in CHES pH 9.0, revealed density that was consistent for both acetate and a weakly bound glycerol (Figure 2.7,B). Furthermore, the electron density was continuous between the acetate, glycerol and S195 which could be interpreted as a tetrahedral intermediate.

2.2.2.3. Conformational Changes

Examination of the unit cell information of crystals after pH jumps revealed a number of structures with a 2 Å extension along the *a* axis and a 0.5 Å extension along the *b* axis (approximate unit cell parameters: *a* = 52 Å, *b* = 58 Å, *c* = 75 Å). This effect was observed in crystals of the PPE:BCM7 complex soaked with both CHES and Tris, in the presence and absence of glycerol. The unit cell deviation was observed much more frequently with NaF crystals (30 out of 51 crystals) compared with NaCl crystals (1 / 71). No unit cell changes were observed in crystals that had not been soaked nor was it observed in crystals in which BCM7 was present in the active site. Further, the unit cell change was not observed in every crystal after deacylation. The change could manifest rapidly (< 5 s) in crystals soaked in a high concentration of pH-jump buffer (1 M Tris pH 9.0).

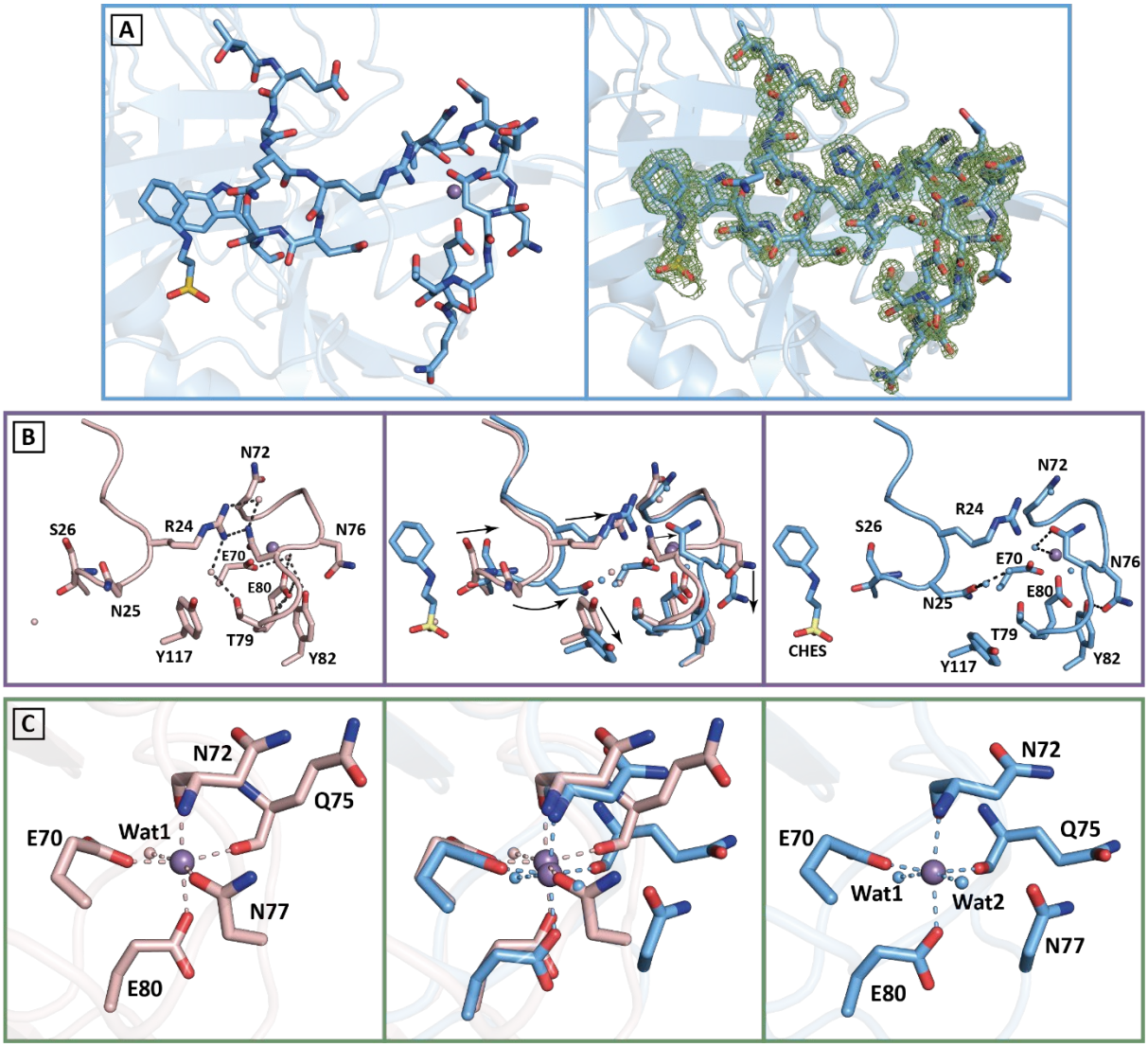


Figure 2.8 | A representative crystal structure showing the conformational changes after pH jump and deacylation of BCM7. A) Overview of a crystal structure (PPE:BCM7 in NaF, soaked in CHES pH 9.0 for 1 min) showing a conformational change; **B)** a view of the two regions exhibiting the conformational change in the PPE:BCM7 complex at pH 5.0 and PPE after soaking in CHES pH 9.0 for 1 min, highlighting residues affected by the conformational change. Dashed lines indicate interactions between residues and/or waters (left) and newly formed interactions after conformational change (right); **C)** views of the metal binding site in the PPE:BCM7 complexes at pH 5.0 and after soaking in CHES pH 9.0 for 55 s highlighting altered coordination of the PPE metal binding site. The carbons and waters (small spheres) of the PPE:BCM7 complex in NaF after 1 min incubation in 0.1 M CHES pH 9.0 are in light blue, the carbons and waters of the apo-PPE structure in NaF at pH 5.0 (Figure 2.3, B) are pink; the Na atom (large sphere) is purple; the Polder omit map in **A** is contoured to 3 σ.

Table 2.3 | Refinement statistics for representative crystal structures with the conformation change. B-factors and real space correlation coefficient (RSCC) are presented as an average for each of the two regions. Region A) G19-W27, region B) E70-Y82.

Condition	Soaking Buffer	Glycerol (20% v/v)	Resolution (Å)	B-Factors			RSCC		
				Overall	A	B	Overall	A	B
NaF	CHES (1 min)	-	1.30	14.92	14.52	17.63	0.981	0.984	0.971
NaF	Tris (55 s)	-	1.50	20.10	20.00	21.86	0.975	0.977	0.971
NaCl	Tris (45 s)	+	1.80	25.77	25.30	29.66	0.972	0.970	0.966

Inspection of three crystal structures from different conditions exhibiting unit cell extensions, revealed main chain deviations in two loop regions (A. G19-W27; B. E70-Y82; Table 2.3), which are represented by the structure in Figure 2.8. Each of the three crystals exhibited the same conformational change and aligned well (< 0.27 Å RMSD between the three structures). In each case, no density remained for the conventional conformation and the real space correlation coefficient (RSCC) (> 0.95) for each of the two loop regions indicated that the modelled conformation fit the density well. In the CHES structure, a CHES molecule was bound within a cleft within H-bond distance to the backbone carbonyl oxygen of S26 (2.9 Å). The Cα of the same residue was shifted 1.4 Å away from CHES compared with the PPE:BCM7 structure at pH 5.0. The backbone conformation of the residues between S26 and A22 was altered, resulting in complete reorientation of the sidechain of N25 (Cγ deviation 6.1 Å) towards region B and within H-bonding distance of two waters (Figure 2.8, B). The second backbone conformation change occurred in the metal binding region (region B), between E70 and Y82. In particular, N77 which is typically involved in coordination of the metal ion, was shifted away from the metal (Cα = 1.9 Å, Cγ = 3.1 Å). The metal remained approximately octahedrally coordinated as a water (Wat2) replaced and H-bonded to the sidechain of N77 (Figure 2.8, C). The shifts within the two regions resulted in the formation of four new intramolecular interactions (Figure 2.8, B).

The conformational changes also altered existing, or introduced new crystal packing interactions, which may relate to the changes to the unit cell parameters (Figure 2.9). N25, R24 and T79 formed indirect interactions with the guanidino group of R188A (using chymotrypsinogen residue numbering – see Appendix A) of a symmetry related molecule, through two bridging waters. The guanidino group of R188A was rotated about the Cγ-Cδ bond, compared with the pH 5 structure (Figure 2.9, A–C). The carbonyl of G78 also makes an indirect interaction with T20, from the same symmetry related molecule as above, through a bridging water, each exhibiting a minor shift in position. Finally, a rotation about the Cβ of Y117, shifting the hydroxy group by 3.4 Å, placed it within position to interact with the carbonyl oxygen of G19 (2.8 Å) of the same symmetry related molecule as above.

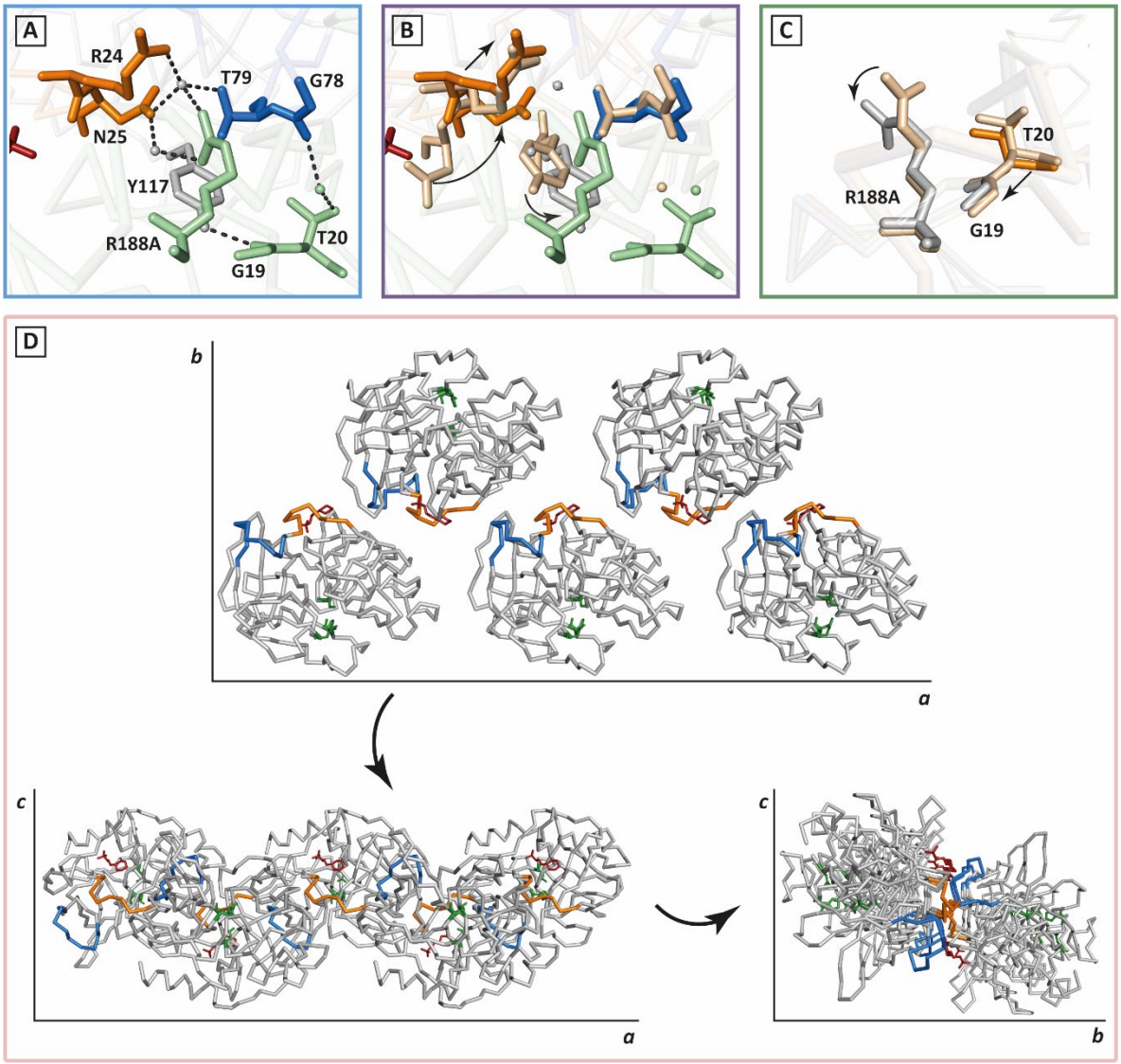


Figure 2.9 | Analysis of conformational changes effects on crystallographic symmetry partners. A-D) Views from the PPE:BMC7 structure in NaF, soaked in CHES pH 9.0 for 1 min; **A)** interactions between residues exhibiting a conformational change with those of a symmetry mate, highlighting H-bond interactions; **B)** superimposition of residues from **A** with the PPE:BCM7 complex in NaF at pH 5.0; **C)** conformational changes of residues involved in symmetry interactions from the symmetry mate; **D)** overview of symmetry interactions between symmetry mates along each of the three crystallographic axes. Residues from region A are orange; residues from region B are blue; residues from the other regions of PPE are grey; residues from the symmetry mate are pale green, and residues from the PPE:BCM7 pH 5.0 structure are beige. In **D** the PPE catalytic triad residues are dark green.

Analysis of the position of the two regions involved in conformational changes in respect to the crystal planes revealed that they form a line along the *a* axis (Figure 2.9, D). Each symmetry molecule along the axis interacts with the next through the intermolecular interaction of several residues discussed above. Of note, T20 is located near the 'origin' of movement in region A, and interacts with the 'terminus' of movement in region B (G78), forming an apparently continuous chain of interactions between symmetry partners along the *a* axis (Figure 2.9).

As mentioned above, electron density, refined as a CHES molecule, was associated within a surface cleft close to the 'origin' of conformational change. As such binding of CHES is a possible cause of the conformational changes due to the interaction with S26 (Figure 2.10, A). However, identical conformational changes were observed after performing soaks with crystals in Tris buffer at pH 9.0. In these structures, binding of Tris was not observed in the cleft where CHES was located (Figure 2.10, B & C). Under both conditions soaked with Tris, two waters were modelled in place of CHES. One water was positioned in place of the CHES nitrogen and was within H-bonding distance of the carbonyl oxygen of S26. The other superimposed with a CHES sulphate oxygen, contributing to a H-bond network of waters that connected H200 (Nδ), V122 (N) and the carbonyl oxygens of L120 and S29. The sulphate oxygens of CHES made two indirect interactions with the amide and carbonyl oxygen of Y207, through bridging waters, that were absent in the Tris-soaked structures (Figure 2.10, A). Comparison of the cleft at pH 5 indicated that most of the waters were bound in similar positions as observed in crystals soaked with Tris (Figure 2.10, D & E). It therefore seems unlikely that the conformational changes are caused directly by CHES binding in the cleft.

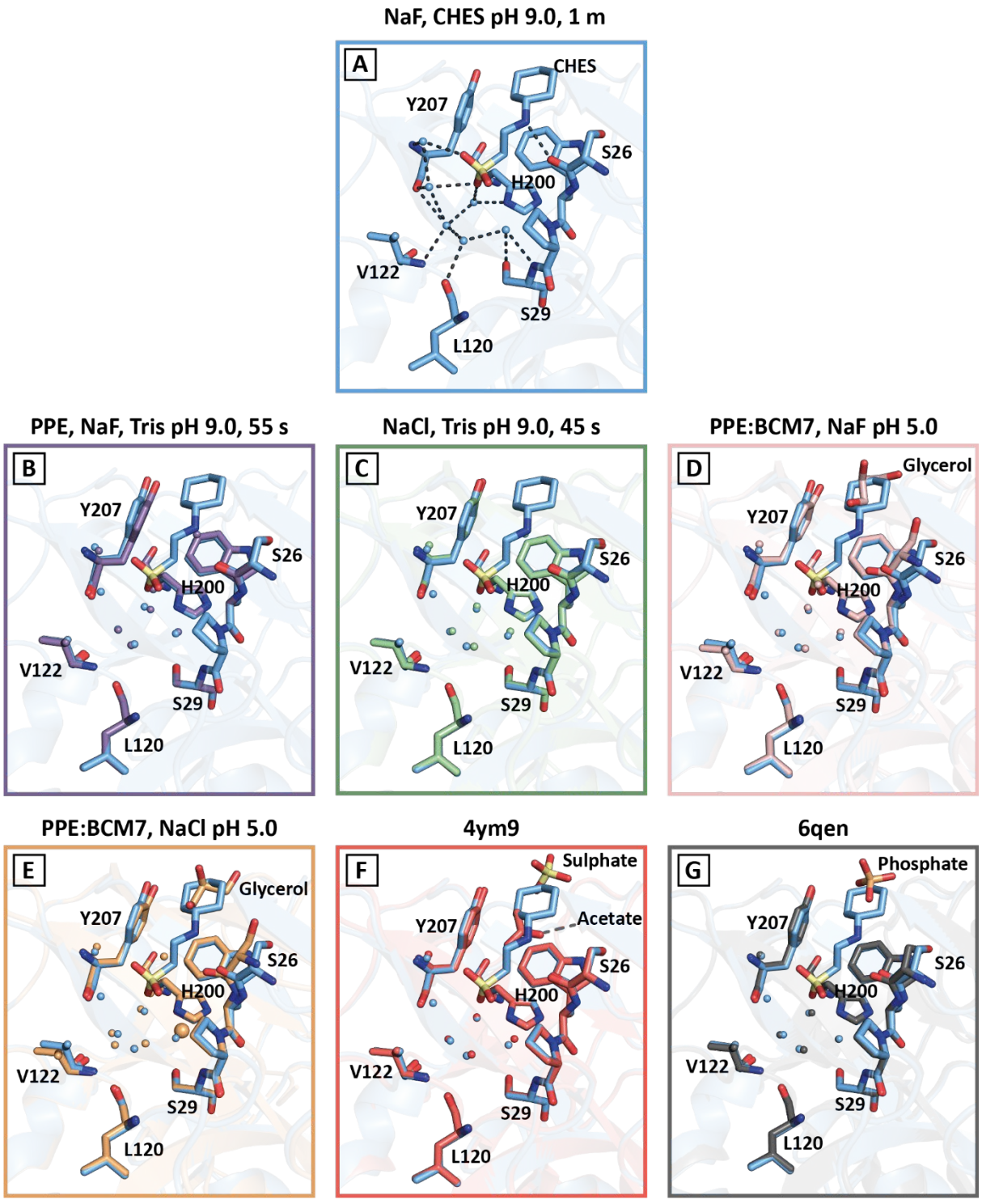


Figure 2.10 | Views from the CHES binding cleft in PPE crystal structures. A) PPE:BCM7 NaF soaked for 1 min in 0.1 M CHES pH 9.0 (light blue cabrons). Dashed lines indicate distances of <3.5 Å; **B)** PPE:BCM7 NaF soaked for 55 s in 0.1 M Tris-HCl pH 9.0 exhibiting the conformational changes (purple carbons) aligned with (**A**); **C)** PPE:BCM7 NaCl soaked for 45 s in 0.1 M Tris-HCl pH 9.0 with 20% v/v glycerol exhibiting the conformational changes (green carbons) aligned with (**A**); **D)** PPE:BCM7 NaF in acetate buffer pH 5.0 (pink carbons) aligned with (**A**); **E)** PPE:BCM7 NaCl in acetate buffer pH 5.0 (orange carbons) aligned with (**A**); **F)** PPE:JM102 complex structure which exhibits the conformational changes (PDB: 4YM9; red carbons) aligned with (**A**); **G)** PPE:LMC240 complex structure which exhibits the conformational changes (PDB: 6QEN; black carbons) aligned with (**A**). Waters are coloured according to the structure.

As the conformational changes were only observed in structures after soaking in buffer at pH 9, experiments were performed to investigate whether it was a pH effect. Apo-PPE was crystallised in both precipitants (NaCl and NaF) with various buffers (Tris pH 7.15, 8 and 8.85; CHES pH 9.0 and 10.0; glycine pH 10.0). Use of NaF as a precipitant led to crystallisation under all pH conditions, while NaCl only led to production of crystals in Tris 7.15 and 8.0, and CHES 9.0. Unexpectedly, whilst conformational changes were more common in NaF precipitant following pH-jumps with crystals of the PPE:BCM7 complex, such changes were only observed in crystals of apo-PPE produced with NaCl precipitant at high pH. Identical conformation changes to those described above were observed in Tris pH 8.0 and CHES 9.0 (NaCl), but only in 56% of the crystals (2/4 crystals in Tris pH 8.0, 3/5 crystals in CHES pH 9.0; data not shown). Crystals of apo-PPE produced at pH 5.0 (acetate buffer) were then soaked in each of the high pH buffers. Similarly, those crystallised with NaF did not exhibit conformational changes with any buffer. However, two crystals crystallised with NaCl precipitant exhibited identical conformational changes (Tris pH 8.85, glycine pH 10.0). The combined results suggest that whilst raising the pH can make the observed conformational change more likely to occur, it does not guarantee a change and it is therefore likely not a purely pH-induced effect.

Conformational changes in the rigid backbone of PPE are unusual. The PDB was thus searched for structures deposited with specific unit cell parameters (a = 51-53 Å; b = 57-59 Å; c = 74-76 Å). 43 of the 45 returned structures were PPE and of these, two were found to have identical conformational changes to those described here. Both were complexes of PPE with oxo-β-lactam inhibitors, JM102 (PDB: 4YM9; Figure 2.10, F)[229] and LMC240 (PDB: 6QEN; Figure 2.10, G)[230]. In each case PPE was crystallised under distinct conditions; 4YM9 was crystallised under standard conditions with sulphate and acetate at pH 5.1 (glutaraldehyde crosslinked; unit cell: a = 51.03 Å; b = 58.34 Å; c = 74.92 Å)[229] while 6QEN was crystallised with 2-methyl-2,4-pentanediol (MPD) and phosphate buffer at pH 5.9 (unit cell: a = 51.13 Å; b = 57.71 Å; c = 74.64 Å)[230]. In both the structures (4YM9 and 6QEN), the conformation of the CHES binding cleft resembled that of the pH 5 structures (Figure 2.10, D & E), with

glycerol being replaced by sulphate, or phosphate and acetate (Figure 2.10, F & G). As the crystals were both produced at relatively low pH, the conformational changes likely are not solely caused by pH.

2.2.3. Ambient Temperature Serial Crystallography with BCM7

2.2.3.1. Optimisation of Microcrystal Slurries

To prepare for time-resolved serial crystallography experiments, an ambient temperature structure of apo-PPE crystallised with sulphate and acetate at pH 5.0 was collected at PAL-XFEL. Crystals were grown in 96-well plate batch format with seeding (Section 7.2.4.1). The crystals adhered to the bottom of the well and were only released by mechanical stimulation and repetitive pipetting of the crystallisation solution. The resulting crystal slurry was non-homogenous (Figure 2.11, A). As the aim was to collect an apo-structure, the non-homogenous character was considered sufficient for initial data collection by serial fixed-target crystallography[231-233].

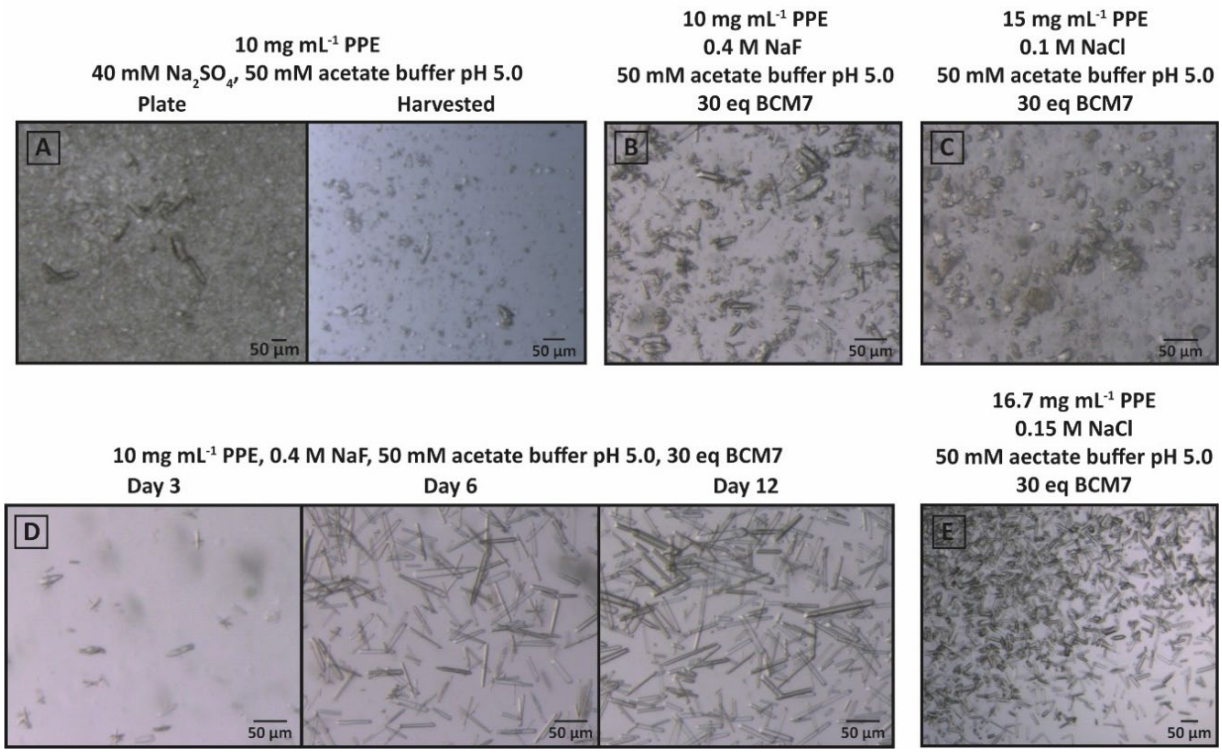


Figure 2.11 | Photographic Images of PPE microcrystal optimisation. A) Images of batch (plate) crystallisation of apo-PPE in Na_2SO_4 precipitant, left) before harvesting, right) after harvesting; **B)** an image of a seeded batch (plate) crystallisation of PPE:BCM7 in NaF; **C)** an image of a seeded batch (plate) crystallisation of PPE:BCM7 in NaCl; **D)** images of a seeded batch (tube, 18 rpm rotation) crystallisation of PPE:BCM7 in NaF over 12 days; **E)** image of hanging drop vapour diffusion crystallisation of apo-PPE in NaCl (seeded).

Microcrystal slurries of PPE:BCM7 were also optimised. As with experiments in single crystals, both the NaCl and NaF conditions were examined. Initial trials for both conditions in 96-well batch format, with seeding, yielded poor results. Crystals in both conditions formed under a range of precipitant concentrations, however, adherence to the plate was again observed. The resultant samples were non-homogenous with substantial clumping, which are not suitable for time-resolved crystallography mixing experiments (Figure 2.11, B & C). Agitation during crystallisation did not improve the homogeneity, nor reduce the observed adherence to the plate.

Subsequently, crystallisation was tested in tubes to examine whether the use of an end-over-end rotor could improve the suspension of the crystals. Under these conditions, use of NaCl did not give rise to crystallisation (0.05–0.5 M NaCl, 10–15 mg mL^{-1} PPE). However, use of NaF produced crystals and exhibited reduced adherence to the tube. The resultant slurry was of an improved quality after harvesting, although a completely homogenous sample was not achieved. The optimal condition (10 mg mL^{-1} PPE, 0.4 M NaF, 0.1 M acetate buffer pH 5.0, with seeding, 250 µL volume in a 1.5 mL tube, 18 rpm) produced long, thin crystals (Figure 2.11, D). Crystals formed after three days but the density was too low to be useful for serial crystallography experiments. The crystal density and size continued to increase for two weeks. The optimal growth time was approximately one week, which yielded a high crystal density with crystals of ranging from 30–70 µm along the long axis and 2–7 µm in width. These crystals would generally be considered too large for most time-resolved experiments, partially due to the risk of clogging sample delivery systems, but mainly due to the increased diffusion times for substrates to reach the centre of the crystal which would likely result in non-uniform reaction initiation throughout the crystal. As the planned approach for collection of time-resolved data was fixed-target, there was no risk of clogging the equipment. Further, due to the long, thin morphology, it was envisaged that diffusion through the crystal would be sufficiently fast to homogenously initiate the reaction.

Further optimisation with the NaCl sample using the seeded hanging-drop vapour diffusion method yielded crystals of a similar size and quality to NaF in tube. The optimal condition (12.5 mg mL^{-1} PPE, 0.3 M NaCl, 0.1 M acetate buffer pH 5.0) produced PPE:BCM7 crystals that grew to approximately 50 µm along the long axis over 5 days; however, there was often variation between wells as the sample was split across many wells during crystallisation. Whilst PPE, BCM7, and precipitant solution were mixed prior to plating, each droplet will have experienced slightly different environments. Factors that were difficult to control include sealing of the well and the surface area to volume ratio of the droplet, based on the degree to which it spread across the coverslip. Further, the crystal density of the drops was low, forming a single layer at the bottom of the drop. The resulting crystal slurries often required concentration by more than 2-fold after harvesting. The fixed-target method used at Diamond Light Source requires 80–120 µL of slurry per chip and a full data set (10,000–15,000 indexed images) can require 2–3 chips. Therefore, a single dataset required at least 240 µL of concentrated slurry (500–750 µL of crystallisation solution or crystal slurry from two 24-well plates (3 × 6 µL droplets per well)). As a result of the extremely labour-intensive preparation of the NaCl crystals, initial time-resolved experiments focussed on the NaF condition.

2.2.3.2. Collection of Resting State Ambient Structures

The first ambient temperature, serial crystallography structure of PPE collected at PAL-XFEL was apo-PPE crystallised with Na_2SO_4 and acetate buffer at pH 5.0 (Section 7.2.8). The 1.35 Å resolution structure exhibited identical secondary structure to apo-PPE under various cryogenic conditions (Figure 2.12, A-E). No substantial changes in backbone conformation were observed (backbone RMSD varied by less than 0.26 $Å^2$) and only surface residue sidechains were observed in alternate conformers. The catalytic triad residues overlaid identically to those under cryogenic conditions and as previously observed, a sulphate ion was associated with S195 and H57 (Figure 2.12, A & B).

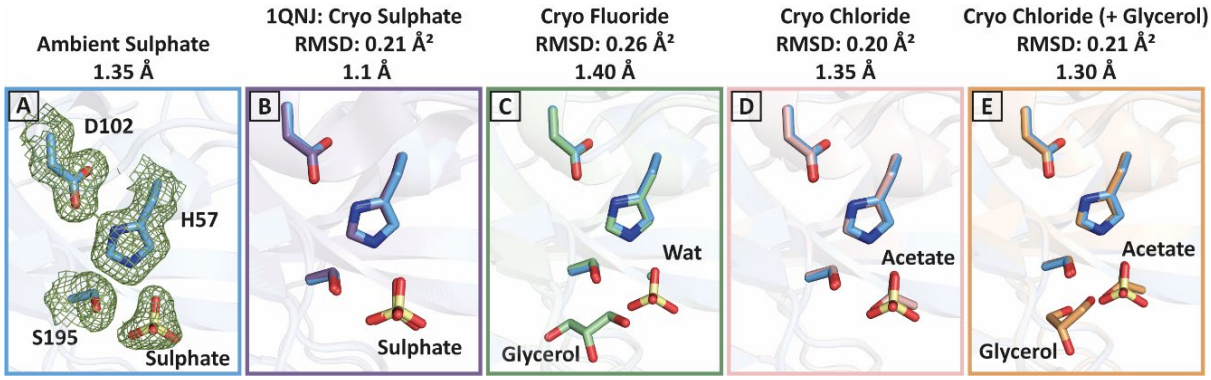


Figure 2.12 | Comparison of ambient temperature XFEL structure of apo-PPE with cryogenic temperature structures. A) A view of the active site of apo-PPE crystallised in Na_2SO_4 precipitant obtained at ambient temperature by SFX (light blue carbons). Green density represents a Polder omit map contoured to 3 σ; **B)** superimposition of **A** with a reported structure of apo-PPE crystallised in Na_2SO_4 precipitant obtained at cryogenic temperature by single crystal X-ray diffraction (PDB: 1QNJ; purple carbons); **C)** superimposition of **A** with apo-PPE crystallised in NaF precipitant (+20% v/v glycerol; Figure 2.3, B) obtained at cryogenic temperature by single crystal X-ray diffraction (green carbons); **D)** superimposition of **A** with a structure of apo-PPE crystallised in NaCl precipitant (Figure 2.4, B) obtained at cryogenic temperature by single crystal X-ray diffraction (pink carbons); **E)** superimposition of **A** with a structure of apo-PPE crystallised in NaCl precipitant (+20% v/v glycerol) obtained at cryogenic temperature by single crystal X-ray diffraction (orange carbons).

An ambient temperature XFEL structure was also collected of the PPE:BCM7 complex crystallised in NaF precipitant and acetate buffer at pH 5.0 (1.55 Å resolution; Figure 2.13). As observed with the apo-PPE ambient temperature structure, there were no substantial deviations in backbone structure and the RMSD against three cryo-temperature structures of the PPE:BCM7 complex in different precipitants was less than 0.21 Å (Figure 2.13). Continuous density was observed from Ile7 to S195, and was consistent with formation of an acyl-enzyme complex. The orientation of the catalytic triad and BCM7 residues correlated with the cryogenic structures (Figure 2.13, A-D), except for the sidechain of Glu5 which was confidently modelled into the density in the ambient structure (Figure 2.13, A). There was a minor deviation in the position of the hydrolytic water compared with other structures, but it was still in position to react with the ester carbon (2.8 Å from both the carbonyl carbon of Ile7 and H57 Nε2).

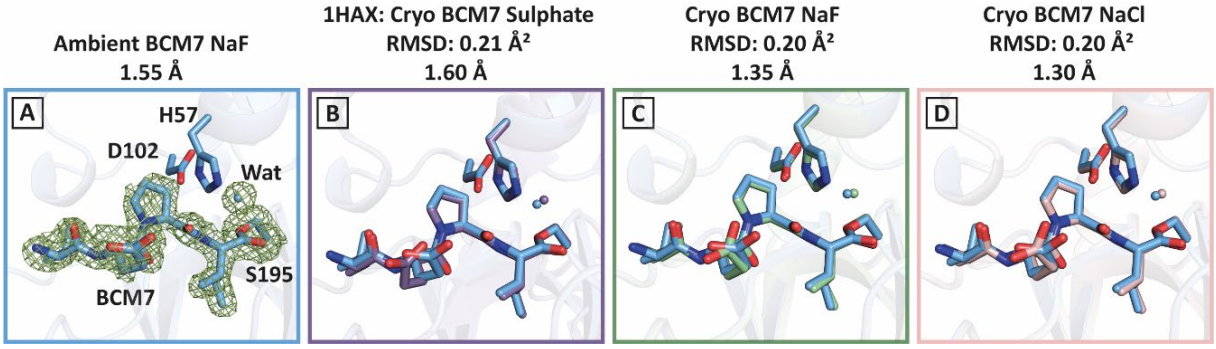


Figure 2.13 | Comparison of ambient temperature XFEL structure of PPE:BCM7 with cryo-temperature structures. A) A view of the active site of the PPE:BCM7 complex (30 eq) crystallised in NaF precipitant obtained at ambient temperature by SFX (light blue carbons). The Polder omit map (green) is contoured to 3 σ; **B)** superimposition of **A** with the reported structure of the PPE:BCM7 complex crystallised in Na_2SO_4 precipitant, obtained at cryogenic temperature by single crystal X-ray diffraction (PDB: 1HAX; purple carbons); **C)** superimposition of **A** with the PPE:BCM7 complex crystallised in NaF precipitant, obtained at cryogenic temperature by single crystal X-ray diffraction (Figure 2.5, B; green carbons); **D)** superimposition of **A** with the PPE:BCM7 complex crystallised in NaCl precipitant, obtained at cryogenic temperature by single crystal X-ray diffraction (Figure 2.5, C; pink carbons) aligned with (A).

2.2.3.3. Drop-on-Chip Fixed-Target Serial Synchrotron Crystallography

The drop-on-chip time-resolved (DoC-Tr) SSX methodology at Diamond Light Source (I24 Microfocus Beamline) was used for collection of ambient temperature data with PPE[231-233]. This method, in analogy to the "liquid application method for time-resolved analyses" (LAMA)[234], combines the fixed-target method, with an ADE system. In short, chips were loaded with crystals in the 'resting' state (PPE:BCM7 acyl-enzyme complex, pH 5.0). The chips were prepared as for fixed-target data collection (Section 7.2.9), but the loading face of the chip was not sealed with Mylar®, to facilitate mixing. At the beamline, the ejector, loaded with the reaction initiation solution (i.e., high pH buffer), was aligned with the beam by ejecting solution onto the Mylar® in the calibration region. Droplet ejection was synchronised with the beam repetition rate and the mixing time was varied through the selection of raster scan options. For example, a short mixing time of 12 ms was achieved by ejecting a droplet onto the chip (2 ms) and immediately collecting X-ray data (10 ms exposure). Longer time points were achieved through the excite and visit again (EAVA) strategy[235] where droplets were ejected onto two rows of crystals (40 wells) and data was then collected from the same two rows. Running at a 60 Hz repetition rate in this manner enabled collection of data at 560 ms intervals.

The first experiments conducted with PPE:BCM7 (NaF) explored reaction initiation by addition of 0.2 M CHES pH 9.0 (Figure 2.14). Three structures were collected at ambient temperature (resting, 1.12 s and 5.60 s). In the resting state, there was strong continuous electron density for all four *C*–terminal residues of PPE and the ester bond (Figure 2.14, A). The hydrolytic water was also confidently modelled into the structure. This structure closely resembled the ambient temperature XFEL structure (backbone RMSD: 0.16 Å). Density for BCM7 gradually diminished over the time course with strong density remaining after 1.12 s and disjointed density after 5.6s (Figure 2.14, B & C). Negative density gradually appeared in the isomorphous difference maps suggesting that fewer active sites were occupied by BCM7 as the mixing time increased (Figure 2.14, D & E).

Structures were refined based on the assumption that the B-factors of the carbonyl carbon of Ile7 of BCM7 were not substantially different than the Oγ of S214 (< 15 % difference). Using this method, the occupancy for BCM7 was examined (Figure 2.14, F). The occupancy of BCM7 decreased as mixing time increased while the average B-factors remained approximately unchanged suggesting that deacylation was occurring, albeit incompletely. The B-factors also increased from Ile7 to Val4 in each of the structures. DoC-Tr SSX was therefore deemed to be a viable method for initiating reaction in PPE, however, further optimisation was required to improve homogeneity and speed of the reaction.

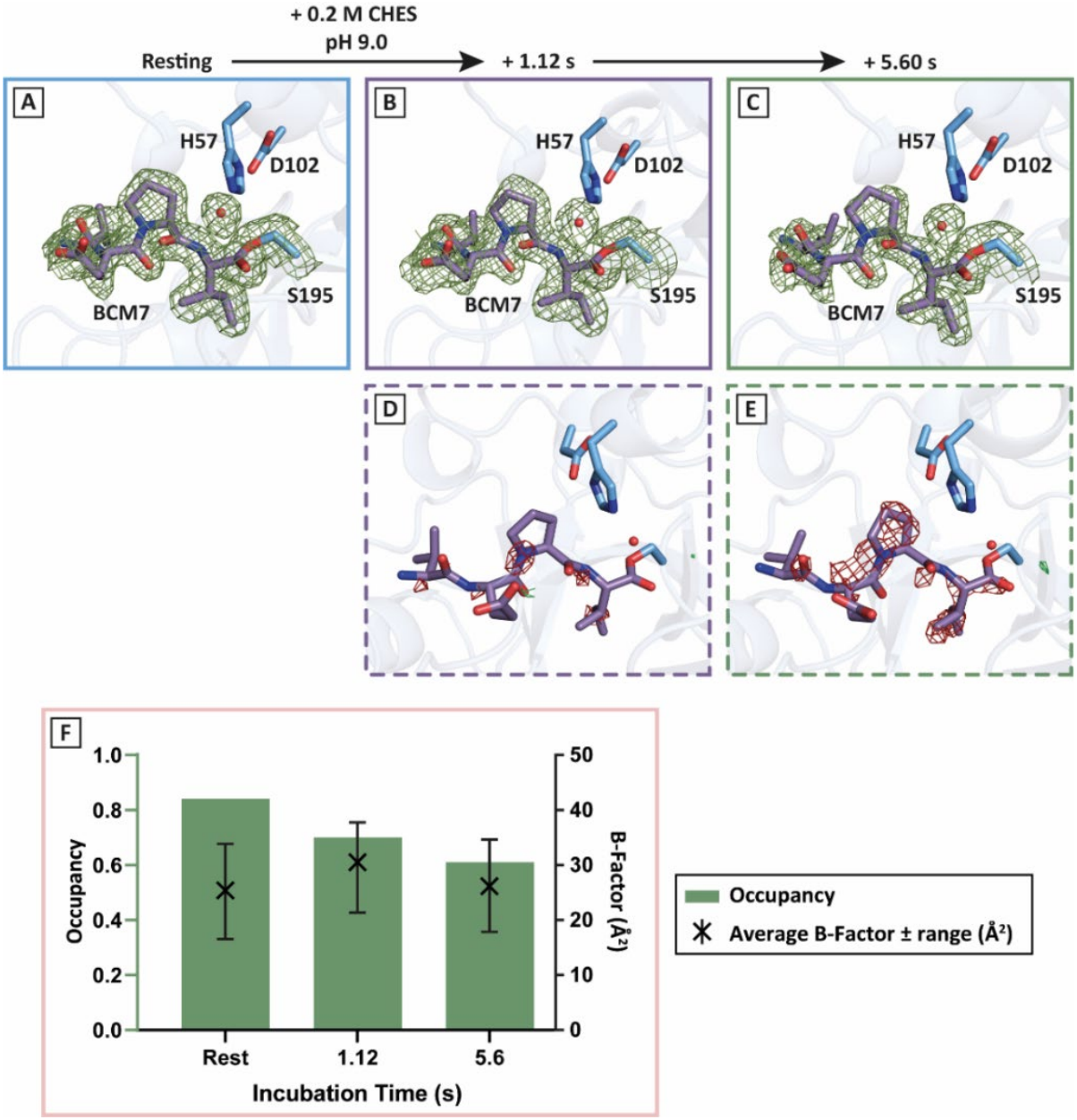


Figure 2.14 | Drop-on-Fixed-Target ambient temperature pH jumps on PPE:BCM7 crystals (NaF) with 0.2 M CHES pH 9.0. **A)** An active site view from a resting state structure of PPE:BCM7 in NaF precipitant at pH 5.0 with a Polder omit map; **B)** an active site view from a resting state structure of PPE:BCM7 1.12 s after ejection of 0.2 M CHES pH 9.0 with a Polder omit map; **C)** an active site view from a resting state structure of PPE:BCM7 5.60 s after ejection of 0.2 M CHES pH 9.0 with a Polder omit map; **D)** 1.12 s structure (**B**) with isomorphous difference map subtracting the resting state structure (**A**); **E)** 5.60 s structure (**C**) with isomorphous difference map subtracting the resting state structure (**A**); **A-E)** the catalytic triad carbons are light blue, BCM7 carbons are purple; **A-C)** the Polder omit maps in green are contoured to 3 σ; **D-E)** positive (green) and negative (red) density of the isomorphous difference maps, is contoured to 3 σ and -3 σ, respectively; **F)** a graph of the occupancy (green bars) and average B-Factors ± range (black cross) of BCM7 in each of the above structures.

To examine whether a smaller buffer component could diffuse through the crystals faster than the relatively large CHES molecule, use of Tris (0.2 M, pH 9.0) was explored. Single crystal pH jumps had previously shown that Tris was able to initiate reaction and that deacylation of BCM7 could occur within 10 s (Section 2.2.2.2). Three further structures with NaF precipitant were collected using the DoC-Tr SSX method. The resting state structure exhibited strong density and good refinement statistics for BCM7 (Figure 2.15, A). After 1.12 s of mixing, strong density remained for BCM7, negative density was present over BCM7 in the isomorphous difference map, and the average B-factors were slightly raised (Figure 2.15, B, D, & F). After 5.6 s, the continuous density began to break around BCM7 in the 2m*F*o-D*F*c map, although strong density was still observed in the Polder omit map (Figure 2.15, C). The amount of negative density also increased in the isomorphous difference map (Figure 2.15, E). As observed previously (Figure 2.14), the refined occupancy decreased as mixing time increased while the average B-factor remained approximately the same. As observed with CHES, the data suggests that BCM7 is less present as mixing times increased, however, deacylation was not complete under these conditions.

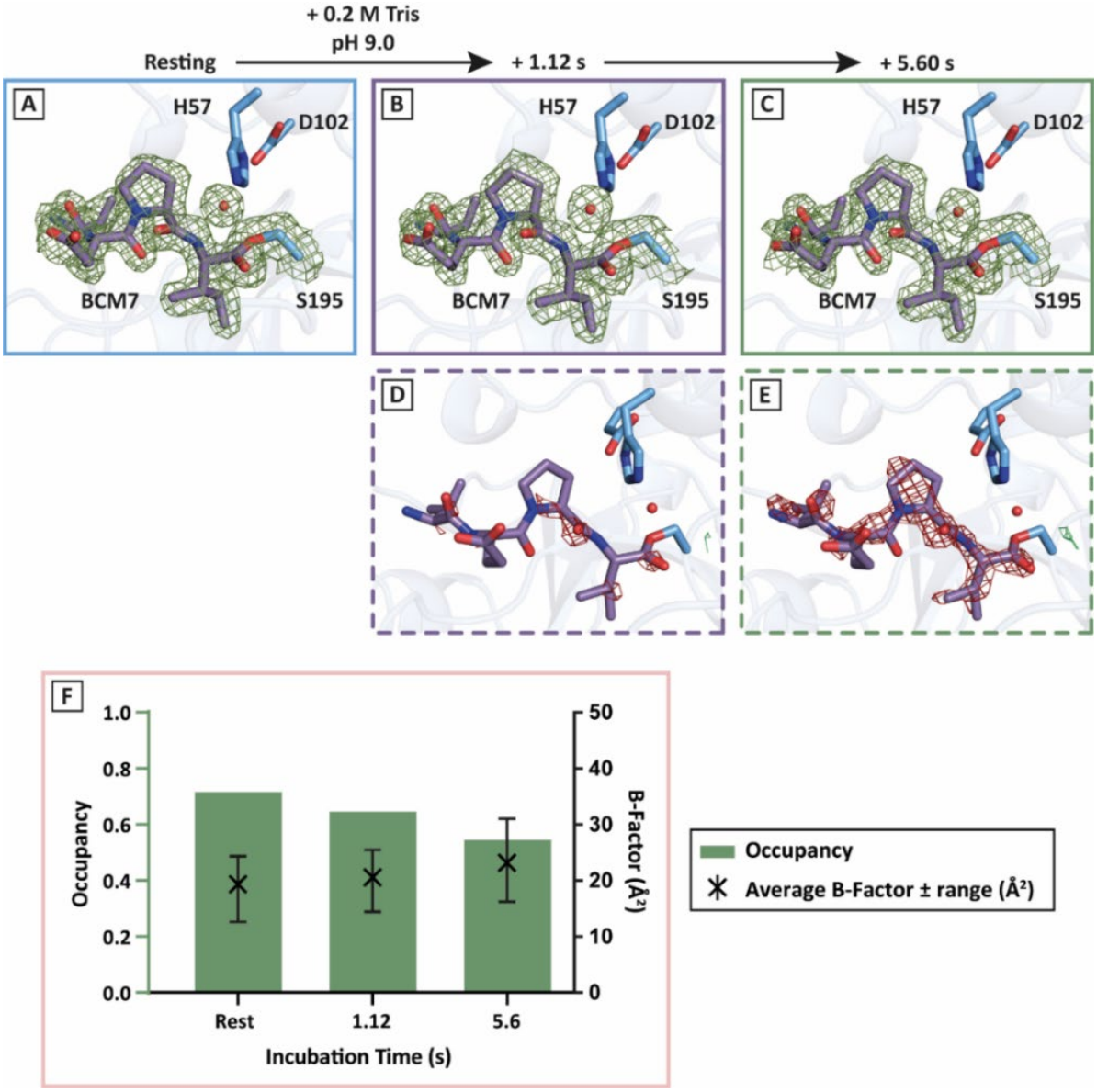


Figure 2.15 | Drop-on-Fixed-Target ambient temperature pH jumps on PPE:BCM7 crystals (NaF) with 0.2 M Tris pH 9.0. A) An active site view from a resting state structure of PPE:BCM7 in NaF precipitant at pH 5.0 with a Polder omit map; **B)** an active site view from a structure of PPE:BCM7 1.12 s after ejection of 0.2 M Tris pH 9.0 with a Polder omit map; **C)** an active site view from a structure of PPE:BCM7 5.60 s after ejection of 0.2 M Tris pH 9.0 with a Polder omit map; **D)** 1.12 s structure (**B**) with an isomorphous difference map subtracting the resting state structure (**A**); **E)** the 5.60 s structure (**C**) with an isomorphous difference map subtracting the resting state structure (**A**); **A-E)** the catalytic triad carbons are light blue, BCM7 carbons are purple; **A-C)** the Polder omit maps in green are contoured to 3 σ; **D-E)** positive (green) and negative (red) density of the isomorphous difference maps, is contoured to 3 σ and -3 σ, respectively; **F)** A graph of the occupancy (green bars) and average B-Factors ± range (black cross) of BCM7 in each of the above structures.

To compare the NaF and NaCl conditions, an ambient temperature resting structure and a pH-jumped structure of PPE crystallised in NaCl (seeded hanging-drop vapour diffusion) were collected (Figure 2.16). As with the fluoride condition, there was continuous density and strong refinement statistics for BCM7 in the resting state (Figure 2.16, A & D). After addition of 0.2 M Tris pH 9.0 and incubation for 1.12 s, negative density appeared for BCM7 and the occupancy decreased, although continuous density for BCM7 persisted in both 2m*F*o-D*F*c and Polder omit maps (Figure 2.16, B, C & E). The results closely matched those collected after similar pH jumps with PPE crystallised with fluoride (Figure 2.15).

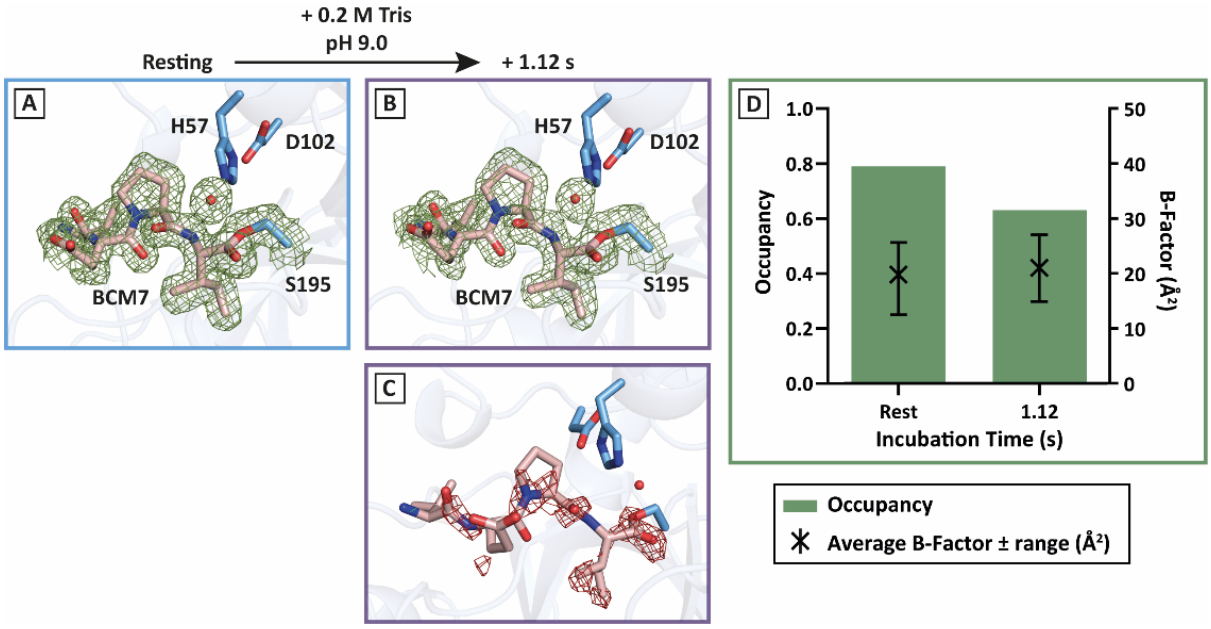


Figure 2.16 | Drop-on-Fixed-Target ambient temperature pH-jumps on PPE:BCM7 crystals (NaCl) with 0.2 M Tris pH 9.0. A) An active site view from a resting state structure of PPE:BCM7 in NaCl at pH 5.0 with a Polder omit map; **B)** an active site view from a structure of PPE:BCM7 1.12 s after ejection of 0.2 M Tris pH 9.0 with Polder omit map; **C)** 1.12 s structure (**B**) with an isomorphous difference map subtracting the resting state structure (**A**); **A-C)** the catalytic triad carbons are light blue, BCM7 carbons are purple; **A & B)** the Polder omit maps are in green and contoured to 3 σ; **C)** positive (green) and negative (red) density of the isomorphous difference maps, is contoured to 3 σ and -3 σ, respectively; **D)** a graph of the occupancy (green bars) and average B-Factors ± range (black cross) of BCM7 in each of the above structures.

With the aim of increasing the rate of reaction across the crystal, higher concentrations of pH-jump buffer were tested. Structures of PPE in NaCl, were collected after pH jumps with 0.5 and 1 M Tris pH 9.0 and incubation times of 1.12 and 5.60 s (Figure 2.17). After mixing with both concentrations of buffer for 1.12 s, electron density for BCM7 diminished in both the 2m*F*o-D*F*c and Polder omit maps (Figure 2.17, A & C). R217A could not be confidently modelled into either apo- or BCM7 bound conformations with 0.5 M Tris but was modelled into the apo-conformation after soaking with 1 M Tris suggesting that it was predominantly deacylated. Negative density was present in both conditions, covering the entire peptide chain with 1 M Tris pH 9.0 (Figure 2.17, E & F). The occupancy of BCM7 dropped to approximately 50% in both conditions, however, in contrast to previous experiments the average B-factors for BCM7 rose compared with the resting structure (Figure 2.17, H). After 5.6 s mixing with 500 mM Tris pH 9.0, the electron density continued to diminish for BCM7. In the 2m*F*o-D*F*c map little density for BCM7 was present and more closely resembled that of a resting structure with several waters present (data not shown). However, broken density remained in the Polder omit map (Figure 2.17, D). As with 1 M Tris at 1.12 s, R217A was modelled into the apo-conformation. Negative density also covered the entire BCM7 chain after 5.6 s and the catalytic water. The occupancy for BCM7 remained at approximately 50%, while the B-factors for BCM7 continued to rise. The rise in average B-factors was caused by a concomitant rise in the B-factors of the Oγ and of Cβ of S195, indicating some flexibility in these atoms during deacylation. This rise in B-factor is also likely masking a further drop in occupancy, as the additional negative density, broken electron density, and conformation of R217A suggest. The combined results suggest that PPE was predominantly deacetylated within 5.60 s using 500 mM Tris pH 9.0 and 1 M Tris after 1.12 s, however, deacylation remained incomplete. Interestingly, no conformational changes were observed in any of the SSX or SFX datasets.

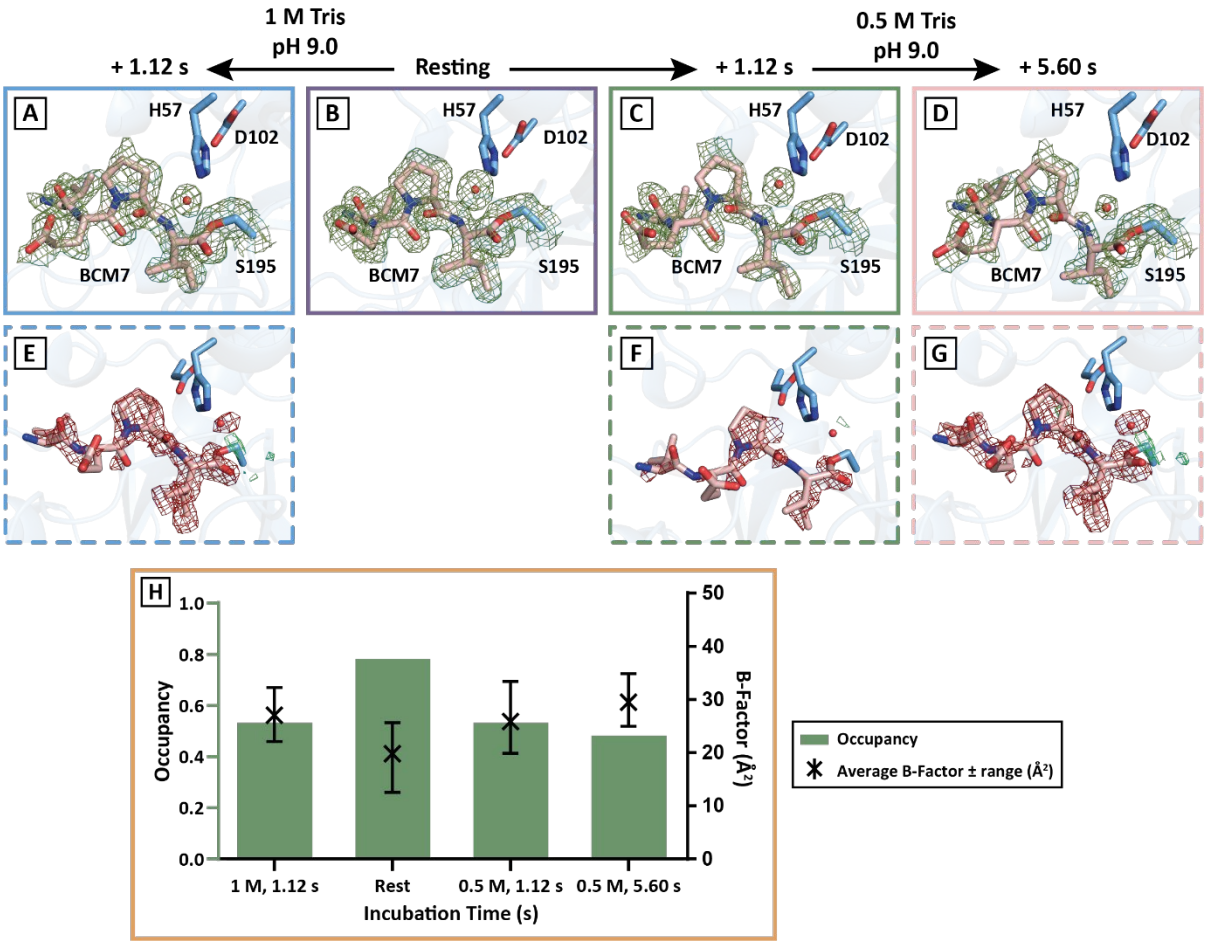


Figure 2.17 | Drop-on-Fixed-Target ambient temperature pH jumps on PPE:BCM7 crystals (NaCl) with 0.5 and 1 M Tris pH 9.0. A) An active site view from a structure of PPE:BCM7 1.12 s after ejection of 1 M Tris pH 9.0 with a Polder omit map; **B)** an active site view from a resting state structure of PPE:BCM7 in NaCl at pH 5.0 with a Polder omit map (Figure 2.16, A); **C)** an active site view from a structure of PPE:BCM7 1.12 s after ejection of 0.5 M Tris pH 9.0 with a Polder omit map; **D)** an active site view from a structure of PPE:BCM7 5.60 s after ejection of 0.5 M Tris pH 9.0 with a Polder omit map; **E)** 1.12 s, 1 M structure (**A**) with an isomorphous difference map subtracting the resting state structure (**B**); **F)** 1.12 s, 0.5 M structure (**C**) with an isomorphous difference map subtracting the resting state structure (**B**); **G)** 5.60 s, 0.5 M structure (**D**) with an isomorphous difference map subtracting the resting state structure (**B**); **A-G)** the catalytic triad carbons are light blue, BCM7 carbons are purple; **A-D)** the Polder omit maps are in green and contoured to 3 σ; **E-G)** positive (green) and negative (red) density of the isomorphous difference maps, is contoured to 3 σ and -3 σ, respectively; **H)** a graph of the occupancy (green bars) and average B-Factors ± range (black cross) of BCM7 in each of the above structures.

2.3. Discussion and Perspectives

The work presented in this chapter aimed to re-evaluate the data collected in previous 'time-resolved' crystallography experiments performed with PPE[107,170,236]. The reported experiments exposed crystals of the PPE:BCM7 complex (pH 5.0) to a pH-jump solution (CHES pH 9.0) which induced the deacylation reaction. One of the resultant crystal structures exhibited electron density apparently resembling a tetrahedral intermediate bound to S195 (Figure 2.1). However, due to the presence of sulphate in the precipitant, the results were questioned[118].

Novel crystallisation conditions were sought for PPE that did not contain sulphate. Crystallisation screens of apo-PPE revealed a novel condition (NaF and acetate buffer pH 5.0) that appeared to have only waters (or fluoride ions) and a glycerol molecule that was derived from the cryoprotectant solution, present at the active site (Figure 2.2, F). Subsequent examination of PPE crystallised with NaCl and acetate buffer (pH 5.0) without cryoprotectant revealed that an acetate ion could associate with the active site in place of the assigned hydrolytic water (Figure 2.4).

Both NaCl and NaF were found to be capable of crystallising the PPE:BCM7 complex that diffracted to high resolution (1.30 Å; Figure 2.5). pH-jumps with crystals formed in both conditions were reproduced, initially with cryo-protectant (20% glycerol). One structure (NaCl precipitant) frozen after a 1 min 30 s incubation with 0.1 M CHES pH 9.0 exhibited additional electron density within the active site that closely matched that of the assigned tetrahedral intermediate (Figure 2.6; PDB: 1HAZ)[170]. However, with the understanding that both glycerol and acetate can bind within the active site the assignment of an intermediate was re-evaluated. Considering the binding of the two precipitant molecules, the likely binding of three waters and a glycerol molecule in the P_2-P_4 positions, and the conformation of R217A, the electron density is more consistent with the active site of apo-/deacetylated PPE, with glycerol and acetate (partially) occupying the region surrounding S195. Furthermore, when crystals of PPE:BCM7 were soaked in buffer without glycerol, analogous density was not reproduced in any of the 87 structures obtained. Importantly, analogous density to that

observed in 1HAZ and Figure 2.6, was observed after performing pH-jumps with NaCl derived apo-PPE, which demonstrates that analogous density to that observed in the presence of BCM7 can also be observed in the absence of BCM7 (Figure 2.7, B).

It is possible that substrate hydrolysis by SPs proceeds via a concerted mechanism, as is observed with good leaving groups such as *p*-nitrophenol, where the tetrahedral species is essentially a TS[118,237]. Attempts to directly observe SP tetrahedral intermediates by spectroscopic methods have thus far been unsuccessful. For example, stopped-flow studies with PPE and *p*-nitroanilide substrates found no evidence for the accumulation of a tetrahedral intermediate, even at low temperatures[238]. It has been suggested that while a tetrahedral intermediate likely does exist, it is short-lived and does not accumulate[239,240]. Theoretical studies on the acylation[241] and deacylation[242] reactions of SPs support that formation of the tetrahedral intermediate is rate limiting, and thus it does not accumulate. Interestingly, the theoretical study regarding the deacylation reaction was based on the experimental data of the PPE:BCM7 acyl-enzyme complex and the pH-jump studies[170,242]. While these studies do not necessarily mean that a tetrahedral intermediate cannot be observed, it seems unlikely that will be the case under efficient catalysis conditions, particularly where monitoring is on the second time-frame.

It is important to note that X-ray crystal structures are averaged over the crystal volume. While diffusion of small substrates is fast (e.g., O_2 = 2 µm ms^{-1} at 310 K)[192], the large crystal size required for efficient manipulation (>100 × 50 × 50 µm) combined with the rapid catalytic rate exhibited by SPs (k_{cat} = 100–>1000 s^{-1})[105,243,244], likely means that the active sites across the crystal volume are not uniform in regard to the stage of reaction. Furthermore, glycerol can affect the rate of reaction during mixing experiments, likely by reducing the rate of diffusion through the crystal[205]. Thus, the accumulation of a short-lived tetrahedral intermediate to the levels observed here and in the previously assigned structure[170] under these conditions seems unlikely.

Future studies aiming to study the tetrahedral intermediate could make use of the pump-probe SFX methodology. To achieve this, a photocaged substrate, which contains a photolabile group which prevents association with the enzyme active site, is pre-soaked with the enzyme. The photolabile group can be removed using a 'pump' exposure to light, which permits association of the substrate with the enzyme active site. Subsequently, diffraction data is collected during the 'probe' exposure. As discussed in Section 1.5, this approach promotes uniform reaction initiation and fine temporal control over the reaction. Tosha *et al.* (2017) used this approach to study a nitric oxide (NO) reductase with a photocaged NO, and following photolysis, observed NO binding to the active site haem group[245].

Photocaged peptide substrates therefore present an opportunity to study the reaction intermediates of SPs. As the S_1 subsite is dominant in substrate specificity, the P_1 residue is likely the best position to functionalise with a photolabile group. There are a number of commercially available photocaged amino acids with fluorenylmethyloxycarbonyl (Fmoc) protection groups for solid phase peptide synthesis. The majority of studies on photocaged amino acids have thus far explored amino acids with polar sidechain functional groups[246]. As elastase-like SPs preferentially bind to small and hydrophobic residues, the commercially available photocaged amino acids are of limited use for SPs with elastase-like specificity, although the photocaged residue could also be positioned further upstream or downstream from the scissile bond. Photocaged Lys (e.g., Fmoc-L-Lys(Nvoc)-OH; Merck Cat No. 914339) is commercially available. As Lys is preferentially bound by the TR S_1 subsite, substrates with a photocaged P_1 Lys could therefore be used to investigate the TR mechanism[54]. Alternatively, the backbone amide of a CTR substrate (Ac-SIINF▼GHKL-NH_2) was selectively labelled using a boronic acid compound that was directed by the adjacent substrate His residue[247]. The backbone amide of the Gly residue of the CTR substrate was selectively labelled and was unreactive towards CTR-α, prior to irradiation with 365 nm light[247].

It is interesting to note that from all of the datasets collected here, there was no clear evidence of an enzyme:product complex between the carboxylate of the peptide and S195. It is generally accepted

that product dissociation is rapid, based on ~3-fold isotope effects, which have been interpreted as evidence for a chemical rate-limiting step rather than a diffusion-controlled rate-limiting step[248]. The lack of evidence for an enzyme:product complex is therefore consistent with rapid dissociation of the product.

In many of the PPE:BCM7 pH-jumped structures a 2 Å extension along the *a* axis was observed. This extension correlated to a conformational change in two regions of PPE including the metal binding site. The conformational changes involved backbone alterations and substantial movements of sidechain atoms (up to 6 Å) that led to altered coordination of the metal. Conformational changes involving backbone residues in PPE are rare and even large peptide inhibitors have rarely been reported to induce backbone movements[249,250]. Although a CHES molecule associated near what appeared to be an 'origin' of the conformational change, the changes were also produced by soaking crystals with Tris (apo-PPE and PPE:BCM7) and glycine (apo-PPE) which did not appear to associate in the same position. It is therefore unlikely that the changes are due to steric effects caused by binding of buffer components. Raising the pH of the crystal did not always a promote a conformational change, hence the changes are therefore unlikely to be the result of a purely pH effect. The conformational changes were also observed after pH-jumps in the presence and the absence of substrate and thus likely is not linked to substrate binding or release. Due to extensive symmetry interactions between the two regions, it cannot be ruled out that the observed conformational change is a crystallographic artefact. Importantly, the changes are distant to the active site (~16 Å between S195 Cα and W27 Cα) and may therefore not be relevant to catalysis. However, as the metal binding region is important in protein stability, the observations may warrant further study[251].

For serial crystallography experiments, attempts were made to optimise microcrystal slurries of PPE (apo- and PPE:BCM7). Difficulties with adherence to the batch plates and production of non-isomorphous slurries persisted in each condition tested. True microcrystal slurries (< 10-20 µm) were not produced. However, slurries of sufficient density and size (Figure 2.11) were produced to examine

the deacylation of the PPE:BCM7 complex using the serial drop-on-fixed-target methodology[231-234]. While this methodology was appropriate for the larger crystals, other methods that require sample delivery through capillaries would likely encounter blockages and result in poor use of the available beamtime. Smaller crystals were produced for apo-PPE by hanging-drop vapour-diffusion method, but the method is much more labour intensive. To fully prepare for further SSX/SFX experiments, more optimisation is required to form a more robust crystallisation method for PPE microcrystal slurries.

DoC-SSX experiments with PPE:BCM7 (NaF and NaCl) were initiated. pH-jump solutions with lower concentrations (0.2 M) of buffer were found to initiate reaction too slowly for the methodology (> 5.60 s). In single crystal experiments, 0.1 M Tris pH 9.0 was found capable of initiating reaction within 5 s; however, it is difficult to estimate the final concentration of buffer added to each crystal using the DoC methodology. The pyramidic wells of the chips had a volume of approximately 170 pL[235]. However, once the crystallisation solution was drawn through the chip, it was not possible to give an accurate estimate of the volume remaining in the well, and from visual inspection there was variation across the chip. Nonetheless, higher concentrations of buffer (0.5 M and 1 M Tris pH 9.0) exhibited improved reaction times. 1 M Tris resulted in ejection difficulties; a common issue due to the low humidity of the experimental hutch resulting in precipitation of the solution on the ejector tip (personal communication with Dr Pierre Aller and Johan Glerup, Diamond Light source)[235]. Addition of 0.5 M Tris, pH 9.0, was found to be capable of inducing complete deacylation of PPE between 1.12 s and 5.60 s. While this represents progress, as mentioned above this timescale likely remains too slow to accumulate the short-lived intermediate across the crystal. The crystals used were large, and the rapid rate of reaction means that any tetrahedral intermediate present, if any tetrahedral intermediate is in fact present, will likely be a minor fraction and therefore averaged out. Furthermore, while ejection of buffer was optimised to at least a 95% hit rate prior to experiments, there is currently no method to monitor ejection in real-time. Thus, merged structures likely represent a mixture of states (no mixing,

mixing for the correct time, and mixing for an incorrect time). It is unlikely that under these conditions, short-lived intermediates would accumulate to sufficient levels to observe *in crystallo*.

Chapter 3: Alternate Inhibitors for Serial Crystallography with PPE

3.1. Introduction

3.1.1. Derivatives of BCM7

To develop an understanding of the interaction between PPE and BCM7, the kinetics of *N*–terminal truncation BCM7 variants and P_3 mutants have been investigated (Table 3.1)[107,252]. Truncation from the *N*–terminus of BCM7 (1-5) reduced the inhibitory potency of the peptides as determined by an increase in K_i. However, *N*–acetylation of the truncated peptides restored the inhibitory potency of certain peptides. Removal of the *N*–terminal positive charge, which may be poorly tolerated within the substrate binding groove of PPE, likely caused the restoration of inhibitory activity of the peptides. The potency of *N*–acetyl-Val-Glu-Pro-Ile (Ac-VEPI) was restored to that of BCM7 (K_i = 0.76 mM for both peptides), while *N*–acetyl-Glu-Pro-Ile (Ac-EPI) exhibited only a slight decrease in inhibitory potency compared with BCM7 (K_i = 0.97 mM)[107]. This observation demonstrated that strong inhibition of PPE could be achieved through the use of small tetra- and tri-peptides[107].

Due to the proposed interaction between the P_3 residue of BCM7 and Q192 of PPE, kinetic analyses were performed on various P_3 substituted variants of the *N*–acetylated tetra- and tri-peptides (Table 3.1)[107,252,253]. The presence of a P_3 ornithine or glutamine produced the most pronounced effects, increasing the potency by an order of magnitude in the tripeptide series (K_i = 0.018 mM, 0.095 mM, and 0.97 mM for *N*–acetyl-Orn-Pro-Ile (Ac-OPI), *N*–acetyl-Gln-Pro-Ile (Ac-QPI), and Ac-EPI, respectively). This observation suggests that both the side chain length (α–γ for Asp, α–ε Gln and Orn) and the presence of a nitrogen containing side chain group can lead to improved potency. Thus, the work supported that the P_3 residue was important for the formation of the stable acyl-enzyme complex, at least in the *N*–terminally truncated peptides.

Table 3.1 | Kinetic analysis of BCM7 truncations and P3 mutants. The table was adapted from Wright *et al*. (2001)[107] and Katona *et al*. (2002)[252]. All peptides had a *C*–terminal carboxylic acid. Ac = *N*–acetyl; Hse = homoserine; Orn = ornithine.

Sequence	Ki (mM)	IC_{50} (mM)
BCM7 *N*–terminal truncation[107]		
NH_2-YPFVEPI	0.76	
Ac-PFVEPI	1.8	
NH_2-PFVEPI	2.1	
Ac-FVEPI	2.7	
NH_2-FVEPI	1.4	
Ac-VEPI	0.76	
NH_2-VEPI	1.2	
Ac-EPI	0.97	
NH_2-EPI	1.3	
NH_2-PI	No inhibition	
P3 mutagenesis[252]		
	Ac-V**X**PI	
Ac-V**E**PI	0.76	2.4
Ac-V**Y**PI		6.4
Ac-V**H**PI	1.7	2.6
Ac-V**K**PI	1.3	
Ac-V**M**PI		2.4
Ac-V**D**PI		2.5
Ac-V**S**PI	2.7	
Ac-V**N**PI	0.87	2.0
Ac-V**Q**PI	0.42	1.1
Ac-V-**Hse**-PI		2.4
Ac-V-**Orn**-PI		0.72
	Ac-**X**PI	
Ac-**E**PI	0.97	3.1
Ac-**D**PI		3.3
Ac-**S**PI	0.34	1.2
Ac-**N**PI	0.89	1.2
Ac-**Q**PI	0.095	0.34
Ac-**Hse**-PI	0.11	0.42
Ac-**Orn**-PI	0.018	0.24

Crystal structures of PPE in complex with Ac-VEPI and Ac-NPI have been reported[107,252,253]. The Ac-NPI structure diffracted to atomic resolution (0.95 Å) and provided valuable information regarding the complex. Ac-NPI bound analogously to BCM7, however, additional density was observed in the S' subsites in the Ac-NPI structure. The apparent occupancy of the peptide was low (59%) and the additional density in the S_1' position was proposed to be a sulphate ion in an alternate conformation

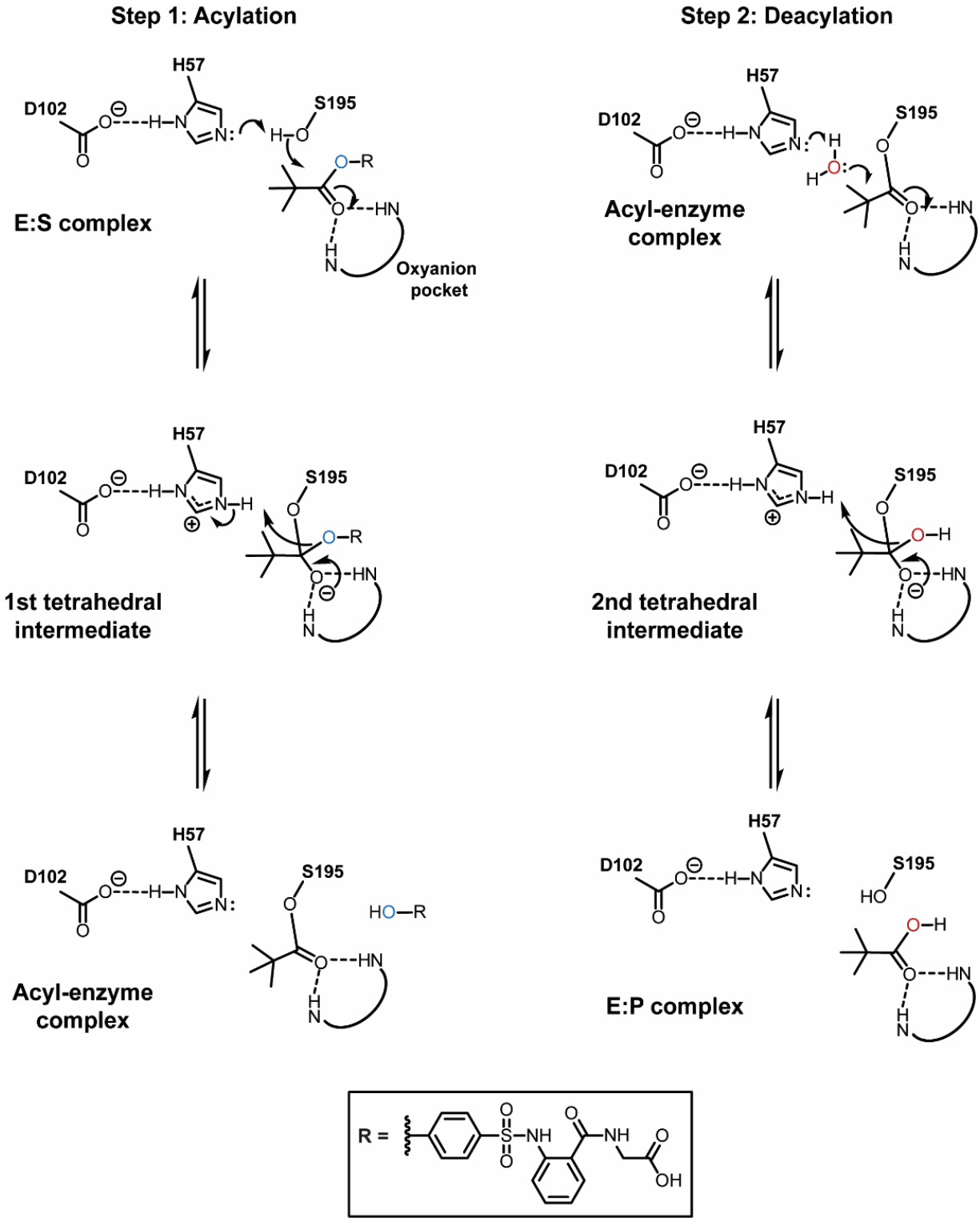
to the acyl-enzyme[106]. The structure provided evidence for the planar geometry of the ester bond and potential positioning of the hydrogen atom between H57 and D102[252]. However, attempts to capture crystal structures of PPE with peptides exhibiting higher inhibitory potency compared with BCM7 were unsuccessful (e.g., Ac-OPI and Ac-QPI). This may be indicative of a conformational change occurring upon binding of the more potent inhibitors. Thus, it is of interest to obtain crystal structures of PPE with the mutated P_3 variant peptide inhibitors.

3.1.2. Sivelestat

The work presented thus far in this thesis has predominantly focused on the deacylation reaction of the PPE:BCM7 complex through pH-jump methodology. However, alternative approaches to time-resolved crystallography have proven successful, including the soaking of inhibitors into crystals of apo-enzyme[254,255]. One such inhibitor that is potentially useful for studying the PPE mechanism is the acyl-enzyme forming inhibitor Sivelestat (ONO-5046)[256,257]. The pivaloyloxy benzene derivative, Sivelestat, is one of only two clinically approved inhibitors of human neutrophil elastase (HNE)[258]. The other inhibitor is sold as Prolastin® or Prolastin-C®. The active component is purified alpha$_1$-antitrypsin (α_1-antitrypsin or α_1-proteinase inhibitor (α_1-PI)) an endogenously expressed serpin that regulates the activity of HNE[259]. Sivelestat (Elaspol® 100) is therefore the only approved small molecule inhibitor of HNE, though to date it has only been approved for use in Japan and the Republic of Korea. Negative results from the 180-day mortality rates of the Sivelestat Trial in Acute Lung Injury (ALI) Patients Requiring Mechanical Ventilation (STRIVE) clinical trial led to cessation of the trial in the USA, Canada, Belgium, Spain, Australia, and New Zealand[260,261]. Sivelestat was nonetheless the first approved drug for the treatment of ALI associated with systemic inflammatory response syndrome (SIRS)[262].

Sivelestat is a potent inhibitor of HNE (IC_{50} = 44 nM)[263]. HNE is proposed to hydrolyse the ester bond of Sivelestat with the reaction likely proceeding through a tetrahedral intermediate[264]. Collapse of the intermediate leads to formation of an acyl-enzyme complex comprising a *tertiary*-butyl ester (*t*Bu-ester) with concomitant release of a fragment of Sivelestat (Scheme 3.1). Nucleophilic attack of a water

on the carbonyl carbon of the acyl-enzyme complex results in formation of a second tetrahedral intermediate. Subsequent collapse of this intermediate leads to release of pivalic acid and regeneration of PPE to the native state (Scheme 3.1)[264]. This proposed mechanism is supported by LC-MS experiments whereby an 84 Da adduct was bound to each of the four isozymes of human neutrophil elastase (HNE) immediately after addition of Sivelestat. The rapid acylation of HNE was consistent with formation of a *t*Bu-ester (calculated mass adduct = 84 Da). The enzyme remained fully acylated after 10 min but was fully deacylated by 20 minutes. Thus, the soaking of Sivelestat into crystals of apo-PPE offers an opportunity to gain insight into the mechanism of the acylation and deacylation reactions.



Scheme 3.1 | The Proposed mechanism of inhibition for Sivelestat. The figure was adapted from Nakayama *et al.* 2002[264].

3.1.3. Aims

The aims of the work presented in this chapter were to explore alternative acyl-enzyme complex forming inhibitors that have potential utility for time-resolved experiments with PPE. First, studies on the crystallisation of PPE with derivatives of human BCM7 are discussed. Secondly, studies on the non-peptide inhibitor, Sivelestat, are described.

3.2. Results

3.2.1. Alternative Peptide Inhibitors

Following the work on BCM7 (Chapter 2). Alternative peptide inhibitors were explored for viability in time-resolved crystallography experiments (Figure 3.1). It was envisaged that those with stronger binding affinities might result in different results during pH-jumps, with such inhibitors being less likely to be ejected from the active site. Note that many of the previously investigated BCM7 derivatives did not co-crystallise with PPE and led to rapid cracking of crystals while soaking, indicating that they may induce a conformational change in PPE[253].

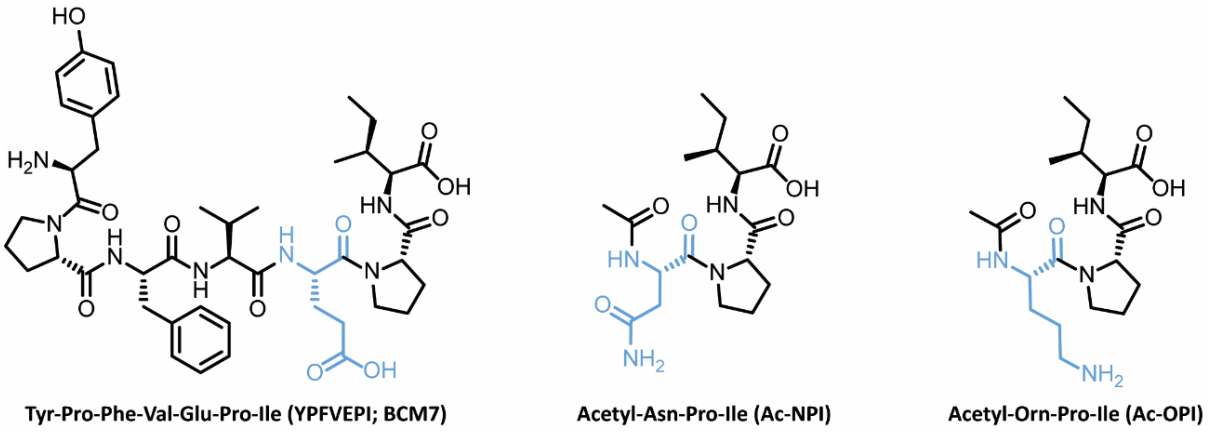


Figure 3.1 | BCM7 and two derivatives targeted for PPE pH-jump studies. Left) human BCM7; **middle)** *N*–terminally acetylated NPI; **right)** *N*–terminally acetylated ornithine-PI. The P_3 residue of each peptide is highlighted in blue.

Co-crystallisation of PPE with Ac-NPI produced crystals under NaF and NaCl precipitant conditions (Figure 3.2, A & B). The resulting crystal structures demonstrated weaker binding of Ac-NPI compared with BCM7, on the basis of refined occupancy. While the occupancy of BCM7 was refined to 75–90% at 30 eq, Ac-NPI at 40 eq was not refined to more than 77% occupancy (Figure 3.2, C). In the 2m*F*o-D*F*c map there was continuous density corresponding to the ester bond between Ile7 and S195, which is suggestive of an acyl-enzyme complex. However, only weak electron density was present between the Cα and the carbonyl carbon of Ile7. This may be due to the lower occupancy of the peptide caused by the weaker binding affinity of Ac-NPI compared with BCM7. Similarly to the reported Ac-NPI structure (PDB: 1GVK)[252], additional density was present in the S' region. The additional density in the S' region

superimposed with the acetate ion observed in the apo-PPE structure (Figure 3.2, C). Due to lower occupancy of Ac-NPI, this electron density likely represents a composite of two states across the crystal, with Ac-NPI and the proposed hydrolytic water[106] bound in one state and acetate bound in the other.

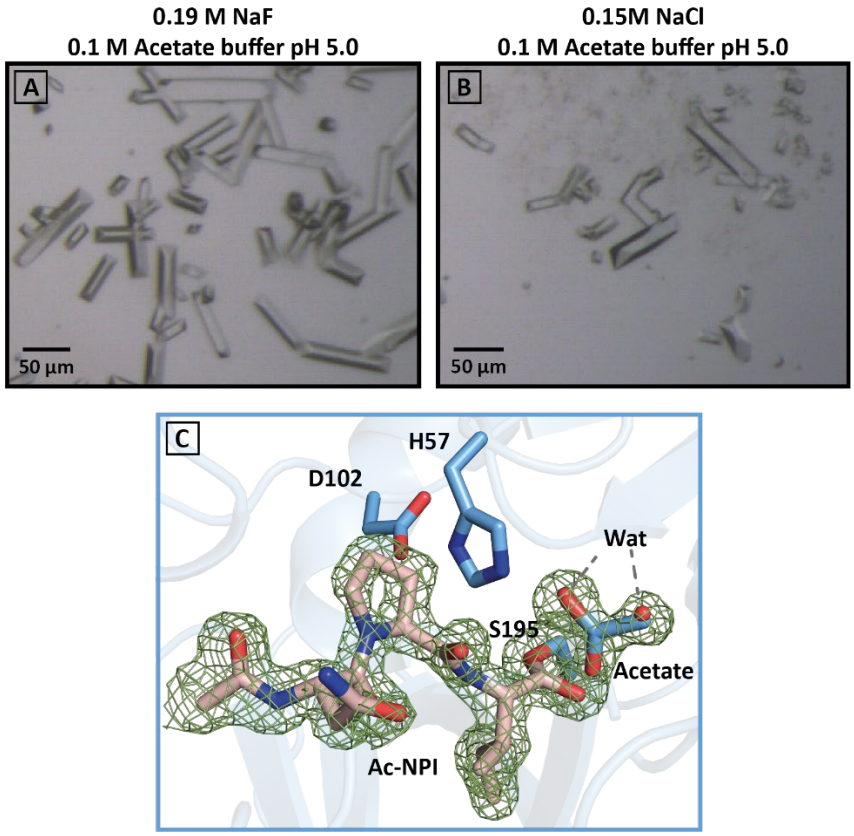


Figure 3.2 | PPE:Ac-NPI crystallography. A) Image of PPE crystallised in the presence of 40 eq Ac-NPI using NaF as a precipitant; **B)** Image of PPE crystallised in the presence of 40 eq Ac-NPI using NaCl as a precipitant; **C)** Crystal structure of PPE:Ac-NPI complex. The catalytic triad and acetate carbons are in light blue, Ac-NPI carbons are pink, and the Polder omit map shown as green mesh is contoured to 3 σ.

Previous attempts to crystallise more potent BCM7 derivatives with PPE had been unsuccessful using the traditional sulphate precipitant[253]. Attempts to co-crystallise or soak Ac-OPI, the most potent BCM7 derivative, in NaF and NaCl precipitants were similarly unsuccessful. Thus, PPE:Ac-OPI co-crystallisation was explored using a range of commercial screens (SaltRx, Nextal Anions, Nextal Cations, Pact Premier, Morpheus, SG1, JCSG Plus, BCS, and Crystal Screen). Crystals formed in four conditions within four days, but only one produced X-ray diffraction consistent with a protein crystal. The crystal processed to the *P* 6_4 space group (2.1 Å) and was non-isomorphous with the common unit cell parameters for

PPE (Figure 3.3, left and Table 3.2). There were two monomers in the asymmetric unit and no major conformational changes compared with PPE in the common *P* 2_1 2_1 2_1 space group. The crystal was buffered to a pH of 8.5 and thus was not envisaged to be favourable for formation of a stable acyl-enzyme complex[106]. Additional density was present in the active sites of both monomers in the asymmetric unit, but the observed electron density was different for each monomer. In neither case was the density consistent with formation of an acyl-enzyme complex, and neither could be modelled with the Ac-OPI in any orientation (data not shown). To improve the likelihood of formation of an acyl-enzyme complex, crystallisation of PPE:Ac-OPI was investigated in acetate buffer (pH 4.1, 4.5, 5.0, and 5.4), using KSCN as a precipitant. Large crystals with hexagonal morphology formed in each of the four pH conditions within two days (Figure 3.3). While the crystals formed in acetate were much larger than those from the original screen, the resolution decreased (≥ 2.8 Å). The resultant X-ray crystal structures also processed to *P* 6_4 and were approximately isomorphous to the unit cell parameters of the crystal at pH 8.5 (Table 3.2). Structures of PPE crystallised at both pH 4.1 and 5.4 revealed similar electron density within the active site that was again not consistent with an acyl-enzyme complex and which could not be modelled convincingly (data not shown). Due to this, only initial rounds of refinement were performed on these structures.

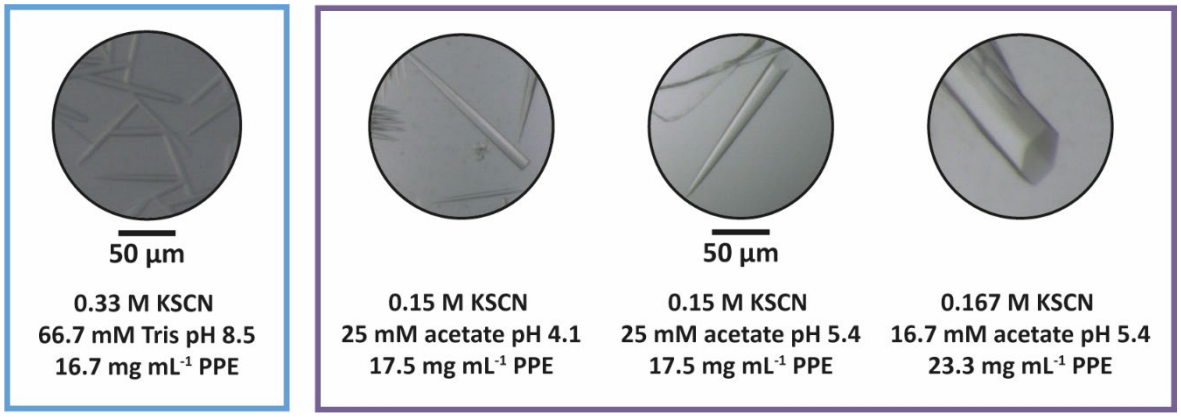


Figure 3.3 | Images of PPE:Ac-OPI crystallisation screen and optimisation. Blue box) Image of crystals from the initial screen; purple box) images of crystal from an optimisation screen using acetate buffer (pH 4.1-–5.4). The conditions of each crystal are underneath each image.

There is precedent for crystallisation of PPE in space groups other than the common $P\,2_1\,2_1\,2_1$. A search of the PDB for PPE in alternate space groups returned six structures (Table 3.2). All six were PPE:inhibitor complexes, of which most were large peptide inhibitors. The highest resolution of these structures was 1.80 Å, which could be sufficient for time-resolved studies. However, most of these structures, including the novel structures reported here, did not diffract to the atomic resolution achievable by PPE in the orthorhombic space group[89,252]. Even if Ac-OPI could be observed within the active site in the novel crystal forms, it would be unlikely that any useful mechanistic data would be attainable at such low resolution. Therefore, further investigation with PPE:Ac-OPI under these conditions was not pursued.

Table 3.2 | Crystals structures of PPE in alternate space groups to $P\,2_1\,2_1\,2_1$.

Inhibitor	Resolution (Å)	Space group	Unit cell	PDB Accession
Ac-OPI (Tris pH 8.5)	2.10	$P\,6_4$	a=b=95.92 Å, c=113.89 Å α=β=90°, γ=120°	N/A
Ac-OPI (acetate buffer pH 4.1)	2.81	$P\,6_4$	a=b=96.48 Å, c=113.46 Å α=β=90°, γ=120°	N/A
Ac-OPI (acetate buffer pH 5.4)	2.80	$P\,6_4$	a=b=96.92 Å, c=113.58 Å α=β=90°, γ=120°	N/A
Ascaris chymotrypsin/elastase inhibitor	2.40	$P\,3_2\,2\,1$	a=b=82 Å, c = 191 Å α=β=90°, γ=120°	IEAI
MDL 101, 146	1.80	$P\,3_1$	a=b=69 Å, c=51 Å α=β=90°, γ=120°	1B0E
Scyptolin A	2.80	$P\,6\,2\,2$	a=b=156 Å, c=91 Å α=β=90°, γ=120°	1OKX
Elafin	1.90	$P\,1\,2_1\,1$	a=38 Å, b=73 Å, c=49 Å α=γ=90°, β=105.4°	1FLE
α_1-antitrypsin	3.30	$C\,1\,2\,1$	a=110 Å, b=85 Å, c=76 Å α=γ=90°, β=121°	2D26
Kunitz-type protease inhibitor ShPi-1 (K13L mutant)	2.00	$C\,1\,2\,1$	a=133 Å, b=47 Å, c=43 Å α=γ=90°, β=100.07°	2UOU

3.2.2. Non-Peptidic Elastase Inhibitor: Sivelestat

The inhibitory potency of Sivelestat towards PPE was first measured using an absorbance assay (Section 5.1.2.1). The dose-response curve demonstrated that Sivelestat is a potent inhibitor of PPE with a pIC_{50} of 6.37 ± 0.06 (IC_{50} = 0.43 µM; Figure 3.4, A) but is 10-fold less potent compared to HNE (IC_{50} = 44 nM)[264]. The reaction between PPE and Sivelestat was also monitored by protein observed LC-MS (Figure 3.4, B). Under control conditions (Tris pH 7.5) a single ion peak corresponding to 25,892 Da was observed (Figure 3.4, B). Similarly to HNE[264], after a 10 min incubation with Sivelestat, the predominant peak corresponded to 25,976 Da (+ 84 Da), supporting formation of an acyl-enzyme complex comprising a *t*Bu-ester. After 15 min, only the deacylated enzyme was observed. However, at no time was PPE fully acylated by Sivelestat. Furthermore, after 5 min incubation the major peak corresponded to apo-PPE with only a minor peak exhibiting acylation. This observation contrasts the data collected with HNE which exhibited complete acylation immediately after addition of Sivelestat and may reflect the slightly lower inhibitory potential towards PPE.

To further confirm the identity of the complex with PPE and Sivelestat, attempts were made to obtain the complex *in crystallo*. Co-crystallisation proved unsuccessful, as previously observed[264]. Crystals of apo-PPE (NaCl) were instead soaked with Sivelestat (20 mM) for up to 5 min. Most of the crystals revealed continuous density from S195 to a small adduct that was confidently modelled as a *t*Bu-ester (Figure 3.4, C). The adduct refined to high occupancy (96 %) and the B-factors were low (< 20 $Å^2$). The shorter reaction time *in crystallo* compared with LC-MS may reflect the higher concentration of Sivelestat used for soaking compared with LC-MS (20 mM vs 150 µM, respectively). However, it is not possible to give an accurate value of the ratio of PPE to Sivelestat *in crystallo* as crystals were transferred to a drop containing Sivelestat. Nonetheless, the results were promising. Although a 1 min incubation would be infeasible for time-resolved SSX/SFX experiments, the use of smaller crystals and a higher concentration of Sivelestat has the potential to further reduce the time.

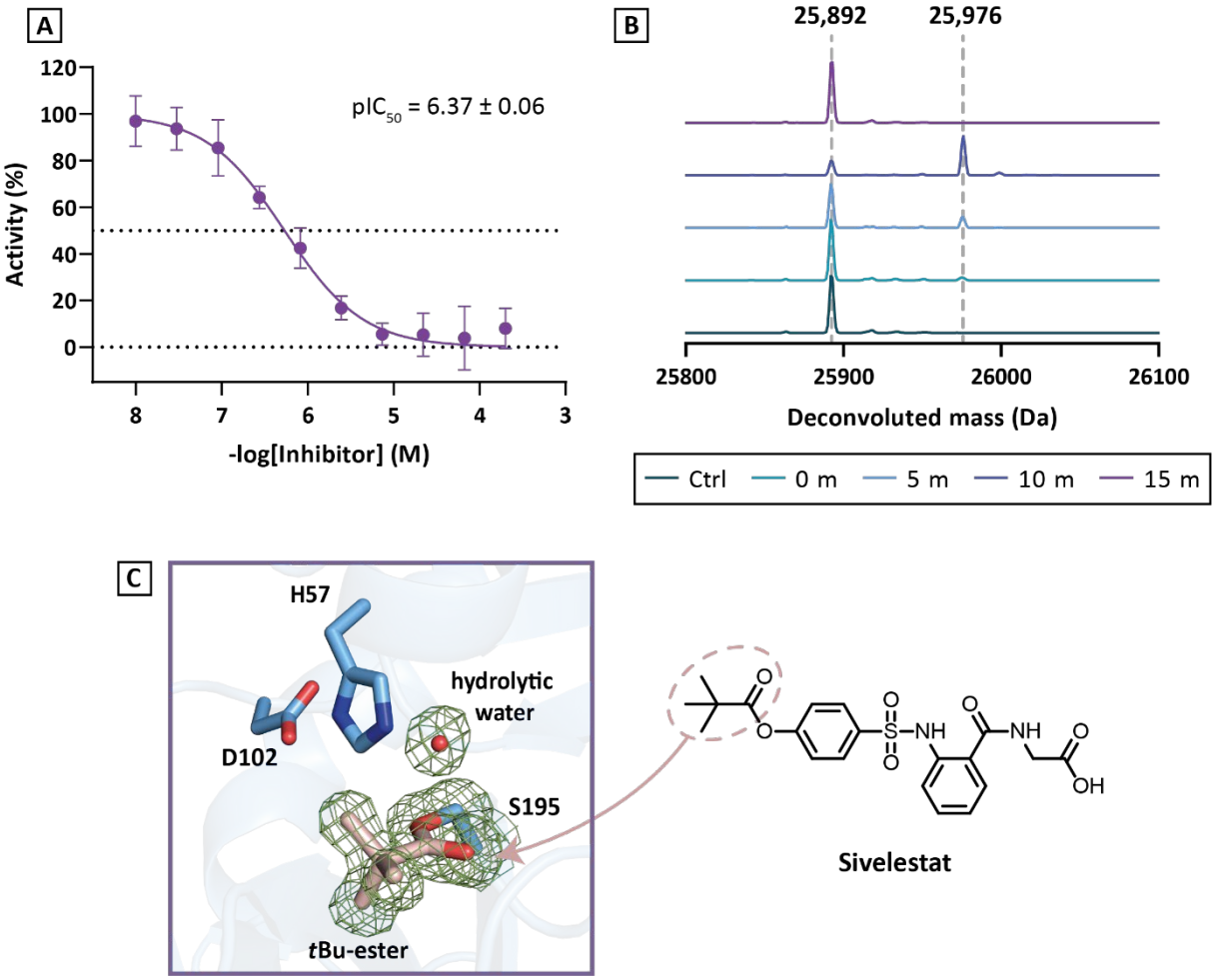


Figure 3.4 | Reaction of Sivelestat with PPE. A) A dose-response curve for PPE and Sivelestat. Data are shown as 2 independent repeats, each with 4 experimental repeats (mean ± SD, n=8); **B)** LC-MS monitoring the reaction of PPE and Sivelestat in Tris buffer pH 7.5 over 15 min; **C)** a view from a crystal structure of PPE soaked with Sivelestat. The catalytic triad carbons are light blue, *t*Bu-ester adduct carbons are pink, and the Polder omit map shown as green mesh is contoured to 3 σ.

To prepare for soaking experiments with Sivelestat, microcrystalline slurries of apo-PPE in NaCl precipitant were optimised using the hanging drop vapour diffusion method (Figure 3.5). However, as discussed previously (Section 2.2.3.1), due to the large volume required for time-resolved DoC-SSX experiments, the preparation of slurry was extremely labour intensive and was prone to well-to-well differences. Nonetheless, an initial test was performed using the DoC-SSX method at Diamond Light Source. Sivelestat was found to be insoluble in water, unless sodium hydroxide (NaOH) was added to a ratio of 1:2 (Sivelestat:NaOH), but was soluble in DMSO. Ejection of Sivelestat at 400 mM was optimised prior to data collection at Diamond Light Source in both DMSO and water (+NaOH). However, only Sivelestat in DMSO was found to reliably eject (personal communication with Dr Pierre

Aller and Johan Glerup, Diamond Light Source). Therefore, 400 mM Sivelestat in DMSO was used for the initial test.

Data was collected from a single chip after mixing with Sivelestat for 5.6 s. The DoC-SSX method of data collection was discussed in Section 2.2.3.3. The sample produced strong protein diffraction, indicating that the crystals did not dissolve within the mixing time. However, there was no additional electron density within the active site of PPE for Sivelestat or a *t*Bu-ester adduct. Furthermore, upon visual inspection of the chip after data collection there was substantial leakage of solution across the chip. Similar leakage of solution was not observed using Tris buffer after pH jumps with PPE, and likely resulted from the use of DMSO solvent. The data collection method involves ejecting solution onto the chip in a raster pattern followed by data collection on the same wells after a time delay (Chapter 2). The leakage may therefore contaminate wells below with the inhibitor prior to ejecting a droplet into that well. Thus, the actual mixing time experienced by each crystal may not be consistent. It would therefore be unwise to continue with this approach. Furthermore, Sivelestat did not react within the mixing time and while slightly longer mixing times are available (5–10 s), a wiser approach may be to identify another inhibitor that reacts within this window.

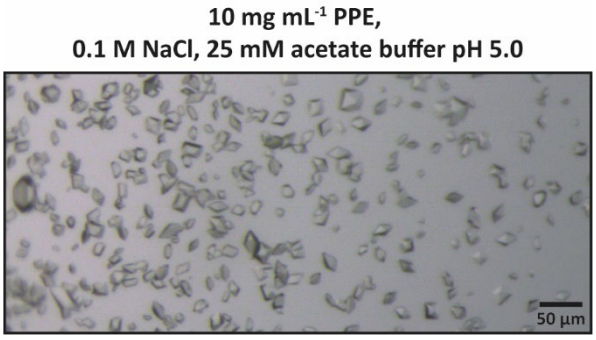


Figure 3.5 | Image of the native PPE microcrystal slurry in NaCl precipitant. Crystals were grown at ambient temperature by hanging drop diffusion method.

3.3. Discussion and Perspectives

To explore other options for time-resolved crystallography with PPE, crystallisation with alternative inhibitors was investigated. Ac-NPI was initially explored due to the atomic resolution diffraction afforded by the PPE:Ac-NPI complex (0.95 Å)[252]. While Ac-NPI was successfully co-crystallised with PPE in NaCl and NaF, the occupancy remained relatively low and additional electron density in the S' region would likely complicate subsequent time-resolved experiments. Instead, the most potent BCM7 derivative was examined. Under traditional conditions using sulphate precipitant the complex did not co-crystallise[253]. Soaking Ac-OPI into crystals previously led to rapid cracking, which is suggestive of movement within the crystal lattice. Similar observations were made for PPE in both NaCl and NaF precipitants by co-crystallisation and soaking methodology. Co-crystallisation of PPE and Ac-OPI yielded a novel crystallisation condition for PPE with KSCN as the precipitant. The resulting crystals processed to a novel space group ($P\,6_4$), although the resolution afforded by crystals of the complex in this condition was poor. Furthermore, there was no evidence of acylation of S195 by Ac-OPI and no conformational changes were observed. No other conditions yielded protein crystals and therefore it is unlikely that the PPE:Ac-OPI complex will be useful for time-resolved studies. However, several other BCM7 derivatives with improved potencies were identified in the initial study (Table 3.1). In particular, the *N*–acetylated tripeptides with glutamine, homoserine, or serine, and the *N*–acetylated tetrapeptides with glutamine or ornithine in the P_3 position were all better inhibitors of PPE compared with BCM7. In the future it may be worth pursuing these inhibitors for crystallographic analyses.

Sivelestat, the only approved small molecule inhibitor of HNE, has a generally accepted mechanism of reaction involving formation of an acyl-enzyme complex. However, to date the mechanism has not been confirmed *in crystallo*. Here, the first crystal structure of an intermediate of the Sivelestat reaction with PPE was reported. The data showed that following ester bond hydrolysis, a *t*Bu-acyl-enzyme complex was formed between PPE and the *t*Bu sidechain of Sivelestat, supporting solution studies[264]. However, while Sivelestat was able to form a complex with PPE *in crystallo*, the timeframe

for acylation was found to be insufficient for time-resolved SSX/SFX studies. Further optimisation of the crystal conditions, such as crystallisation at pH 7 or above rather than pH 5, which would increase the activity of PPE, may reduce the required mixing time. Alternatively, studies have been dedicated towards the production of Sivelestat derivatives, several of which display improved potency towards HNE and improved stability in aqueous buffers[258,265,266]. These inhibitors may offer an opportunity to target PPE for SSX/SFX experiments, although they are generally not commercially available and the selectivity towards PPE requires assessment.

Chapter 4: Recombinant Expression and Purification of PPE

4.1. Introduction

Proteins are dynamic structures, but until recently it has not been possible to study fast protein dynamics by crystallographic means. The use of XFEL radiation sources combined with SFX methods has already started to provide examples of protein dynamics during catalysis that have previously not been observed[193-195,199,202,205], as discussed in Section 1.5. The active sites of PPE and related CTR-like peptidases have generally been considered to be rigid[118]. The binding of substrates and inhibitors has rarely been demonstrated to promote conformational changes, at least that are observable by X-ray crystallography at cryogenic temperatures, although examples of subtle conformational changes have been observed with certain tetrahedral inhibitor complexes[118,267,268]. Future SFX studies with PPE could reveal protein motions associated with catalysis, including those of the catalytic residues, S195 and H57, during proton transfer steps[153]. Additionally, the active site of PPE and related SPs contains a network of hydrogen bonds that connect the catalytic residues to distal regions, including the S_1 subsite and the residues involved in formation of the substrate antiparallel β-sheet[88,118,170]. The network has therefore been proposed to act as link a between catalysis and substrate binding[53,118]. While SFX studies may reveal the extent of motions associated with catalysis in the future, it is important to verify the importance of any observed motions using alternative methodologies.

Site-directed mutagenesis combined with biophysical and kinetic studies, has been utilised to investigate the importance of the catalytic triad and oxyanion residues of SPs, as discussed in Section 1.4.3. Protein engineering studies have also explored SPs, with aims of varying the specificity of the SPs, e.g., converting the specificity of TR to CTR-like[269-271]. The differences between the S_1 subsites of CTR, TR, and PPE are often described simply in terms of residues 189, 216, and 226 (Figure 1.4)[54]. However, simply exchanging D189 of TR for S189 of CTR was not sufficient to reproduce CTR specificity, instead the resultant mutant was relatively nonspecific and a poor catalyst[269]. The CTR-

like activity of TR was improved by additional mutations in the S_1 subsite (Q192M, I138T, and $T219_{insertion}$), and exchange of the TR residues in two loops (185–188 and 221–224) for the analogous CTR residues[270]. Interestingly, the two loops are not involved directly in formation of the S_1 pocket, nor in the other enzyme subsites. The rationale behind the improvements was therefore not definitively identified, but may have been related to structural flexibility around the S_1 pocket[270]. Finally, an additional Y172W mutation, with the previously mentioned S_1 and double loop mutations, increased the catalytic activity to 15% of the wild-type CTR enzyme[271].

Following from the above described work, attempts were made to impart TR-like specificity onto CTR, using the reverse S_1, double loop, and 172 mutations[272]. The reverse mutations did not impart TR-like specificity and the catalytic activity of the mutants towards both TR and CTR peptide substrates was poor[272]. Similarly, when TR was mutated with the analogous residues from PPE, it did not impart elastase-like specificity onto the mutant TRs[273]. Thus it was proposed that it is not simply the tertiary arrangement of the S_1 pocket that determines specificity and that each SP has a unique set of contributing interactions[273].

Site-directed mutagenesis studies, analogous to the above, have thus far not been performed in PPE. The aim of the work presented in this Chapter, was to optimise soluble expression and purification of catalytically active PPE from *E. coli*. The *E. coli* expression system was selected due to the widespread availability of the system, the relatively low cost, and the scalability.

4.2. Results

4.2.1. His_6-ProPPE Expression Trials

4.2.1.1. Gene Design and Cloning

The inherent proteolytic properties of PPE and its structure necessitated consideration of the construct design: the presence of four disulphide bonds could cause issues within the reducing environment of the *E. coli* cytoplasm[274], and the *N*–terminal initiator methionine residue could impair folding of the active enzyme (Figure 1.8). Thus, PPE expression in a zymogen form, which included the native *N*–terminal propeptide, was investigated (Figure 4.1). Expression of PPE in the zymogen form would also likely reduce the toxic proteolytic effect of PPE towards the host cell. It was envisaged that subsequent cleavage of the propeptide by TR during purification would reveal the active form.

A codon-optimised gene encoding for proelastase (ProPPE) was subcloned into the pNIC-28_Bsa4 (pNIC) vector by ligation independent cloning (LIC). A His_6-ProPPE gene was also subcloned into a modified pCOLD-I (pCOLD) vector. As the 3C cleavage site was deemed to be unnecessary due to the presence of the activation peptide, the pCOLD vector was further modified during Gibson Assembly, that is by removing DNA encoding for the standard His_6-tag and 3C cleavage site and inserting DNA encoding for the His_6-ProPPE gene immediately following the start codon (Figure 4.1).

Both constructs were transformed into two *E. coli* strains for expression trials, BL21(DE3) and SHuffle T7 (DE3). SHuffle T7 (DE3) cells were selected as PPE contains four disulphide bonds. While the reducing environment of the *E. coli* cytoplasm may impair folding, the SHuffle strain expresses a cytoplasmic disulphide bond isomerase (DsbC) that promotes formation of disulphide bonds within the cytoplasm[274]. From here on, these two strains will respectively be referred to as BL21 and SHuffle for brevity.

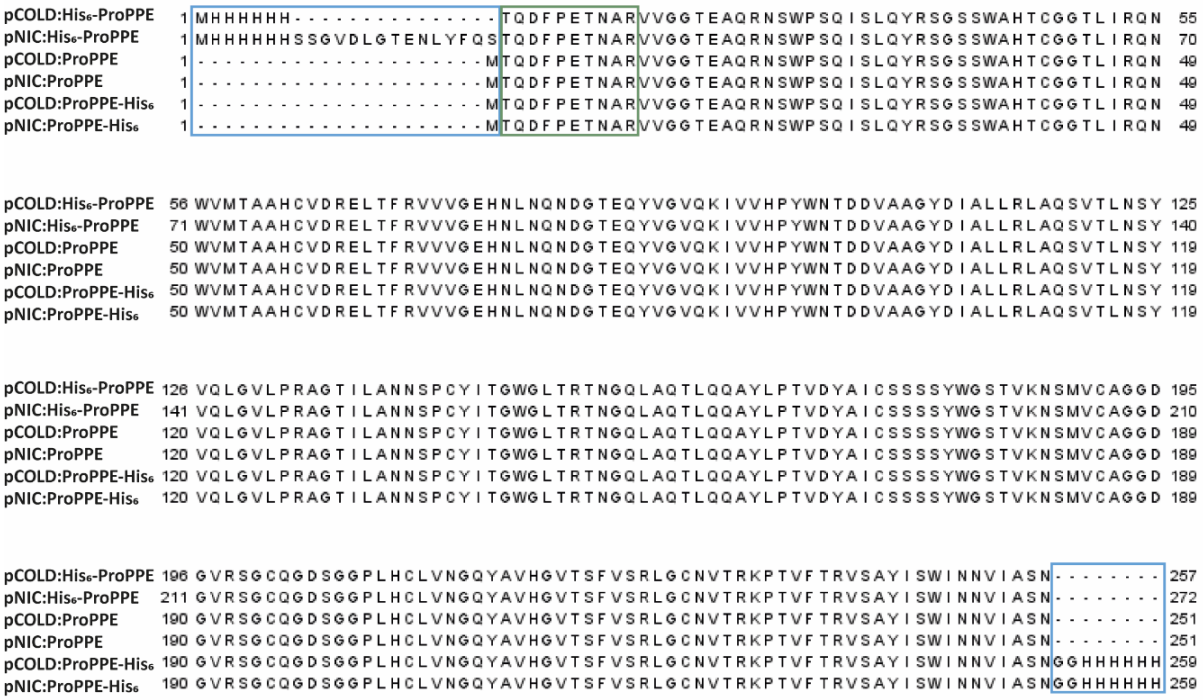


Figure 4.1 | Protein sequence alignment of His_6-ProPPE, ProPPE, and ProPPE-His_6. *N*– and *C*–terminal alterations are highlighted in blue; the propeptide sequence is highlighted in green.

4.2.1.2. His_6-ProPPE Expression Trials

Cells were cultured in 2×YT media at 30–37 °C, with agitation (200 rpm) until an optical density (OD_{600}) of 0.5–0.7 was achieved. Gene expression was induced through addition of 0.5 mM isopropylthio-β-galactoside (IPTG; for pNIC vector) or by cold shock at 4 °C for 15 min (for pCOLD vector). The cultures were then split into 10 mL fractions and incubated at various temperatures for 16 h, with agitation (200 rpm). The cell pellets were lysed using detergent (BugBuster® supplemented with lysozyme and DNase I). As PPE is a SP, no protease inhibitors were added to the lysis buffer. Soluble and insoluble expression was analysed by SDS-PAGE for overexpression of a 28 kDa protein (Figure 4.2, A). The results demonstrated that the majority of, if not all, conditions resulted in insoluble recombinant PPE production.

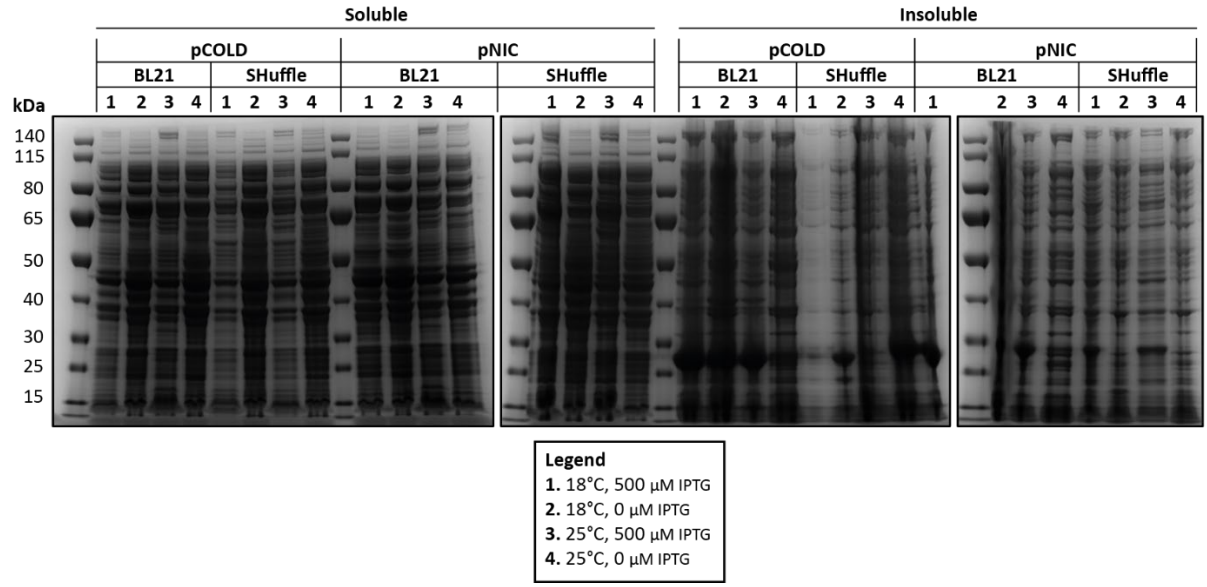


Figure 4.2 | Expression trials with His_6-ProPPE. SDS-PAGE gel of soluble and insoluble fractions from His_6-ProPPE expression trials in BL21 and SHuffle cells.

4.2.1.3. PPE: N-Terminal Signal Peptides

The presence of multiple disulphide bonds within the PPE structure was considered to be a potential cause of the observed insolubility. Signal peptide sequences were thus inserted into the *N*–terminus of the vectors. The presence of specific *N*–terminal sequences can direct unfolded proteins from the reducing cytoplasm of *E. coli* into the oxidising periplasmic space immediately following translation, where disulphide bond formation is promoted[275-277].

Three signal peptides from DsbA, PhoA, and OmpA were inserted into the pNIC vector and one into the pCOLD vector (DsbA) at the *N*–termini by insertion PCR (Table 4.1)[278]. Expression trials were conducted with the four constructs in BL21 cells under conditions stated in Section 7.5.2. The periplasmic fraction was isolated by osmotic shock in Tris:ethylenediaminetetraacetic acid (EDTA):sucrose (TES) solution, followed by ultrapure water supplemented with $MgCl_2$ (hypotonic solution). The hypotonic fractions were analysed by SDS-PAGE (Figure 4.3)[279]. In all the samples, bands were observed at approximately the correct molecular weight (MW) for His_6-ProPPE (28 kDa), but these bands were not intense. To verify the protein identity, samples were purified by Ni^{2+} immobilised metal affinity chromatography (IMAC). No protein was observed by SDS-PAGE after affinity

purification, suggesting that either the band was not His_6-ProPPE or that the His_6-tag was not solvent exposed.

Table 4.1 | N-terminal signal peptide sequences

Peptide	Sequence
DsbA	MKKIWLALAGLVLAFSASA
OmpA	MKKTAIAIAVALAGFATVAQA
PhoA	MKQSTIALALLPLLFTPVTKA

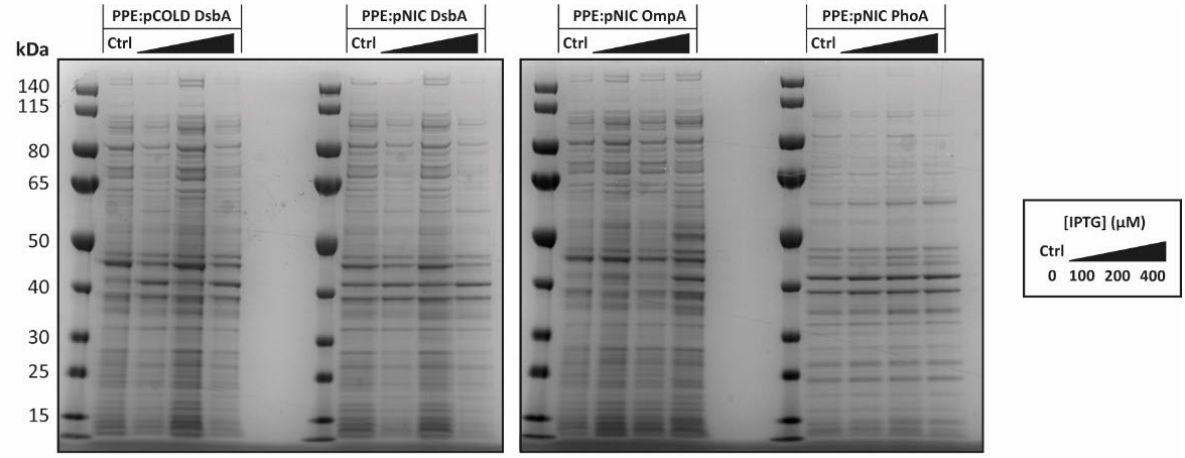


Figure 4.3 | Expression trials with His_6-ProPPE constructs with *N*–terminal signal peptides. SDS-PAGE gel of soluble fractions from His_6-ProPPE signal peptide expression trials in BL21 cells.

4.2.2. ProPPE-His_6 and *N*–Terminal Solubility Tags Expression Trials

4.2.2.1. Cloning

The exact mechanism by which the propeptide inactivates PPE is currently unknown. It is possible that the *N*–terminal His_6-tag may sterically or electronically interfere with the ProPPE secondary structure, ultimately leading to enzyme destabilisation. Therefore, the *N*–terminal His_6-tag was moved from the *N*– to the *C*–terminus (ProPPE-His_6), that is after the *C*–terminal asparagine residue in both the pCOLD and pNIC constructs (Figure 4.1). As the carboxyl group of N266 is solvent exposed, a short linker (Gly-Gly) was inserted between the enzyme and the tag (Figure 4.4). In an attempt to improve protein solubility, the His_6-tagged constructs were fused to *N*–terminal solubility tags: maltose binding protein (MBP), small ubiquitin modifier protein (SUMO), glutathione *S*–transferase (GST) and thioredoxin A

(Trx)[280]. For brevity, these constructs will be referred to as the abbreviations of the fusion partner (MBP, SUMO, GST, or Trx) and vector (pCOLD or pNIC).

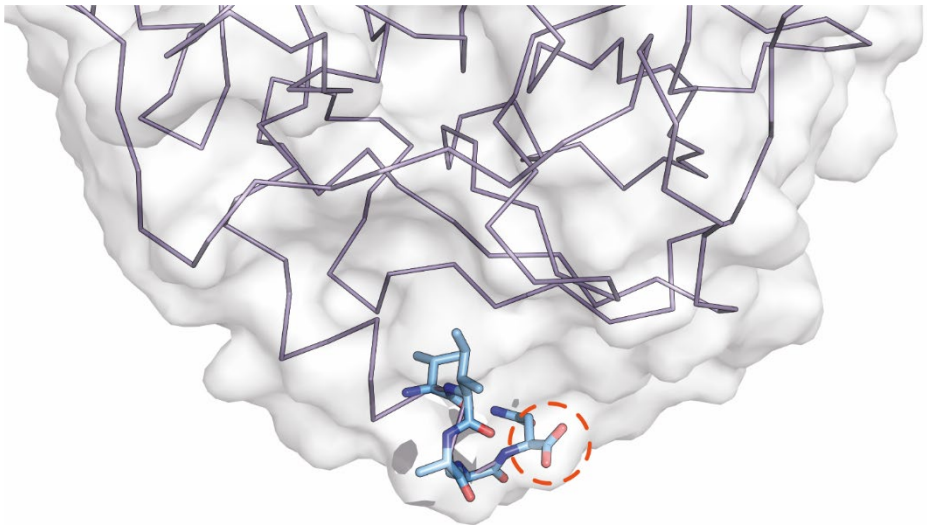


Figure 4.4 | A view of the *C*–terminus from a crystal structure of wildtype PPE (Figure 2.5). The five *C*–terminal residues are shown in light blue with the surface exposed carboxyl group of N266 highlighted in orange.

4.2.2.2. Expression Trial Screening

Plasmids containing ProPPE with an *N*–terminal His_6-tag, *C*–terminal His_6-tag, and *C*–terminal His_6-tag with each of the four *N*–terminal fusion tags were transformed into BL21 or SHuffle cells. Small scale expression trials were conducted in 96-deepwell blocks. In brief, 20 mL of each construct was cultured in 2×YT media at 30 or 37 °C (for SHuffle and BL21, respectively), with agitation (180 rpm) to an OD_{600} of 0.4-0.6. Cultures were then split into 1 mL fractions in a 96-deepwell block. Expression in pCOLD samples was induced by cold shock for 15 min, whilst expression in pNIC samples was induced by addition of 0-400 μM IPTG. The plates were incubated at 18 °C and 25 °C for 16–18 h, with agitation (180 rpm). The cell pellets were lysed using detergent (BugBuster® supplemented with lysozyme and DNase I). Soluble and insoluble fractions were subsequently isolated and analysed for ProPPE-His_6 expression by dot blot with a His_6-tag specific primary antibody (Figure 4.5).

While no soluble expression was observed with the *N*–His_6-tagged samples, at least minimal levels of apparent soluble expression were observed for all other constructs. It was suspected that the His_6-tag

was responsible for the complete insoluble expression observed previously, as simply moving the tag from the *N*– to *C*–terminus promoted soluble expression. However, the poor soluble expression was not exclusively due to steric hinderance as the constructs harbouring other *N*–terminal fusion tags also resulted in soluble expression. The strongest His_6 signal was observed with the MBP:pCOLD vector, at 18 and 25 °C (Figure 4.5, lane 3), although GST fusion constructs also produced signal from both vectors (Figure 4.5, lane 6). All pNIC samples that produced signal also exhibited leaky expression that was often similar to or stronger than in the induced samples. Overall, the BL21 cells expressed more soluble recombinant protein compared to the SHuffle cells. While the results were promising, only a qualitative analysis could be produced, as sample pelleting and insoluble fraction isolation were carried out under relatively low centrifugal force (3,370 × *g*). Thus, cell pellets may have been lost before lysis and the soluble fractions may have been contaminated by the insoluble fractions. To address this, in subsequent trials cultures were transferred to Eppendorf tubes and isolated using a higher gravitational force equivalent.

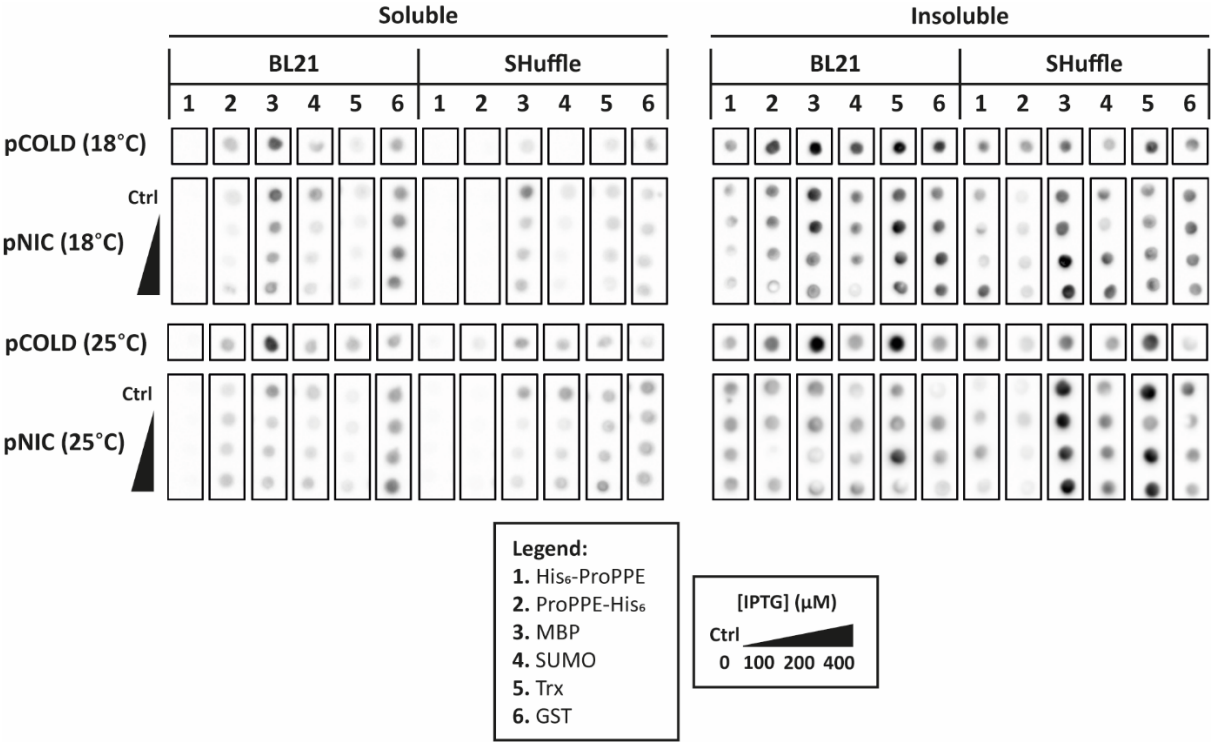


Figure 4.5 | Dot blot analysis of ProPPE 96-deepwell expression trials. Samples of ProPPE with an *N*–terminal His_6tag, *C*–terminal His_6-tag, and *C*–terminal His_6-tag with each of the four *N*–terminal fusion tags were expressed at 18 or 25 °C in a 96-deepwell block. Dot blot analysis was performed on soluble and insoluble fractions. An anti-His_6 primary antibody was used for expression level analysis.

Following the previous dot blot analysis, each construct (barring His_6-ProPPE which produced no soluble expression) was further analysed by SDS-PAGE and western blotting. Whilst preparing and analysing the previous dot blots, the samples had been stored at 4 °C for more than 24 hours. Western blot analysis revealed signs of degradation, with multiple protein bands for all samples (Figure 4.6). This was likely the result of proteolytic degradation due to the absence of protease inhibitors in the lysis medium and the extended time between lysis and analysis (over 24 h at 4 °C). Regardless of the observed degradation, bands were observed at the approximate expected molecular weight for each intact construct. A band at approximately 28 kDa, the expected molecular weight of intact ProPPE-His_6, was also observed in all samples with a fusion partner, except for the SUMO tagged PPE (Figure 4.6). As SUMO was the only construct without a TEV cleavage site between the fusion tag and PPE, this may indicate that this sequence is vulnerable to cleavage by native *E. coli* proteases.

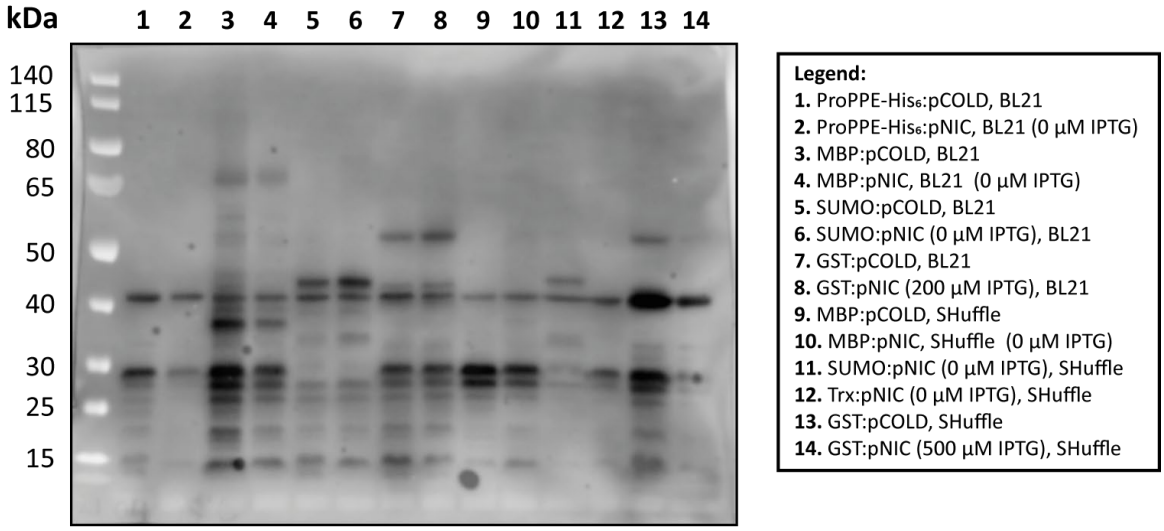


Figure 4.6 | Western blot analysis of recombinant ProPPE 96-deepwell expression trials. Samples of ProPPE with a *C*–terminal His_6-tag, and *C*–terminal His_6-tag with each of the four *N*–terminal fusion tags were expressed at 18˚C in a 96-deepwell block. The calculated molecular weights are: ProPPE-His_6 (28.1 kDa), MBP-ProPPE-His_6 (69.6 kDa), SUMO-ProPPE-His_6 (39.3 kDa), Trx-ProPPE-His_6 (40.8 kDa), GST-ProPPE-His_6 (54.7 kDa). An anti-His_6 primary antibody was used for expression level analysis.

4.2.2.3. Small-Scale Expression of C-Terminally His_6-Tagged ProPPE

Further small-scale expression trials were performed to examine whether the soluble recombinant protein was of the correct molecular weight and whether it was resistant to protease degradation over shorter timeframes. Constructs that produced soluble protein (PPE-His_6, MBP-PPE-His_6, GST-PPE-His_6, SUMO-PPE-His_6, Trx-PPE-His_6) at both 18 and 25 °C were expressed on a larger scale. After expression, the cultures were split into two fractions and lysed using detergent (Bugbuster® Protein Extraction Reagent). For each condition, one fraction was lysed in the presence of protease inhibitor (1 mM PMSF in isopropanol [final] = 1% v/v) and the other in the absence of PMSF (supplemented with 1% v/v isopropanol). After clarification of cell lysate, the His_6-tagged proteins were isolated using HisPur Ni-NTA spin columns and analysed by western blotting (Figure 4.7). Clearly defined bands were observed at the approximate molecular weight for each construct. The strongest signal was observed with PPE-His_6 expressed at 18 °C (Figure 4.7, lane 1). At both 18 and 25 °C, PPE-His_6 was free of prominent lower molecular weight bands suggesting that ProPPE-His_6 is resistant to protease degradation in the absence of protease inhibitors (Figure 4.7, lanes 1 & 2). Trx also exhibited strong soluble expression, however, at both 18 and 25 °C, clear bands were visible at the same molecular

weight as PPE-His_6 suggesting that the linker region between Trx and ProPPE-His_6 was vulnerable to proteolytic attack (Figure 4.7, lanes 9 & 10). Bands potentially correlating to ProPPE-His_6 (~28.1 kDa) were much stronger in the absence of PMSF, with most of the soluble protein observed at this molecular weight in samples without inhibitors. Both MBP and GST exhibited relatively low expression, and as with Trx, both were vulnerable to proteolytic cleavage in the absence of PMSF (Figure 4.7, lanes 3-6). An uncharacterised weak band was also observed at 40 kDa for all GST samples. Relatively strong expression was also observed with the SUMO construct which again appeared to be resistant to cleavage even in the absence of PMSF (Figure 4.7, lanes 7 & 8). Due to the observed soluble expression level and its resistance to protease degradation, ProPPE-His_6:pCOLD with expression conducted at 18 °C in the BL21 strain was selected for purification trials.

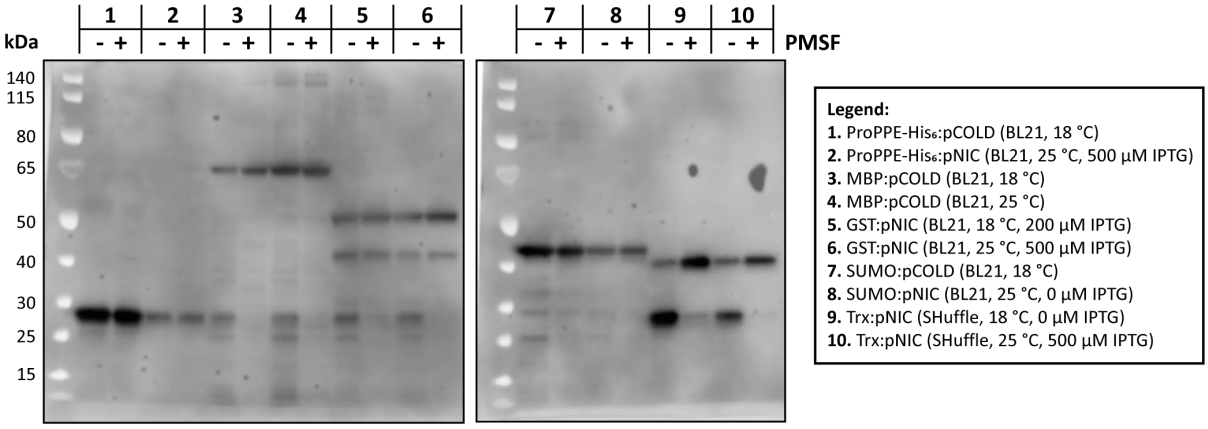


Figure 4.7 | Western blot analysis of small-scale expression trials with ProPPE. Samples of ProPPE with a *C*–terminal His_6-tag, and *C*–terminal His_6-tag with each of the four *N*–terminal fusion tags were expressed at 18 or 25 °C. Calculated molecular weights: ProPPE-His_6 (28.1 kDa), MBP-ProPPE-His_6 (69.6 kDa), SUMO-ProPPE-His_6 (39.3 kDa), Trx-ProPPE-His_6 (40.8 kDa), GST-ProPPE-His_6 (54.7 kDa). An anti-His_6 primary antibody was used for expression level analysis.

4.2.3. ProPPE-His_6 Purification Trials

4.2.3.1. Initial Purification Trial with PPE-His_6 (Purification 1)

A cell pellet from 3.6 L of BL21 strain culture expressing pCOLD:ProPPE-His_6 at 18 °C was resuspended 4:1 (buffer [mL]:cell pellet [g]) in binding buffer and lysed by high-pressure homogenisation. After clarification and filtration, proteins were purified by Ni^{2+} IMAC using a linear elution gradient (Figure 4.8, A & B). Protein eluted as a single peak between approximately 22 mM and 82 mM imidazole. It was envisaged that the purification step would concentrate ProPPE-His_6 to a level

sufficient to observe by SDS-PAGE analysis. However, the two strongest protein bands observed by SDS-PAGE were of a slightly lower molecular weight than expected for ProPPE-His_6. As faint bands were observed in the expected region, all peak fractions were pooled. To ensure that peak fractions contained recombinant PPE in subsequent purification steps, western blot analysis was performed. The pooled fractions were incubated with bovine pancreatic TR (TR:sample approx. 1:25 w/w) at pH 7.5, during concomitant dialysis against buffer lacking imidazole. Complete cleavage of the ProPPE propeptide was achieved within 5 min as observed by western blot analysis (Figure 4.8, C). After 1–3 h of incubation, the intensity of many contaminating protein bands weakened, likely as a result of the catalytic activity of both TR and recombinant PPE.

In order to assess the activity of the isolated enzyme, an absorbance assay measuring the hydrolytic cleavage of a PPE substrate (succinyl-Ala-Ala-Pro-Ala-paranitroanilide (AAPA))[105] was performed. In brief, samples before and after activation (1 mg mL^{-1}), and a TR control (1 mg mL^{-1}), were incubated with AAPA at room temperature. PPE cleaves between the *C*–terminal Ala and the *para*-nitroanilide (pNA) moiety, which subsequently exhibits an absorbance maxima shift. The release of pNA was monitored by absorbance spectroscopy at 405 nm[106]. While an observed absorbance increase confirmed that PPE was active only after cleavage of the propeptide, the catalytic activity was much lower than would be expected at a concentration of 1 mg mL^{-1} likely due to the poor purity of the sample (Figure 4.8, D). TR exhibited minor activity in the control reaction. However, the TR concentration was 25-fold lower in the purified sample, suggesting the observed activity was not due to TR contamination.

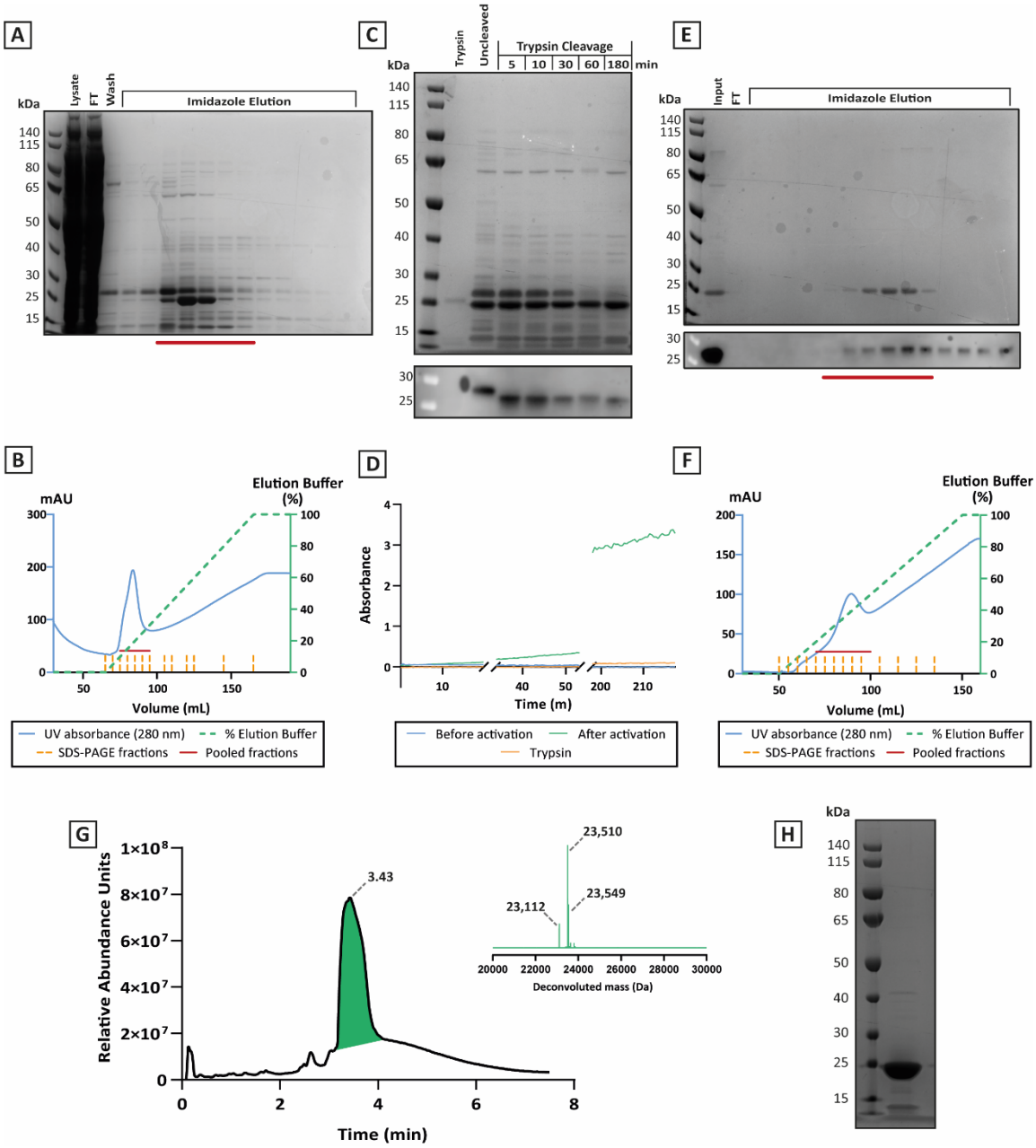


Figure 4.8 | PPE-His$_6$ initial protein purification. A) SDS-PAGE analysis of ProPPE-His$_6$ lysis and IMAC purification (red line indicates pooled fractions); **B)** IMAC UV elution profile of (**A**); **C)** SDS-PAGE (top) and western blot (bottom) analysis of the tryptic activation of ProPPE-His$_6$; **D)** an activity assay comparing a sample from (**C**) before and after activation by trypsin hydrolysis with absorbance measured at 405 nm; **E)** SDS-PAGE (top) and western blot (bottom) analysis of IMAC purification after tryptic activation (red line indicates pooled fractions); **F)** an IMAC UV elution profile of (**E**); **G)** protein LC-MS analysis of purified and activated PPE-His$_6$ (predicted mass = 26,845 Da); **H)** SDS-PAGE analysis of the final purified sample. An anti-His$_6$ primary antibody was used for expression level analysis.

A final purification step was performed with the aim of removing TR from the sample. Due to almost identical physiochemical properties of PPE and TR, it was considered unlikely that TR could be isolated by ion exchange chromatography (theoretical isoelectric point (pI) = 8.45 and 8.3 for PPE-His$_6$ and TR, respectively) or gel filtration (MW = 26.8 kDa and 23.5 kDa for PPE-His$_6$ and TR, respectively). To

separate PPE from TR, which did not contain a His_6-tag, the sample was again purified by IMAC (Figure 4.8, E & F). As before, the protein eluted in a single peak but required a higher concentration of imidazole (approximately 82-130 mM imidazole; Figure 4.8, F). By the time that the samples had been prepared for SDS-PAGE analysis, much of the contaminating protein had degraded, except for a strong band at 25 kDa and weaker bands at approximately 60 and 85 kDa (Figure 4.8, E), possibly due to the hydrolytic activity of both TR and PPE. The IMAC elution fractions contained a band at slightly below 25 kDa and a minor amount of the 85 kDa contaminant, but no bands at the predicted molecular weight (Figure 4.8, E, top). Western blot analysis indicated that PPE-His_6 was present in the peak fractions, although PPE-His_6 also continued to elute beyond the observable UV peak. (Figure 4.8, E, bottom). The concentration of PPE-His_6 in each fraction was also low, as evidenced by the requirement for longer activation during western blot imaging (16 s activation in Figure 4.8, E vs 1 s activation in Figure 4.8, C). Subsequently, the peak fractions were pooled, concentrated, and dialysed against buffer containing no imidazole. Protein LC-MS analysis of the final product presented a single peak eluting at 3.43 minutes (Figure 3.4, G). This peak likely corresponds to the lower molecular weight peak observed by SDS-PAGE (observed masses: 23,510, 23,549, and 23,112 Da). While the observed mass is close to the molecular weight of TR, the corresponding SDS-PAGE band was present before addition of TR (Figure 4.8, C). The TR LC-MS elution profile also differed and corresponded to a lower mass (Figure 4.9, A; predicted mass: 23,305 Da). Although PPE-His_6 was isolated to poor purity, an activity assay was performed on the final sample at a concentration of 1 μg mL^{-1} and compared with commercial PPE (0.1 and 1 μg mL^{-1}) (Figure 4.10). While the purified sample did exhibit low catalytic activity, the activity was much lower than the commercially available PPE.

Overall, the initial purification trial afforded an impure sample of recombinant PPE-His_6 with low catalytic activity. However, tryptic cleavage of the ProPPE propeptide was achieved to produce active PPE, indicating that, at least some of, the enzyme was folded correctly. Furthermore, TR was successfully removed from the sample following propeptide cleavage.

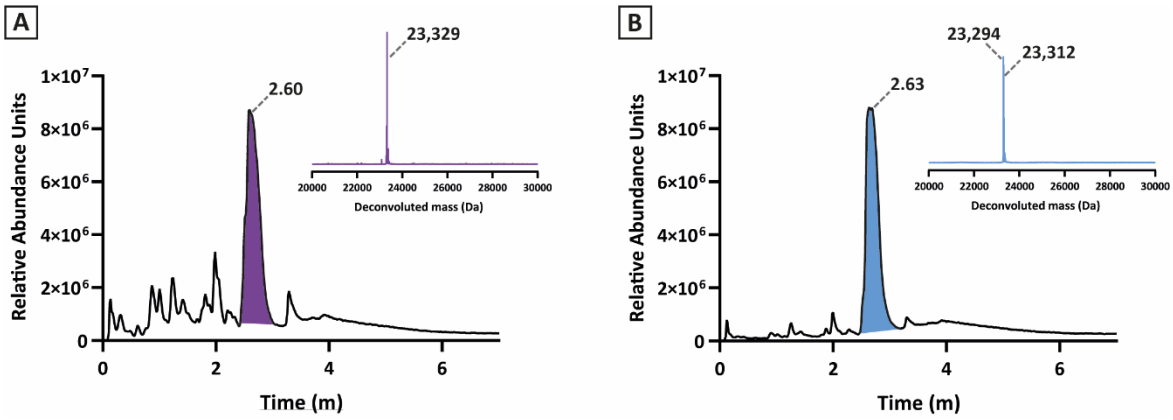


Figure 4.9 | Protein LC-MS analysis of bovine pancreatic trypsin. A) TR dissolved in Tris pH 7.5; **B)** TR dissolved in acetate buffer pH 4.5. Predicted mass = 23,305 Da. The 35 kDa shift is consistent with a chloride (Cl) adduct. TR contains a Cl binding site and Cl is present in the Tris buffer but is not present in the acetate buffer.

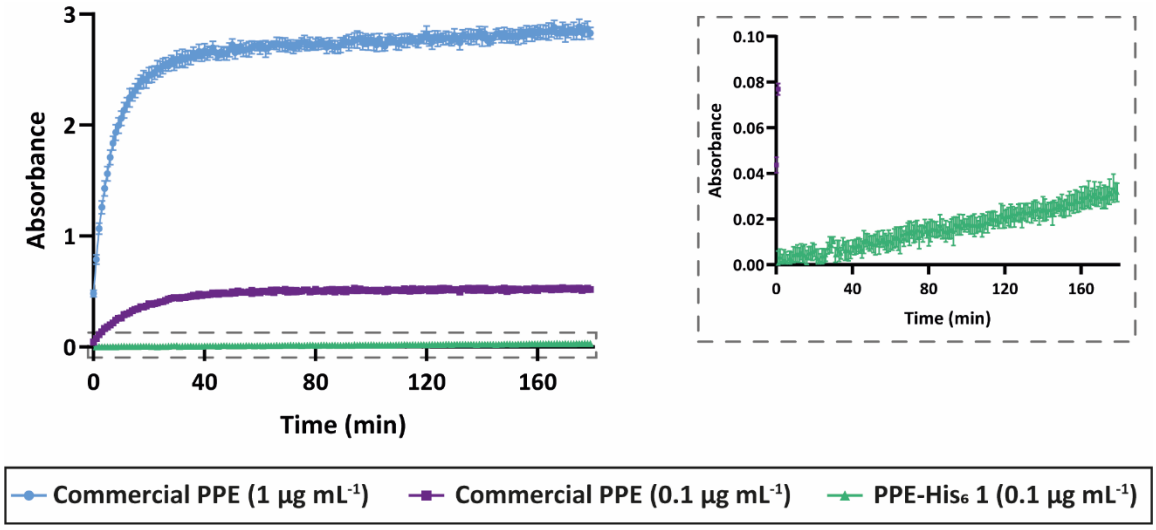


Figure 4.10 | Catalytic activity assay comparing purified PPE to commercial PPE. Absorbance of hydrolysed pNA was measured at 405 nm. Each data point represents the mean four experimental replicates (n = 4 ± SD).

4.2.3.2. PPE-His$_6$ Purification Optimisation (Purification 2)

In an attempt to improve the overall isolated yield, ProPPE-His$_6$ was expressed in BL21 strain on a larger scale (14.4 L) and alterations to the purification method were implemented to improve the purity. Cation exchange chromatography (CEX) was included as a final step. Although extended incubation with TR during activation improved the purity by digesting contaminating bands, there was evidence that the sample was being degraded after extended incubation times (Figure 4.8, C). Therefore, the incubation time was reduced.

The cell pellet was resuspended and lysed as above (Section 4.2.3.1). The clarified soluble fraction was then purified by Ni^{2+} IMAC. Although the amount of clarified lysate loaded onto the column was increased four-fold compared with the previous purification, the resulting UV elution peak was smaller and western blot analysis of the fractions indicated poor capture efficiency, with the majority of recombinant enzyme remaining in the flow through (Figure 4.11, A & B). The colour of the IMAC resin had also changed from blue to pink. After washing with a higher concentration of imidazole (500 mM), the colour of the resin returned to white indicating that the nickel ions had been stripped (Figure 4.11, C).

E. coli cells contain metallophores, which can contribute to metal ion leaching when large volumes of lysate containing poorly abundant protein of interest are applied to the IMAC resin[281]. To investigate removal of the metallophores prior to IMAC, precipitation of the sample by ammonium sulphate was examined. ProPPE harvested from mammalian tissue has previously been partially purified by ammonium sulphate precipitation at 27 g/100 mL[77]. To test whether an analogous precipitation was feasible for purification of recombinant ProPPE-His_6 from a bacterial system, ammonium sulphate was serially titrated into a small volume of the soluble fraction in 10% w/v intervals from 20% up to 50% w/v. After clarification, each pellet was resuspended in buffer containing no ammonium sulphate, to the initial volume. Samples were analysed by western blot before and after affinity chromatography using Ni-NTA spin columns (Figure 4.11, D). Protein of the expected molecular weight was predominantly observed in the 20% and 30% w/v ammonium sulphate fractions before and after IMAC purification. As contaminating proteins were observed in the 40% and 50% w/v fractions, it was clear that ammonium sulphate precipitation could be utilised as an initial purification step before affinity chromatography and could potentially reduce metallophore contamination in downstream chromatography steps (Figure 4.11, D).

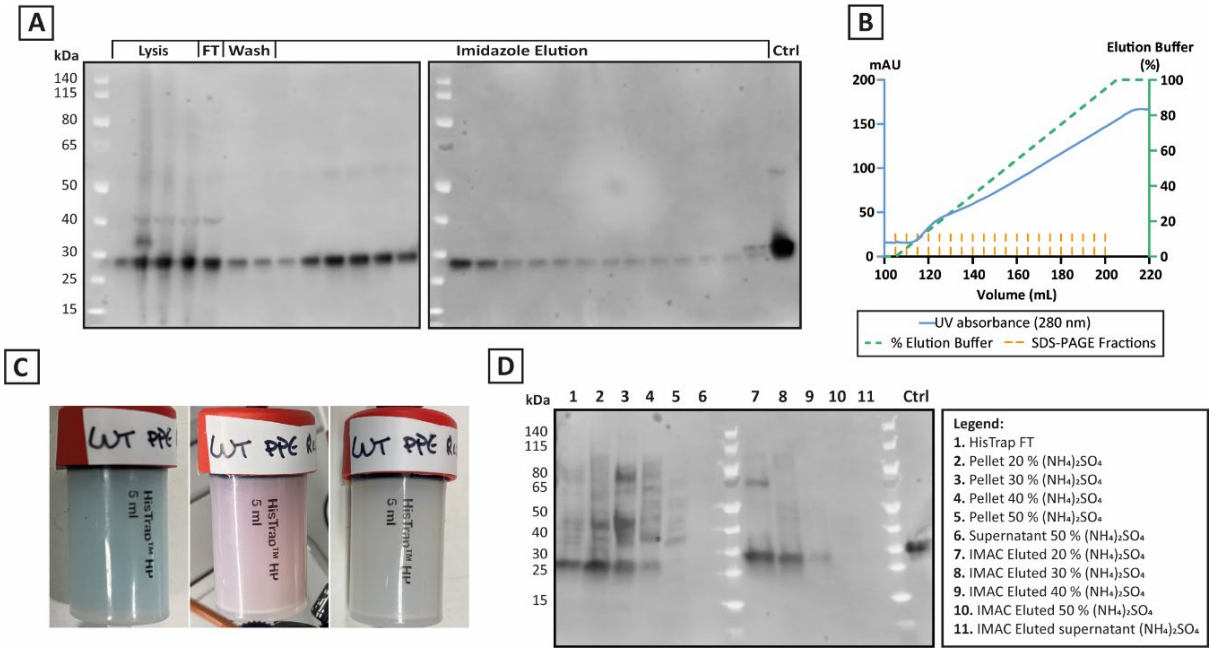


Figure 4.11 | ProPPE-His$_6$ purification trial 2, ammonium sulphate precipitation. A) Western blot analysis of lysis and IMAC purification of ProPPE-His$_6$; **B)** IMAC UV elution profile; **C)** photographic images of the HisTrap column. Left to right: recharged, after elution with 250 mM imidazole, after elution with 500 mM imidazole; **D)** western blot analysis of ProPPE-His$_6$ ammonium sulphate precipitation (1-6. Ammonium sulphate precipitation; 7-11. IMAC purification of precipitated sample). An anti-His$_6$ primary antibody was used for expression level analysis.

The remaining soluble fraction was precipitated in 30% w/v ammonium sulphate, the solution was clarified, and the pellet was resuspended in buffer lacking ammonium sulphate to the original volume. IMAC purification afforded a single peak eluting between 22 mM and 130 mM imidazole (Figure 4.12, A & B). Western blot analysis indicated that ProPPE-His$_6$ eluted within the peak. However, a large portion of ProPPE-His$_6$ was observed within the flow through fraction indicating inefficient binding capability, perhaps as a result of in-optimal protein folding. As the *C*–terminus of active PPE is solvent exposed (Figure 4.4), it was expected that the His$_6$-tag would be freely accessible for interaction with to the IMAC resin. Nonetheless, the fractions containing ProPPE-His$_6$ were pooled, incubated with TR for 1 h and immediately purified by a second round of IMAC (Figure 4.12, C & D). The protein eluted as a main peak with a leading shoulder (Figure 4.12, D), with PPE-His$_6$ present in the main peak and trailing peaks, as observed by western blotting (Figure 4.12, C). Although cleavage of ProPPE-His$_6$ was clearly observed by western blot analysis, there was no clear band shift observed by SDS-PAGE, indicating that recombinant protein still only contributed a small portion of total protein mass (Figure 4.12, C).

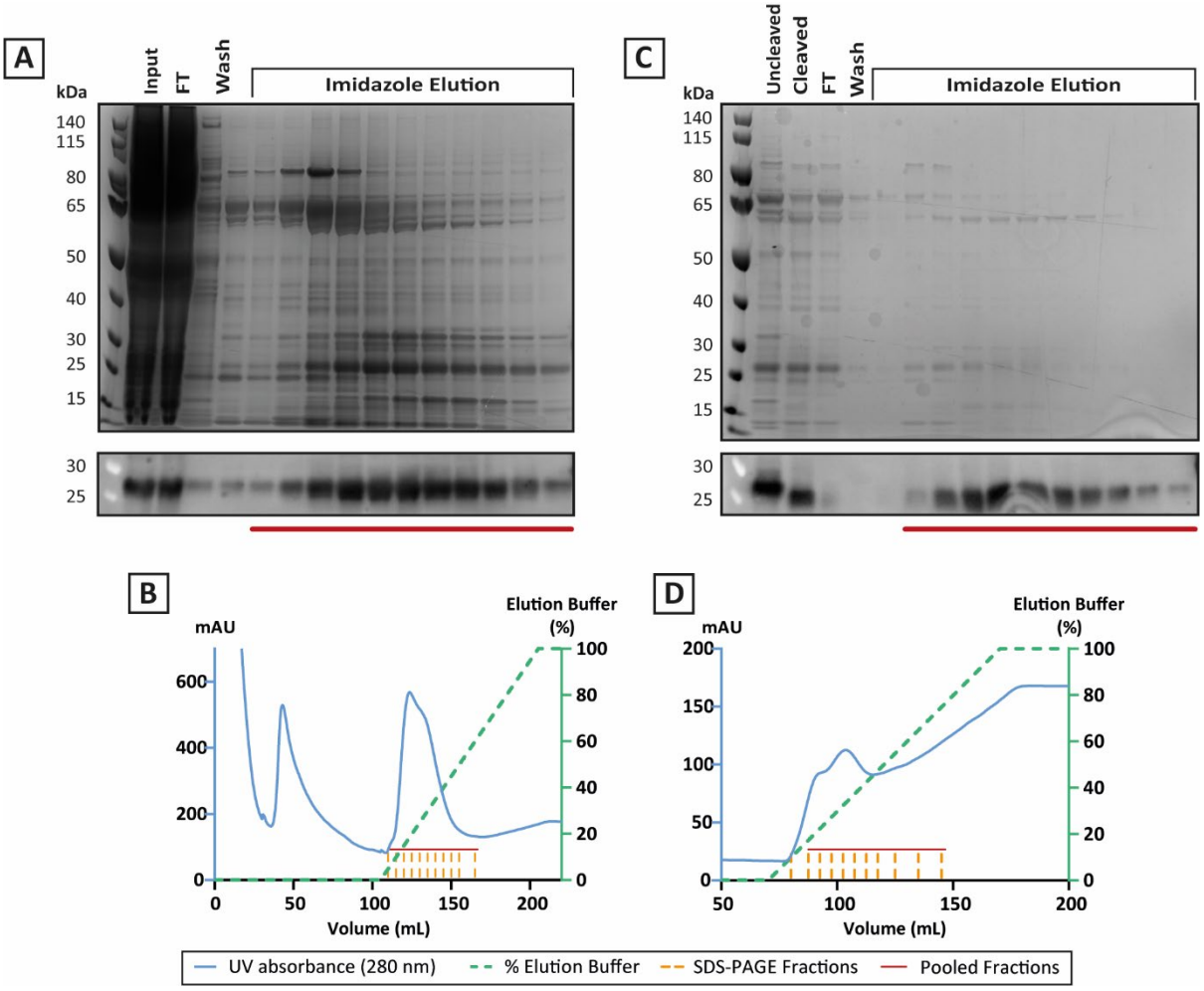


Figure 4.12 | ProPPE-His$_6$ purification trial 2, IMAC purification and tryptic activation. A) SDS-PAGE (top) and western blot (bottom) analysis of IMAC purification of ammonium sulphate precipitated ProPPE-His$_6$ resuspended in Tris buffer pH 7.5; **B)** IMAC UV elution profile of ammonium sulphate precipitated ProPPE-His$_6$ resuspended in Tris buffer pH 7.5; **C)** SDS-PAGE (top) and western blot (bottom) analysis of IMAC purification with PPE-His$_6$ after tryptic activation; **D)** IMAC UV elution profile of PPE-His$_6$ after tryptic activation. An anti-His$_6$ primary antibody was used for expression level analysis.

Pooled fractions containing activated PPE-His$_6$ were finally purified by cation exchange chromatography (CEX; SP resin) after dialysis into HEPES buffer, as Tris buffer is not recommended for CEX (according to the manufacturer's instructions). However, buffer exchange resulted in partial PPE-His$_6$ precipitation, reducing the total yield (Figure 4.13, A). The pH of the buffer used for cation exchange was 7.2, at least one pH unit below the theoretical isoelectric point of PPE-His$_6$ (8.45). However, a substantial portion of PPE-His$_6$ remained within the flow through fraction, again indicating inefficient binding. Nevertheless, PPE-His$_6$ was present within the eluted fractions. Three main bands were observed by SDS-PAGE analysis at approximately 25 kDa, 15 kDa, and at lower molecular weight (< 15 kDa). The 25 kDa band aligned well with the PPE-His$_6$ band observed by western blot analysis

(Figure 4.13, A). LC-MS analysis of the sample presented two main peaks eluting at 2.60 minutes and 3.16 minutes, with two smaller peaks eluting later (Figure 4.13, C). The first eluted peak (2.60 min) corresponded to two mass peaks of 16.1 and 16.8 kDa, likely representing the observed protein band located at approximately 15 kDa by SDS-PAGE (Figure 4.13, A & D). The peak eluting at 3.16 minutes predominantly corresponded to a mass of 26,839 Da (Figure 3.8, B), in good agreement with the calculated mass for PPE-His_6 (26,845 Da). Combined SDS-PAGE and LC-MS analysis indicates a purity of < 50% for the final sample. The total yield of protein was 200 µg, suggesting a recombinant PPE-His_6 yield of < 100 µg.

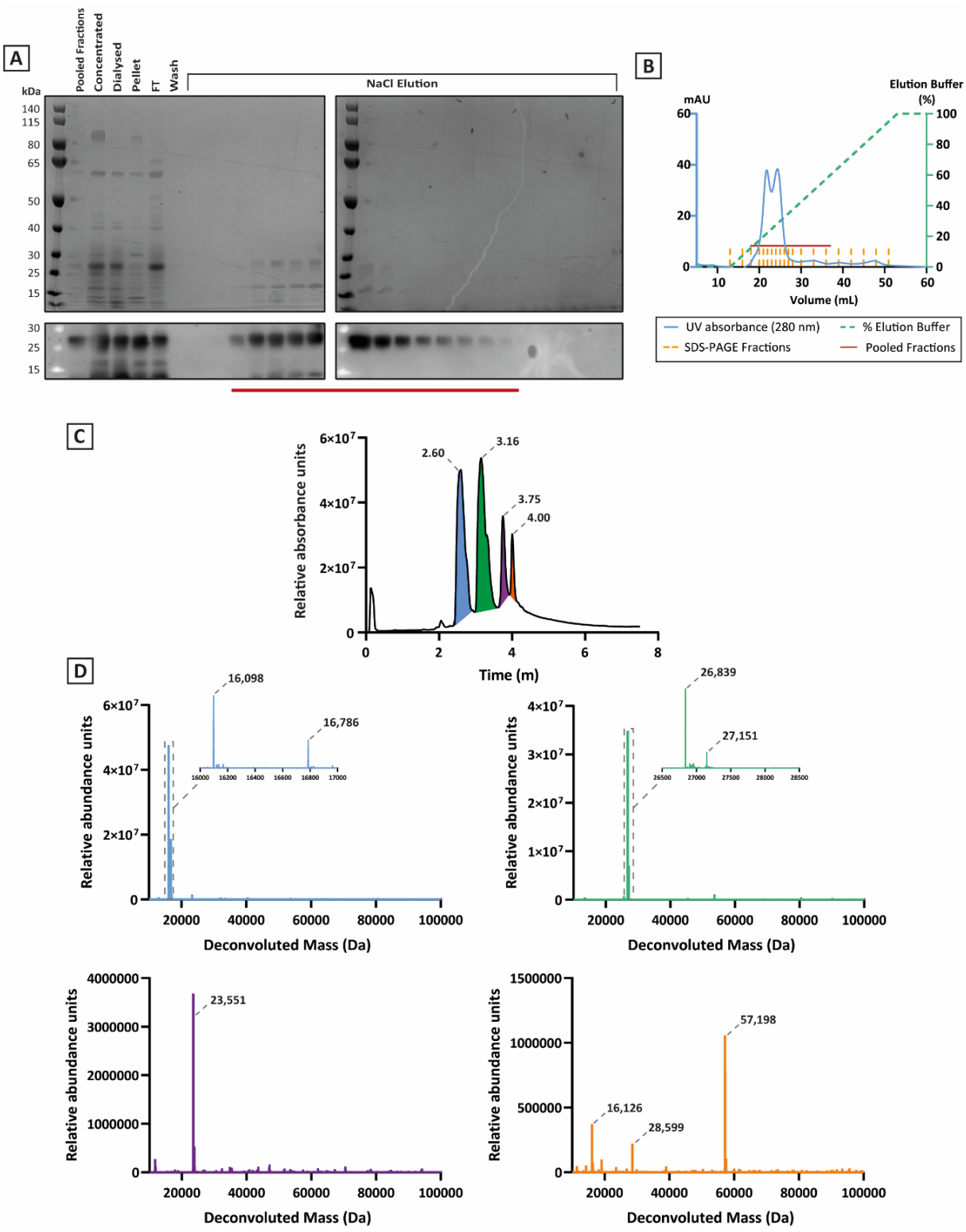


Figure 4.13 | ProPPE-His$_6$ purification trial 2, CEX purification and LC-MS analysis of the final product. A) SDS-PAGE (top) and western blot (bottom) analysis of CEX purification of PPE-His$_6$; **B)** CEX UV elution profile of PPE-His$_6$; **C)** LC-MS elution profile of purified and activated PPE-His$_6$; **D)** LC-MS analysis of purified and activated PPE-His$_6$. Mass spectrum colours correspond to peak colours in (**C**). Predicted mass of PPE-His$_6$ 26,845 Da. An anti-His$_6$ primary antibody was used for expression level analysis.

The catalytic activity of the recombinant PPE from the second purification was analysed at 1 µg mL^{-1} and compared to commercial PPE at 0.1 and 1 µg mL^{-1}. As recombinant PPE-His_6 was shown to be less than 50% pure by SDS-PAGE analysis, its catalytic activity at 1 µg mL^{-1} was predicted to fall within the range of the two commercial PPE concentrations. However, the observed activity of PPE-His_6 was more than 10-fold lower than the commercial sample, although the recombinant enzyme was more stable than the commercial counterpart. While commercial PPE activity stalled after 30 minutes at 0.1 µg mL^{-1}, the recombinant sample remained active for six times longer and ultimately hydrolysed as much substrate as the 0.1 µg mL^{-1} sample (Figure 3.17, B). All three samples were run using identical concentrations of substrate (250 µM) which was shown to exhibit an absorbance value of approximately 2.5 after complete hydrolysis (data not shown). Therefore, the observed stalling of the 0.1 µg mL^{-1} commercial sample is not due to complete hydrolysis of the substrate.

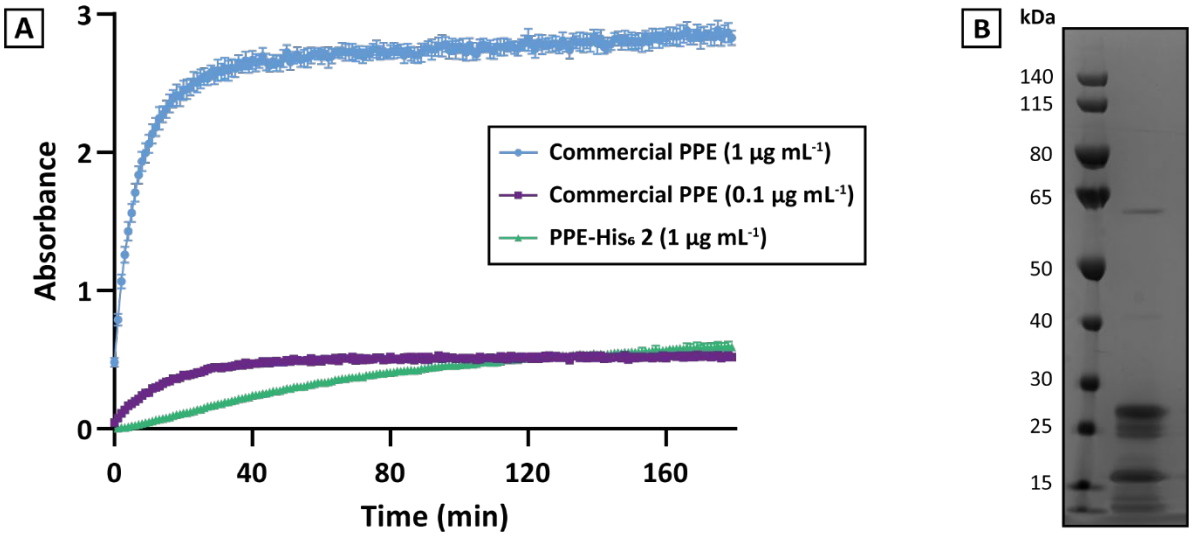


Figure 4.14 |ProPPE-His_6 purification trial 2, catalytic activity assay and SDS-PAGE analysis of the final product. A) Results from a catalytic activity assay comparing recombinant PPE-His_6 to commercial PPE. The absorbance was measured at 405 nm; **B)** SDS-PAGE analysis of PPE-His_6 2 at 1 mg mL^{-1}.

While the above-described results demonstrate that recombinant PPE-His_6 could be produced and partially purified, the activity and yield remained relatively poor. Improvements to the yield would likely require further optimisation of expression or redesign of the construct, such as increasing the linker length and introducing additional histidine residues into the tag (e.g., His_8-/His_{10}-tag). An alternative buffering option for CEX may be beneficial as PPE-His_6 precipitated during buffer exchange

and the sample was captured poorly on the CEX column. Acetate buffer (pH 4.5–5.0) is compatible with PPE and has the added benefit of reducing the proteolytic activity of PPE[106]. As the pH would also be further away from the theoretical pI for PPE-His_6, there should be a higher degree of PPE ionisation which could improve capture of PPE-His_6 on the CEX column. The observed poor activity may be explained, at least in part, by the low purity at the point of tryptic activation, as active PPE-His_6 would immediately begin hydrolysing the peptide bonds of contaminating proteins. Consequently, a fraction of the total sample may have reached the total turnover number (TTON) resulting in a lower final activity[282]. Additionally, a further reduction in the contact time with TR may be beneficial as TR may exhibit degradative effects towards PPE-His_6 on longer timescales. Therefore, subsequent purification trials would attempt to address these issues.

4.2.3.3. Purification at lower pH

To address the issue of purity at the point of activation, the CEX purification step was carried out immediately after ammonium sulphate precipitation, in an attempt to reduce the amount of contaminating protein exposed to active PPE-His_6. Furthermore, CEX would be carried out at pH 5.0 in acetate buffer, to strengthen the interaction between PPE-His_6 and the ion exchange resin.

PPE-His_6 was lysed in low pH buffer (100 mM acetate buffer pH 5.0) and was precipitated using ammonium sulphate, as described above, but on a smaller scale (3.6 L of culture). Before resuspension, the pellet was washed twice with 30% w/v ammonium sulphate with aim of further purifying the recombinant protein. However, the wash steps did not appear to substantially improve the purity of the sample, as determined by SDS-PAGE analysis, and are therefore likely unnecessary (Figure 4.15, A, lanes 7 & 9). Resuspension of the pellet in acetate buffer lacking ammonium sulphate produced a turbid solution and dilution to a total of two times the original volume with vigorous mixing was insufficient to resuspend the observed precipitate. After clarification, a substantial portion of recombinant ProPPE-His_6 was located within the pellet as determined by western blot analysis (Figure 4.15, A, lane 10). Despite this, the supernatant was purified by CEX. Other than a large peak at

approximately 350 mL of elution, caused by a system error leading to a stall of the protocol, no peaks of more than 100 mAU were observed (Figure 4.15, C). Analysis of the eluted fractions revealed that the majority of the ProPPE-His_6 was located within the flow through; to clearly observe the recombinant protein within the eluted fractions by western blot, an extended exposure time of 8 s was required (Figure 4.15, B, middle) versus 1 s for the lysis step (Figure 4.15, A).

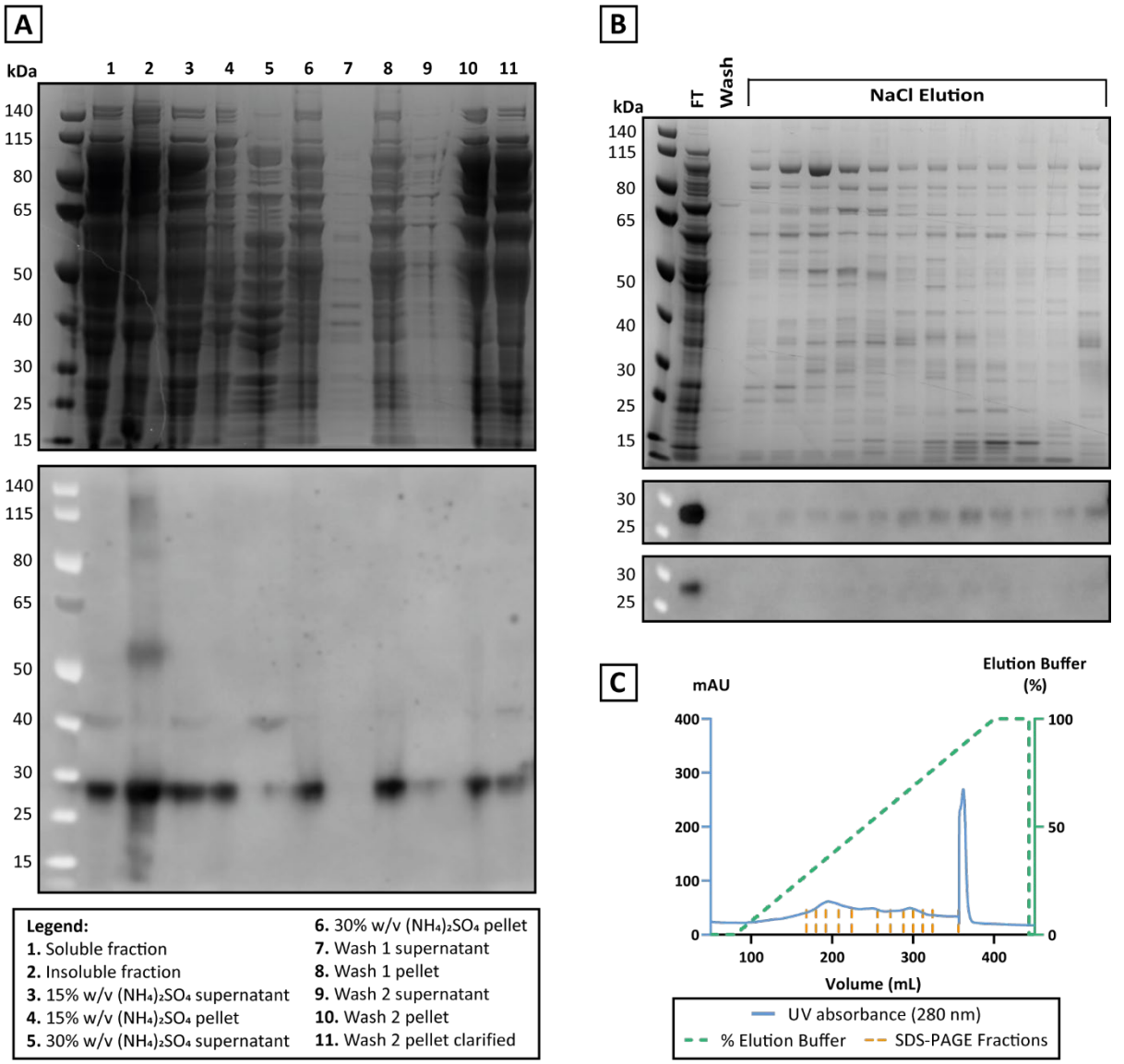


Figure 4.15 | PPE-His_6 purification at low pH. A) SDS-PAGE (top) and western blot (bottom) analysis of lysis, ammonium sulphate precipitation and resuspension of ProPPE-His_6; **B)** SDS-PAGE (top) and western blot analysis (middle and bottom) of CEX purification with 8 s exposure (middle) and 1 s exposure (bottom); **C)** CEX UV elution profile. An anti-His_6 primary antibody was used for expression level analysis.

As it was evident that acetate buffer pH 5.0 was not an acceptable resuspension buffer, several other buffers were tested for their ability to resuspend the sample and for their compatibility as buffers for CEX with ProPPE-His_6. A fresh ammonium sulphate precipitated sample was clarified, and the pellets resuspended in various buffers (Figure 4.16, A). Dot blot analysis of resuspended sample and reclarified supernatant, indicated that most buffers, except for acetate buffer, were capable of efficient resuspension of ProPPE-His_6 (Figure 4.16, A). Four of the samples (in 50 mM HEPES pH 7.2, 10 mM carbonate, 50 mM phosphate buffer pH 7.0, and 50 mM Tris pH 7.5; from here on referred to as HEPES, carbonate, phosphate, and Tris, for brevity) were purified by CEX (Figure 4.16, B & C). Each of the four buffering systems produced different elution profiles indicating distinct binding patterns (Figure 4.16, D). However, western blot analysis again revealed poor binding of ProPPE-His_6 to the cation exchange resin in all four of the buffers. Of the four buffers, Tris resulted in the most efficient binding, with ProPPE-His_6 eluting in both the 250 mM and 500 mM NaCl fractions, however, the majority of ProPPE-His_6 was observed to remain within the flow through (Figure 4.16, C). ProPPE-His_6 resuspended in phosphate and carbonate also exhibited binding but to a lesser degree compared with Tris. No binding was observed with ProPPE-His_6 resuspended in HEPES. These results were unexpected as Tris, due to the presence of its primary amino group, is not a recommended buffer for CEX. Tris was also buffered the closest to the theoretical pI of ProPPE-His_6 (8.14). The reason for the poor interaction between ProPPE-His_6 and the CEX resin is unknown, but it was clear that CEX at this stage of the purification was unsuitable, and thus a re-evaluation of the protocol was necessary.

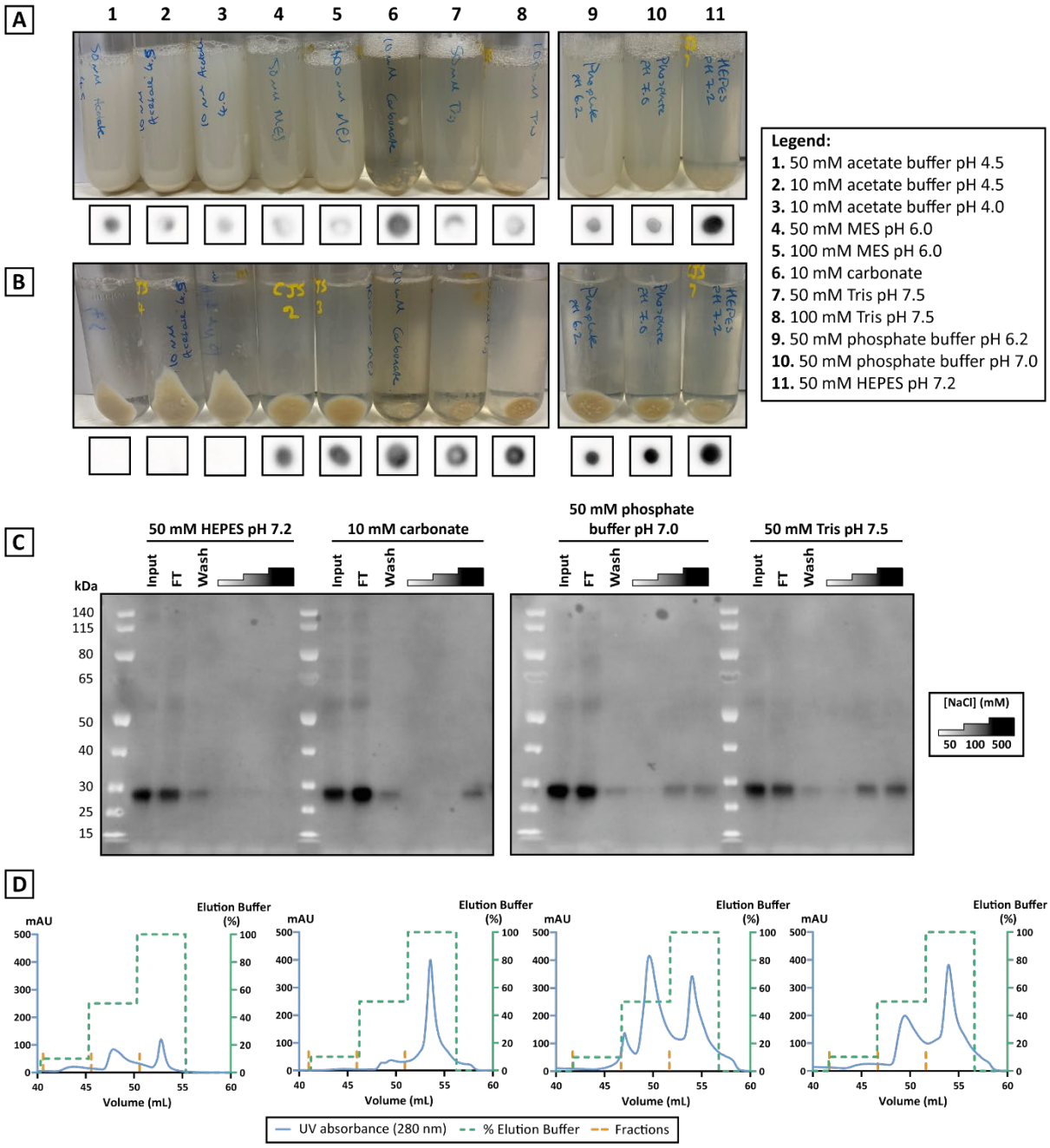


Figure 4.16 | Ammonium sulphate resuspension and CEX tests. A) Photographic images (top) and corresponding dot blots (bottom) of ProPPE-His_6 ammonium sulphate pellets resuspended in various buffers; **B)** photographic images (top) and corresponding dot blots (bottom) of ProPPE-His_6 clarified soluble fractions after resuspension of ammonium sulphate pellets in various buffers; **C)** western blot analyses of CEX purification of ProPPE-His_6 in various buffers; **D)** UV elution profiles from CEX purifications corresponding to the western blots in (**C**). **C**-**D**) The buffers used to resuspend the pellets and for CEX from left to right are: HEPES, carbonate eluted with HEPES, phosphate buffer, Tris. An anti-His_6 primary antibody was used for expression level analysis.

4.2.4. Further ProPPE-His_6 Expression Optimisation

Before further purification trials were undertaken, attempts were made to improve ProPPE-His_6 expression levels.

4.2.4.1. Expression Media and Lysis Additives

Initial trials tested whether a richer media could produce a greater yield of soluble recombinant ProPPE-His_6. It has previously been shown that addition of chemical additives during lysis can improve the yield of recombinant proteins by stabilising proteins in the soluble form[283]. Hence, the lysis buffer was supplemented with a selection of 11 chemical additives and the soluble fractions were examined for soluble production of ProPPE-His_6 after lysis and again after IMAC purification using the HisPur Ni-NTA Spin Columns (Figure 4.17, A & B).

The most significant difference was observed between the two types of media. Soluble expression in the richer Terrific Broth (TB) media was substantially diminished compared to the use of 2×YT media in most conditions after lysis (Figure 4.17, B, top). The trial did not account for differences in dry pellet after cell harvesting, instead resuspending all samples in an equal volume of lysis buffer. As, cultures grown in TB medium generally produce a larger cell pellet due to higher optical density (OD_{600} = 0.8–1.0 and 0.4–0.6 for TB and 2×YT, respectively) at the point of induction, the observed differences in expression were even more pronounced. Following IMAC purification, little soluble protein was eluted from the resin in the TB samples, while there was only a slight reduction in signal strength for the 2×YT samples (Figure 4.17, B, bottom). The substantial loss of soluble protein following IMAC in TB conditions could potentially be attributed to the increased concentration of metallophores as a result of the higher cell density. While metallophore contamination would not be an issue in large-scale purifications due to pre-treatment with ammonium sulphate, using the richer TB medium provided no benefits over the 2xYT medium in small scale trials.

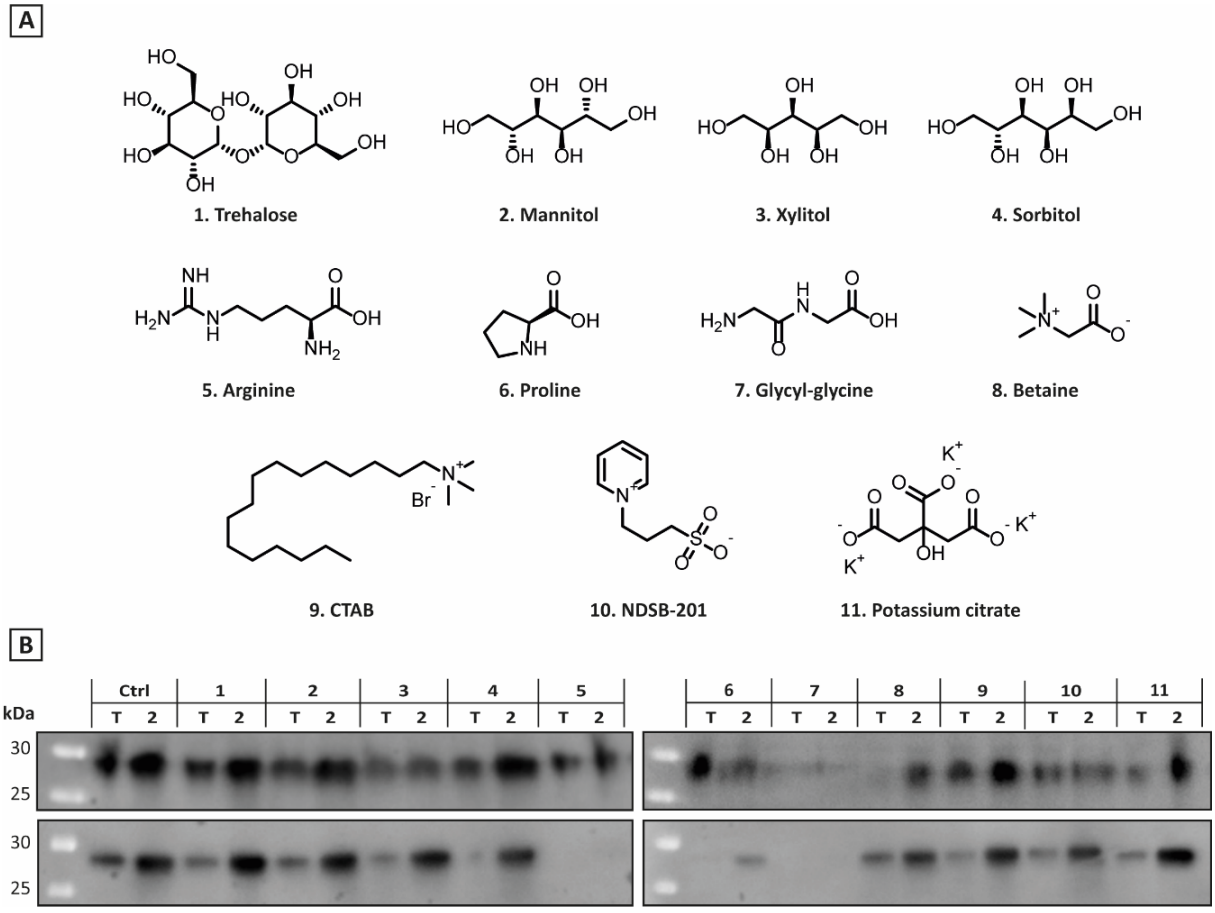


Figure 4.17 | Expression trials examining the effect of expression media and lysis buffer additives. A) Chemical structures of lysis buffer additives; **B)** western blot analyses of soluble (top) and IMAC purified (bottom) fractions from expression trials, varying the expression media and lysis buffer composition. T, Terrific Broth; 2, 2×YT. An anti-His_6 primary antibody was used for expression level analysis.

In terms of lysis additives, there was no single tested condition that substantially improved soluble production or IMAC retention over the no additive control (Figure 4.17, B). Soluble yields after lysis in the presence of osmolytes trehalose, mannitol, and sorbitol were comparable to the no additive control. After IMAC purification, the results with the four sugar/sugar alcohols (1-4) and potassium citrate matched, but did not clearly outperform, the no additive control. Unexpectedly, glycyl-glycine effectively reduced the soluble recombinant production level to zero. While soluble production of ProPPE-His_6 was observed with Arg and Pro (5 and 6), both appear to have interfered with the interaction between ProPPE-His_6 and the resin, leading to poor retention on the IMAC column. As none of the tested conditions clearly improved soluble expression or retention to the IMAC column, none of the conditions were carried forward to further expression trials.

4.2.4.2. Variation of Expression Temperature, Time and Tuneable Induction

Next, variations in post-induction conditions including temperature, time, and transcription rate were examined. Expression trials were performed using the ProPPE-His_6 construct expressed from both the pCOLD and pNIC vectors. Trials were carried out using BL21 cells at 18–37 °C, with expression halted at varying times (3/5/18 h) to examine whether a potential threshold was reached before aggregation or degradation began reducing the total soluble yield (Figure 4.18, A & B). An additional *E. coli* strain, Lemo21(DE3) (from here on referred to Lemo for brevity), was tested with the pNIC vector (Figure 4.18, A & B). The Lemo strain contains a plasmid encoding the natural inhibitor of T7 RNA polymerase (T7 lysozyme or LysY)[284]. LysY is under the control of the P_{RhaBAD} promoter which can be tightly modulated by addition of L-rhamnose, contrasting the essentially binary IPTG-responsive P_{lacUV5} promoter. Increasing the concentration of L-rhamnose leads to amplified expression of LysY and therefore reduced activity of T7 RNA polymerase, effectively creating a tuneable IPTG induction system[284]. It was envisaged that the total soluble yield could be increased by reducing the rate of transcription, and consequently, the rate of translation.

A lower footprint trial design facilitated sample handling in a tube format as opposed to plate format, enabling clarification at higher relative centrifugal force (3,270 × *g* vs 21,000 × *g* for plate and tube, respectively) and reducing the likelihood of incidental soluble fraction contamination with pelleted material. Induction in BL21 strain was carried out as before (Section 4.2.1.2), whilst gene expression in Lemo cells was induced by the addition of 400 µM IPTG in the presence of varying concentrations of L-rhamnose. pNIC expression in BL21 cells produced poorly soluble recombinant protein, with highest expression observed in the no IPTG controls (Figure 4.18, A), demonstrating leaky expression and suggesting that rapid overexpression of ProPPE-His_6 could lead to aggregation and a poor soluble yield. The pCOLD samples exhibited improved soluble expression. The strongest signal was observed from samples incubated at 37 °C for 3 or 5 h, although 18 °C for 5 h also comparable expression (Figure 4.18, A). The results with Lemo cells were varied, but expression at 18 °C for 18 h produced the

most soluble recombinant PPE-His_6, irrespective of the concentration of L-rhamnose, and appeared to be as efficacious as the highest expressing pCOLD condition in BL21 strain (Figure 4.18, A). The two earlier time points at 18 °C exhibited virtually no signal suggesting that the accumulation of soluble protein at the lower temperature occurred slowly. As the temperature for expression was increased, the delay in accumulation of soluble recombinant protein decreased, with ProPPE-His_6 being observed after 5 h at 25 °C and after 3 h above 30 °C. At 25 °C and 30 °C, samples with lower concentrations of L-rhamnose (125–500 µM) exhibited a stronger ProPPE-His_6 signal in the earlier timepoints, but soluble accumulation stalled or diminished after 18 h. At these temperatures, use of a higher concentration of L-rhamnose (1–2 mM) resulted in a weaker ProPPE-His_6 signal at earlier timepoints, but a greater ProPPE-His_6 signal after 18 hours. At 37 °C, higher concentrations of L-rhamnose yielded stronger signal throughout all timepoints (Figure 4.18, A). The results suggest that a reduction in the rate of expression, by means of modulating the rate of transcription and the temperature, could improve the soluble ProPPE-His_6 yield.

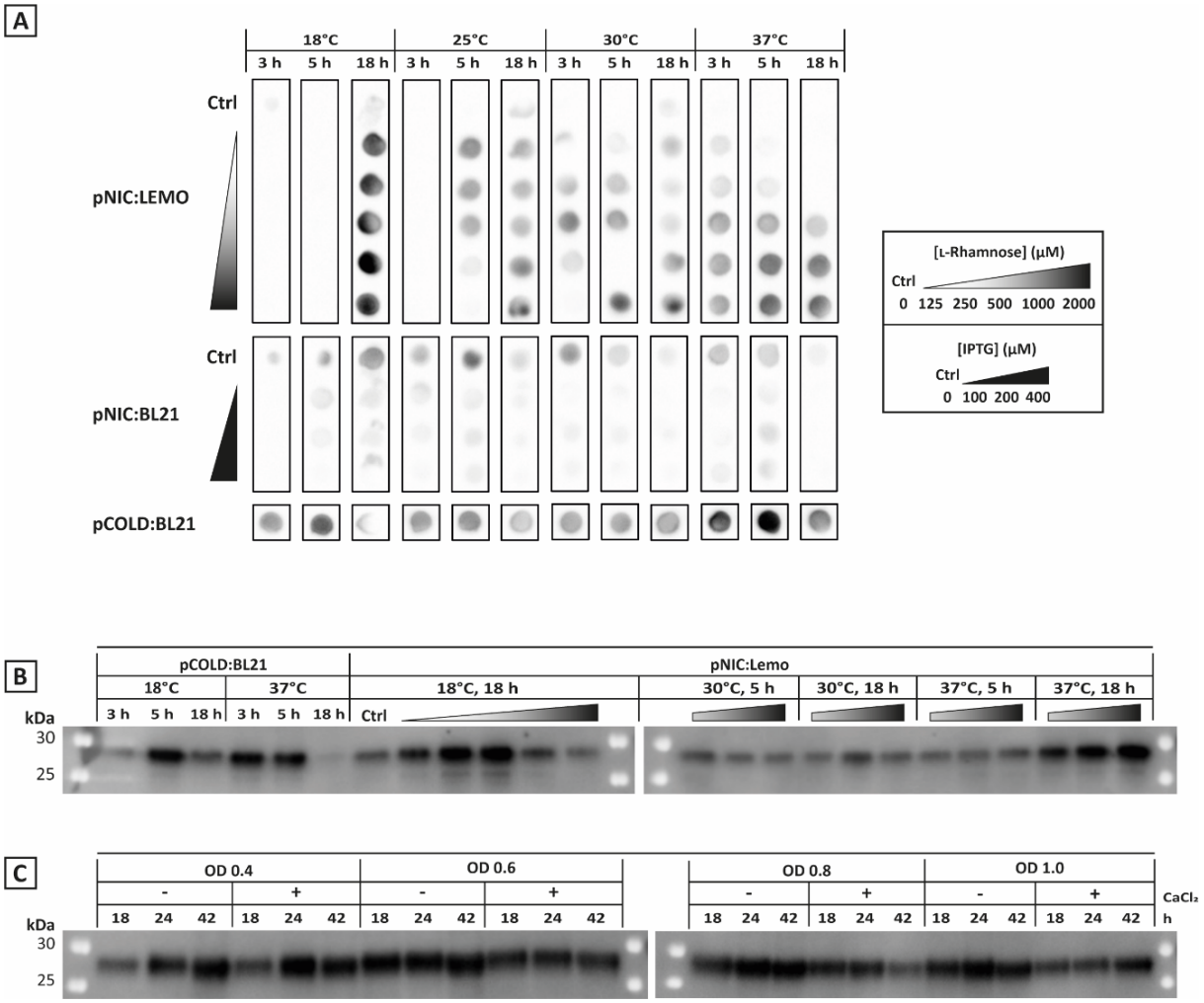


Figure 4.18 | Expression trials of ProPPE-His$_6$ using various expression strains, temperatures, and lengths, and the effects of $CaCl_2$. A) Dot blot analyses of soluble fraction from expression trials of ProPPE-His$_6$, varying expression temperature and length, and using an tuneable induction system; **B)** western blot analyses of the soluble fraction from expression trials of ProPPE-His$_6$, varying expression temperature and length, and using a tuneable induction system; **C)** western blot analysis of the soluble fraction from expression trials of ProPPE-His$_6$, varying expression length, cell density at induction and supplementation with $CaCl_2$. An anti-His$_6$ primary antibody was used for expression level analysis.

To validate the results of the dot blot analyses, conditions that produced the greatest soluble yield of ProPPE-His$_6$ were reproduced and analysed by western blot for better clarity of results (Figure 4.18, B). Of the conditions tested, several exhibited an improved soluble yield of ProPPE-His$_6$ over the condition previously used for purification (pCOLD vector, BL21 cells, 18 °C, 18 h). One condition that resulted in improved soluble yield of ProPPE-His$_6$ was BL21 cells, 18 °C, for 5h, which demonstrates that rapid expression can in fact produce good yields of soluble ProPPE-His$_6$, but longer incubation periods can lead to a decrease in the soluble yield, perhaps as a result of aggregation and inclusion body formation (Figure 4.18, B). Use of the pCOLD construct expressed at 37 °C for short time intervals (3 or 5 h) in

BL21 also produced a strong ProPPE-His_6 signal. The most efficacious expression was observed at 18 °C in Lemo cells with mid-range L-rhamnose concentrations (250 and 500 µM), followed closely by the 37 °C, 18 h, 2 mM L-rhamnose condition. As there appeared to be slightly stronger signal from the 500 µM, 18 °C, 18 h expression sample, this condition was carried forward for further expression trials.

4.2.4.3. Addition of $CaCl_2$ and optimisation of cell density over extended expression times

As PPE contains a metal binding site, which preferentially binds to calcium (Ca^{2+}) and that binding of Ca^{2+} stabilises the structure for the related SPs: TR and CTR[251], it was investigated whether supplementation of expression media with calcium chloride ($CaCl_2$) would improve the soluble yield by stabilising the structure of recombinant ProPPE-His_6 (Figure 4.18, C). Additionally, the optimum cell density for induction and extended expression time lengths were examined. Expression was carried out in Lemo cells at 18 °C in the presence of 500 µM L-rhamnose. Lysis and clarification steps were performed at high relative centrifugal force (21,000 × *g*), as above, and the soluble fractions were analysed by western blot.

Lemo cells supplemented with Ca^{2+} exhibited faster growth up to an OD_{600} of 0.6, compared with those samples that were not supplemented. After this point growth in the presence of $CaCl_2$ slowed and both samples reached an OD_{600} of 0.8 and 1.0 at comparable times. This effect has previously been observed in *E. coli*[285]. However, it was clear from the western blot that addition of 10 mM $CaCl_2$ had a negative effect on soluble production of ProPPE-His_6, with these conditions exhibiting a reduced soluble yield compared with the no $CaCl_2$ counterparts (Figure 4.18, C). For the no $CaCl_2$ conditions, the optimum cell density was 0.6; extending the expression time beyond 18 h did not affect the levels of soluble ProPPE-His_6 in this condition. Induction at lower cell density (OD_{600} = 0.4) was capable of producing levels of soluble ProPPE-His_6 comparable to those conditions with induction at OD_{600} = 0.6, but only after longer expression intervals (42 h and 24 h for the no $CaCl_2$ and plus $CaCl_2$ conditions, respectively). Induction at higher cell densities (OD_{600} = 0.8–1.0) resulted in reduced soluble yield (Figure 4.18, C). As extended post-induction incubation times or $CaCl_2$ supplementation did not afford substantial

improvements in soluble ProPPE-His_6 levels, induction at OD_{600} = 0.6 and expression at 18 °C for 18 h without $CaCl_2$ supplementation were considered as the optimal conditions for purification trials.

4.2.5. Purification of ProPPE-His_6 Examining $CaCl_2$ Stabilisation (Purification 3)

ProPPE-His_6 was produced using Lemo21 cells in 2xYT media supplemented with 0.5 mM L-rhamnose (14.4 L scale). After lysis by high-pressure homogenisation, the lysates were split into equal volumes and 10 mM $CaCl_2$ was added to one sample. For brevity, the samples in this purification will from here on be referred to as PPE+ or PPE- for samples purified with buffers supplemented with or without $CaCl_2$, respectively. Each sample was purified using identical steps, except for $CaCl_2$ supplementation. The lysates were initially treated with a low concentration of ammonium sulphate (10% w/v) to precipitate contaminants (Figure 4.19, A). While contaminants were separated in the 10% $(NH_4)_2SO_4$ pellet, the pellets also contained some of the recombinant ProPPE-His_6, as determined by western blotting (Figure 4.19, A). Nonetheless the pellets formed after precipitation with 30% w/v ammonium sulphate contained a viable amount of recombinant ProPPE-His_6 (Figure 4.19, A). Following resuspension of the precipitated pellets in Tris buffer, the samples were purified by IMAC (Figure 4.19, B, C). Both PPE+ and PPE- samples eluted as two distinct peaks; $CaCl_2$ supplementation did not affect the elution profiles (Figure 4.19, B). The PPE+ sample produced a slightly stronger UV signal upon initial protein elution, and SDS-PAGE analysis of the elution peak revealed highly abundant contaminating protein bands between 65–80 kDa (Figure 4.19, C). Strong bands of similar mass were also observed in the PPE- sample, but were localised to the wash fraction (Figure 4.19, C), indicating that $CaCl_2$ stabilised the interaction between the contaminant and either ProPPE-His_6, the resin itself, or another protein that non-specifically bound to the IMAC resin. Western blot analysis indicated, as before, that ProPPE-His_6 eluted across most of the elution gradient and across both UV peaks.

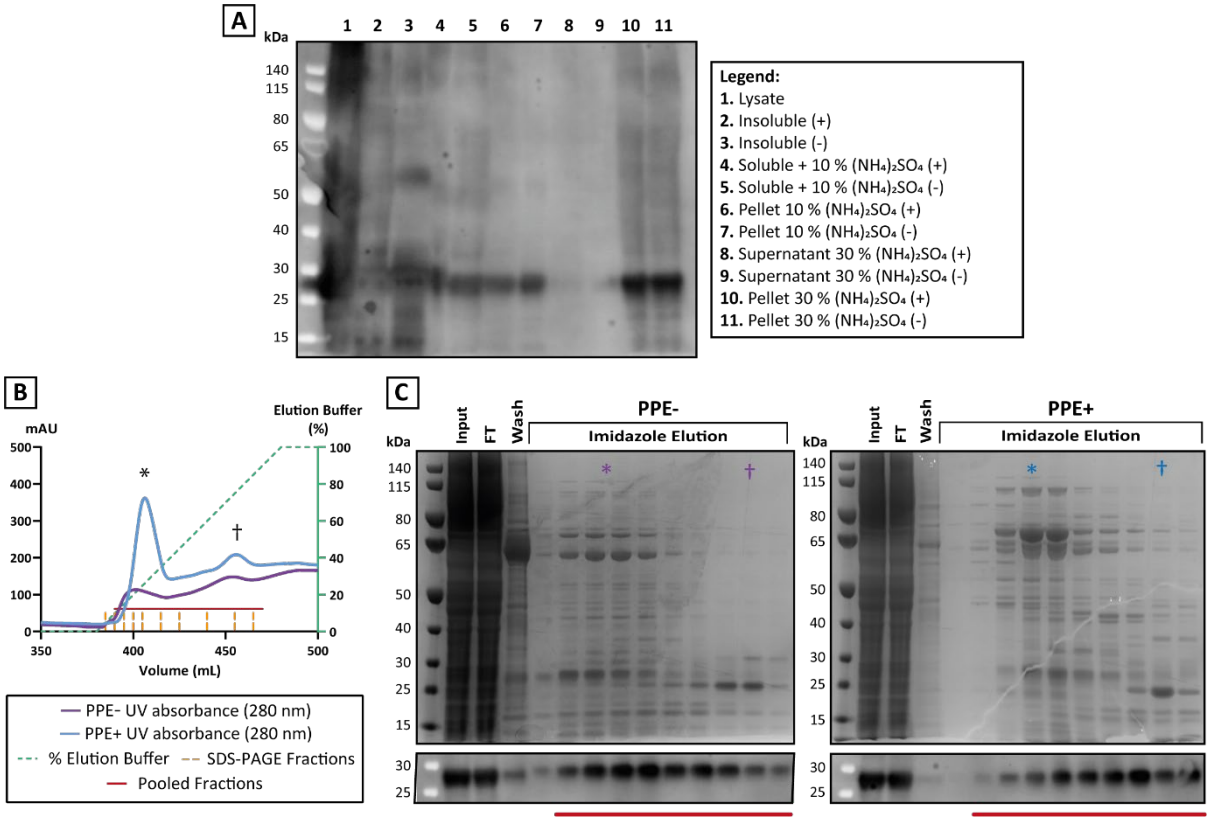


Figure 4.19 | ProPPE-His$_6$ purification 3, lysis, precipitation and IMAC purification. A) Western blot analyses of lysate, ammonium sulphate precipitation, and resuspension of recombinant ProPPE-His$_6$; **B)** IMAC 280 nm UV elution profile; **C)** SDS-PAGE (top) and western blot (bottom) analyses of IMAC purification of ProPPE-His$_6$. * and † correspond to the elution profile peaks in (**B**) and the red line indicates pooled fractions. An anti-His$_6$ primary antibody was used for expression level analysis.

Before tryptic activation of ProPPE-His$_6$, a further purification step was performed to reduce the amount of contaminating protein exposed to the activated PPE-His$_6$. As previous purification trials had indicated that CEX was not a feasible purification technique for recombinant ProPPE-His$_6$ due to poor retention (Section 4.2.3.3), the samples were purified by gel filtration using S75 Sephadex resin. As the amount of ProPPE-His$_6$ was low, purification was initially envisaged as using a 25 mL S75 column. This required concentration of each sample to a volume of less than 250 µL to reduce peak broadening. However, protein precipitation was already evident at a sample volume of approximately 2.5 mL. To expedite the purification of a larger sample volume, a 120 mL S75 column was used with 1.2 mL of protein loaded per run (Figure 4.20, A & B). Both samples eluted as two major peaks and the UV traces superimposed well (Figure 4.20, A). The initial and larger peak corresponded to the column void volume, indicating the presence of large macromolecules such as protein aggregates. The void peak

also contained most of the eluted ProPPE-His_6 and was less pure than the subsequently eluted peaks (Figure 4.20, B). As the concentration of recombinant protein was very low, all the peak fractions likely containing ProPPE-His_6 were combined. In the future, isolating each peak may lead to further improvements in activity.

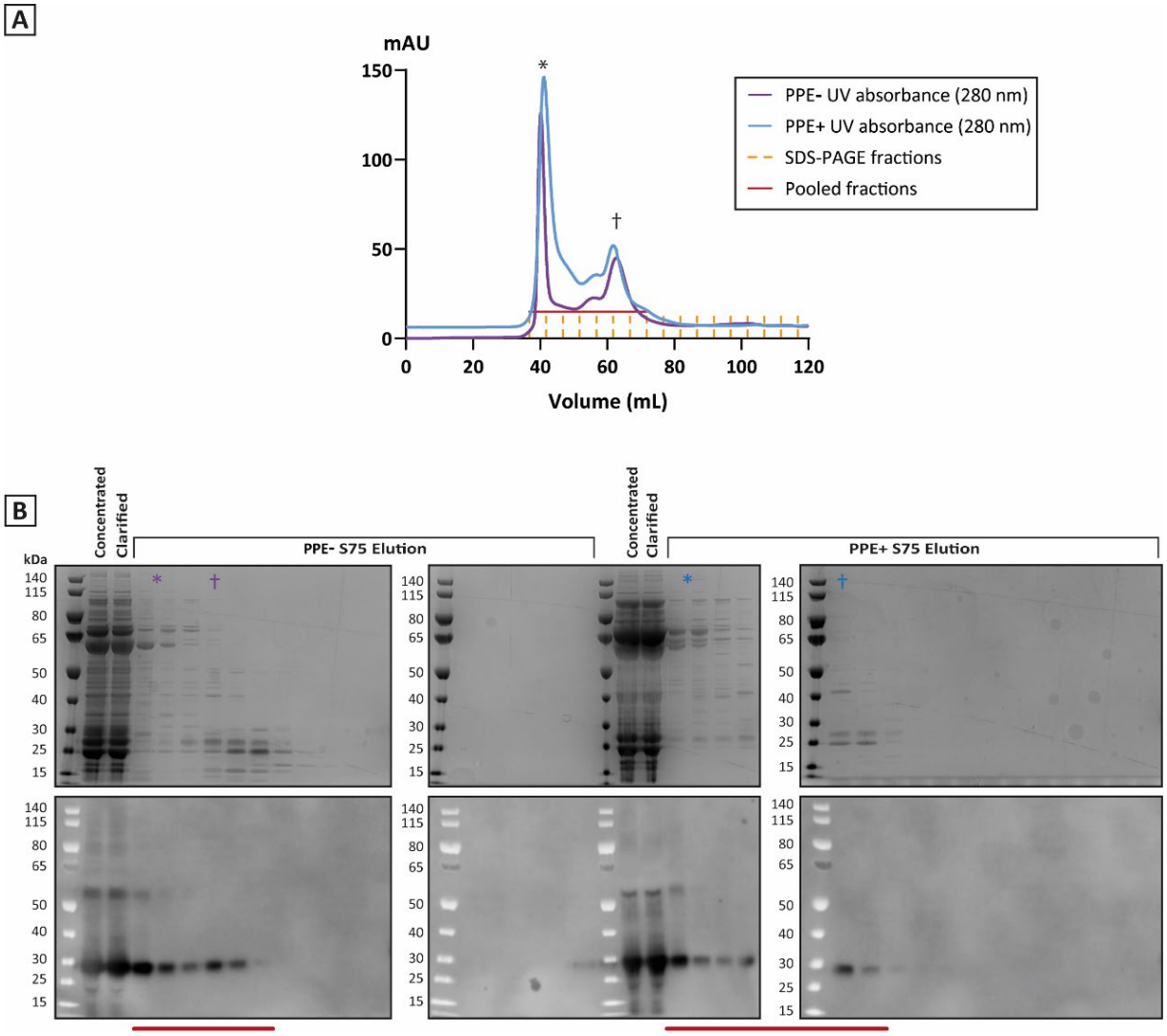


Figure 4.20 | ProPPE-His_6 purification 3, size exclusion chromatography. A) SEC UV elution profile of recombinant ProPPE-His_6; **B)** SDS-PAGE (top) and western blot (bottom) analysis of size exclusion chromatography purification of ProPPE-His_6. * and † correspond to the elution profile peaks in (**A**) and the red line indicates pooled fractions. An anti-His_6 primary antibody was used for expression level analysis.

To assess the benefits of reducing the duration of incubation with TR, on-column (IMAC) activation by TR was explored. Previously, an incubation time of five minutes had been found to be sufficient to fully cleave the propeptide from ProPPE-His_6 (Figure 4.8, C). However, it was unknown whether the resin would hinder the interaction between enzymes. Therefore, a longer incubation time of 15 min was

selected. After activation using TR, the column was extensively washed in an effort to remove residual TR before a three-step elution protocol (22 mM, 130 mM, and 250 mM imidazole) (Figure 4.21, A-D). The PPE- sample was slightly purer prior to activation compared with the PPE+ sample (Figure 4.21, A & B). Eluting with 22 mM imidazole produced a minor elution peak for both samples but no protein was observed by SDS-PAGE or western blot analysis (Figure 4.12). The strongest peak eluted at 130 mM imidazole for both PPE- and PPE+ samples. This peak contained contaminating proteins and some recombinant PPE-His_6 (Figure 4.12). Whilst the PPE+ sample had apparently been fully activated by TR (Figure 4.21, B), a double band was observed for the PPE- sample by western blot analysis suggesting that activation of ProPPE-His_6 was incomplete (Figure 4.21, A). Close inspection of the sample handling procedure revealed that the activation time for PPE- was up to 50% shorter than projected due to improper preparation of the liquid handling system (ÄKTA). However, even considering the reduction in contact time to 7.5 min, on-column tryptic activation remained slower compared to in-solution activation, suggesting that immobilisation of ProPPE-His_6 may introduce a degree of steric hinderance. A second smaller protein peak eluted at 250 mM imidazole for both samples (Figure 4.21, C & D). This peak was apparently purer and a contained a greater concentration of PPE-His_6 for both PPE+ and PPE- samples. Western blot analysis indicated that the PPE- 250 mM imidazole sample only contained one band that aligned with the lower molecular weight band from the 130 mM peak, and therefore likely corresponded to activated PPE-His_6 (Figure 4.21, A). As this was the final stage of purification, fractions from each peak were pooled separately. To improve the stability of the samples, the buffer was exchanged to 50 mM acetate buffer pH 4.5 to reduce the catalytic activity as standard for storage of related SPs. Each sample was then concentrated to 0.3–0.6 mg mL^{-1} (Table 4.2).

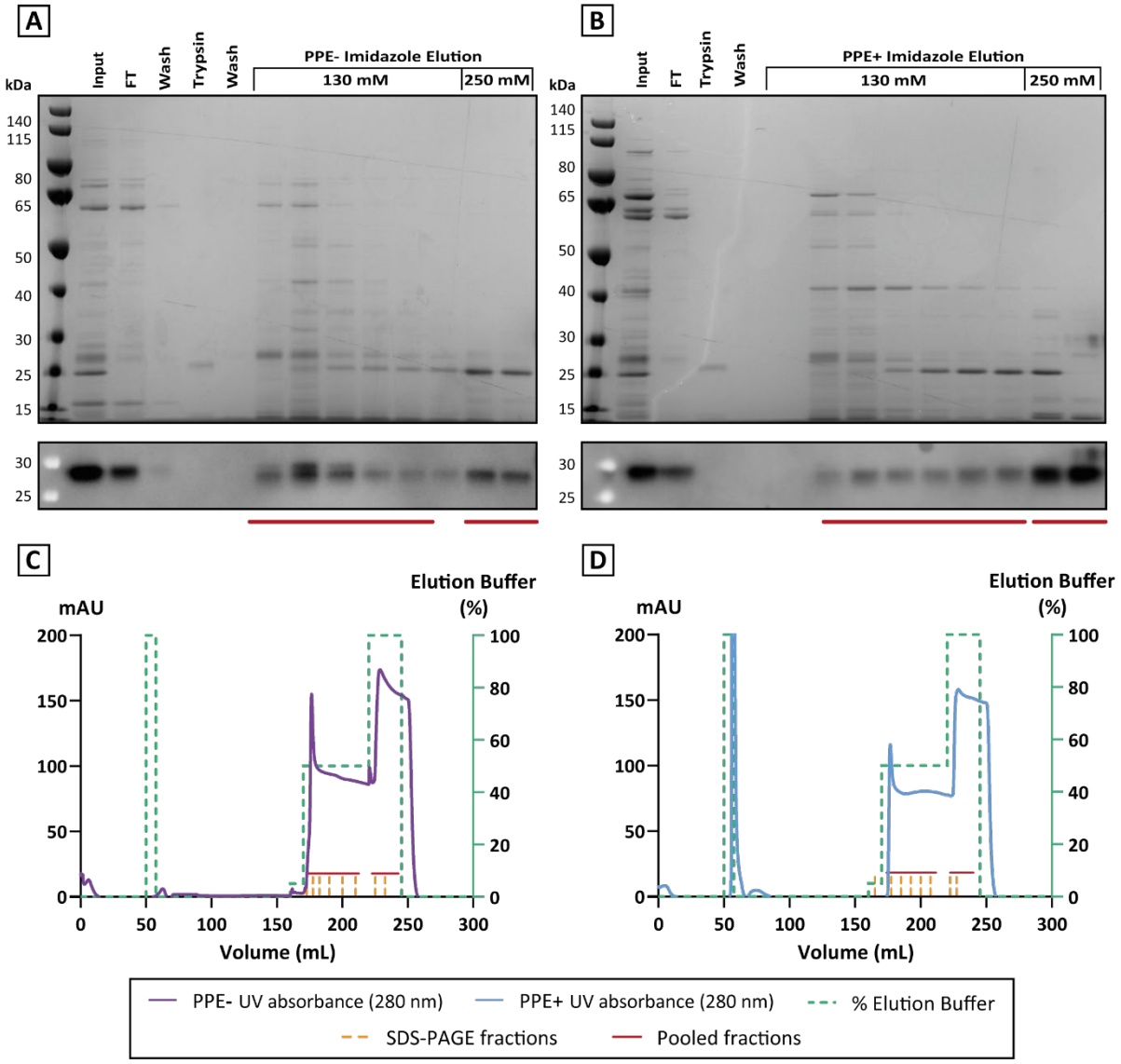


Figure 4.21 | ProPPE-His$_6$ purification 3, on-column TR activation. A) SDS-PAGE (top) and western blot (bottom) analysis of IMAC on-column activation of PPE-His$_6$ in the absence of $CaCl_2$; **B)** SDS-PAGE (top) and western blot (bottom) analysis of IMAC on-column activation of PPE-His$_6$ in the presence of $CaCl_2$; **C)** UV absorbance profile of IMAC on-column activation of PPE-His$_6$ in the absence of $CaCl_2$; **D)** UV absorbance profile of IMAC on-column activation of PPE-His$_6$ in the presence of $CaCl_2$. The red line indicates pooled fractions. An anti-His$_6$ primary antibody was used for expression level analysis.

Table 4.2 | Final purity and concentration of PPE-His$_6$ samples from purification 3.

Sample	[Imidazole] (mM)	Final [C] (mg/mL)	A260/A280	Final mass (mg)	Mass/culture volume (µg/L)
PPE+	130	0.502	0.56	0.351	48.8
	250	0.356	0.51	0.206	28.6
PPE-	130	0.435	0.55	0.157	21.8
	250	0.530	0.51	0.201	27.9

Subsequently, the purity of each sample was analysed by both SDS-PAGE and LC-MS; the results show the low purity of the samples (Figure 4.22, A-E). The samples eluted at 130 mM imidazole were much less pure than those eluted at 250 mM. The strongest band in all four samples likely corresponds to the 23,510 Da peak observed by LC-MS (Figure 4.8, G & H). This band was present in previous purification trials and was only removed by CEX as a final purification step (Figure 4.13). As the concentration of recombinant protein was already low in this purification, CEX was not performed on the PPE+ and PPE- samples, but remains a viable option for a final purification step. LC-MS analysis confirmed the presence of PPE-His_6 in all samples except PPE- eluted with 130 mM imidazole (Figure 4.22, B-E).

The catalytic activity of the PPE+ and PPE- samples was compared to that of commercial PPE (Figure 4.23). All the samples from PPE+ and PPE- purifications performed at least as well as the purer sample from the purification 2 (Figure 4.14). This observation suggests that on-column activation combined with minimal incubation with TR is capable of improving the final specific activity. Importantly, PPE- (250 mM imidazole; PPE- 250*/†; Figure 4.23) exhibited substantially improved catalytic activity over all other purified samples. The PPE- samples eluted at 250 mM imidazole demonstrated comparable initial catalytic activity to the 0.1 μg mL^{-1} commercial sample and again appeared to have improved stability, effecting an almost complete hydrolysis of the substrate after three hours (Figure 4.23). Considering the poor purity of the sample, the observed proteolytic activity of PPE- (250 mM) is promising, as its purity could likely be improved by further optimisation of the purification protocol, for example with a final CEX stage. While PPE- was less pure than PPE+, all the PPE- conditions exhibited improved catalytic activity over the PPE+ counterparts, highlighting the detrimental effects of 10 mM $CaCl_2$ supplementation during purification. Furthermore, supplementation of the assay buffer with $CaCl_2$ (10 mM) reduced the activity in all conditions except for PPE+ (250 mM).

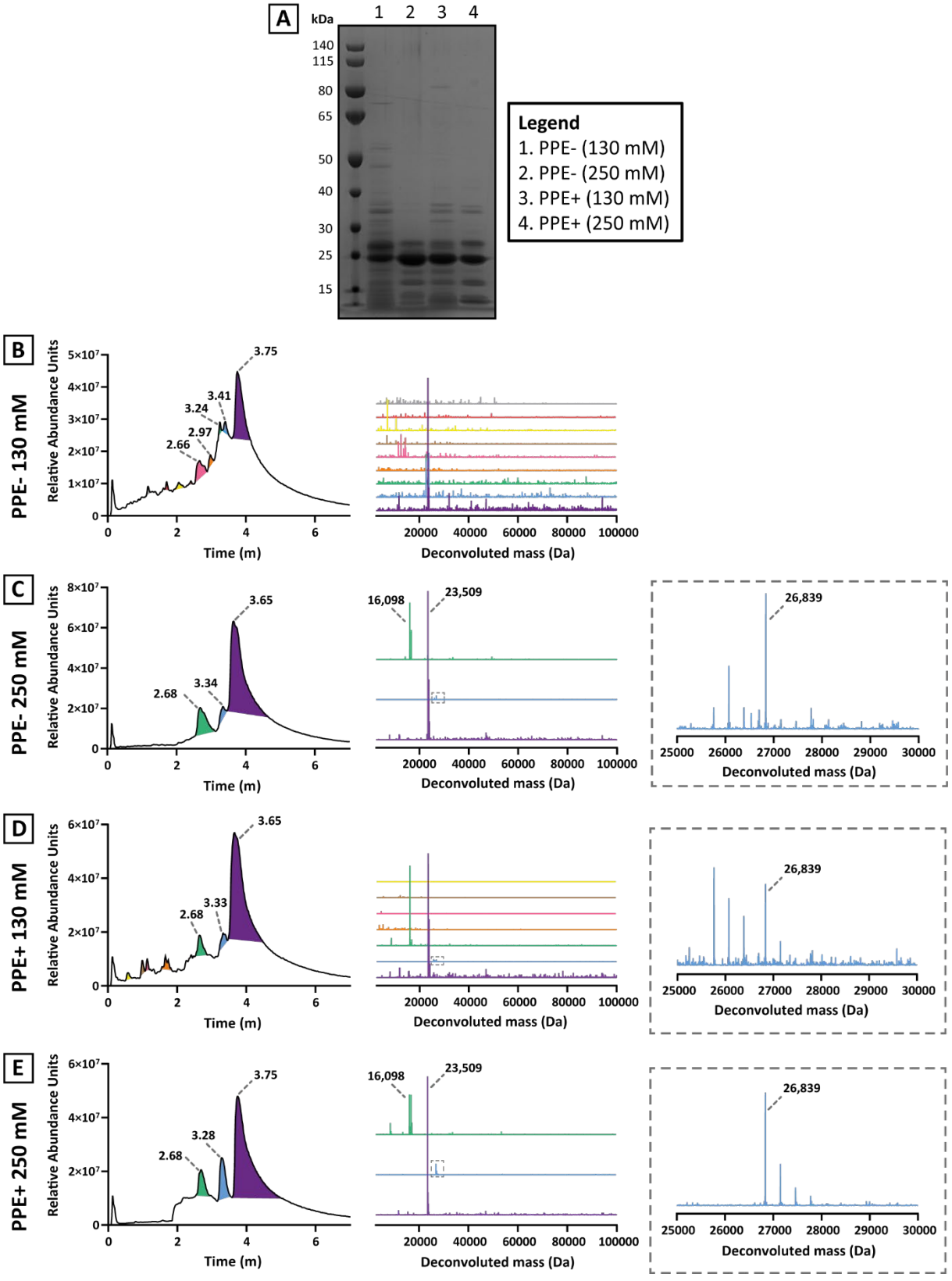


Figure 4.22 | Purity analysis of the final samples of PPE+ and PPE- from purification 3. A) SDS-PAGE analysis of the samples after buffer exchange and concentration; **B-E)** LC-MS analysis of the samples; elution chromatogram (left), mass spectra with colours corresponding to the peak colours (middle), mass spectra indicating the potential PPE-His_6 peak (right); **B)** PPE- 130 mM; **C)** PPE- 250 mM; **D)** PPE+ 130 mM; **E)** PPE+ 250 mM. Calculated mass for PPE-His_6 = 26,845 Da.

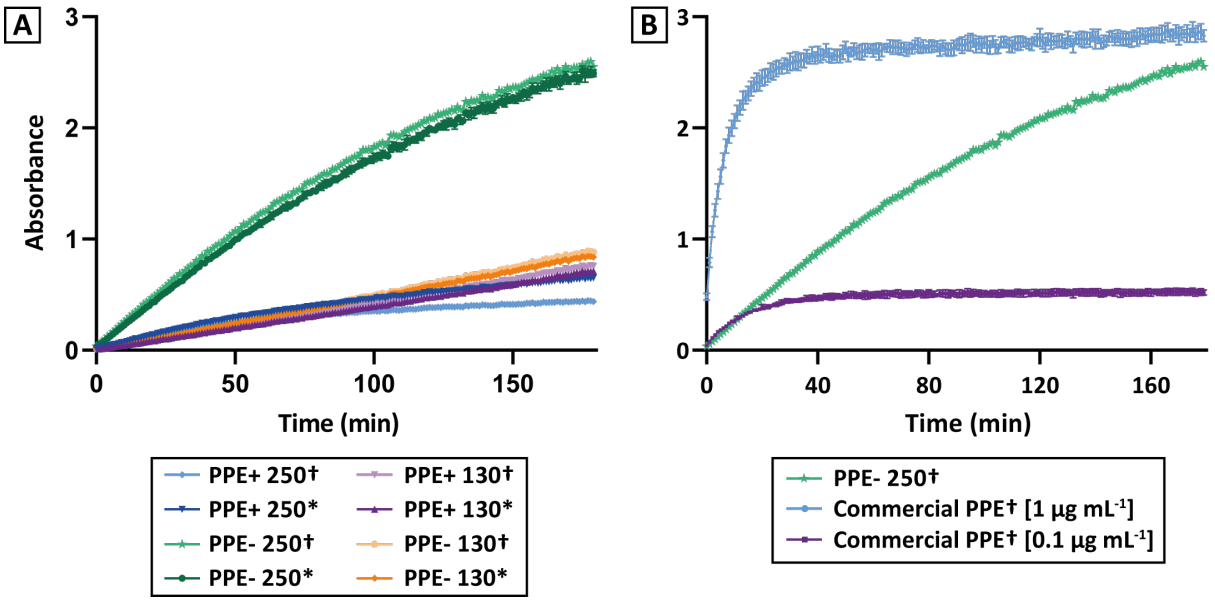


Figure 4.23 | Catalytic activity assays of the final samples of PPE-His$_6$ from PPE+ and PPE- purification. A) Comparison of all purified PPE-His$_6$ samples from the PPE+ and PPE- purification; **B)** comparison of commercial PPE with PPE- 250*. All assays with PPE-His$_6$ were performed at a protein concentration of 1 µg mL^{-1}. Each data point represents the mean of four experimental repeats (n = 4 ± SD). 130 and 250 indicate the concentration of imidazole (mM) at which the sample eluted after on-column tryptic activation. * Indicates activity assays in which the buffer was supplemented with 10 mM $CaCl_2$; † indicates activity assays performed without $CaCl_2$ supplementation.

4.3. Discussion

The initial design of the expression construct for recombinant PPE required careful consideration to bypass potential cytotoxicity issues in the bacterial expression host due to proteolytic promiscuity of PPE. Further, the *N*–terminus of active PPE folds into the protein structure, with V16 as the *N*–terminal residue suggesting that presence of an initiator methionine at the *N*–terminus may adversely affect protein folding (Section 1.3.2; Figure 1.7 and Figure 1.8). As such, it was proposed that the PPE construct must be expressed as a zymogen, harbouring the natural *N*–terminal propeptide with a TR cleavage site. This strategy should reduce the potential toxicity and place the correct amino acid (Val16) at the *N*–terminus upon tryptic hydrolysis, enabling proper folding.

ProPPE was initially cloned into two expression vectors containing *N*–terminal His_6-tags (pNIC28 and a modified pCOLD-I). Expression trials with these constructs in BL21 cells yielded no observable soluble His_6-ProPPE expression (Figure 4.2). Analysis of the insoluble fractions indicated that expression had occurred but the vast majority of ProPPE was produced as inclusion bodies. This was initially suspected to be an issue of expression in a reducing environment that was unfavourable for formation of disulphide bonds. Neither the use of SHuffle cells, which favour disulphide bond formation within the cytoplasm, nor targeting expression to the periplasmic space through insertion of *N*–terminal signal peptides were able to achieve soluble recombinant His_6ProPPE production (Figure 4.3; Figure 4.5)[275-277,279]. It was therefore predicted that the observed insolubility of ProPPE was not solely due to the inability to form disulphide bonds.

Re-evaluation of the constructs led to the hypothesis that the *N*–terminal His_6-tag may have caused steric or electronic hinderance to the folding of the ProPPE zymogen. As such, the tag was repositioned to the *C*–terminus. Various solubility tags were also fused to the *N*–terminus of the proenzyme to improve protein folding[280]. The results of expression trials with these constructs were promising. Soluble expression was observed in the majority of constructs without an *N*–terminal His_6-tag (Figure 4.5). While this somewhat supported the theory that the *N*–terminal His_6-tag was the issue in

previous expression attempts, it was not a simple case of steric hinderance, as soluble expression was observed with constructs harbouring larger N-terminal fusion proteins. Further trials demonstrated that, while recombinant ProPPE-His_6 appeared to be resistant to proteolytic degradation after lysis, the linker between ProPPE and the fusion tags containing a TEV cleavage site was labile (Figure 4.7). As a result, ProPPE-His_6:pCOLD expressed in BL21 cells without a fusion partner was selected for larger scale purification trials. Of interest, recombinant trypsinogen has been recombinantly expressed from *E. coli* with the signal peptide from alkaline phosphatase[269,286]. While this strategy was unsuccessful for constructs with *N*–terminal, it may be of interest to explore periplasmic expression of ProPPE-His_6.

The initial attempt at ProPPE-His_6 purification comprised of two rounds of IMAC purification, with activation by tryptic cleavage performed in-between to assess the viability of enzyme activation with TR. While the purity and yield were extremely poor, these results show that active PPE-His_6 can be produced by tryptic cleavage of the propeptide (Figure 4.8). A second attempt at purification demonstrated that CEX successfully removed the main contaminant protein observed in the first purification, but the yield and catalytic activity remained poor (Figure 4.13; Figure 4.14). Further issues were also identified with the purification method, including Ni^{2+} ions being stripped from the IMAC column potentially due to small molecule metallophores produced by *E. coli* (Figure 4.11)[281]. To remove the suspected metallophores, ammonium sulphate precipitation was examined and found to work well (Figure 4.11). Subsequent purification of the precipitated sample by IMAC did not result in column stripping. Another issue observed after analysis of the western blots was the poor capture efficiency of the IMAC column, possibly due to poor His_6-tag solvent accessibility. Inspection of the crystal structure of active PPE shows that the C-terminus is essentially solvent exposed, suggesting that a short linker would be sufficient to enable the interaction between the tag and the resin. However, this was not observed empirically. Extending the linker and adding extra histidine residues would likely be a viable strategy for increasing the yield by reducing the loss at each IMAC purification stage. Furthermore, adding a cleavage sequence between the His_6-tag and PPE may provide another method

to improve the purity by enabling an additional reverse IMAC step after affinity tag cleavage. However, the cleavage site would need to be examined to ensure that it is resistant to both hydrolysis by *E. coli* proteases and TR. Alternatively, other tags, such as the strep-II tag which has high specificity, may offer improvements in purity[280].

A subsequent purification trial examined whether performing a low pH (5.0) CEX step prior to IMAC, to reduce the contact with contaminating proteins after tryptic activation would improve the catalytic activity of purified PPE-His_6. ProPPE-His_6 was found to be insoluble in acetate buffer after ammonium sulphate precipitation (Figure 4.16). Furthermore, while ProPPE-His_6 was found be soluble in various buffers (pH 6.0–7.5) after ammonium sulphate precipitation, the subsequent capture on the CEX column was extremely poor. The reasons for the poor capture are again unknown, especially considering at these pH values the His_6-tag would be positively charged, but it was clear that CEX was not viable as an early purification method.

To address the poor yield, further expression trials were performed using the ProPPE-His_6 construct. Expression trials first sought to improve the yield by culturing in a nutritionally richer media and aimed to stabilise the soluble protein during lysis through the addition of chemical additives (Figure 4.17)[283]. Neither of these approaches improved the soluble yield. In particular, the richer media led to a substantial reduction in soluble expression, leading to the hypothesis that a reduction in the rate of expression could improve the total soluble yield. Expression trials with the construct were repeated in BL21 *E. coli* cells with shorter expression times and a greater variety in expression temperatures. These trials also examined whether expression in Lemo *E. coli* cells, which have a tuneable expression induction system that is tightly controlled by titration of L-rhamnose, would be capable of improving the yield (Figure 4.18)[284]. The results were promising, providing several conditions with improved soluble expression over the previous condition. Conditions facilitating a faster rate of transcription, such as higher temperature or lower L-rhamnose concentration, accumulated soluble ProPPE-His_6 faster on shorter time scales but resulted in overall reduced levels of soluble enzyme after longer

expression times, supporting the previously raised hypothesis. From these experiments a new candidate condition was carried forward to the final expression trials (pNIC:ProPPE-His_6, 18 h, 18 °C, expressed in Lemo cells with 0.5 mM L-rhamnose).

The final expression trials examined the optimal cell density at the time of induction and whether longer expression times would improve the yield (Figure 4.18). Additionally, the effect of Ca^{2+}, which is known to stabilise the structure of related SPs, was examined[251]. No substantial improvement was observed diverging from the standard induction cell density (OD_{600} = 0.6) or by extending the expression time beyond 18 h. Additionally, supplementation with $CaCl_2$ diminished the level of soluble recombinant protein production. The previously mentioned condition was therefore carried forward for a final attempt at purification.

For the final purification, the stabilising effects of Ca^{2+} were examined during purification. While supplementation of the expression medium with Ca^{2+} proved to be detrimental, this may have been an effect on the cell rather than the enzyme itself. The effects of Ca^{2+} on prokaryotes are less well understood compared with eukaryotic counterparts. It is well known that cytosolic free Ca^{2+} levels in *E. coli* are tightly controlled irrespective of the extracellular levels[287] and that Ca^{2+} supplementation promotes up-/down-regulation of numerous genes[288]. It is possible that alterations to expression patterns in response to Ca^{2+} are detrimental to ProPPE-His_6 expression. Hence, the effects of Ca^{2+} on ProPPE-His_6 during purification were investigated. Furthermore, to minimise the contact with contaminating protein, a gel filtration step before tryptic activation was introduced. However, this step proved problematic due to precipitation whilst concentrating the sample. ProPPE-His_6 purity was not markedly improved post-gel filtration, thus the only tangible benefit of the gel filtration step was a buffer exchange in preparation for the second round of IMAC. Therefore, re-evaluation of this step is necessary, although few alternatives were available.

To evaluate whether further reduction in contact time with TR could improve the final activity, TR activation was performed on-column. Substantial loss of enzyme due to poor capture efficiency by IMAC was again observed (Figure 4.21). However, on-column activation proved successful. The PPE-His_6 enzyme subsequently eluted at two concentrations of imidazole. A user error led to a reduced contact time between TR and the no $CaCl_2$ sample (PPE-) and resulted in incomplete cleavage of the zymogen. All remaining ProPPE-His_6 eluted within the lower imidazole concentration peak which is indicative of weaker interaction with the resin. This observation is consistent with the proposal that the His_6-tag is less exposed in the zymogen, although there is no evidence for this with the related chymotrypsinogen[94]. The samples eluting at a greater concentration of imidazole were purer, but PPE-His_6 still contributed to only a minor fraction of the total protein mass (Figure 4.21). Nonetheless, assessment of the catalytic activity indicated that all samples in the latest purification were at least as active as the previously purified PPE-His_6. Importantly, PPE-His_6 isolated from the PPE- condition (250 mM imidazole elution), showed comparable activity at 1 μg mL^{-1} to commercially available PPE at 0.1 μg mL^{-1} and was substantially more stable over the course of 3 h (Figure 4.23).

The main contaminant observed throughout every purification performed was identified by LC-MS to be a 23,510 Da protein. This may be the *E.coli* cAMP-regulatory protein (CRP) which is known as a common contaminant that can interact with IMAC resin due to the presence of several surface exposed histidine residues[289]. The calculated mass of this enzyme without the initiator methionine is 23,509 Da. This contaminant was only separated from PPE-His_6 by CEX after tryptic activation (Figure 4.13). Overall, the combination both of on-column activation for a minimal time, isolation of the peak eluting at higher imidazole concentration, and a final purification step of CEX to remove the major contaminant, constitute a promising purification strategy for future attempts.

The yield was poor throughout all procedures explored. If the previously discussed attempts to improve the column retention are insufficient, then the expression system will likely require reassessment. While the current yields would be sufficient for kinetic assessment of PPE, it would not

permit structural analysis. Trypsinogen and chymotrypsinogen have both been produced using yeast cells[270,272,290]. Thus, yeast being a eukaryotic organism may prove to be advantageous for the expression of mammalian ProPPE[291]. Human trypsinogen has also been expressed in *E. coli* and refolded[292]. The insoluble expression of ProPPE was often strong in the conditions tested here (Figure 4.2) and thus this approach may offer a better route to improving the yield obtained from bacteria. However, this would require substantial optimisation of the methodology which usually cannot be inferred from the methods used for other proteins, and the final activity can be poor[292].

Nonetheless, this work, with further optimisation to the purification method, may open new avenues to perform mutagenetic studies of PPE in order to complement the time-resolved studies reported in Chapter 2.

Chapter 5: PPE Inhibition Studies

5.1. Part 1: Bicyclic γ-Lactams

5.1.1. Introduction

5.1.1.1. Human Neutrophil Elastase in Disease

The gene for HNE is expressed by several types of human cells[60,293], including neutrophils and macrophages[59,61]. In combination with additional SPs (proteinase 3 (PR3), cathepsin G (CG), and neutrophil serine proteinase 4 (NSP4)) and other immune factors (e.g., metalloproteinases and reactive oxygen species (ROS)), HNE provides neutrophils with antibacterial, antifungal, and antiviral activity[294]. Neutrophils 'store' HNE and the other SPs in acidic azurophilic granules[61]. In response to an infection stimulus, HNE is released from the granules into the phagosome or extracellular space where it effects a range of antimicrobial functions. HNE, specifically, has been directly linked to antibacterial activity against Gram-negative bacteria (*E. coli*)[295,296], Gram-positive bacteria (*Streptococcus pneumoniae*)[297,298], and antifungal activity (*Candida albicans* and *Aspergillus fumigatus*)[299,300]. HNE also activates antimicrobial peptides[301], attenuates infections by cleaving the virulence factors of various enterobacteria[302], and plays an essential role in the formation and release of neutrophil extracellular traps (NETs) in a process termed NETosis[303]. NETs containing decondensed chromatin, proteolytic enzymes, and other cytosolic and granular enzymes, are released into the extracellular space. Released NETs trap both the infectious agents, to prevent dissemination of infection, and various antimicrobial agents, to ensure that these destructive antimicrobial agents remain localised at the site of infection and at sufficiently high concentrations[304].

In addition to the functions described above, HNE plays a role in modulating the inflammatory response[305,306]. The excessive and sustained release of the HNE or NETs into the extracellular space contributes to inflammatory disorders and tissue remodelling[307]. HNE has been implicated in the pathogenesis of diseases such as, chronic obstructive pulmonary disease (COPD)[308,309], pulmonary

arterial hypertension (PAH)[310], chronic kidney disease (CKD)[294], and cancers[307,311,312]. Accordingly, significant effort has been committed to the development of HNE inhibitors.

5.1.1.2. γ-Lactams

Since the discovery of penicillin in the late 1920s[313], β-lactam antibiotics have consistently remained amongst the most commonly prescribed antibiotic classes. β-Lactams comprised 65% of injectable antibiotics in the USA between 2004 and 2014[314], whilst penicillins alone accounted for 50% of antibiotic prescriptions in England between 2013 and 2015 (a sample of 7% of the general population)[315]. However, microbial resistance to β-lactams has been a persistent and growing concern that was raised even before the widespread public availability of the antibiotics[316]. To combat this resistance, novel scaffolds and derivatives of known β-lactams have been explored, resulting in the synthesis of many classes and thousands of individual compounds containing a β-lactam moiety (Figure 5.1).

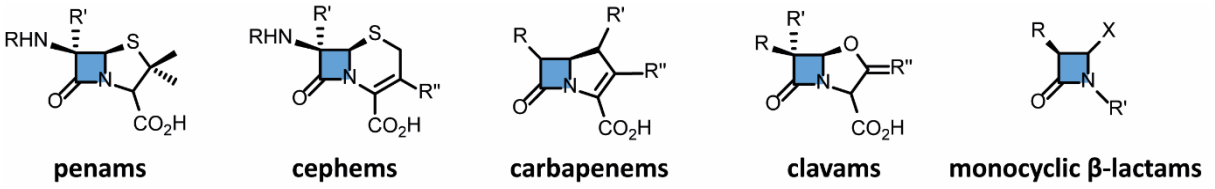


Figure 5.1 | β-lactam classes.

γ-Lactam analogues of β-lactam antibiotics, which possess an expanded five-membered lactam ring, were prepared as early as 1949 but did not possess biological activity[317-319]. More than three decades later, the first γ-lactams with antibacterial activity were reported, independently, by research groups from Oxford and Eli Lilly[320-322]. Analogously to the β-lactams, many γ-lactam containing compounds have been isolated from natural sources (fungi[323], plants[324], and bacteria[325]) or produced by synthetic means[326,327]. These compounds have not only found utility as antibacterial agents but also as anti-inflammatory, anticancer, antifungal, and antimalarial agents[327].

Lactam compounds exhibit antibacterial activity through the inhibition of penicillin binding proteins (PBPs) which are responsible for peptidoglycan transpeptidation, resulting in improper cell wall biosynthesis [328]. Lactams demonstrate antimicrobial activity by acylating catalytic serine residues of PBPs[328]. It is therefore perhaps unsurprising that β- and γ-lactams are capable of acylating the nucleophilic serine residues in other classes of enzymes, including SPs[329-331]. Early crystallographic analyses with monocyclic β-lactams and PPE revealed that once the acyl-enzyme complex had formed, the ester bond could rotate 120° out of the oxyanion pocket (Figure 5.2, A)[332]. In this conformation the proposed hydrolytic water[106] was displaced and out of plane with the ester, potentially explaining the hydrolytic stability of the acyl-enzyme complex[332]. Analogous monocyclic γ-lactams were shown to be weaker inhibitors of PPE[333,334]. The poor potency of the γ-lactams likely derives from the relatively low ring strain release after the hydrolysis of the five-membered lactam ring, and more productive orientation of the acyl-enzyme complex. Nonetheless, inhibition of PPE by monocyclic γ-lactams was achieved through the selective positioning of the additional lactam methylene group relative to the lactam nitrogen and the substituents (Figure 5.2, B)[334].

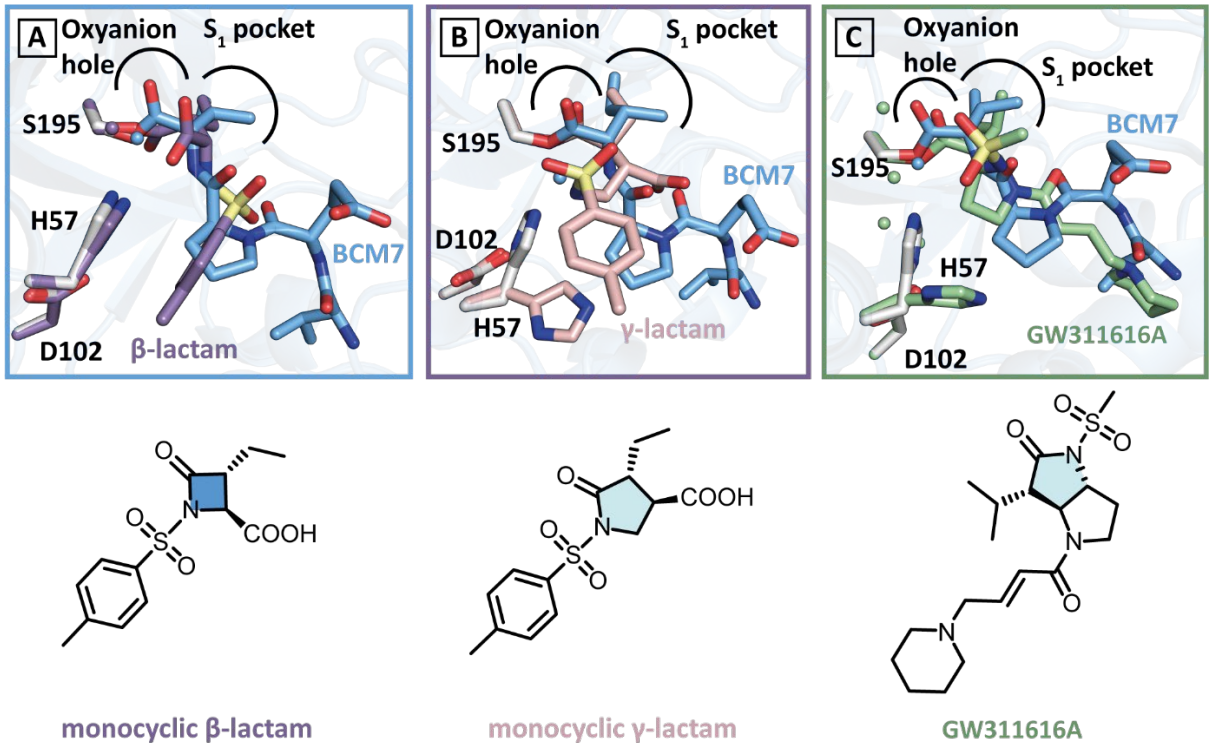


Figure 5.2 | Comparison of binding modes of lactam inhibitors of PPE with the productive inhibitor BCM7. A) Superimposition of the PPE:BCM7 complex (light blue carbons; Figure 2.5, D) and a PPE:monocyclic β-lactam acyl-enzyme complex (purple carbons; PDB: 1BTU)[332]; **B)** superimposition of the PPE:BCM7 complex and a PPE:monocyclic γ-lactam acyl-enzyme complex (pink carbons; PDB: 1QGF)[334]; **C)** superimposition of the PPE:BCM7 complex and a 5,5-*trans*-fused γ-lactam (GW311616A) acyl-enzyme complex (green carbons; PDB: 1HZ7)[335]. PPE:BCM7 catalytic triad carbons are in white and waters are shown as spheres coloured according to the ligand colour. The structures of the unreacted lactam inhibitors are shown below each of the crystal structures. The β-lactam ring is coloured dark blue, and the γ-lactam ring is coloured light blue.

In 1999 and 2001, potent 5'5-*trans*-fused γ-lactams that inhibited HNE were reported (Figure 5.2, C). These inhibitors exemplified that lactam compounds could exhibit strong inhibition against SPs[335,336]. However, the synthesis of such compounds was complex. In an attempt to simplify the synthesis, but retain the inhibitory potency, Migaud *et al.* explored 5,5-fused thiophene γ-lactams[331]. The reasoning behind the potency of the compounds was that resonance between the thiophene ring and the amine lone pair following cleavage of the lactam ring, stabilising the acyl-enzyme complex. Four inhibitors were produced with various C4-substituents (Figure 5.3). All four inhibited HNE with IC_{50} values ranging from 1 µM (**4**) to >500 µM (**3**). Crystallographic analysis of PPE soaked with **2** (monoalkylated C4) was also reported[331]. The inhibitor bound in a productive manner with the ester bond oriented as observed in the PPE:BCM7 complex[106]. The C4 ethyl group was located within the S_1 specificity pocket. Thus, the poor inhibition of (**3**) was ascribed to steric hinderance of the bis-allylated C4 within the S_1 pocket. Importantly, the proposed hydrolytic water was absent from the structure meaning that hydrolysis of

the acyl-enzyme complex was unlikely. Thus, the resonance and displacement of the hydrolytic water rationalised the potency demonstrated by the 5'5-fused thiophene γ-lactams. The proposed reaction mechanism is shown in Scheme 5.1.

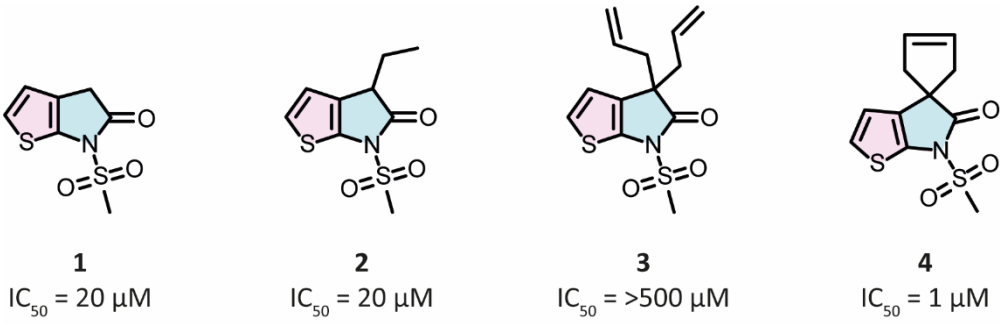
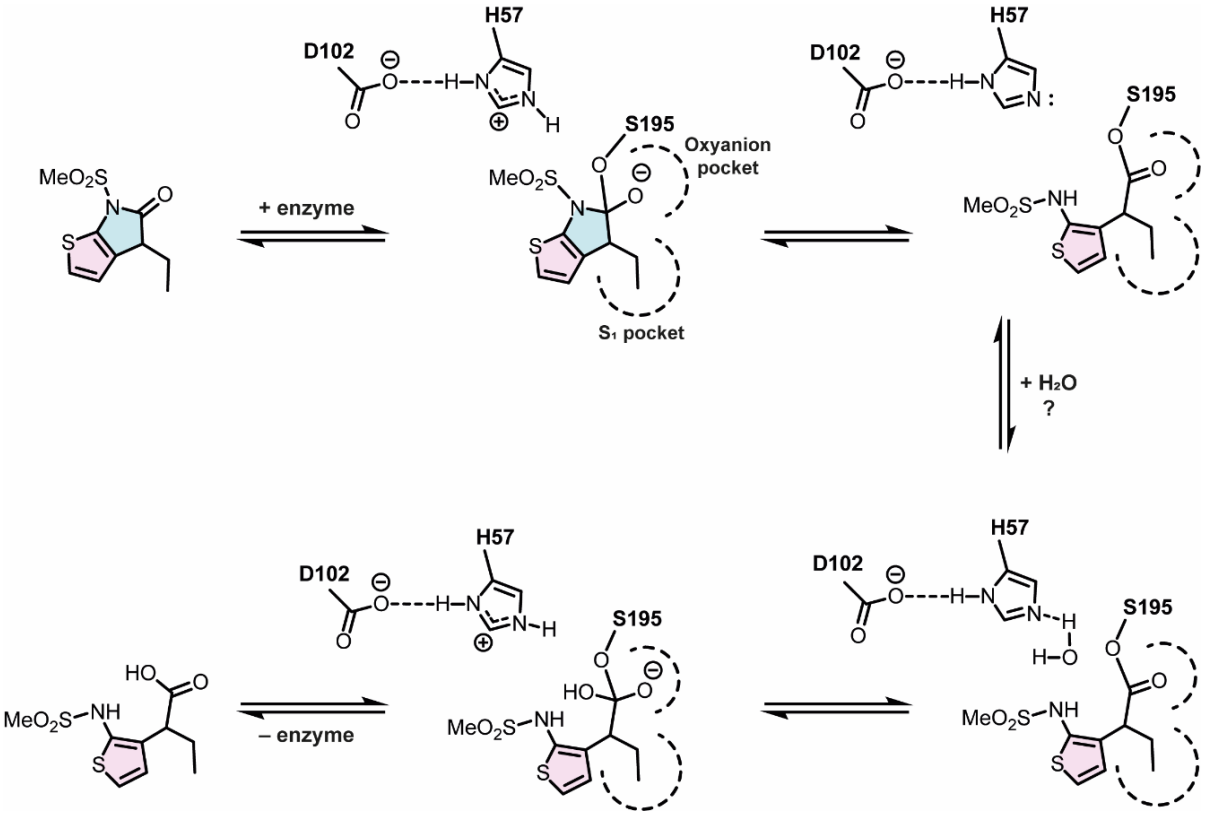


Figure 5.3 | Previously reported 5,5-fused thiophene γ-lactams. IC_{50} values reported for HNE[331]. The γ-lactam ring is coloured blue; the thiophene ring is coloured pink.

Time-resolved studies of γ-Lactams have previously been performed with PPE, using the methodology discussed in Section 2.1.3. Crystal structures of PPE in complex with monocyclic γ-lactams demonstrated that the binding of the compounds could 'force' H57 out of position, *in crystallo* at pH 5.0, where it would not be able to effect deacylation[334]. Subsequently, soaking in high pH buffer (CHES pH 9.0) for 1–2 min promoted H57 to return (predominantly) to the 'normal' position[337]. There is interest in investigating the hydrolytic rection of lactam antibiotics with β-lactamase enzymes from a drug design perspective. Indeed, the reaction of lactam antibiotics with β-lactamases has been explored using both SSX and SFX methodology[338,339]. Thus, the bicyclic γ-lactams are a potential target for time-resolved studies with PPE and perhaps with γ-lactamases[340].



Scheme 5.1 | The proposed mechanism for 5'5-fused thiophene γ-lactam (2) hydrolysis by SPs. The figure was adapted from Wright *et al*. 2000[337].

5.1.1.3. Aims

The aims of the work presented in this chapter were to build upon the fused thiophene γ-lactam scaffold exploring further functionalisation at the C4 position. Dose-response experiments, protein LC-MS, and crystallographic analyses were undertaken with a range of compounds. Additionally, a second bicyclic γ-lactam scaffold (oxindole) was explored for the inhibitory potential against PPE. The work focuses on PPE as a model for HNE due to the active site similarity, the relatively low cost of PPE, and the robust crystallisation conditions established for it.

5.1.2. Results

5.1.2.1. Assay Optimisation

A 384-well spectrophotometric assay based on reported absorbance assays was optimised for PPE[105,341]. The catalytic activity of PPE was assessed by hydrolytic release of *para*-nitroanilide (pNA) from the *C*–terminus of the tetrapeptide substrate: succinyl-Ala-Ala-Pro-Ala-pNA (AAPA). A standard curve for pNA absorption was generated to identify the linear response range (Figure 5.4, A) followed by optimisation of the enzyme concentration produced a linear enzymatic activity window between 0–3 min (Figure 5.4, B). The binding affinity (K_m = 204.1 ± 42.2 µM), turnover number (k_{cat} = 3656 ± 206 min^{-1}), and catalytic efficiency (k_{cat}/K_m = 17.9 min^{-1} μM^{-1}) (Figure 5.4, C) were in agreement with the reported figures for PPE with AAPA (190 µM, 3180 min^{-1}, and 16.7 min^{-1} μM^{-1}, respectively)[105].

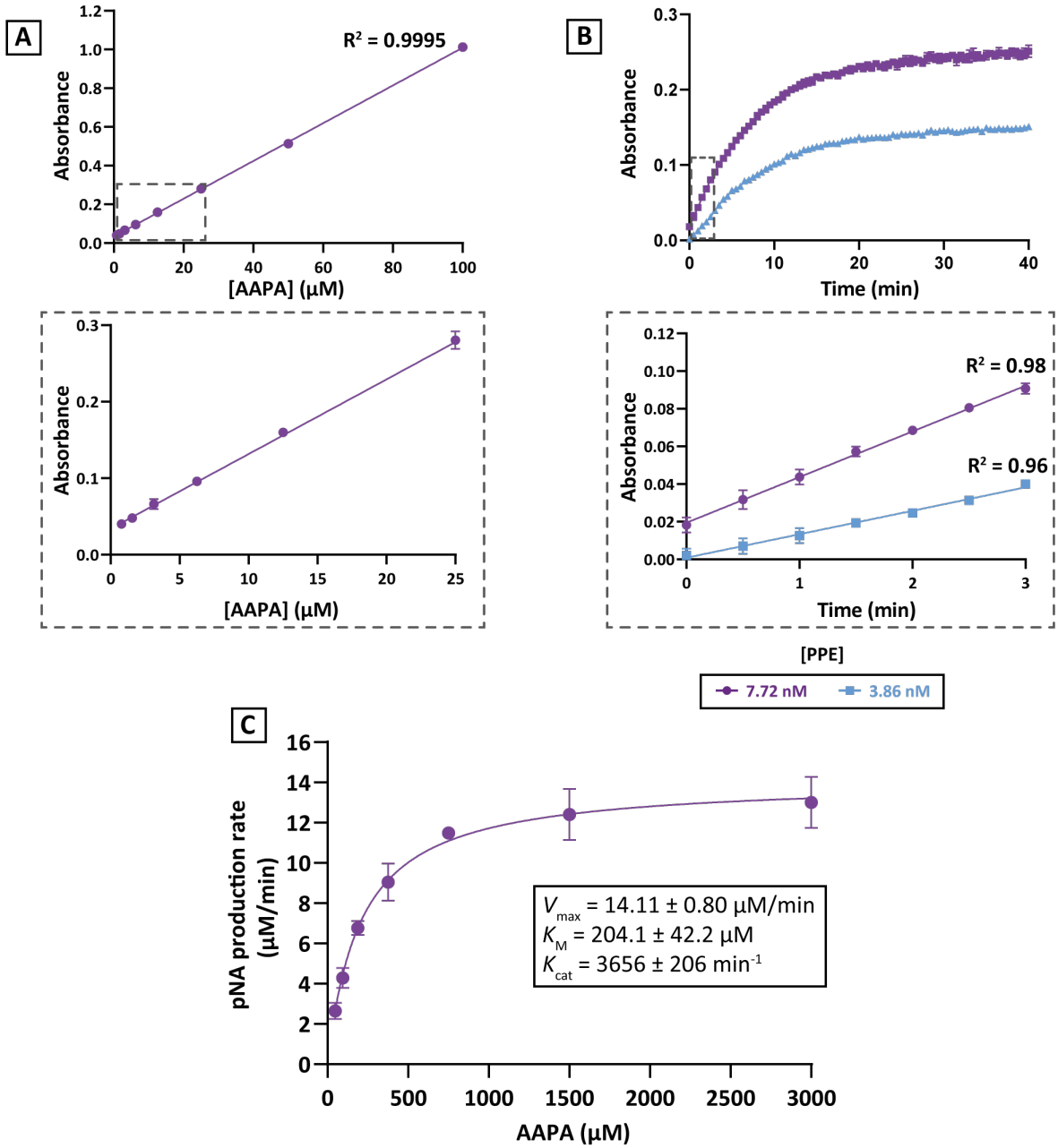


Figure 5.4 | Determination of enzyme concentration and linear enzymatic activity window. A) A standard curve of pNA absorbance measured at 405 nm. The data are shown as the mean of four experimental replicates (n = 4; mean ± SD); **B)** a catalytic activity assay at two concentrations of PPE with AAPA (50 µM), to optimise the enzyme concentration and the linear region. Top shows the full activity curve; bottom shows an initial portion (0-3 min). The data are shown as the mean of four experimental replicates (n = 4; mean ± SD); **C)** PPE Michaelis-Menten substrate saturation curve. Conditions: 3.86 nM PPE. The data are shown as the mean of three independent repeats (n = 3; mean ± SD).

Due to the short linear enzymatic activity period (<3 min), which is shorter than the time it would take to measure a full 384-well plate, reaction quenching using 0.9% (v/v) formic acid was explored (Figure 5.5, A). Addition of formic acid was sufficient to stop the reaction and did not appear to affect the absorbance values and was thus used to quench the reactions in subsequent dose-response assays. The tolerance of PPE to DMSO was also examined as all inhibitors used in subsequent sections required

DMSO as a solubilising vehicle (Figure 5.5, B). Within the linear enzymatic region (< 3 min), there was no substantial difference with 0–5% v/v DMSO. Use of 10% v/v DMSO produced a minor increase in signal, which may be suggestive of absorbance interference at higher DMSO concentrations. Over longer timeframes, each incremental increase in DMSO concentration led to an extended activity period for PPE, suggesting that DMSO may stabilise PPE. 4% v/v DMSO was selected as the optimal concentration for subsequent experiments.

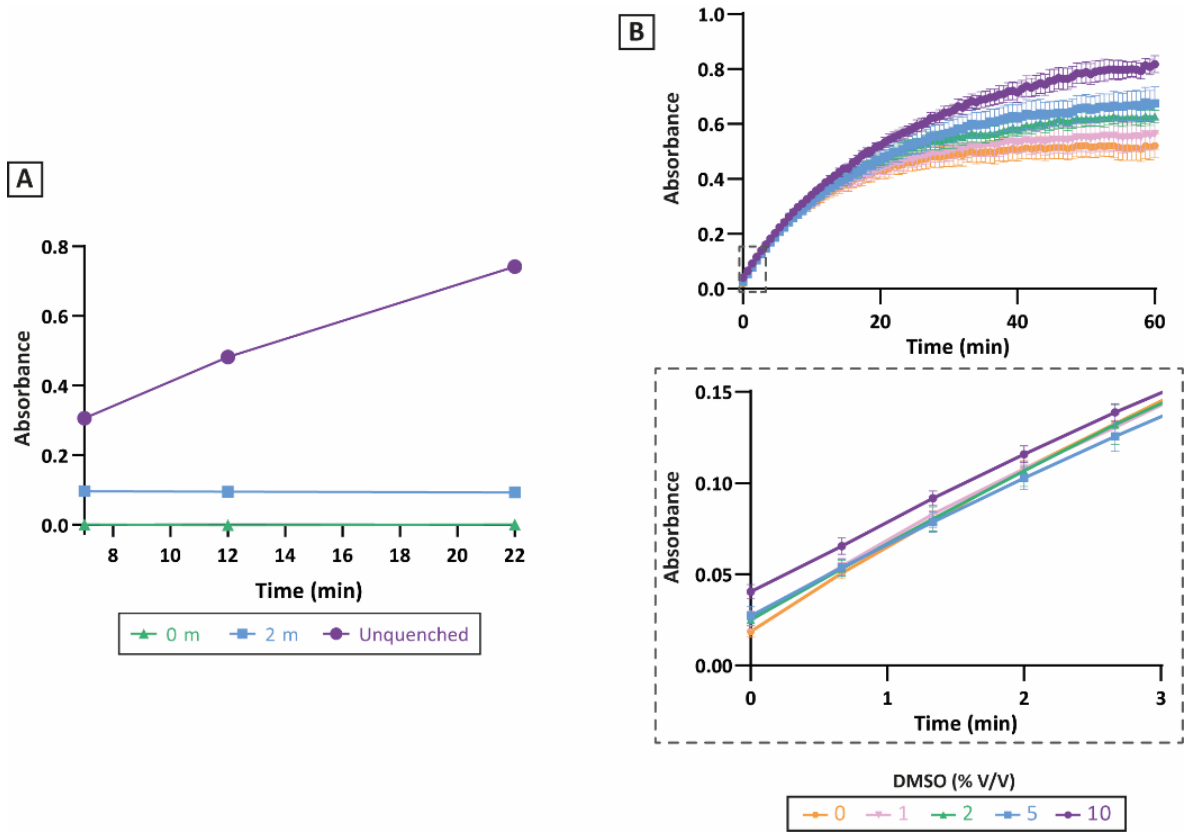


Figure 5.5 | Optimisation of PPE assay conditions. A) Reaction quench assay with formic acid. Activity was quenched at 0 or 2 min by the addition of formic acid (final [C] = 0.9% v/v). A control sample was not quenched. The absorbance was measured for 5–20 minutes after quenching. Conditions: 3.86 nM PPE, 250 µM AAPA, absorbance measured at 405 nm. The data are shown as the mean of eight experimental replicates (n = 8; mean ± SD); **B)** activity of PPE in DMSO (0–10% v/v). Conditions: 3.86 nM PPE, 250 µM AAPA, absorbance measured at 405 nm. The data are shown as the mean of eight experimental replicates (n = 8; mean ± SD).

5.1.2.2. C4 Substituted 5,5-Fused Thiophene γ-Lactams

The four reported thiophene γ-lactams (**1**-**4**) were resynthesised by Gayatri (unpublished data; Schofield group, University of Oxford). As observed for HNE[331], all four compounds inhibited PPE (Figure 5.6). The spirocycle **4** was found to be the best inhibitor of PPE (pIC_{50} = 6.25; IC_{50} = 0.56 µM).

The monoalkylated lactam **2** also exhibited strong inhibition (pIC_{50} = 5.30; IC_{50} = 5.01 µM). The bis-allylated γ-lactam **3** and unmodified γ-lactam **1** did not show strong inhibition with incomplete inhibition at 200 µM inhibitor.

Protein LC-MS analyses with all the compounds were consistent with covalent reaction to form acyl-enzyme complexes, except for **1** (Figure 5.6). **2** and **4** completely acylated PPE after 10 min and remained so for more than 5 h. However, whilst PPE remained fully acylated by **4** after 7 h, **2** had been deacylated. As the analyses were performed using protein LC-MS, it is not yet possible to say if the deacylation is due to the reverse reaction or due to hydrolysis. **3** was predominantly acylated after 10 min and was fully acylated after 7 h. However, **1** demonstrated only a minor acylation peak after 10 min, despite being a slightly stronger inhibitor compared to **3**.

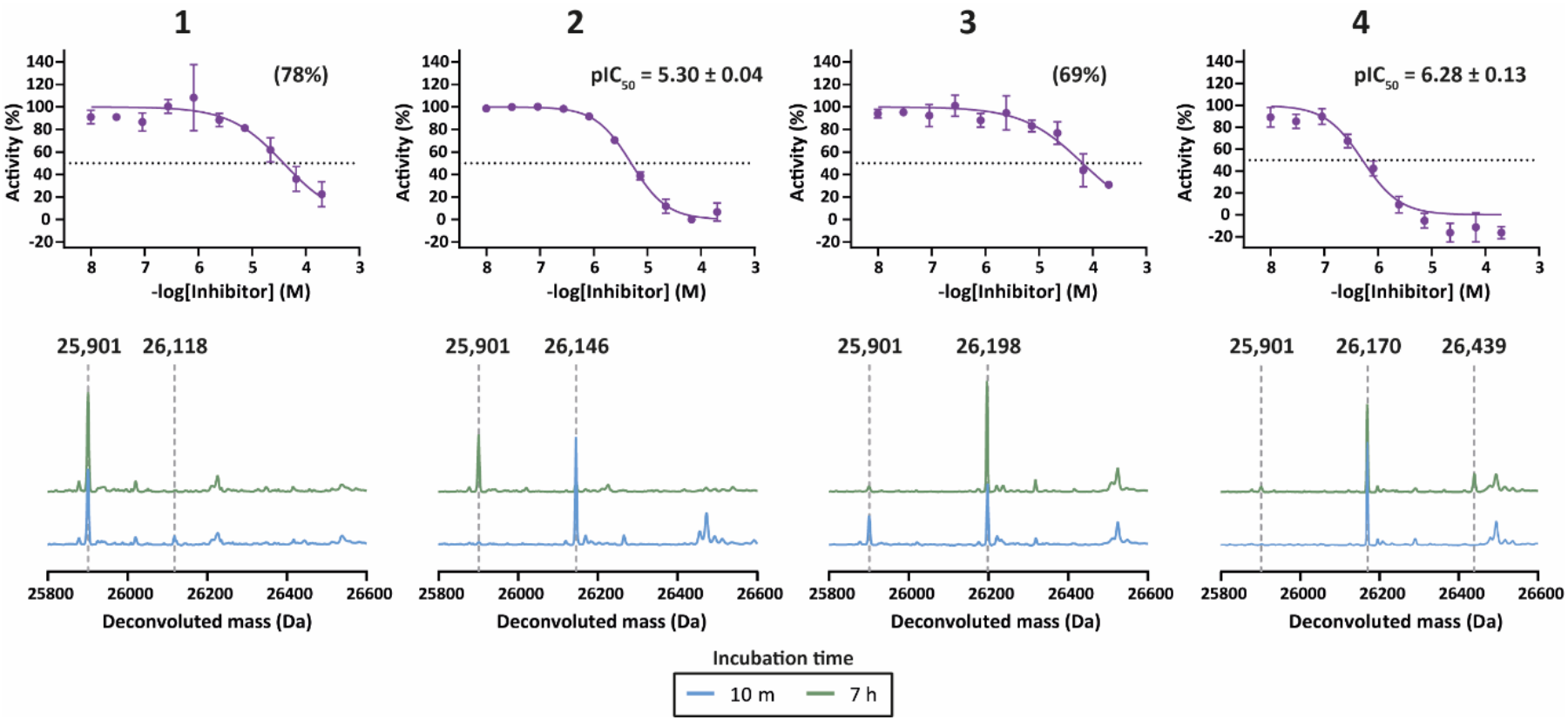


Figure 5.6 | Dose-response curves and protein LC-MS of reported thiophene γ-lactams. Top) Dose-response curves. Conditions: 3.86 nm PPE, 250 µM AAPA, inhibitor: 0.01–200 µM (final [DMSO] = 4% v/v), quenched after 2 min with formic acid (0.9% v/v), absorbance measured at 405 nm. Data are shown as the mean of four experimental repeats (n = 4; mean ± SD); **bottom)** deconvoluted protein LC-MS analysis after 10 min and 7 h incubation times. Conditions: 2.5 µM PPE, 100 µM inhibitor. † Corresponds to a peak that appeared after 7 h which is consistent with binding of a second inhibitor (+538 Da compared to the main apo-PPE).

In addition to the main apo-PPE LC-MS peak centred around 25,901 Da, there were several other smaller peaks observed in the mass spectrum (Figure 5.7). These smaller peaks exhibited analogous mass shifts relative to the main PPE peak after incubation with the compounds that formed acyl-enzyme complexes, suggesting that the peaks were enzyme. The MW of one such peak was 23 Da less than apo-PPE which is consistent with PPE without a Na ion bound at the metal binding site (Figure 5.7). The nature of the other peaks (+120 and +326 from the main apo-PPE) is currently unknown. In addition to the peaks observed with apo-PPE, after 7 h incubation with compounds **4** (Figure 5.6) and **10** (Figure 5.10) a new peak was observed at a MW that was consistent with binding of a second inhibitor molecule (+538 Da or +518 Da from the main apo-PPE peak, for **4** and **10**, respectively).

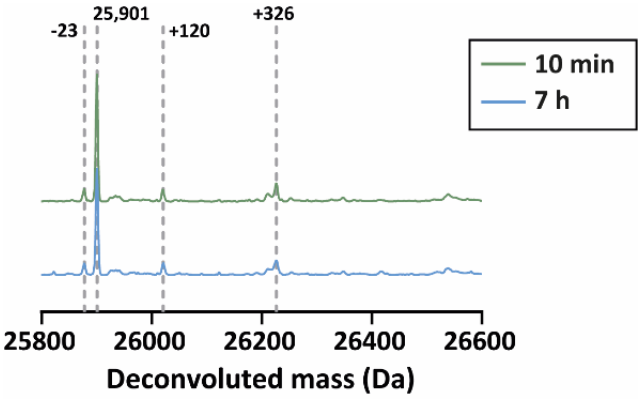


Figure 5.7 | Protein LC-MS with apo-PPE. Conditions: 2.5 µM PPE, 100 µM inhibitor.

Nine novel thiophene γ-lactams were further synthesised by Gayatri (unpublished data; Figure 5.8). A regioisomer of the thiophene ring was prepared (**5**). As the poor inhibition of **3** was attributed to the bulkiness of the bis-allylated C4, two analogues (**6** & **7**) with shorter alkyl groups were synthesised. The monosubstituted C4 group was located within the S_1 specificity pocket of PPE *in crystallo*[331]. It was therefore envisaged that further variation at the C4 position would lead to altered potency. Thus, six additional monosubstituted lactams with hydrophobic groups were synthesised. Three compounds were monoalkylated derivatives (**8**-**10**), including methyl (**8**) and isopropyl (**10**) groups mimicking amino acid sidechains of alanine and valine, respectively. Additionally, an allyl group (**11**), a

propionitrile group (**12**) and a benzyl group (**13**), mimicking the phenylalanine sidechain, were substituted at the C4 position.

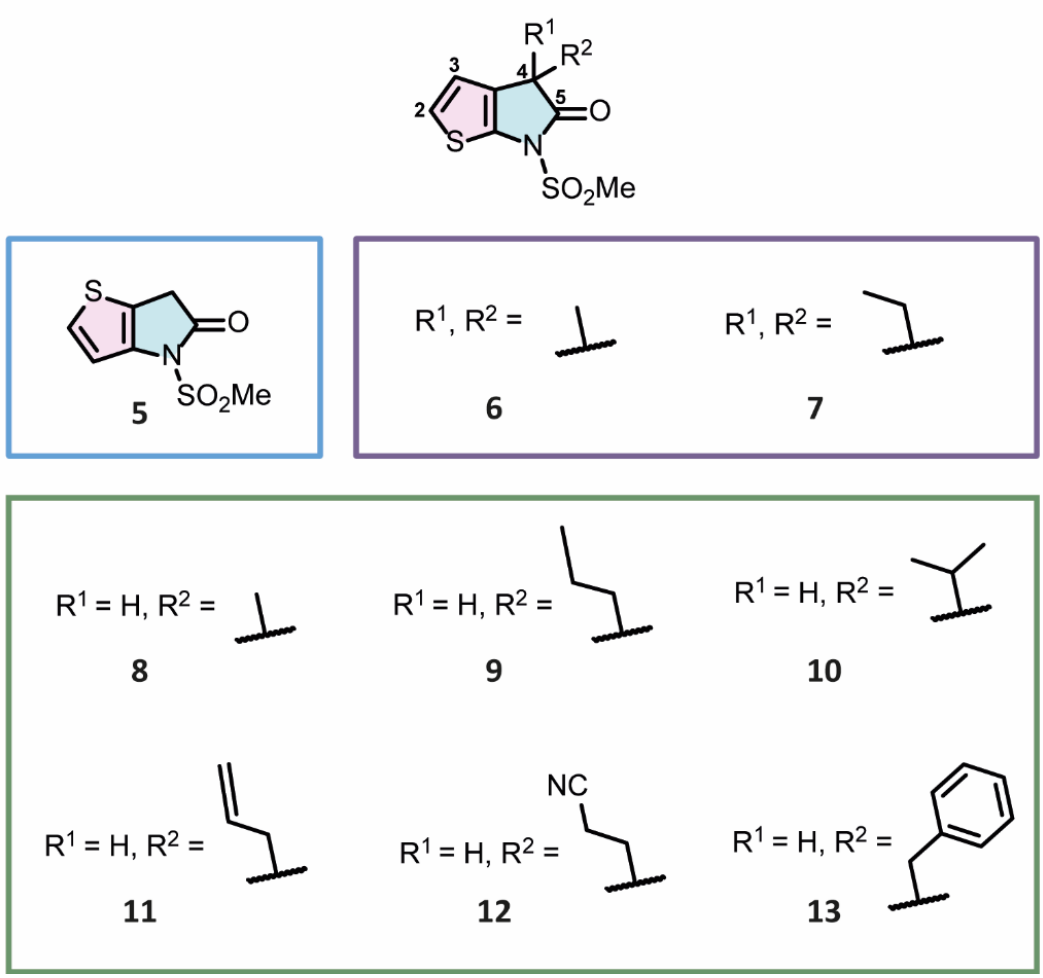


Figure 5.8 | Structures of bicyclic thiophene γ-lactams. The thiophene regioisomer of **1** is in the blue box, the bis-alkylated compounds are in the purple box, and the racemic mono-alkylated compounds are in the green box. All fused thiophene γ-lactams were synthesised by Gayatri.

The thiophene regioisomer **5** was a weaker inhibitor of PPE compared with the reported unsubstituted compound **1**, suggesting that positioning of the sulphur with respect to the lactam amide is important for activation of the amide (Figure 5.9). Similarly, **5** did not acylate PPE, as determined by protein LC-MS. The two novel bis-alkylated lactams **6** and **7** were weaker inhibitors compared with the bis-allylated lactam **3**. **7** demonstrated minimal inhibition (17% at 67 μM). **6** was more potent than **7**, but was still a relatively poor inhibitor reducing the activity of PPE only by 43% at 200 μM. There was no acylation of PPE by **6** as determined by LC-MS. After 10 min, PPE was not acylated by **7**, however,

after 7 h, the enzyme was fully acylated demonstrating slow binding as determined by protein LC-MS. **3** also demonstrated slow binding by protein LC-MS, although to a much lesser extent as there was substantial acylation at 10 min (Figure 5.6). The inhibition assay included a 10 min incubation of PPE and the inhibitor prior to initiation of reaction. Therefore, while **7** was a poor inhibitor under these conditions, better inhibitory results may be observed after a longer incubation. Overall, the results suggests that the of inhibitory potency demonstrated by the disubstituted thiophene γ-lactams is not simply due to steric hinderance with the S_1 specificity pocket as the compound with the largest substituents **3** was the most potent.

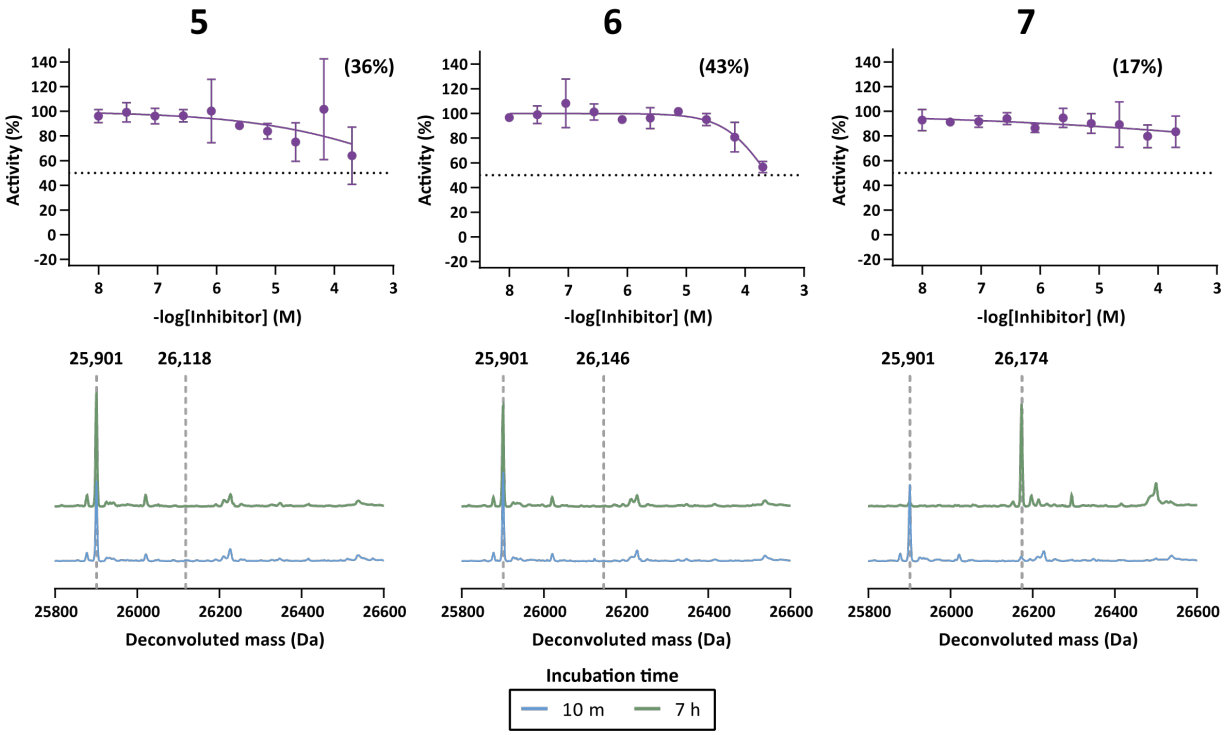


Figure 5.9 | Dose-response curves and protein LC-MS of PPE with the thiophene regioisomer 5 and bisalkylated thiophene γ-lactams 6–7. Top) Dose-response curves. Conditions: 3.86 nm PPE, 250 µM AAPA, inhibitor: 0.01–200 µM (final [DMSO] = 4% v/v), quenched after 2 min with formic acid (0.9% v/v), absorbance measured at 405 nm. Data are shown as the mean of four experimental repeats (n = 4; mean ± SD); **bottom)** deconvoluted protein LC-MS analysis after 10 min and 7 h incubation times. Conditions: 2.5 µM PPE, 100 µM inhibitor.

Compound **8** with a C4 methyl group demonstrated similar potency to the unsubstituted lactam **1** (Figure 5.10). **1** and **8** both reduced the activity of PPE by approximately 75% at 200 µM; however, **8** partially acylated PPE after 10 min as determined by LC-MS; after 7 h PPE was deacylated. Based on the crystal structure with **2**[331], the methyl group of **8** would likely be positioned shallowly in the S_1

pocket, which does not appear to provide any additional benefit over that unsubstituted lactam **1**, in terms of inhibitory potency. The lactams with C4 propyl **9** and isopropyl **10** substituents exhibited stronger inhibition (Figure 5.10). **9** was the second most potent inhibitor of PPE (pIC_{50} = 5.86; IC_{50} = 1.38 µM). **10** was 5-fold less potent than **9** (pIC_{50} = 5.13; IC_{50} = 7.41 µM). Both compounds fully acylated PPE after 10 min, but in contrast, **9** was fully deacylated after 7 h while **10** remained fully acylated, as observed by protein LC-MS. The S_1 pocket of PPE can accommodate isoleucine residues as was observed with the acyl-enzyme complex between PPE and BCM7 which has a P_1 Ile residue[106]. Thus, it was envisaged that inhibitors **9**, **10**, and **11** would be good inhibitors as the substituents would be well tolerated within the S_1 pocket.

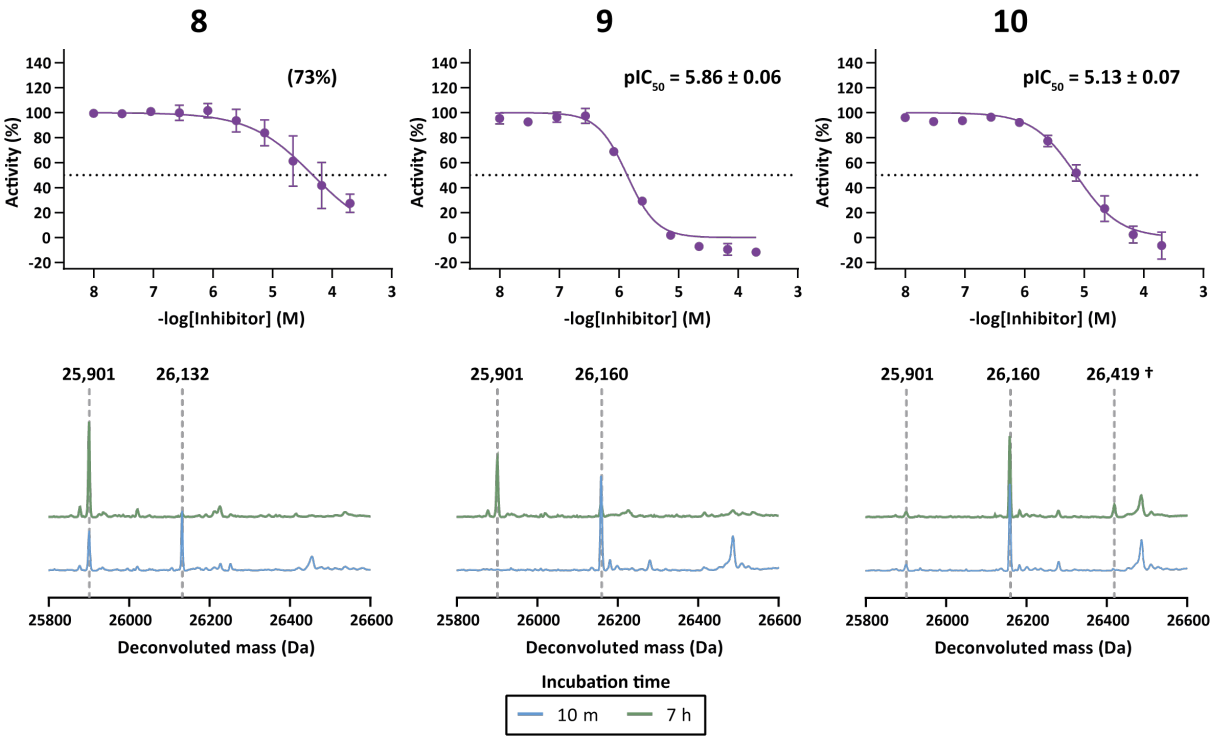


Figure 5.10 | Dose-response curves and protein LC-MS of PPE with monoalkylated thiophene γ-lactams 8–9. Top) Dose-response curves. Conditions: 3.86 nm PPE, 250 µM AAPA, inhibitor: 0.01–200 µM (final [DMSO] = 4% v/v), quenched after 2 min with formic acid (0.9% v/v), absorbance measured at 405 nm. Data are shown as the mean of four experimental repeats (n = 4; mean ± SD); **bottom)** deconvoluted protein LC-MS analysis after 10 min and 7 h incubation times. Conditions: 2.5 µM PPE, 100 µM inhibitor. † Corresponds to a peak that appeared after 7 h which is consistent with binding of a second inhibitor (+518 Da compared to the main apo-PPE).

The monoallylated lactam **11** was a potent inhibitor of PPE (pIC_{50} = 5.76; IC_{50} = 1.74 µM; Figure 5.11), although slightly weaker compared with the propyl analogue **9**. The protein LC-MS results with **11** were almost identical to those with **9**, i.e., full acylation of PPE after 10 min, but deacylated after 7 h. In

contrast, results with the propionitrile compound **12** were comparable to **1** and **8**; 74% inhibition at 200 µM, and no acylation of PPE as observed by protein LC-MS (Figure 5.11). The lactam with the benzyl group **13**, which mimics a phenylalanine sidechain, was envisaged to be a poorer inhibitor of PPE and was predominantly synthesised with aims of targeting SPs with CTR-like specificity. Surprisingly, **13** demonstrated good inhibition of PPE (pIC_{50} = 5.35; IC_{50} = 4.47 µM; Figure 5.11) and fully acylated PPE after 10 min, as determined by protein LC-MS. At 7 h, PPE was fully deacylated, analogous to **2**, **9**, and **11**. HNE has a slightly larger S_1 pocket[53] which may therefore accommodate compounds **12** and **13** better than PPE, but this remains to be determined. All pIC_{50} values are reported in Table 5.1.

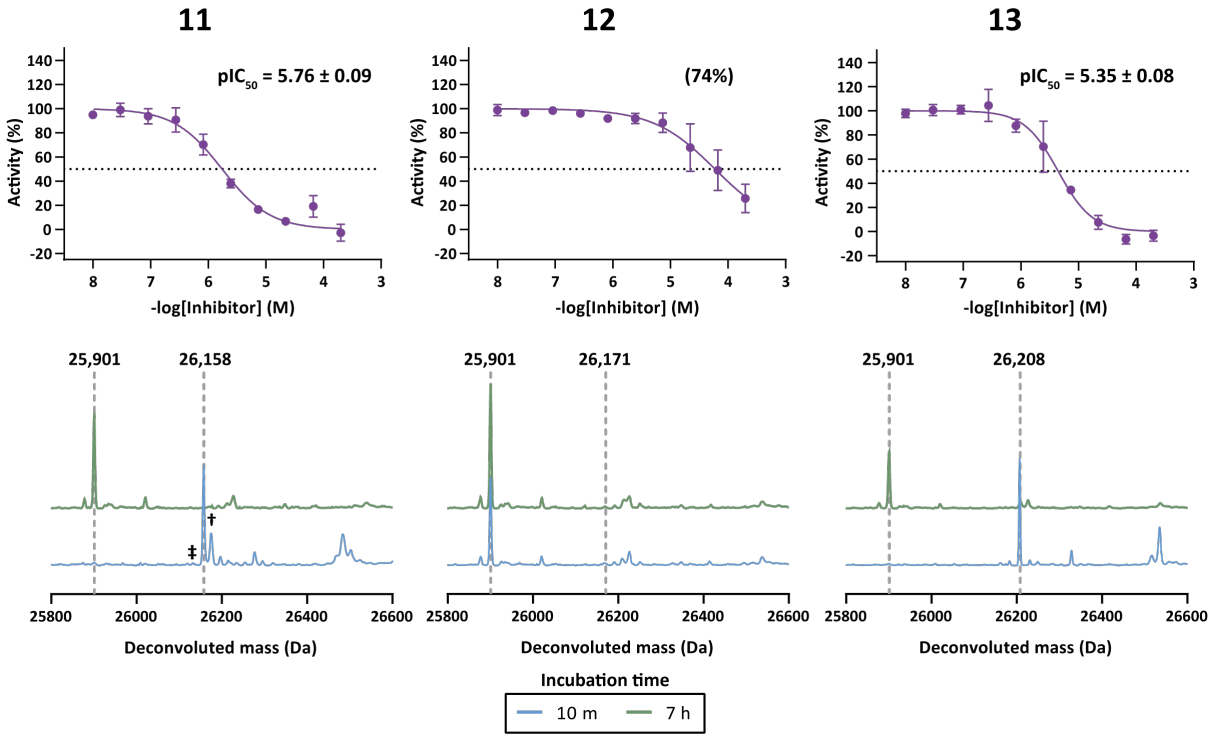


Figure 5.11 | Dose-response curves and protein LC-MS of PPE with monosubstitued thiophene γ-lactams 11–13. Top) Dose-response curves. Conditions: 3.86 nm PPE, 250 µM AAPA, inhibitor: 0.01–200 µM (final [DMSO] = 4% v/v), quenched after 2 min with formic acid (0.9% v/v), absorbance measured at 405 nm. Data are shown as the mean of four experimental repeats (n = 4; mean ± SD); **bottom)** deconvoluted protein LC-MS analysis after 10 min and 7 h incubation times. Conditions: 2.5 µM PPE, 100 µM inhibitor. † Corresponds to a peak that appeared after 10 min which is consistent with binding of a second inhibitor (+17 Da compared to the main PPE); ‡ corresponds to a missing peak that was present in other measurements, including after 7 h (-23 Da compared to the main PPE peak).

Table 5.1 | pIC_{50} and IC_{50} values of fused thiophene γ-lactams against PPE. Incomplete dose-response curves are reported as the % inhibition at 200 μM.

Inhibitor	pIC_{50} ± SD (% inhibition)	IC_{50} (μM)
1	(78)	–
2	5.30 ± 0.04	5.01
3	(69)	–
4	6.28 ± 0.13	0.52
5	(36)	–
6	(43)	–
7	(17)	–
8	(73)	–
9	5.86 ± 0.06	1.38
10	5.13 ± 0.07	7.41
11	5.76 ± 0.09	1.74
12	(74)	–
13	5.35 ± 0.08	4.47

5.1.2.3. Crystallographic Analyses of the Thiophene γ-Lactams

To rationalise the structure-activity relationship observed herein, PPE crystallisation trials were carried out in the presence of the thiophene γ-lactams. Most of the monosubstituted compounds (**1**, **2**, **4**, and **8**–**12**) were successfully soaked into PPE crystals and were consistent with formation of acyl-enzyme complexes. However, **13**, which demonstrated strong inhibition and full acylation of PPE by protein observed LC-MS, was not successfully soaked nor co-crystallised with PPE, while **1** and **12**, which did not acylate PPE as determined by protein LC-MS, were successfully soaked into PPE. None of the disubstituted lactams (**3**, **6**, and **7**), nor the regioisomer **5** produced additional density in the active site of PPE after soaking. Attempts to co-crystallise PPE with **3**, **6, 7**, and **13** were unsuccessful.

Representative views from the crystal structures of three PPE:inhibitor complexes are shown in Figure 5.12, A-C. In all the complex structures, the lactam ring was opened and the ester bond between S195 of PPE and the inhibitor superimposed well with that of the PPE:BCM7 complex with the productively bound peptide inhibitor (Figure 5.12, A & B)[106]. The carbonyl oxygens were orientated within the oxyanion pocket with O–HN distances of < 3.1 Å. Most of the inhibitors bound in an analogous manner, which is exemplified by compound **4** (Figure 5.12, A). However, in the structures of PPE with **10** and **11**, the inhibitor was predominantly bound in an alternate orientation compared

to that of **4**, with the thiophene ring rotated towards the face of H57 (Figure 5.12, B). This also reorientated the amide of the sulphonamide group. The remainder of the methanesulphonamide group remained in approximately the same position, except for one of the sulphate oxygen atoms which was rotated towards the position of the proposed hydrolytic water in the PPE:BCM7 complex and was within H-bonding distance of H57 (2.8 Å; Figure 5.12, B).

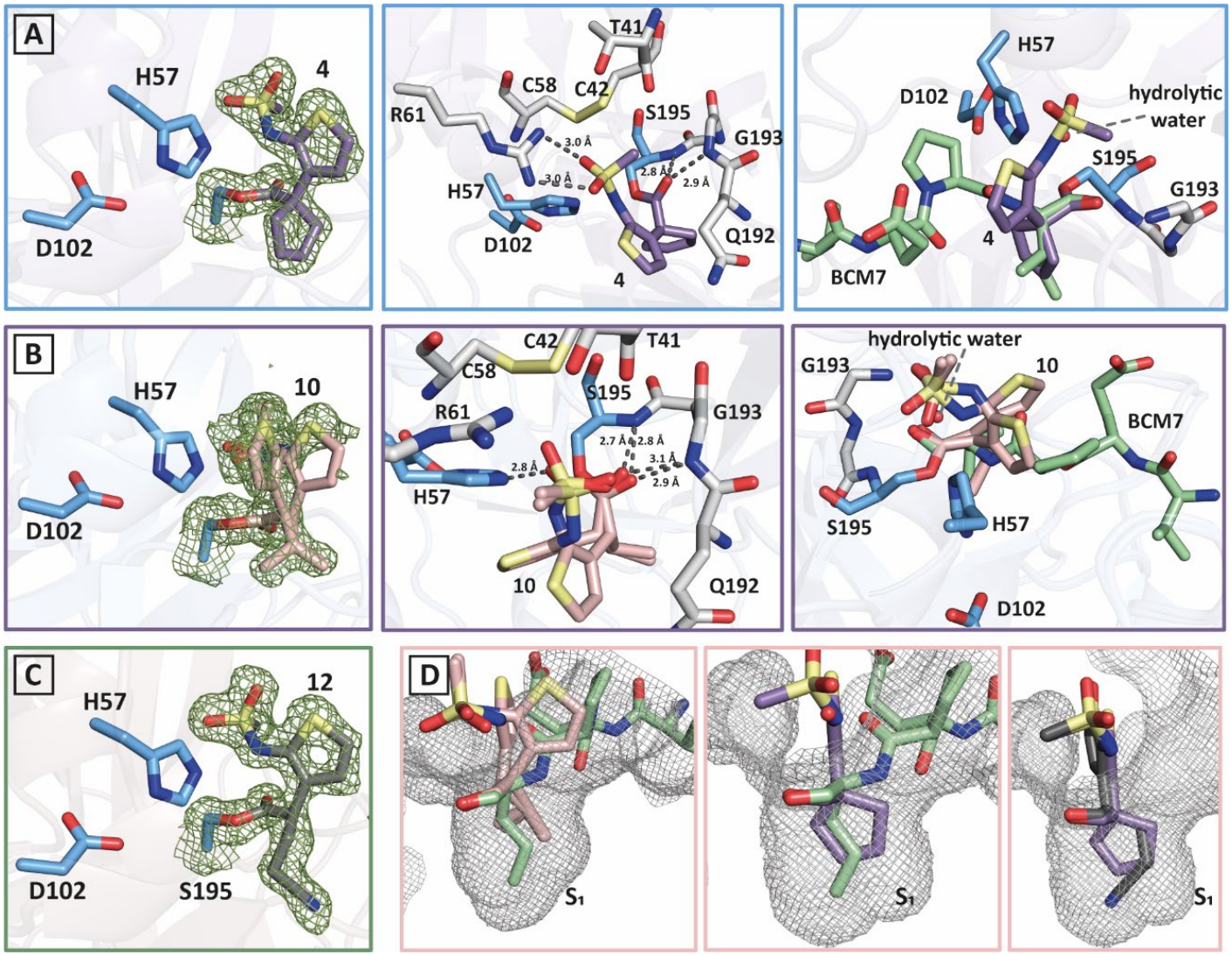


Figure 5.12 | Views from crystal structures of PPE soaked with thiophene γ-lactams. A) Active site views from a 1.40 Å resolution crystal structure of a PPE:**4** complex: a Polder omit map shown (green mesh) contoured to 3 σ (left), polar contacts between PPE and **4** (middle), and a superimposition of the complex with the PPE:BCM7 complex (Figure 2.5, D) with the proposed hydrolytic water (right); **B)** active site views from a 1.40 Å resolution crystal structure of a PPE:**10** complex: a Polder omit map shown (green mesh) and contoured to 3 σ (left), polar contacts between PPE and **4** (middle), and a superimposition of the complex with the PPE:BCM7 complex (Figure 2.5, D) with the proposed hydrolytic water (right); **C)** an active site view from a 1.40 Å resolution crystal structure of a PPE:**12** complex and a Polder omit map (green mesh) contoured to 3 σ; **D)** superimpositions of crystal structures of **A-C** showing the S_1 pocket surface (grey mesh): PPE:**10** with PPE:BCM7 (left), PPE:**4** with PPE:BCM7 (middle), and PPE:**4** with PPE:**12**. The catalytic triad carbons are light blue, all other PPE carbons are white, **4** carbons are purple, **10** carbons are pink, **12** carbons are black, BCM7 carbons are green, and the hydrolytic water is shown as a red sphere.

All the inhibitor C4 substituents were directed towards the S_1 pocket. The substituents of **4** and **10** fit snuggly into the centre of the S_1 pocket; however, superimposition with BCM7 demonstrated that there was more room to accommodate slightly larger substituents (Figure 5.12, D). Indeed, the largest C4 substituent captured *in crystallo* **12**, demonstrated that the C4 substituent can 'wrap' around the edges of pocket following a similar path to the ring of **4**. **12** only deviated from this path at the rigid nitrile group (Figure 5.12, D). While larger substituents may be accommodated within the S_1 pocket, it does not necessarily equate to improved potency, as **12** was a relatively weak inhibitor compared with compounds with smaller C4 substituents such as **4**. There was very little backbone conformational change of the active site, but the solvent exposed sidechains of Q192 and R61, which contribute to formation the active site, were modelled into several different orientations. These generally did not interact with the inhibitor, except perhaps hydrophobic interactions between the sidechain carbons of Q192 and the inhibitor. However, in the complex with **4**, the guanidino group of R61 was in position to interact with the methanesulphonamide oxygen atoms (heteroatom distances of 3.0 and 3.0 Å; Figure 5.12, A), although the electron density for the sidechain was weak and thus likely was not the sole conformation. However, it is possible that this interaction contributes to the stability of **4**, as it was not observed in this conformation in the other structures.

There were backbone conformational changes observed in several of the soaked inhibitor complexes with PPE. The changes are represented by the structures Figure 5.13. However, the residues affected by the conformational changes were not associated with the active site or binding of the inhibitor. As the changes were all surface loops these, may be representative of a crystallographic artefact but this may deserve investigation.

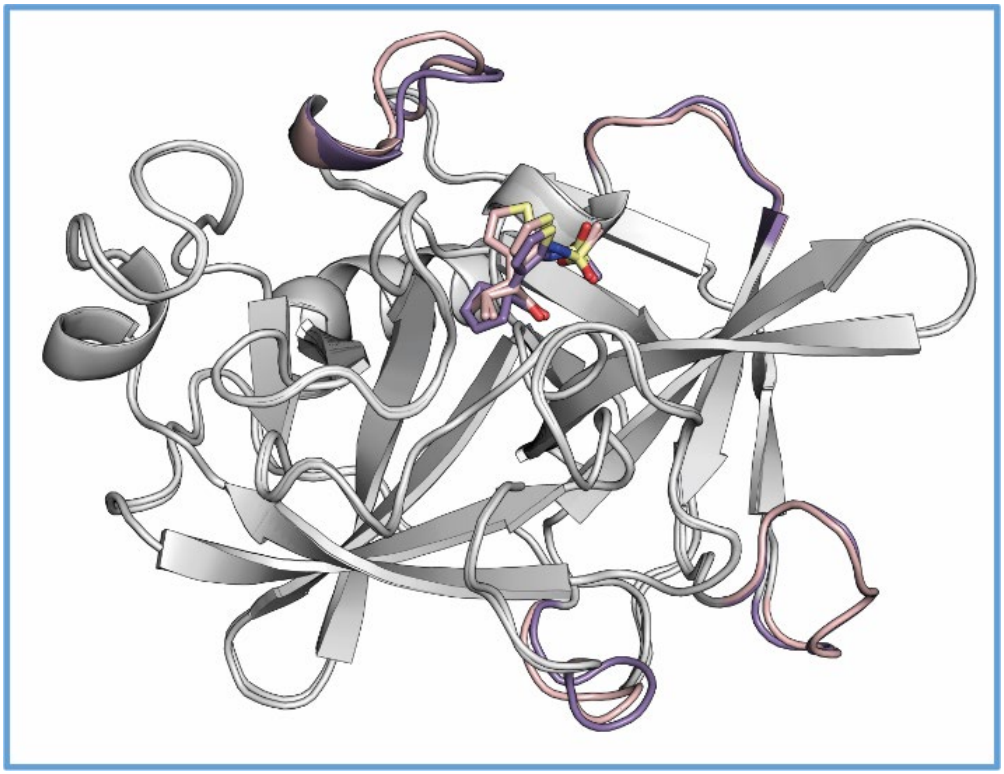


Figure 5.13 | Conformational changes observed after soaking with thiophene γ-lactams. The regions exhibiting conformational changes are shown in the colour corresponding to the inhibitor. The PPE:**4** complex is in purple, and the PPE:**10** complex is in pink.

5.1.2.4. C3 Substituted Oxindole γ-Lactams

A second bicyclic γ-lactam scaffold was also explored for the inhibitory potential against PPE. The thiophene and benzene rings are generally considered to be isosteric, due to the comparable size, reactivity, and aromaticity of the rings[342]. Thus, it was envisaged that the oxindole scaffold would exhibit similar reactivity towards PPE as observed with the thiophene scaffold. Four oxindole compounds were synthesised by Dr Lewis Ibbotson (unpublished data, Schofield group; Figure 5.14). All the oxindole compounds contained a methanesulphonamide substituent, analogous with the thiophene γ-lactam scaffold. **14** was unsubstituted at the C3 position and is the oxindole analogue of **1**. **15**–**17** were all derivatives of **14**, containing a spirocycle (**15**) or a disubstituted C3 (**16** and **17**).

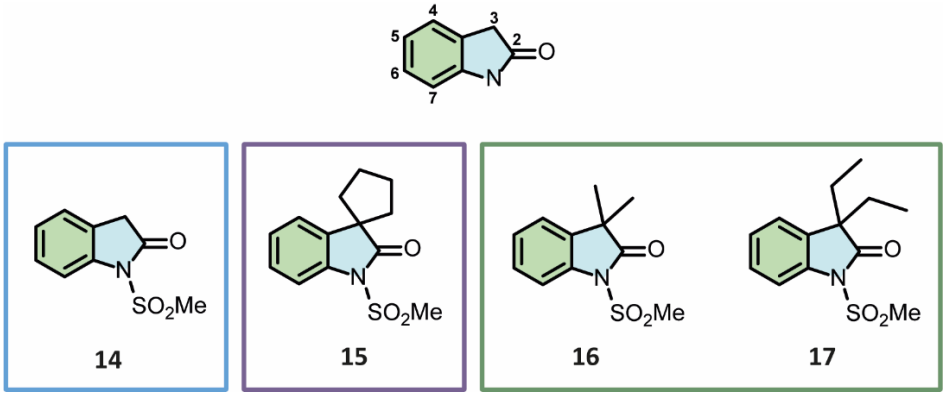


Figure 5.14 | Structures of oxindoles. Unsubstituted C3 (blue box), oxindole with a C3 spirocycle (purple box) bisalkylated oxindoles (green box). All oxindoles were synthesised by Dr. Lewis Ibbotson.

Dose-response curves against PPE were generated for all four compounds (Figure 5.15). **14**, with the unsubstituted C3 position, was the best oxindole inhibitor, reducing the activity of PPE by 46% at 200 µM. All three compounds with C3 substituents exhibited weak inhibition of PPE (< 30%). While this was somewhat expected for the disubstituted compounds, the spirocycle was envisaged to be a potent inhibitor, based on the thiophene γ-lactam results with **4**. Although the nature of the spirocycle substituents differ, i.e., cyclopentane **15** versus cyclopentene **4**, this was not envisaged to cause such a substantial decrease in potency.

To develop an understanding of why the oxindoles were poor inhibitors of PPE, protein LC-MS and crystallographic analyses were performed with each of the four compounds (Figure 5.15). None of the compounds were found to acylate PPE by protein LC-MS and none of the compounds were successfully soaked into PPE crystals nor co-crystallised with PPE, supporting the poor dose-response results.

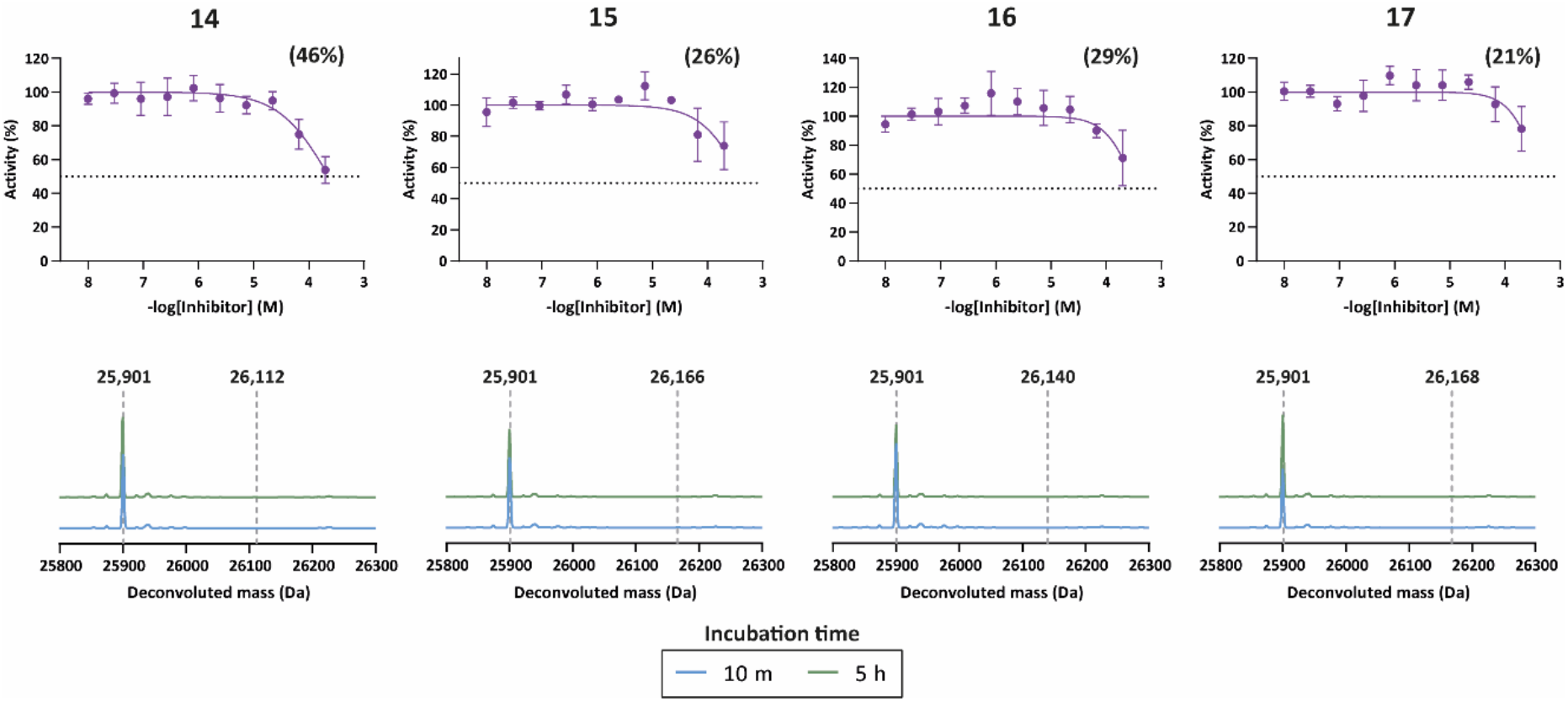


Figure 5.15 | Dose-response curves and protein LC-MS of PPE with oxindole γ-lactams. Top) Dose-response curves. Conditions: 3.86 nm PPE, 250 µM AAPA, inhibitor: 0.01–200 µM (final [DMSO] = 4% v/v), quenched after 2 min with formic acid (0.9% v/v), absorbance measured at 405 nm. Data are shown as the mean of four experimental repeats (n = 4; mean ± SD); **bottom)** deconvoluted protein LC-MS analysis after 10 min and 7 h incubation times. Conditions: 10 µM PPE, 400 µM inhibitor.

5.1.3. Discussion and Perspectives

To build upon the previous work of Migaud *et al*. and the 5,5-fused thiophene γ-lactam scaffold[331], several novel compounds with alternate C4-substiuents were characterised. Four of the compounds (**7**, **9**-**11**) demonstrated single digit micromolar activity towards PPE, although, none were more potent than the previously reported spirocycle **4**[331].

While PPE is a useful model for HNE activity, there are differences in specificities. PPE and HNE both preferentially bind hydrophobic residues, but HNE can accommodate larger sidechains, such as Leu and Met, whilst PPE preferentially binds to smaller sidechains (Ala, Val, and Ile)[53]. Thus, future experiments should be directed towards assessment of the series towards HNE and other SPs (e.g., thrombin, subtilisin, CTR, TR, etc.). While the spirocycle **4** was the most potent compound against PPE, the alternate substituents reported here may confer better specificity or selectivity, but this remains to be determined.

Interestingly, the thiophene regioisomer **5** exhibited reduced activity compared with **1**, indicating that positioning of the sulphur atom within the ring system is important for inhibitory activity. Similarly, all the oxindole inhibitors displayed poor activity against PPE and were not able to form acyl-enzyme complexes as determined by protein observed LC-MS. Instead, it suggests that the thiophene ring plays a role in activation of the lactam amide moiety, which the benzene ring is unable to replicate.

While thiophene and benzene are generally considered to be isosteric, there are differences between the rings: thiophene is less aromatic due to the higher electronegativity of sulphur compared with carbon, and the thiophene is comparatively electron rich (6 π e^- shared by 5 thiophene atoms Vs 6 π e^- shared by 5 benzene atoms). As thiophene is more electron rich it could be suggested that it would be less likely to withdraw electrons from the lactam amide, thus reducing reactivity; however, this was not observed empirically. The electronegativity of the sulphur may therefore be of importance. Replacement of the thiophene to furan or pyrrole (which are both more electronegative, and thus less

aromatic, heterocycles) may further activate the lactam moiety, but this requires investigation. Alternative functionalisation of the oxindole compounds could be directed towards better activation of the nitrogen, perhaps through the substitution of additional electron withdrawing groups to make the amide more electron deficient. For example, the methanesulphonamidyl group could be exchanged with more electronegative groups.

Currently, the reaction products of the thiophene and oxindole compounds explored here have not been monitored by NMR or small molecule LC-MS. Thus, the mechanism by which PPE was reactivated after longer incubations with **1**, **8**, and **12** is unclear. The hydrolytic water was displaced in each of the PPE:thiophene γ-lactam crystal structures collected (Figure 5.8). Therefore, it is more likely that reactivation proceeds via reformation of the lactam ring, although slow hydrolysis of the acyl-enzyme complex is also possible.

The bicyclic γ-lactam scaffold may be of interest for time-resolved crystallographic studies. Here, short crystal soaking procedures were performed and the formation of acyl-enzyme complexes between PPE and the thiophene γ-lactams (**4**, **10**, **12**) *in crystallo* occurred within minutes (see Appendix C for details). However, as observed with Sivelestat, this is likely too long for time-resolved studies (Section 3.2.2). The concentration of inhibitor used for soaking the thiophene γ-lactams was low compared with Sivelestat (< 10 mM Vs 20 mM) due to poor solubility of the γ-lactams in water. The combination of smaller crystals and soaking at higher concentrations of both inhibitor and DMSO may reduce the mixing time, although this would prohibit the use of the DoC methodology due to DMSO leakage. Miguad *et al.* utilised a longer soaking method whereby the crystals were soaked in saturated solution of **2** for 24 h. This method was sufficient to form the acyl-enzyme complex between PPE and **2**, although the precise crystallisation conditions were not reported. Complexes of PPE and thiophene γ-lactam inhibitors formed in this manner could be utilised for pH-jump studies analogous to those reported with the monocyclic γ-latams[170,343]. However, the issues discussed in Section 2.3 would also likely persist using these compounds.

5.2. Part 2: Benzoxaboroles

5.2.1. Introduction

In an analogous manner to the lactam-based inhibitors, boronic acid compounds have found utility against enzyme classes that employ a catalytic nucleophile. Due to the vacant *p* orbital of boronic acids, which confers Lewis acidity, the compounds readily form reversible tetracoordinate complexes with nucleophilic serine residues (Figure 5.16, A)[268]. As such, the resultant complexes often mimic a 'tetrahedral intermediate', although other binding modes are possible. For example, a structure of a peptidyl boronic acid inhibitor in complex with PPE more closely resembled that of an acyl-enzyme complex, where, it was proposed that H57 donated a proton to one of the boron hydroxy groups leading to the formation of water (Figure 5.16, B)[344]. Hence, boron was observed in a trigonal planar geometry. A wide range of boronic acid-containing compounds have been developed to target enzymes from each of the peptidase classes[345].

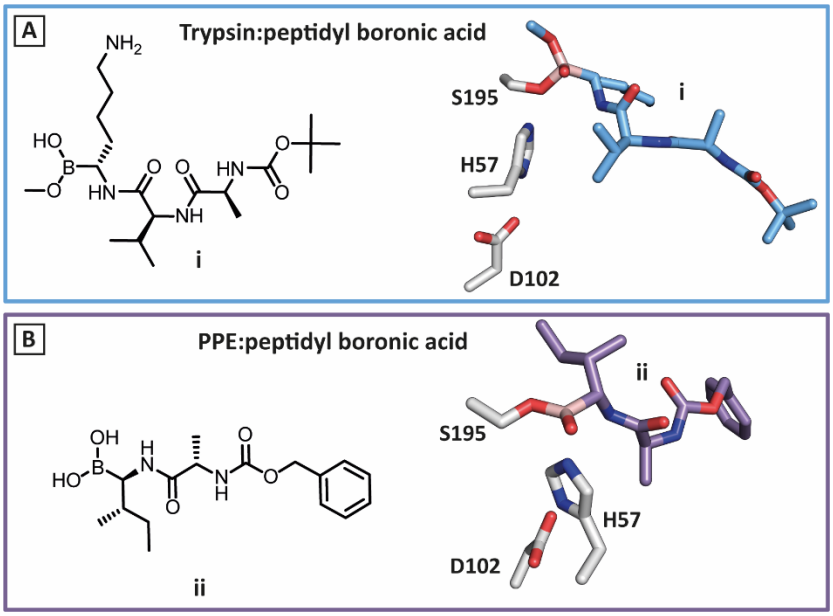


Figure 5.16 | Views from reported crystal structures of serine proteases complexed with peptidyl boronic acid inhibitors. A) A tetrahedral TR:peptidyl boronic acid complex (PDB: 1BTZ) and chemical structure of the peptidyl boronic acid inhibitor (i)[268]; **B)** a trigonal planar PPE:peptidyl boronic acid complex (PDB: 5EST) and the chemical structure of the peptidyl boronic acid inhibitor (ii)[344]. The enzyme carbons are white, the inhibitor carbons are light blue for the tetrahedral complex and purple for the trigonal planar complex, and boron atoms are pink.

A derivative of phenylboronic acid, benzoxaborole, was first prepared in 1957[346]. However, interest in this scaffold did not truly emerge until the potent antifungal activity of AN2690 (tavaborole) was reported in 2006 and 2007 (Figure 5.17, A)[347,348]. Since then, benzoxaboroles with antibacterial[349-352], antiviral[353,354], anti-inflammatory[355], and antiprotozoal[356,357] activity have been reported (Figure 5.17, B-F). Recently, potent inhibitors of both cysteine (SARS-CoV-2 main protease (M^{Pro})) and serine (dengue virus NS2B and NS3) peptidases with (sub-)micromolar potency have been developed, demonstrating the potential and the continued interest in this scaffold[358]. As with the boronic acids, benzoxaboroles can form tetrahedral complexes. Crystal structures of benzoxaboroles with both PBP3 (*Pseudomonas aeruginosa*)[351], carbapenemase OXA-48 (Class D β-lactamase; *Klebsiella pneumoniae*)[352], and AmpC (Class C β-lactamase; *P. aeruginosa*) have confirmed the formation of a tetrahedral complex between the boron and nucleophilic serine residues (Figure 5.17, B & C). While potent boronic acids inhibitors have been developed for CTR-related SPs (including HNE[177], and TR[268]), inhibition of elastase-like SPs by benzoxaboroles has not been demonstrated. Furthermore, the unsubstituted benzoxaborole scaffold has been reported to be a poor inhibitor of HNE (IC_{50} > 100 μM)[359]. Molecular docking experiments from the same report suggested that the unsubstituted benzoxaborole did not form a complex with S195. While the benzoxaborole was located at the surface of the S_1 specificity pocket of HNE, the oxaborole moiety was oriented away from S195, rationalising the poor inhibition[359].

5.2.1.1. Aims

The aims of the work presented in the following section were to investigate the inhibitory potential of a series of C6-substituted benzoxaboroles against PPE. Additionally, crystallographic analyses of several benzoxaborole compounds were undertaken to assess whether the binding mode corroborated the molecular docking studies[359].

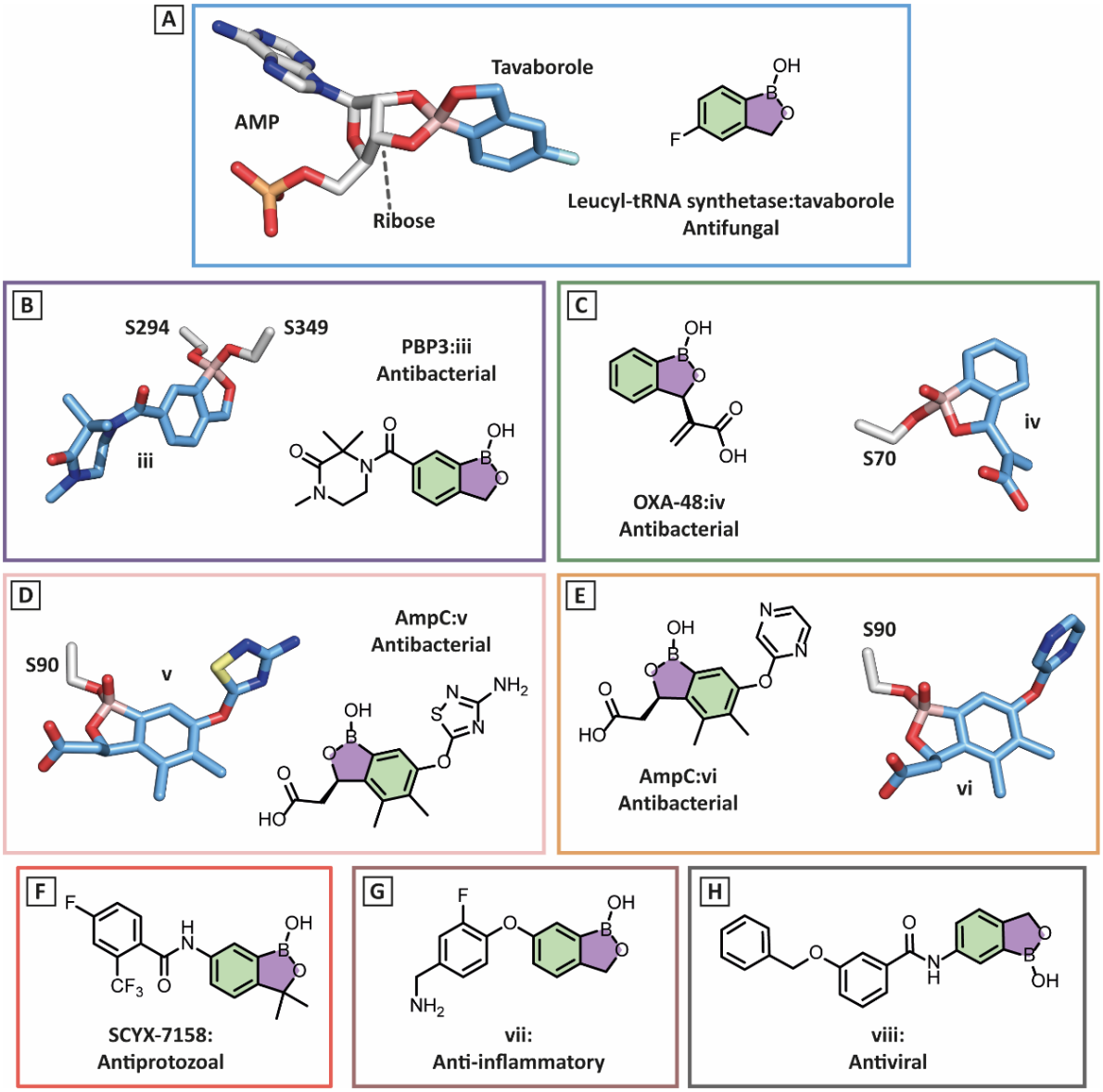


Figure 5.17 | Views from crystal structures of enzyme:benzoxaborole complexes and the chemical structures of reported benzoxaboroles. A) A tetrahedral complex between tavaborole and AMP which is also bound to a leucyl tRNA synthase (*Thermus thermophilus*; PDB: 2V0C)[348]; **B)** a tetrahedral complex between a C6-substituted benzoxaborole (**iii**) and PBP3 (*P. aeruginosa*; PDB: 7ATX)[351]; **C)** a tetrahedral complex between a C3-substituted benzoxaborole (**iv**) and OXA-48 (*K. pneumoniae*; PDB: 7DML) [352]; **D)** a tetrahedral complex between a benzoxaborole (**v**) and AmpC (*P. aeruginosa*; PDB: 4WZ4)[350]; **E)** a tetrahedral complex between a different benzoxaborole (**vi**) and AmpC (*P. aeruginosa*; PDB: 4WYY)[350]; **A-E)** enzyme and AMP carbons are white, inhibitor carbons are light blue, and boron atoms are pink; **F)** the chemical structure of an antiprotozoal C6-carboxamide benzoxaborole (SCYX-7158) that targets *Trypanosoma brucei* (human African trypanosomiasis (HAT))[356]; **G)** the chemical structure of an anti-inflammatory benzoxaborole (**vii**) that targets Rho-activated kinase (ROCK2)[355]; **H**) the chemical structure of a bezoxaborole compound (**vii**) that targets SARS-CoV-2 main protease (M^{Pro}) and dengue virus NS2B/NS3 proteases[358].

5.2.2. Results

5.2.2.1. Inhibition with C6-Substituted Benzoxaboroles

A series of C6-substituted benzoxaboroles (**18**-**45**) were synthesised by Gayatri (unpublished data; Figure 5.18). The inhibitors have been placed into six groups based on the C6-substituent for ease of discussion. Group A (purple box) did not contain a sulphonamide. The C6 position of compounds from group B (blue box), group C (green box), and group D (pink box) was functionalised with a sulphonamide group. The sulphur atoms of the sulphonamide were functionalised with aliphatic groups (group C), alicyclic groups (group C), or aromatics (group D). Groups E (orange box) and F (red box) were functionalised with a sulphamide group. The amines of the sulphamide group were functionalised with either aliphatic groups (group E) or heterocycles (group F).

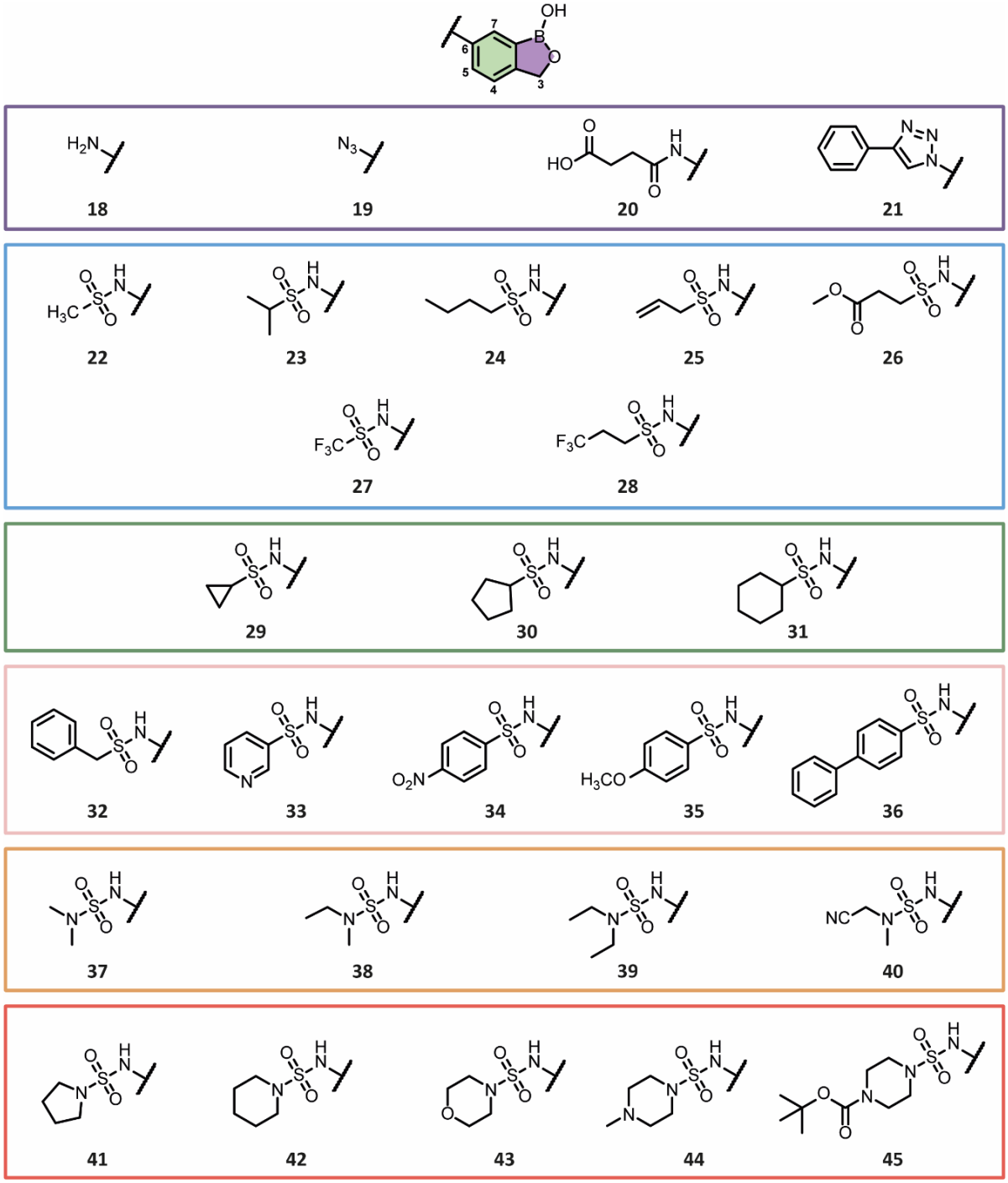


Figure 5.18 | Chemical structures of C6-substituted benzoxaborole inhibitors. Group A (purple box), group B (blue box), group C (green box), group D (pink box), group E (orange box), and group F (red box). All benzoxaboroles were synthesised by Gayatri.

Dose-response curves for PPE were generated for all 28 inhibitors using the absorbance assay described in Section 5.1.2.1. Of concern, several compounds absorbed at 405 nm, complicating analysis. Compounds **21** and **36**, exhibited particularly large increases in absorbance value (reported as % activity) at higher concentrations of the respective inhibitors. Both contained conjugated double aromatic ring substituents which may have caused the observed absorbance increases. Smaller absorbance increases at 67 μM and 200 μM of inhibitor may have also occurred with other compounds (**27**, **28**, **30**, **32**, **34**, **35**, and **40**), although they were generally within the error and typically did not contain conjugated substituents, except for the benzyl ring of the benzoxaborole (**27**, **28**, and **30**). To investigate, the absorbance values of the previously mentioned inhibitors were measured in buffer, without enzyme or substrate, and again after addition of formic acid (Figure 5.19). All the inhibitors tested demonstrated some absorbance increase at higher concentrations of inhibitor which could have affected the apparent activity of PPE during dose-response assessment. Compounds **21** and **36** demonstrated large absorbance increases, but only after addition of formic acid. Upon visual inspection of the wells containing **21** and **36**, it was apparent that addition of formic acid led to precipitation, rationalising the substantial increase in absorbance.

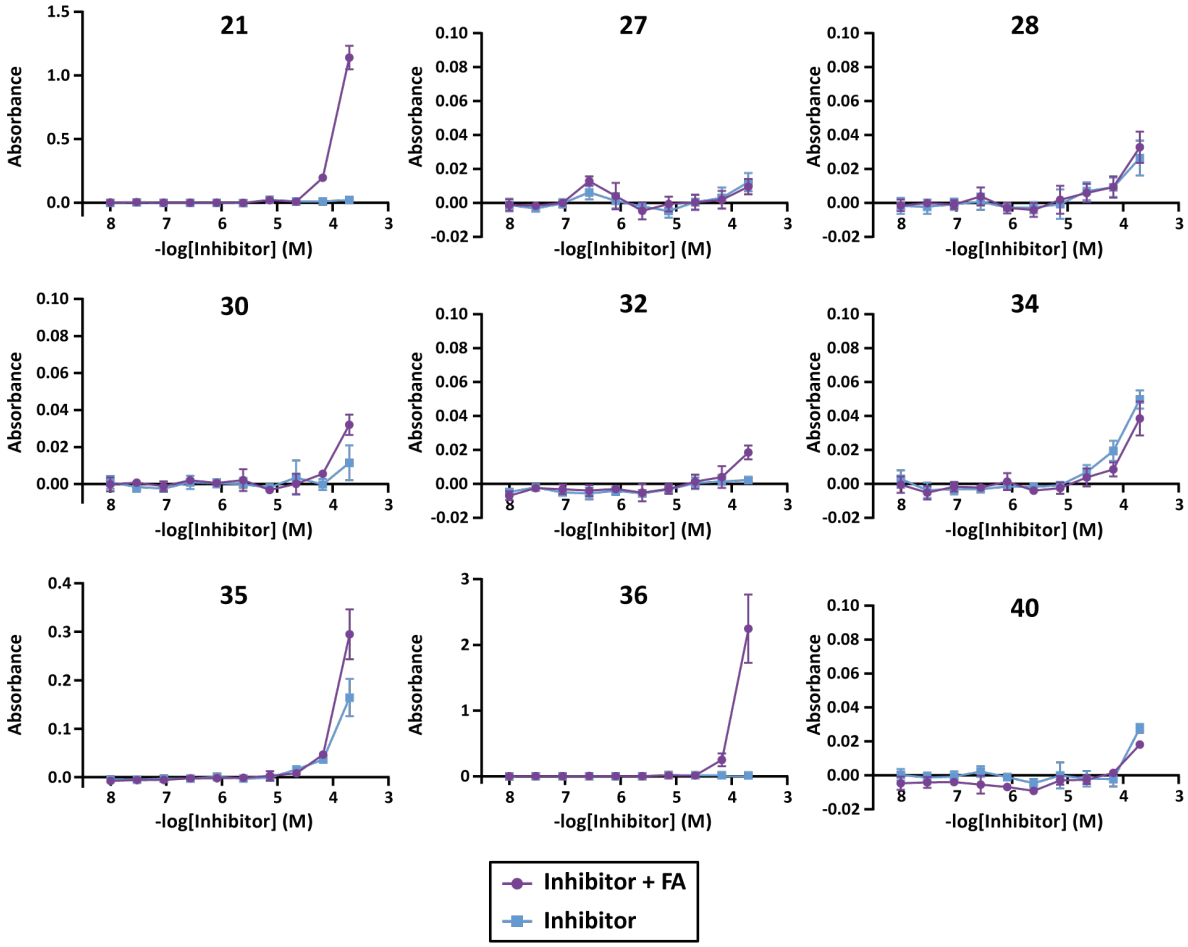


Figure 5.19 | Inhibitor absorbance. Inhibitors were diluted to the appropriate concentration in buffer (100 mM Tris pH 7.5) to a final volume of 50 µL with 4% v/v DMSO, and the absorbance measured at 405 nm. 5 µL of formic acid (10% v/v) was added to each well and the absorbance measured again.

Compounds from group A were generally poor inhibitors of PPE (Figure 5.20). Only **19** and **20** reduced the activity of PPE by more than 10% at 200 µM, and none reduced the activity by more than 40%. As mentioned above, at high concentrations of **21** (67 and 200 µM), the absorbance values increased by an order of magnitude over the positive control. However, no decrease in activity was observed at lower concentrations of inhibitor.

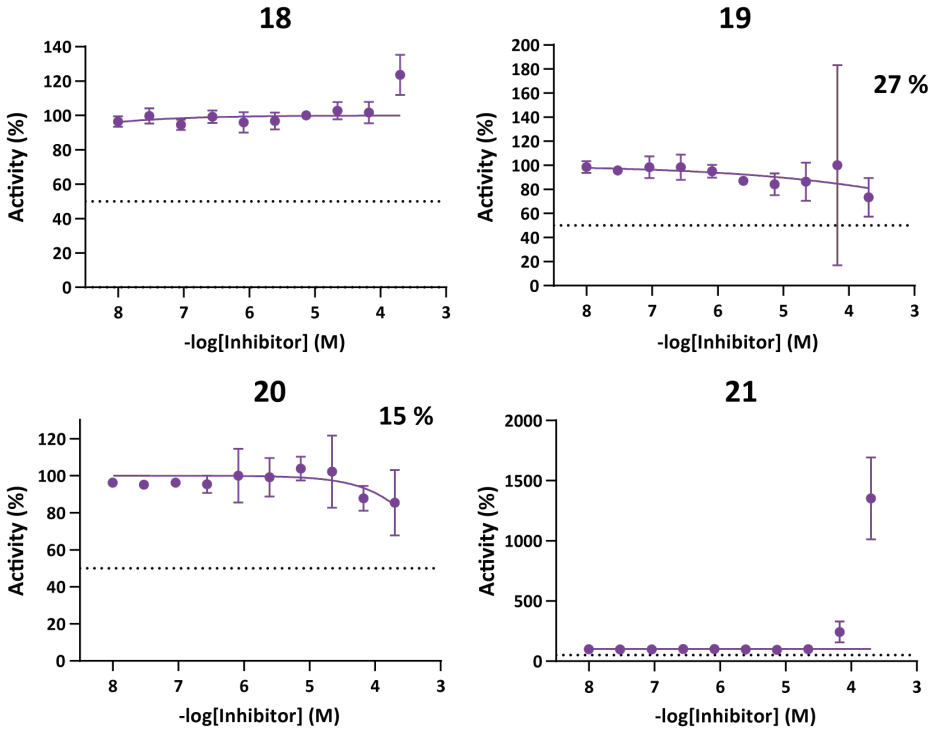


Figure 5.20 | PPE inhibition dose-response curves for group A benzoxaboroles. Conditions: 3.86 nm PPE, 250 µM AAPA, inhibitor: 0.01–200 µM, quenched after two minutes with formic acid (0.9%), absorbance measured at 405 nm. Data are shown as the mean of four experimental repeats (n = 4; mean ± SD).

Inhibitors from group B were better inhibitors of PPE compared with group A, but were still weak inhibitors (< 45 % inhibition at 200 µM; Figure 5.21). While six of the seven compounds exhibited more than 10% inhibition at 200 µM, only **23** and **26**, reduced the activity of PPE by more than 40% at 200 µM. The isopropyl *S*–substituted compound **23** reduced the activity of PPE by 44% at 200 µM, while **28** which has a methyl propionate *S*–substituent, reduced the activity by 40% at 200 µM. **27** which had a trifluoromethyl group exhibited weak inhibition of PPE (16 %) at 22 µM. However, at 67 µM and 200 µM, the observed activity increased due to potential absorbance by the inhibitor.

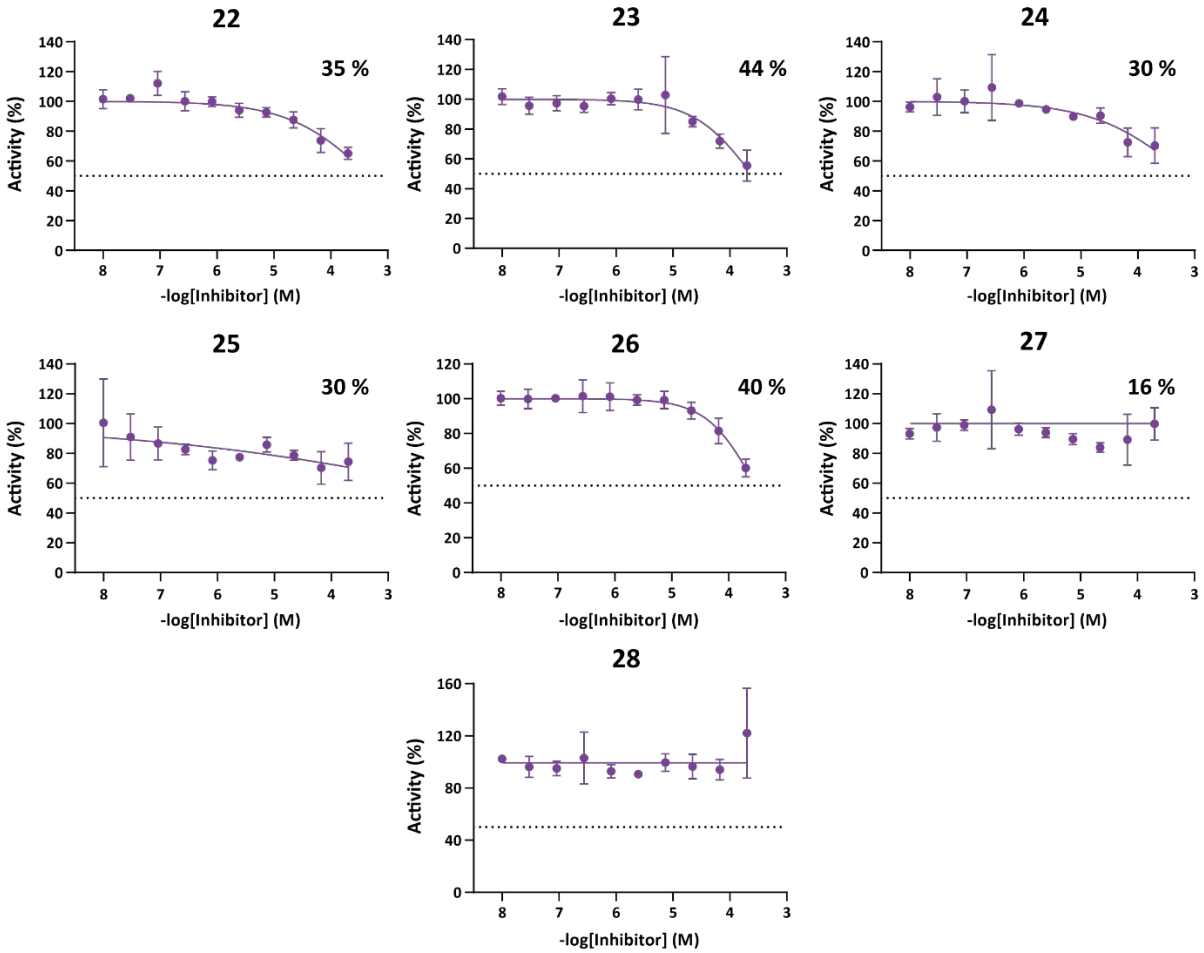


Figure 5.21 | PPE inhibition dose-response curves for group B benzoxaboroles. Conditions: 3.86 nm PPE, 250 µM AAPA, inhibitor: 0.01–200 µM, quenched after two minutes with formic acid (0.9%), absorbance measured at 405 nm. Data are shown as the mean of four experimental repeats (n = 4; mean ± SD).

The potency of group C inhibitors was generally comparable to that of group B (Figure 5.22). Two of the three inhibitors demonstrated some inhibitory potency (> 10% reduction in activity), but only **29** (cyclopropyl *S*–substituent) reduced the activity of PPE by more than 40% at 200 µM. The cyclopropyl group is similar to the isopropyl group of compound **23**, suggesting that the small hydrophobic *S*–substituent in this position is tolerated by PPE.

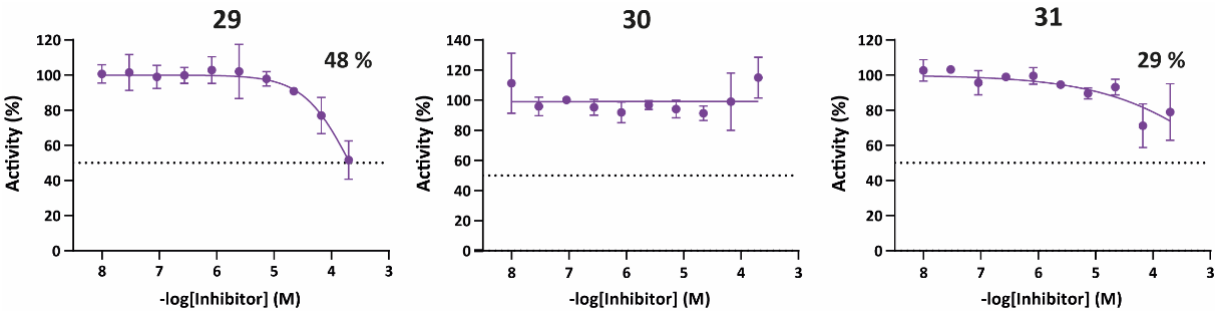


Figure 5.22 | PPE inhibition dose-response curves for group C benzoxaboroles. Conditions: 3.86 nm PPE, 250 µM AAPA, inhibitor: 0.01–200 µM, quenched after two minutes with formic acid (0.9%), absorbance measured at 405 nm. Data are shown as the mean of four experimental repeats (n = 4; mean ± SD).

The aromatic sulphonamide compounds of group D were weaker inhibitors of PPE compared with group B and C (Figure 5.23). However, four of the five compounds (**32**, **34**, **35**, and **36**) were affected by the apparent absorbance interference at 67–200 µM inhibitor. In particular, the apparent activity of **36** rose an order of magnitude over the positive control while that of **35** increased 2-fold. Four of the five compounds reduced the activity of PPE by more the 10% at any concentration but none reduced the activity by more than 40%. Of note, compound **36** reduced the activity of PPE by 31% at 22 µM, but large absorbance increases over the positive control may mask the inhibition at higher concentrations. **36** contained one of the largest substituents tested which suggests that large hydrophobic groups are equally well tolerated as small hydrophobic groups of **23** and **29**.

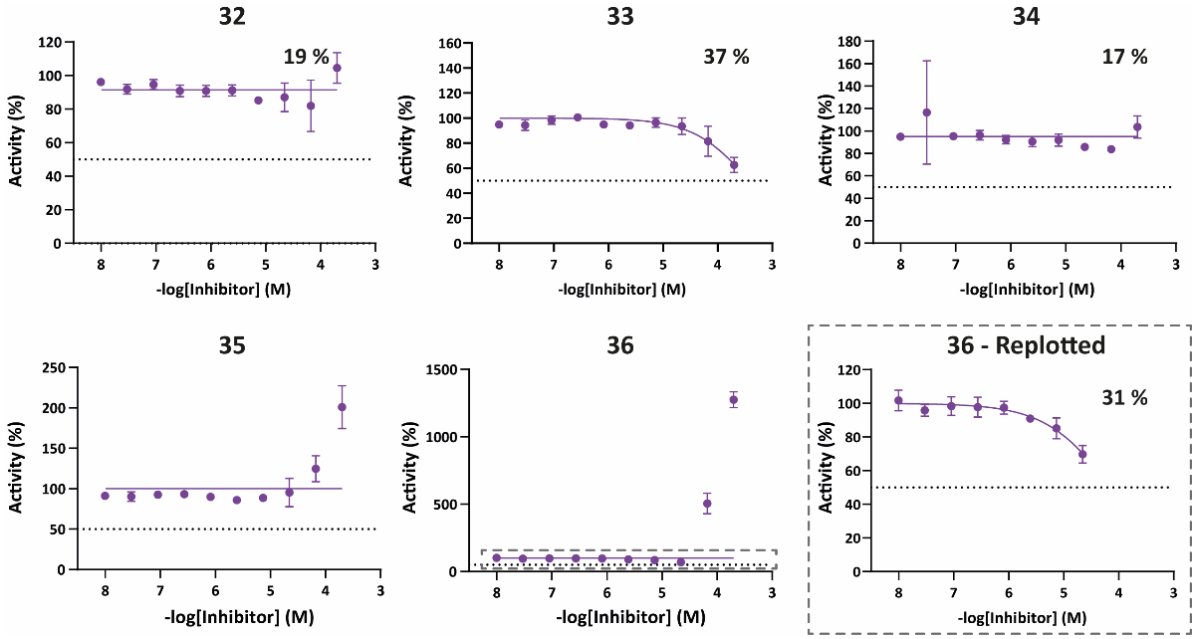


Figure 5.23 | PPE inhibition dose-response curves for group D benzoxaboroles. Conditions: 3.86 nm PPE, 250 µM AAPA, inhibitor: 0.01–200 µM, quenched after two minutes with formic acid (0.9%), absorbance measured at 405 nm. Data are shown as the mean of four experimental repeats (n = 4; mean ± SD).

Compounds from group E (**37**–**40**) were all of a similar chemical structure with only minor differences in the *N*-substituted chain lengths, except **40** which contained a nitrile group (Figure 5.24). Unsurprisingly, all demonstrated a similar degree of inhibition against PPE. Of the group, compound **38** (*N*–ethyl *N*–methyl sulphamide C6–substituent) demonstrated the strongest inhibitory effects (47% at 200 µM).

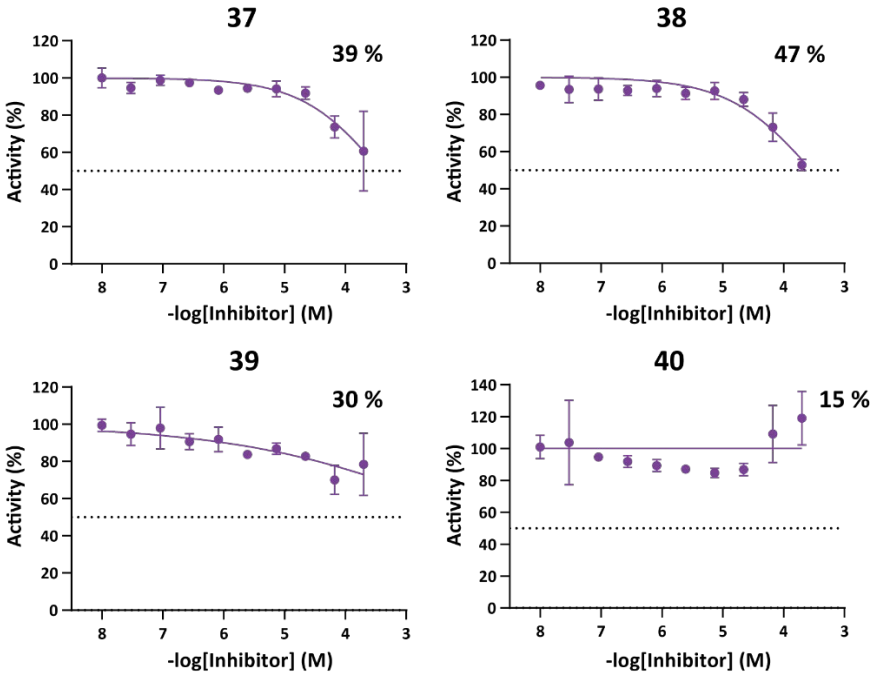


Figure 5.24 | PPE inhibition dose-response curves for group E benzoxaboroles. Conditions: 3.86 nm PPE, 250 µM AAPA, inhibitor: 0.01–200 µM, quenched after two minutes with formic acid (0.9%), absorbance measured at 405 nm. Data are shown as the mean of four experimental repeats (n = 4; mean ± SD).

The compounds in Group F, which all contained a heterocycle and sulphamide, were relatively weak inhibitors of PPE (Figure 5.25). While all exhibited some degree of inhibition against PPE, three of the five reduced the activity of PPE by less than 20%. Only compound **45** with a *tert*-butyloxycarbonyl (BOC) group reduced the activity of PPE by more than 40% (41% at 200 µM). This was the largest substituent explored and supports the result from compound **36** (group D) that large C6 substituents can be tolerated. However, compound **45** is considerably less lipophilic than compound **36**. The results of the dose-response curves for all benzoxaboroles are reported in Table 5.2. Overall, the inhibitory results with the benzoxaborole compounds were poor with no reduction of more than 50% observed with any C6-substituent explored here.

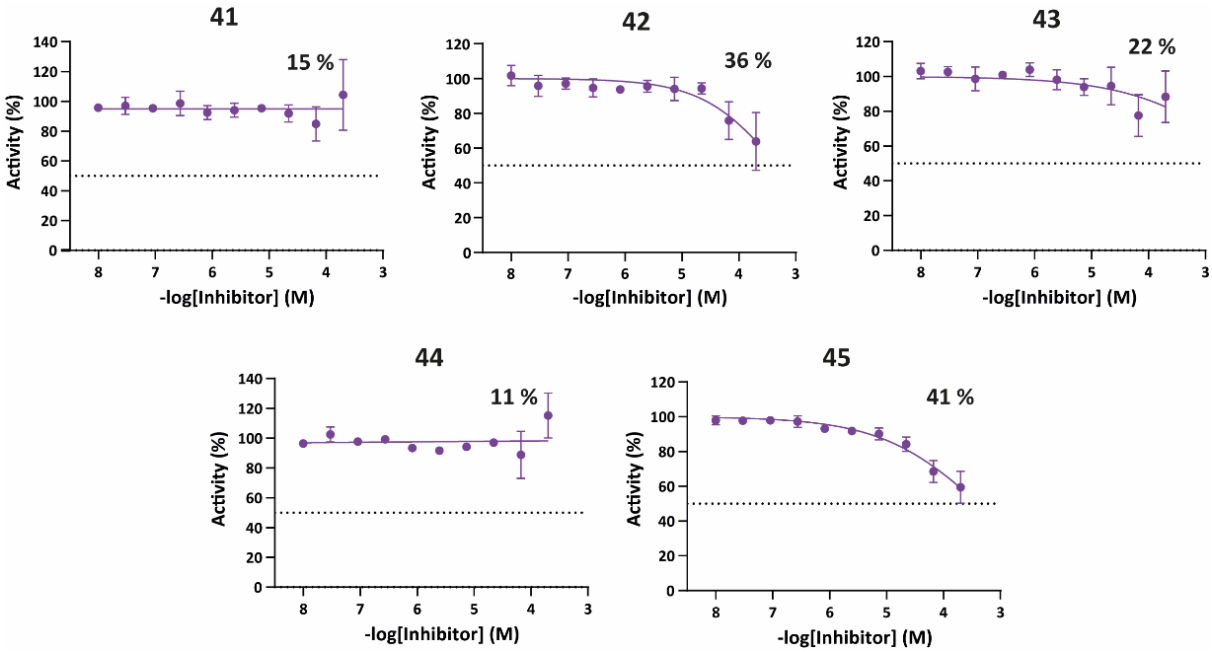


Figure 5.25 | PPE inhibition dose-response curves for group F benzoxaboroles. Conditions: 3.86 nm PPE, 250 µM AAPA, inhibitor: 0.01–200 µM, quenched after two minutes with formic acid (0.9%), absorbance measured at 405 nm. Data are shown as the mean of four experimental repeats (n = 4; mean ± SD).

Table 5.2 | Tabulated PPE inhibition values for C6-substituted benzoxaborole compounds. Values are reported as the mean % inhibition from three or four experimental replicates (n = 4 ± SD). Due to potential absorbance from the compounds the maximum inhibition values are reported with the concentration of inhibitor at which the maximum inhibition was produced. No inhibition = < 10% inhibition.

Group	Compound	% Inhibition	[Inhibitor] (µM)	Group	Compound	% Inhibition	[Inhibitor] (µM)
A	**18**	No Inhibition		D	**32**	19 (± 15)	67
	19	27 (± 16)	200		**33**	37 (± 6)	200
	20	15 (± 18)	200		**34**	17 (± 3)	67
	21	No Inhibition			**35**	No Inhibition	
B	**22**	35 (± 4)	200		**36**	31 (± 5)	22
	23	44 (± 10)	200	E	**37**	39 (± 21)	200
	24	30 (± 12)	200		**38**	47 (± 3)	200
	25	30 (± 11)	67		**39**	30 (± 8)	67
	26	40 (± 5)	200		**40**	15 (± 3)	7.4
	27	16 (± 3)	22	F	**41**	15 (± 11)	67
	28	No Inhibition			**42**	36 (± 17)	200
C	**29**	48 (± 11)	200		**43**	22 (± 12)	67
	30	No Inhibition			**44**	11 (± 15)	67
	31	29 (± 12)	67		**45**	41 (± 9)	200

Subsequently, protein LC-MS analyses with PPE and three of the most potent inhibitors (**23**, **26**, and **29**) were performed. No evidence for covalent modification of PPE with any of the benzoxaborole compounds was observed by protein LC-MS; note that PPE is likely denatured under the LC-MS conditions which may therefore favour dissociation of the complex. However, **23**, **26**, and **29** were successfully soaked into PPE crystals at pH 5.0 (Figure 5.26). The three compounds were observed to bind analogously to one another, with the boron atom associated with S195 of PPE. The geometry of the boron atom in all three structures was tetrahedral, which is indicative of bond formation between S195 and the boron atom (Figure 5.26, A, D, & F). The bicyclic rings of the compounds were positioned at the surface of the S_1 specificity pocket of PPE (Figure 5.26, C). The oxygen atoms of the oxaborole rings were oriented into the oxyanion hole in all three structures and were in position to H-bond with the oxyanion hole amides (Figure 5.26, B, E, & G). As an example, the heteroatom distances of 2.6 and 3.2 Å to S195 and G193, respectively in the PPE:**26** complex (Figure 5.26, B). The sulphur atom of the C6 substituent was orientated into the S_2 pocket in an analogous position to the proline amide of BCM7 (Figure 5.26, C)[106]. One of the sulphur oxygen atoms indirectly interacted with the carbonyl oxygen of V216 via a bridging water (Figure 5.26, B, E, & G). The carbonyl oxygen of the methyl propionate *S*–substituent of **26** also formed an indirect interaction with the carbonyl oxygen of T96 via a bridging water (Figure 5.26, B). The iso- and cyclo-propyl groups of compounds **29** and **23** were orientated towards H57 (Figure 5.26, E & G). The closest contacts were 3.4 Å for the cyclopropyl group to the Cδ1 of His57 and 3.6 Å for the isopropyl group to Cε2 of His 57. Thus, the slightly different orientations adopted by the *S*–substituents explain why a variety of substituents were tolerated in the C6 position.

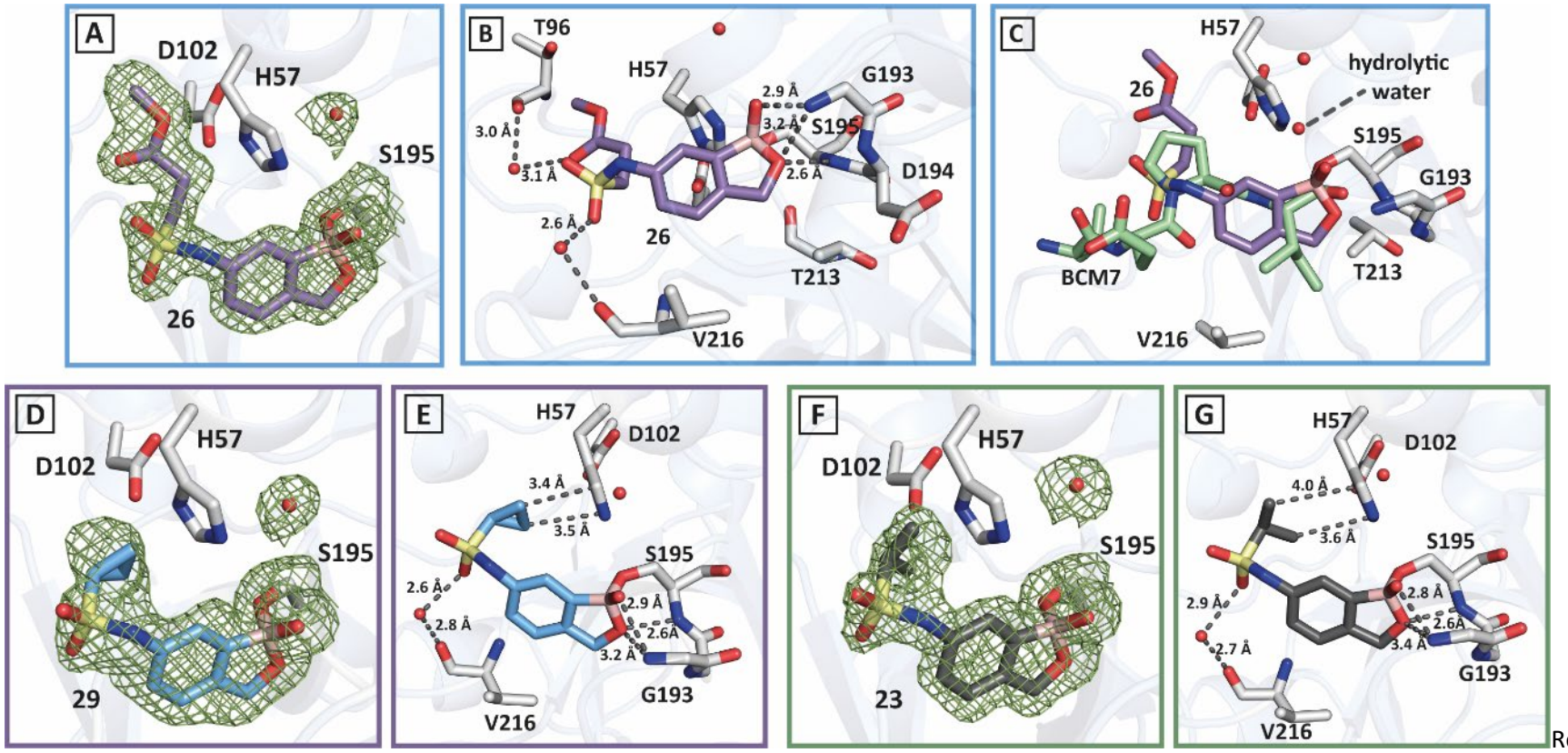


Figure 5.26 | Crystal structures of PPE soaked with bicyclic boronic acid inhibitors. A-C) Active site views from a 1.50 Å resolution crystal structure of the PPE:**26** complex; **A)** the complex with a Polder omit map shown as green mesh contoured to 3 σ; **B)** polar contacts between PPE and **26**; **C)** superimposition of the PPE:**26** complex with the PPE:BCM7 complex; **D & E)** active site views from a 1.50 Å resolution crystal structure of the PPE:**29** complex; **D)** the complex with a Polder omit map shown as green mesh contoured to 3 σ; **E)** polar and hydrophobic contacts between PPE and **29**; **F & G)** active site views from a 1.45 Å resolution crystal structure of the PPE:**23** complex; **F)** the complex with a Polder omit map shown as green mesh contoured to 3 σ; **G)** polar and hydrophobic contacts between PPE and **29**. PPE carbons are white, **26** carbons are purple, **29** carbons are light blue, **23** carbons are black, BCM7 carbons are green, boron atoms are pink, and waters are shown as red spheres.

A proposed reaction mechanism for PPE and the benzoxaborole compounds is shown in Figure 5.27.

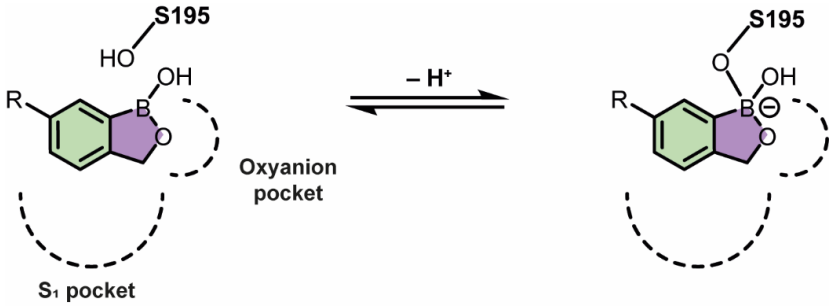


Figure 5.27 | A proposed reaction mechanism for benzoxaborole compounds with PPE. The reaction with a nucleophilic serine orientates the C3 and C4 atoms at the surface of the S_1 specificity pocket, and the oxaborole oxygen in the oxyanion pocket. The formation of the tetrahedral complex results in the loss of a proton[360].

5.2.3. Discussion and Perspectives

Since Baker *et al*. demonstrated the potent antifungal activity of a C5-substituted benzoxaborole[347], the benzoxaborole scaffold has attracted substantial interest. To date, however, the benzoxaborole scaffold has not demonstrated potential as an inhibitor of elastase-like SPs. Thus, a series of C6-substituted benzoxaboroles were investigated as inhibitors of PPE. C6-substituted benzoxaboroles have demonstrated inhibitory potential against a range of enzymes with catalytic serine residues (e.g., Dengue virus SPs (NSB2 and NS3 (**ix**)); β-lactamases (AmpC (**v** and **vi**) and OXA-10 (**v**)) and PBPs (**iii**)). However, the preliminary results from the series against PPE revealed poor inhibition. The five most potent inhibitors reduced the activity of PPE between 40–50% at 200 μM (**23**, **26**, **29**, **38**, and **45**).

Of particular concern was an increase in the apparent hydrolytic activity of PPE in the presence of high concentrations of several of the inhibitors, caused by assay interference due to 405 nm absorption. One such inhibitor that demonstrated an increase in apparent activity by more than an order of magnitude was compound **36**. At lower inhibitor concentrations (22 μM), however, **36** reduced the activity of PPE by 31% – the largest reduction by any benzoxaborole at this concentration (Figure 5.23). While the data presented here were collected using an absorbance-based assay where the pNA moiety produces signal at 405 nm following cleavage from a peptide substrate (AAPA), alternative substrates are commercially available. Similar peptides with 7-amino-4-methylcoumarin (AMC) or 7-amino-4-trifluormethylcoumarin (AFC) moieties which produce fluorescence after hydrolysis of the substrate are commercially available (e.g., MeOSuc-AAPV-AMC). These alternatives may enable better assessment of the inhibitors that were affected by the potential inhibitor absorbance, although similar interference with fluorescent inhibitors is possible. With compounds **21** and **36**, however, the addition of formic acid quench solution was responsible for the absorbance increases, due to precipitation. Therefore, other quenching strategies could be assessed perhaps using alkali or detergents.

Structure activity relationship (SAR) assessment from the benzoxaborole series is difficult to interpret. The functionalisation of the six most potent inhibitors (Figure 5.28) was quite diverse. Although only

four inhibitors were tested without a sulphonamide or sulphamide, these inhibitors were less potent, suggesting a benefit to the presence of the sulphur-containing moieties at the C6-position. This observation can be rationalised by the observation of indirect interactions between one of the sulphonamide oxygen atoms with the carbonyl oxygen of V216, which could stabilise the enzyme:inhibitor complex (Figure 5.26, D). Further *S*– (sulphonamide) or *N*– (sulphamide) functionalisation demonstrated that small, large, polar, and neutral groups were all tolerated. The small polar *S*–substituents of **23** and **29** were directed towards H57 (Figure 5.26, E & G), in a similar manner to the prolyl ring of BCM7[106]. The larger functional group of **26** was also able to form further indirect interactions with the carbonyl oxygen of T96 (Figure 5.26, B). Further crystallographic studies may be of interest to identify key determinants for changes in potency observed within the series but the overall results with the C6-substituted benzoxaboroles explored were poor.

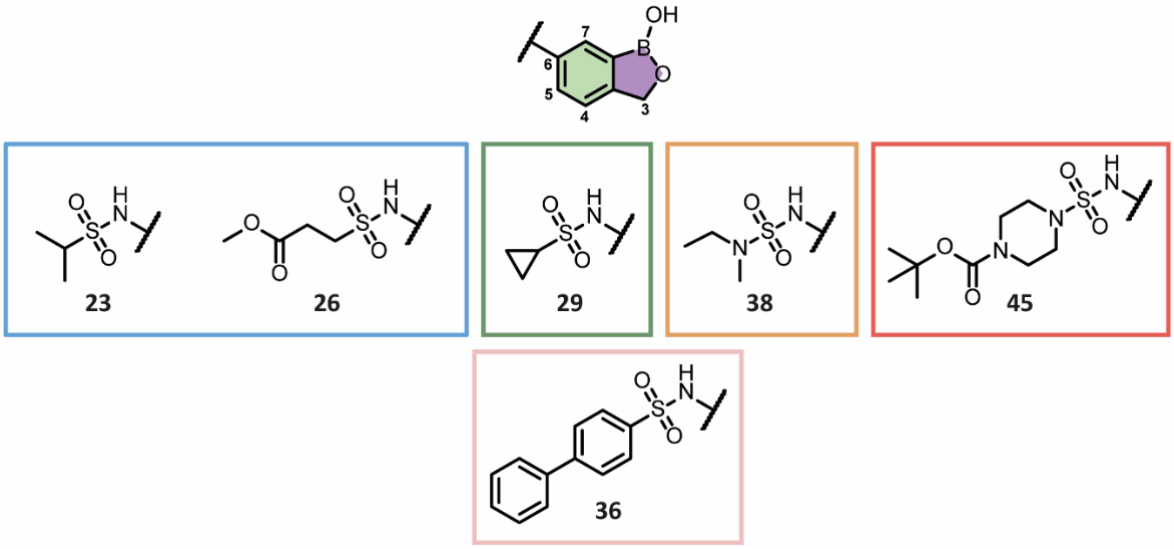


Figure 5.28 | Chemical structures of the most potent benzoxaboroles.

While molecular docking of unsubstituted benzoxaborole and HNE suggested that these compounds do not form complexes with SPs[359], crystallographic analysis of PPE and C6-substituted benzoxaboroles demonstrated that formation of a dative tetrahedral complex was possible. The orientation of the PPE:benzoxaborole complexes positioned the C3 and C4 atoms of the bicyclic rings at the surface of the S_1 specificity pocket. Thus, as observed with the bicyclic thiophene γ-lactams, small and

hydrophobic functional substitution at these positions may improve the potency of the inhibitors. Several benzoxaboroles have been reported with small hydrophobic C3-substituents, where they have been particularly effective antiparasitic compounds[361]. Alternatively, functionalisation of the C5-postion could direct the substituents towards the S_3 specificity pocket, which as discussed in Chapter 2 and 3, offers the potential to improve potency. Polar functional groups that could interact with the sidechain of Q192, may therefore be beneficial in this regard.

While characterisation of a series of C6-substituted benzoxaborole derivatives as PPE inhibitors did not identify any potent inhibitors, the crystallographic analysis reported here provides a base from which to explore alternative functionalisation. As a result, there are clear avenues to explore that offer a strong likelihood of improving the potency of benzoxaborole inhibitors towards PPE and other SPs.

Chapter 6: Summary and Conclusions

Peptidases are one of the largest classes of enzymes and are found ubiquitously throughout life[5]. The action of peptidases has traditionally been ascribed to non-specific degradative roles, partly because the earliest discovered peptidases were digestive enzymes (pepsin, CTR, TR)[15]. However, peptidases are now known to play important roles in many physiological processes where their action is often limited to a single precise hydrolytic reaction[17]. The activity of exogenous peptidases and the aberrant activity of endogenous peptidases is of concern for human health[24]. Accordingly, endogenous peptidases have been linked to 135 hereditary disorders[4], inflammatory diseases[294,308,310], and cancers[307,311]. Proteases are also targets for antiviral drugs of major societal importance including SARS-CoV-2[28].

SPs are a subclass of peptidases which utilise a catalytic serine nucleophile to initiate cleavage of peptide and ester substrates[5]. The mechanism for SPs containing the 'classical' catalytic triad (His, Asp, Ser) was proposed more than 50 years ago[117]. The hydrolytic mechanism is one of the most well-characterised enzyme mechanisms, but questions remain regarding the details. One question concerns the proposed presence of (a) tetrahedral intermediate(s). The existence of such intermediates has been inferred from solution chemistry[176] and structural analyses of complexes of SPs with tetrahedral analogues[109,177], but has not been directly observed by spectroscopic methods in the reaction between a SP and a substrate[118]. There are, however, several proposed *in crystallo* tetrahedral intermediates with SPs[170,171,178], although the results have been questioned.

In 1997, the stable acyl-enzyme complex between PPE and BCM7 was reported[106]. This was amongst the first reports of an acyl-enzyme complex with a SP, *in crystallo*, with a peptide substrate of known composition. The stability of the PPE:BCM7 complex was attributed to a combination of the low pH of the crystallisation solution (pH 5.0), which reduces the catalytic activity of PPE, and interactions between the P_3 substrate residue and S_3 subsite of PPE[106]. A subsequent report demonstrated that

upon raising the pH, the PPE:BCM7 complex was deacylated[170]. One of the crystal structures from the report contained additional electron density around the acyl carbon of Ile7 of BCM7 that was consistent with a tetrahedral intermediate[170]. However, due to the presence of sulphate within the precipitant solution, which has been shown to associate with S195[88,89], the assignment has been met with some scepticism[118].

In **Chapter 2**, crystallographic analyses of the PPE:BCM7 complex were performed to further investigate the proposed tetrahedral intermediate. Novel crystallisation conditions were identified for PPE which did not contain sulphate or phosphate, to eliminate the possibility that the additional density derived from a sulphate ion. Crystal structures of apo-PPE in these novel precipitants (NaF and NaCl) at pH 5.0 (acetate buffer) revealed that, in the absence of sulphate, both acetate and glycerol from the cryoprotectant, could associate with S195. The PPE:BCM7 complex was also successfully reproduced under these conditions with crystals diffracting to high resolution (~1.3 Å)

Subsequently, pH-jumps were reproduced. Crystals produced at pH 5.0 were soaked in high pH buffer (pH 9.0) and flash frozen after a delay. From the resultant crystal structures, one structure revealed additional electron density surrounding the acyl carbon of Ile7, analogous to the proposed tetrahedral intermediate (Figure 2.6)[170]. However, from the evaluation of the electron density (2m*F*o-D*F*c) and difference maps (m*F*o-DFc), the tetrahedral intermediate did not satisfactorily explain the electron density. Instead, an acetate and a glycerol molecule were modelled into the density analogously to the orientations observed in the apo-structure of PPE (Figure 2.4). Additionally, several waters and a glycerol molecule were modelled in place of the other BCM7 residues (Val4-Pro6). These molecules were in identical positions to the apo-structure of PPE and sufficiently explained the observed electron density within this region (Figure 2.6). Subsequently, pH-jumps were performed in the absence of glycerol. From the 87 crystals collected under these conditions, with soaking times ranging between 5 s to 2 m, no analogous electron density was observed. Importantly, analogous electron density was observed after performing pH-jumps on apo-PPE crystals that could have been interpreted as a

tetrahedral intermediate (Figure 2.7). From the combined data, it is proposed that the assigned tetrahedral intermediate between PPE and BCM7 is not, at least solely, a tetrahedral intermediate, but is predominantly formed from the acetate and glycerol molecules which reassociate following deacylation of BCM7 from the PPE active site.

Throughout the course of the above studies, a large conformational change was observed in the metal binding region of PPE in structures that had been subject to 'pH-jumps' (Figure 2.8). Backbone conformational changes are rarely observed in PPE[362] and thus the cause of the changes was investigated. Steric clashes from a CHES buffer molecule binding in a cleft near the 'origin' of the conformational change were ruled out, as identical conformational changes were observed using Tris buffer which was not bound in an analogous position (Figure 2.10). As the conformational changes were only observed following pH-jumps pH effects were investigated. The results from these studies are inconclusive. Several structures of apo-PPE crystallised at high pH or produced at pH 5.0 and subject to analogous pH-jumps exhibited the conformational change. However, this conformational change was not always observed in the structures obtained at high-pH. Therefore, the conformational changes may be more likely to occur at higher pH (> 7) but this does not appear to be the sole cause of the conformational changes, and the motions are likely not linked to substrate binding/release.

Attempts were made to optimise microcrystal slurries of both apo-PPE and the PPE:BCM7 complex in preparation for time-resolved SFX experiments. While this was somewhat achieved for apo-PPE (**Chapter 3**), microcrystals of the PPE:BCM7 complex of sufficiently small size and high density were not obtained and will require further optimisation in the future (Figure 2.11). Should further optimisation prove similarly fruitless, mechanical destruction of the crystals, using vortexing could offer a method to reduce the size of the crystals once grown to an optimal density. Nonetheless, the first ambient temperature SFX structure of PPE was obtained using sulphate as a precipitant (Figure 2.12), and subsequently of the PPE:BCM7 complex in NaF (Figure 2.13). Attempts were also made to perform time-resolved experiments with larger crystals (~50 × 7 × 7 μM) using the DoC

methodology. However, the results were poor. Reaction initiation, as determined by deacylation of the PPE:BCM7 complex, was slow (≥ 1.12 s) and would not be expected to lead to accumulation of a short-lived tetrahedral intermediate. If further optimisation of the crystallisation conditions results in better microcrystal slurries in the future, attempts should be directed towards different SFX methodologies. One approach that has proven successful is the pump-probe method which can obtain structures at extremely short time delays (fs-ms)[193,194]. To achieve this, a photocaged substrate of PPE[246] could be illuminated by the 'pump', releasing the substrate, which would then be free to react with PPE. This approach combined with microcrystals also has the benefit of more uniform reaction initiation which may be necessary to observe short-lived intermediates[192].

The work presented in **Chapter 3** explored alternative inhibitors, to complement BCM7, for time-resolved studies with PPE. First, truncated P_3 variants of BCM7 were examined[252]. The complex between PPE and Ac-NPI has previously afforded atomic resolution diffraction[252] and was thus re-examined. However, poor *in crystallo* occupancy of the peptide and possible presence of an acetate molecule in the active sites unoccupied by Ac-NPI would cause issues with analysis of subsequent time-resolved data and was therefore not pursued (Figure 3.2). The most potent derivative of BCM7 against PPE was Ac-OPI, but attempts to co-crystallise or soak Ac-OPI into crystals of PPE were unsuccesful[252]. Similar was observed with the novel crystallisation conditions for PPE described in Chapter 2. A novel condition was identified from broad crystallisation screens exploring co-crystallisation of PPE with Ac-OPI (Figure 3.3). The crystal was processed in an alternate space group, but the resolution was relatively poor (> 2 Å), and the additional density observed in the active site of PPE was not consistent with an Ac-OPI acyl-enzyme complex. Therefore, this condition was not pursued further. The third inhibitor explored was the only approved synthetic inhibitor of HNE, Sivelestat[256,257]. The mechanism of reaction of Sivelestat with HNE was proposed to proceed through an acyl-enzyme complex with S195 and a *t*Bu-ester adduct. The mechanism was supported by ESI-MS

experiments where an 84 Da adduct was bound to HSE following incubation with Sivelestat[264]. However, previous attempts to obtain the complex *in crystallo* have been unsuccessful[264].

A crystal structure with additional electron density consistent with a *t*Bu-ester acyl enzyme complex was obtained after soaking PPE crystals with Sivelestat (Figure 3.4, C). By LC-MS analysis, Sivelestat demonstrated slow binding where substantial acylation of PPE was not observed until 10 min (Figure 3.4, A). Using a higher concentration of Sivelestat, the time to accumulate the adduct *in crystallo* was reduced to between 2–5 min. However, a preliminary attempt to accumulate this structure *in crystallo* by SSX DoC methodology was unsuccessful. The time delay likely remains too long for accumulation of substrate and is unlikely to be useful for time-resolved studies with PPE.

In the future, co-crystallisation of PPE with additional truncated BCM7 P_3 variants such as Ac-QPI, Ac-SPI, or Ac-HsePI, which were all more potent inhibitors than BCM7, may prove more successful than Ac-OPI[252]. Previous attempts to obtain a crystal structure of these inhibitors was unsuccessful, which may suggest that binding induces a conformational change in PPE that inhibits crystallisation under standard conditions and would thus be of interest[253]. Alternatively, focus could be directed towards further exploration of Sivelestat and other Sivelestat derivatives[258,265,266], many of which have improved aqueous solubility. Should either approach prove successful, time-resolved experiments probing the mechanism of PPE with these inhibitors would be of interest. However, other serial crystallography mixing techniques, such as MISC[202], should be targeted to avoid the disadvantages of the DoC technique.

To complement the time-resolved experimentation, it is essential that additional techniques are used to confirm observations. A common approach combines site-directed mutagenesis with biophysical and kinetic experiments [205]. However, such experiments have not currently been performed with PPE. **Chapter 4** describes the optimisation of expression conditions and subsequent purification of recombinant PPE from an *E. coli* expression vector. Due to the destructive properties of active PPE, the

gene was designed to include the native propeptide of PPE, which would subsequently be removed during purification to reveal active PPE. Initial attempts to express the PPE gene from *E. coli* were unsuccessful using modified pCOLD I and pNIC28 vectors with *N*–terminal His_6 tags, with near complete insoluble expression (Figure 4.2). Exchange of the *N*–terminal His_6 tag for a *C*–terminal His_6 tag resulted in improved soluble expression (Figure 4.5). Inclusion of *N*–terminal fusion tags (MBP, SUMO, GST, or TRX)[280], the use of SHuffle strain[274], or supplementation with chemical additives[283] did not further improve soluble yields of recombinant protein (Figure 4.5; Figure 4.17). However, the use of the Lemo strain[284], and optimisation of the post-transduction incubation time and temperature did afford improvements (Figure 4.18). The optimum expression conditions explored were: ProPPE-$CHis_6$ from a pNIC28 vector, carried out by Lemo cells, induction by supplementation of IPTG, post-induction incubation at 18˚C for 18 h in the presence of 500 mM L-rhamnose.

Initial purification trials using subsequent rounds of IMAC resulted in both low yield, low purity, and low catalytic activity (Figure 4.8). However, the successful activation of ProPPE:$CHis_6$ by TR hydrolysis was demonstrated. Several rounds of purification optimisation were attempted, exploring various purification techniques. In separate purifications, improved purity (≤ 50% by SDS-PAGE analysis) and good catalytic activity were achieved, however, this was not achieved in a single purification. By combining the two methods, an improved method was proposed. Importantly, on-column activation of PPE with minimal contact time with TR and isolation of the later eluting peak during IMAC following activation appear to be important for catalytic activity (Figure 4.23). The poor purity was regularly caused by a contaminant with a molecular weight of 23,509 Da. The contaminant peak was successfully removed from the sample by CEX, after activation by TR hydrolysis (Figure 4.13). Thus, a combination of these three aspects offers a good chance that catalytically active PPE, of good purity, can be obtained. However, issues remain with the purification method. In particular, the lack of retention of ProPPE on the IMAC column resulted in a very low yield.

There are many potential avenues to further optimise PPE expression and purification in the future. Further modification of the construct, perhaps by extending the linker region to enhance the availability of the tag for interaction with the resin, and by increasing the number of His residues in the tag (His_8 or His_{10}) should enhance the binding affinity for the resin and reduce the observed loss of enzyme at each purification step. To improve the soluble yield, periplasmic protein production with the ProPPE-His_6 constructs, and potentially the untagged ProPPE construct, should be explored. Targeting recombinant protein production to the periplasm with the related SP, TR, has previously proven successful[269,286]. While targeting His_6-ProPPE production to the periplasm did not lead to increased soluble yields, improved results may be obtained using the alternate constructs, as was observed with the cytoplasmic production (Section 4.2.2). Alternatively, eukaryotic expression systems may be required to improve the yield. Recently, recombinant HNE, expressed as a zymogen with a fusion protein, was produced using the *Pichia pastoris* (*P. pastoris*) system to high yields[363]. As PPE is used as a model for HNE, which is the more interesting protein from a health perspective, future work could be directed from PPE to recombinant HNE.

The work presented in **Chapter 5** aimed to perform structure-based SAR on three elastase inhibitor scaffolds: C4-substituted 5,5-fused thiophene γ-lactams (prepared by Gayatri), C3-substituted oxindoles (prepared by Dr Lewis Ibbotson), and C6-substituted benzoxaboroles (prepared by Gayatri), with PPE as a model for HNE inhibition. Nine novel thiophene γ-lactam derivatives (**5**–**13**) were explored and compared to four reported compounds (**1**–**4**)[331]. Four of the novel compounds (**9**, **10**, **11**, and **13**) exhibited single digit micromolar IC_{50} values against PPE, all of which contained a monosubstituted C4 (Table 5.1). The best inhibitor remained the spirocycle **4** which was among those previously reported[331]. Protein LC-MS studies with the inhibitors revealed that the disubstituted inhibitors (**3**–**7**) can exhibit slow acylation of PPE, which may explain the poor potency of the compounds under the dose-response conditions, which included only a 10 min incubation. Thus, the disubstituted compounds may exhibit improved IC_{50} values following longer incubation periods.

Crystallographic analyses with the compounds demonstrated that the inhibitors could bind in two orientations within the active site, although both displaced the proposed hydrolytic water and orientated the C4-substituent into the S_1 subsite (Figure 5.12). These analyses also demonstrated that larger C4-substituents, which were not expected to bind well within the small PPE S_1 subsite[53], could curl around the pocket, rationalising how they were tolerated (Figure 5.12). However, a complex between PPE and the compound with the bulkiest substituent **13**, was not successfully obtained despite being a potent inhibitor (IC_{50} = 4.47 μM). Further studies with these compounds should focus on characterisation of potency against HNE and selectivity against a range of other SPs, for example thrombin, TR, and CTR.

Four oxindole compounds (**14**-**17**) were all found to be poor inhibitors of PPE, with the strongest **14** reducing the activity of PPE by only 46% at 200 μM (Figure 5.15). None of the inhibitors were observed to acylate PPE under the LC-MS conditions suggesting that the lactam amide was insufficiently activated by the benzene ring (Figure 5.15). Should further studies be carried out with these compounds, attempts should be made to better activate the lactam amide, perhaps through the addition of electron withdrawing groups to the phenyl ring or the lactam amide. Additionally, other compounds such as furan and pyrrole, which are also isosteric for the benzene and thiophene rings, could be explored. The results from such experiments may also provide a better understanding of the underlying chemistry behind the lack of inhibition exhibited by the oxindole scaffold.

As with the oxindole compounds, all the C6-substituted benzoxaboroles (**18**-**45**) were found to be poor PPE inhibitors; only five compounds (**23**, **26**, **29**, **38**, and **45**) reduced the activity of PPE by more than 40%, but none by more than 50% (Table 5.2). Reported molecular docking studies suggested that the benzoxaborole compound would not make a tetrahedral complex with S195, rationalising the poor inhibitory potency towards HNE (IC_{50} > 100 μM)[359]. However, this was not observed empirically. Crystallographic analyses with three of the most potent compounds (**23**, **26**, and **29**) revealed additional density associated with S195 the was consistent with formation of a tetrahedral complex

with the oxaborole ring (Figure 5.26). The variety of C6-substituents on the five best inhibitors can be explained by alternate orientation of the substituent within the active site, where short hydrophobic substituents (**23** and **29**) are directed towards the imidazole ring of H57, while longer and hydrophilic substituents (**26**) exhibited a minor difference in orientation where they may interact with other PPE residues such as T96 (Figure 5.26). The benzoxaborole rings were located at the surface of the S_1 pocket. To improve the potency of these compounds towards PPE, substitution at the C3 position, and perhaps the C4 and C5 position. The C3 substitution would likely be the most beneficial as such substitutions would direct the C3 substituent into the S_1 pocket, the primary determinant of PPE specificity[118]. Substitution of small hydrophobic groups, such as those that were effective C4 substituents in the thiophene γ-lactam series would likely be beneficial. To target the PPE S_3 subsite with C4 and C5 substitutions, hydrophilic groups may improve the affinity for the PPE active site, as was observed with the truncated P_3 variants of BCM7[252].

Overall, the work described in this thesis informs on the mechanism and inhibition of the model SP, PPE. The results lay the groundwork for future time-resolved studies using microcrystals and XFEL analyses. The first XFEL structures of PPE (apo- and PPE:BCM7) were obtained to high-resolution. The pH-jump results from data collected with crystals of the PPE:BCM7 complex, suggest that a previously assigned tetrahedral species is likely not a true intermediate. The work also provides a step towards the ability to perform site-directed mutagenesis studies with PPE, which would be beneficial for complementation of time-resolved data that is obtained with PPE in the future. Insight into SP inhibition is also provided, both by acylating agents and tetrahedral complex forming benzoxaboroles.

Chapter 7: Materials and Methods

7.1. General Information

Commonly used buffers and chemicals were purchased from either Merck or Thermo Fisher Scientific unless stated. Water used throughout was purified using a Milli-Q® Reference Water Purification System with a Quantum® TEX Polishing Cartridge and a Millipak® Express 40 Filter (MQ water). Crystals were imaged using a LEICA M205C light microscope. All chemical structures and schemes were prepared using ChemDraw 19.1. All crystal structure images were prepared using PyMol[364]. All figures were prepared using Adobe Illustrator.

7.2. General Protein Crystallography Methods

7.2.1. Materials

Lyophilised Porcine Pancreatic Elastase (200 U/mg) was purchased from Serva Electrophoresis GmbH (Cat. No. 20930) and for crystallographic purposes was used without further purification. Crystal screening kits, SaltRx and Crystal Screen were purchased from Hampton Research, NeXtal anions and NeXtal Cations screens were purchased from NeXtal Biotech, and Pact Premier, Morpheus, SG1™ Screen, JCSG Plus, and The BCS Screen were purchased from Molecular Dimensions. BCM7 and Ac-NPI were custom synthesised by Thermo Fisher Scientific, and Ac-OPI was custom synthesised by GL Biochem. Sivelestat was purchased from Activate Scientific. CrystalCap™ SPINE HT crystal mounts with loops between 0.05 and 0.2 µm (Hampton Research) were used to harvest and flash freeze crystals. All chemical solutions used for crystallisation, except those from commercial screening kits, were filtered through 0.2 µm syringe filters prior to use.

7.2.2. Sitting Drop Vapour Diffusion Crystal Screening

7.2.2.1. Apo-PPE

Apo-PPE was screened for crystallisation against various commercial screening kits using the sitting drop vapour diffusion method. Lyophilised PPE was dissolved in acetate buffer (50 mM, pH 5.0) to

50 mg mL^{-1} as determined using a NanoDrop™ One Microvolume UV-Vis Spectrophotometer (Thermo Scientific). Deep well blocks containing the crystallisation screen solutions were prepared from the stock solutions by hand. 25 µL of mother liquor was transferred to 96-well 3-drop low-profile plates (SwissSci; Cat No. UVXPO-3LENSLP) using a VIAFLO 384 electronic multipipette (Integra). Droplets totalling 300 nL in 2:1, 1:1, and 1:2 v/v enzyme:mother liquor ratios, using a Mosquito® crystallization robot (sptlabtech). Plates were sealed with StarSeal clear polyolefin adhesive seals (Starlab) and incubated at room temperature for up to two weeks.

7.2.2.2. PPE:Ac-OPI

PPE:Ac-OPI was screened for crystallisation against various commercial screening kits using the sitting drop vapour diffusion method. Lyophilised PPE was dissolved in acetate buffer (50 mM, pH 5.0) to 50 mg mL^{-1} as above. Lyophilised Ac-OPI was dissolved in acetate buffer (50 mM, pH 5.0) to the desired concentration (100 mM, 50 eq of PPE). PPE and Ac-OPI were mixed 1:1 prior to addition to the sitting drop plate. Swissci 96-well 3-drop low-profile plates were prepared as above with 25 µL mother liquor and 300 nL droplets of PPE and Ac-OPI: mother liquor at 2:1, 1:1, and 1:2 v/v ratios. Plates were sealed with StarSeal clear polyolefin adhesive seals and incubated at room temperature for up to two weeks.

7.2.3. Hanging-Drop Vapour Diffusion

7.2.3.1. Apo-PPE

Lyophilised PPE was dissolved in 50 mM acetate buffer pH 5.0 to the desired concentration (20–50 mg mL^{-1}). This solution was mixed 1:1 v/v with the mother liquor solution (50 mM acetate buffer pH 5.0, and either: 40-400 mM NaF, 50-400 mM NaCl, 50-300 mM NaBr, or 50-200 mM NaI) to a final volume of 4-6 µL. 1–3 drops were added to each plastic cover slips (Opticlear, Hampton Research; Cat No. HR8-082) and the 24-well plates (Hampton Research; Cat No. HR3-171) were incubated at room temperature. Crystals formed within 1-3 days.

7.2.3.2. PPE:BCM7 and PPE:Ac-NPI Co-Crystallisation

Lyophilised PPE was dissolved in acetate buffer (50 mM, pH 5.0) to the desired concentration (50 mg mL^{-1}). Lyophilised inhibitors were dissolved in acetate buffer (50 mM, pH 5.0) to the desired concentration (30–40 eq of PPE). PPE (1 µL) and inhibitor (1 µL) were added to mother liquor solution (50 mM acetate buffer pH 5.0, and either: 300–500 mM NaF, 200-500 mM NaCl) to a final volume of 4–6 µL. 1–3 drops were added to each plastic cover slips and the plates were incubated at room temperature. Crystals formed within 1–7 days. For SSX experiments, 1 µL of dilute microseed solution, prepared using a seed bead™ kit (Hampton research; Cat No. HR2-320) from macrocrystals of PPE:BCM7 in NaCl according to the manufacturer's instructions, was added to the 6 µL drops.

7.2.4. Batch Crystallisation

7.2.4.1. Plate Crystallisation for Apo-PPE SFX Structure

Lyophilised PPE was dissolved to 50 mg mL^{-1} in acetate buffer (50 mM, pH 5.0). 300 µL of PPE was added to 1.2 mL of Na_2SO_4 (final [C] = 50 mM) and acetate buffer (final [C] = 50 mM, pH 5.0). 15 µL of microseeds produced using a sead bead™ kit from crystals prepared using the hanging drop vapour diffusion method in Na_2SO_4 and acetate buffer, were added prior to aliquoting and mixed thoroughly. The solution was added to a 96-well plate (Corning; Cat No. 3635) in 100 µL aliquots, sealed with plastic, and incubated at room temperature for four days. Crystal slurries were pooled into a 1.5 mL Eppendorf tube and left to settle overnight. The following day, supernatant was removed to a final volume of approximately 100 µL, prior to shipment to PAL-XFEL at ambient temperature.

7.2.4.2. In-Tube Crystallisation of PPE:BCM7 for SFX and SSX-DoC Experiments

Lyophilised PPE was dissolved to 50 mg mL^{-1} in acetate buffer (50 mM, pH 5.0). Lyophilised BCM7 was dissolved to 60 mM (30 eq) in acetate buffer (50 mM, pH 5.0). 300 µL of PPE solution and 300 µL of BCM7 solution was added to 900 µL of mother liquor (667 mM NaF and 50 mM acetate buffer pH 5.0). To the final solution, 25 µL of microseeds produced using a sead bead™ kit from macrocrystals of PPE:BCM7 in NaF precipitant by the hanging drop vapour diffusion method. The final concentrations

were: PPE (10 mg mL^{-1}), BCM7 (12 mM), NaF (400 mM), acetate buffer (50 mM, pH 5.0) The solution was aliquoted (250 µL) in 1.5 mL Eppendorf tubes. The tubes were incubated for 7 days at room temperature on an end-over-end rotor (Boekel Scientific) at 18 rpm. Subsequently, the slurries were harvested, pooled, and concentrated to approximately 350 µL.

7.2.5. pH-Jump Methodology

The pH-jump methodology was initially performed as described by Wilmouth *et al*. 2001[170]. Crystals of the PPE:BCM7 complex produced at pH 5.0 by hanging-drop vapour diffusion methodology (Section 7.2.3.2) were transferred to a 2 µL drop of pH-jump solution (50% v/v NaCl or NaF at the same concentration as the mother liquor and 50% v/v CHES or Tris (200 mM, pH 9.0 with 40% v/v glycerol)). Crystals were incubated in the pH-jump solution for the desired time before flash freezing in N_2(l). For latter pH-jumps, the pH-jump solution contained 50% v/v NaCl or NaF at the same concentration as the mother liquor and 50% CHES or Tris (0.2–1.0 M, pH 9.0).

7.2.6. Soaking of Sivelestat, Bicyclic γ-Lactams, and Benzoxaborole Compounds

Crystals of apo-PPE in NaCl precipitant prepared at pH 5.0 by the hanging-drop vapour diffusion method (Section 7.2.3.2) were transferred to a soaking solution (50% relevant mother liquor, 20% v/v DMSO, and the inhibitor at various concentrations: Sivelestat (20 mM), benzoxaboroles (10–20 mM), γ-lactams (3.75–10 mM)). After soaking for the desired time, crystals were flash frozen in N_2(l). The soaking information for each crystal is described in Appendix C.

7.2.7. Data Collection for Single Crystals

All data from single crystals were collected at 100 K from Diamond Light Source synchrotron radiation beamlines I03 and I04. Data was indexed, integrated, and scaled using the automatic Xia2 pipelines[365].

7.2.8. Pal-XFEL SFX Fixed-Target Data Collection

Crystal slurries of apo-PPE (Na_2SO_4) and PPE:BCM7 (NaF) prepared at pH 5.0 were sent to Pal-XFEL for data collection by nylon mesh based fixed-target methodology as described in reference 366. In brief the crystal slurry (20 µL) was loaded onto a nylon mesh and sandwiched between two metal frames which sealed the nylon mesh within a polyimide film. The chips were assembled with sample at PAL-XFEL, and the diffraction data collected, through raster scanning of the chip, by Dr Jaehyun Park, Sehan Park, Gisu Park, and Dr Hyojung Hyun (all PAL-XFEL).

The apo-PPE structure (Figure 2.12), was collected under proposal 2020-2nd-NCI-018. The X-ray beam wavelength was 0.999 Å (12.41 keV). Data was collected at 60 Hz using a Jungfrau 4M detector, with a beam size of 5 µm × 4 µm, pulse duration of 25 fs, and pulse energy of 1.1 mJ.

The PPE:BCM7 structure (Figure 2.13), was collected under proposal 2021-2nd-NCI-023. The X-ray beam wavelength was 0.999 Å (12.41 keV). Data was collected at 60 Hz using a Jungfrau 4M detector, with a beam size of 3 µm × 3 µm, pulse duration of 25 fs, and pulse energy of 1.4 mJ.

7.2.9. Diamond Light Source I24 SSX Fixed-Target Data Collection

Fixed-target SSX data was collected from Diamond Light Source beamline I24 using the fixed-target methodology[232]. Within a high humidity chamber, microcrystal slurries (80-120 µL) were added to one face of a silicon chip with 25,600 wells with apertures between 5 and 12 µM. The crystallisation solution was pulled through the chip using a vacuum to collect the crystals in the wells with minimal crystallisation solution remaining. The chip was loaded into a metal frame which is sealed using two screws. The frame sealed the chip between two mylar® sheets (6 µm thickness). The frame with the chip was attached to a stage, through magnetic contacts, at the beamline. The stage can move in the X, Y and Z directions to enable alignment and focusing of the chip using a camera. Data was collected from the chip in a raster pattern (20 wells × 20 wells per block, and 8 blocks × 8 blocks per chip). Data was collected over three beamtimes (mx25260-14, mx25260-25, and mx25260-26) where the data

collection was performed under similar conditions (0.9999 Å wavelength (12.4 keV), with a beamwidth of 7 µm × 7 µm, and 10 ms pulse length. The data was collected by Dr Pierre Aller, Johan Glerup, Dr Anastasiia Shilova, Dr Nicholas Devenish, Dr Tiankun Zhou, Dr Sam Horrell, Dr Robin Owen, Dr Danny Axford, and Dr Graeme Winter (all Diamond Light Source).

7.2.10. Diamond Light Source I24 SSX Fixed-Target DoC Data Collection

Data for fixed-target drop-on-chip SSX experiments were collected at Diamond Light Source beamline I24 using a piezoelectric injector and excite and visit again strategy (EAVA)[235]. Chips of microcrystal slurries were prepared as for fixed-target experiments above. However, when the chip was placed in the frame, the face of the chip upon which crystals were loaded is left unsealed. Loaded chips in the frame were kept in a sealed box with wet paper towels to keep the relative humidity high on the chip. At the beamline, the frame was loaded onto a stage but the open face of the chip is covered by a plastic cover to reduce drying. A hole in the plastic cover facilitates ejection of buffer solution (CHES or Tris pH 9.0, 0.2–1.0 M). Ejection of the solution was tested offline prior to data collection to determine ejection conditions (performed by Dr Pierre Aller and Johan Glerup, Orville group, Diamond Light Source). The chip was first aligned with the beam and then ejection was aligned relative to the beam. Data was collected in a raster pattern as described above and time delays were produced by altering the ejection and data collection strategy. Data was collected over three beamtimes (mx25260-14, mx25260-25, and mx25260-26) where the data collection was performed under similar conditions (0.9999 Å wavelength (12.4 keV), with a beamwidth of 7 µm × 7 µm, and 10 ms pulse length. The data was collected by Dr Pierre Aller, Johan Glerup, Dr Anastasiia Shilova, Dr Nicholas Devenish, Dr Tiankun Zhou, Dr Sam Horrell, Dr Robin Owen, Dr Danny Axford, and Dr Graeme Winter (all Diamond Light Source).

7.2.11. SSX and SFX Data Processing

The SSX and SFX diffraction data was processed by performed by Dr Pierre Aller, Dr Tiankun Zhou, Dr Agata Butryn, Dr Jos Kamps, Dr Robert Bosman, and Dr Aaron Brewster, according to the method

described in reference 205 (and references therein). Data acquisition was tracked using the cctbx.xfel graphical user interface[367]. Data indexing and integration were performed using dials.stills_process program[365,368-371]. The detector position was estimated using a diffraction pattern obtained from cerium oxide (CeO_2). The unit cell parameters of the reference model used for each structure are shown in the Appendix C. Final scaling and merging were performed using the cctbx.xfel.merge program[367].

7.2.12. Model Building and Refinement Strategy

The structures of PPE were solved by molecular replacement using the PHASER programme in PHENIX[372]. The search models used for molecular replacement of each structure are detailed in Appendix C. The structures were the iteratively modelled and refined using Coot[373] and phenix.refine[372], respectively. For the PPE:BCM7 and PPE:Ac-NPI structures, the Ile7 of BCM7 and S195 or PPE were modelled and refined in alternate conformations to allow Ile7 to adopt a conformation consistent with the electron density. The final rounds of refinement were performed using REFMAC in the CCP4i2 suite[374], where Ile7 and S195 were put into the same conformation and a link added between Ile7 (C) and S195 (Oγ). All refinement validation statistics were calculated with PHENIX to enable comparison. Gradewebserver[375]. The merging and refinement statistics are shown in Appendix C.

7.3. General Molecular Biology and Microbiology Methods

7.3.1. Materials

Trizma® base and Tween® 20 were also purchased from Merck. NuPAGE™ 4-12% Bis-Tris, 1.0 mm, 10- or 15-well precast gels, 4× NuPAGE™ LDL Sample Buffer, and NuPAGE™ MOPS SDS Running Buffer were purchased from Invitrogen. PageRuler™ Prestained Protein Ladder was purchased from Thermo Fisher Scientific. QuickBlue Protein Stain was purchased from Stratech Scientific. Amersham™ ECL™ Prime Western Blotting Detection Reagents (Cat No. RPN2232), Amersham™ Protran® 0.2 µm

nitrocellulose membrane (Cat No. 10600001), and Whatman Grade 3MM Chr Cellulose Chromatography Paper were purchased from Cytiva. Tesco Instant Dried Skimmed Milk was purchased from Tesco. Ponceau S was purchased from VWR. Anti-His_6 IgG1 mouse antibody was purchased from Proteintech (Cat No. 66005-1-Ig). Anti-mouse IgG horse, HRP-conjugated antibody was purchased from Cell Signalling (Cat No. 7076P2). Premixed TB phosphate buffered powder was purchased from Formedium.

Competent cells: XL10-Gold Ultracompetent Cells (Agilent, Cat No. 200317), BL21(DE3) Competent Cells (Thermo Fisher Scientific, Cat No. EC0114), NEB® 5-α Competent *E. coli* (High Efficiency; Cat No. C2987H), SHuffle® T7 Competent *E. coli* (Cat No. C3026J), and Lemo21(DE3) Competent *E. coli* (Cat No. C2528J) were all purchased from NEB and stored at -80 °C. SOC Outgrowth Medium was also purchased from NEB.

7.3.2. Preparation of Media and Agar Plates

2×YT media was prepared according to the recipe in Table 7.1 or provided by Mr Manso Jubier. 2×YT agar was prepared by addition of 1.5 % w/v agar (Table 7.1). Premixed TB powder was used to prepare TB media, according to the manufacturer's instructions. The media, 100 mL or 600 mL in Erlenmeyer flasks with a foam stopper and tin foil cover, and agar, 100-200 mL in bottles, were autoclaved for 20 min at 121 °C. To prepare agar plates, the autoclaved agar was cooled to approximately 50 °C. In a sterile laminar flow hood, the appropriate antibiotics (1,000× stocks) were added to the agar, and 15–20 mL of the agar added to Petri Dishes and allowed to set. The agar plates were sealed with parafilm and stored at 4 °C.

Table 7.1 | 2×YT media recipe prepared with MQ water.

Component	Mass (g) / L	
	Media	Agar Plate
Tryptone	16	16
Yeast Extract	10	10
NaCl	5	5
Agar	—	15

7.3.3. Agarose Gel Electrophoresis

Agarose gels were prepared from, and the electrophoresis carried out in, Tris-acetate-EDTA (TAE) buffer. The 10× stock solution prepared from the recipe shown in Table 7.2 was diluted 10-fold in water and 1% w/v agarose added to the buffer. The agarose TAE mixture was microwaved (800 W) until the agarose was fully dissolved. The solution was allowed to cool before adding SYBR Safe DNA Gel Stain (final 1×). The agarose solution was transferred to a gel tray (Alpha Laboratories) and allowed to solidify at room temperature with a comb. The gel and tray were then transferred to a running chamber (Alpha Laboratories) and covered with 1× TAE buffer. DNA samples were mixed 5:1 with 6× DNA Gel Loading Dye and loaded into the wells. GeneRuler 1 kb DNA ladder was also loaded, and the samples were separated by electrophoresis for 40 min at 100 V. The gels were imaged using a BIO-RAD Universal Hood III with UV illumination.

Table 7.2 | 10× TAE stock recipe prepared with MQ water.

Component	Amount / L H_2O
Tris base	48.4 g
Glacial acetic acid	11.4 mL
EDTA	3.7 g

7.3.4. SDS-PAGE

All SDS-PAGE analyses were performed in NuPAGE™ 4-12% Bis-Tris, 1.0 mm, 10- or 15-well precast gels. A 2× protein loading dye was prepared from the NuPAGE™ 4× LDL Sample Buffer with supplementation of 5% β-mercaptoethanol. Protein samples were mixed 1:1 with the 2× loading buffer and denatured at 95 °C for 5 min. 7.5–10 µL of sample was loaded per well. 5 µL of PageRuler™ Prestained Protein Ladder was also added to a well. Protein samples were separated by electrophoresis at 180 V for 40-50 min (until the dye front reached the bottom of the gel). Electrophoresis was performed in NuPAGE™ MOPS SDS Running Buffer. Gels were stained for 30 min using QuickBlue Protein Stain, at room temperature on a 3D orbital shaker.

7.3.5. Western Blotting

Samples for western blotting were separated by SDS-PAGE as detailed in Section 7.3.4. The western blot was performed under wet conditions and the sandwich was arranged as follows: sponge, three layers of Whatman paper, polyacrylamide gel from SDS-PAGE, nitrocellulose membrane, three layers of Whatman paper, and sponge. All components were pre-soaked in transfer buffer prior to assembly and bubbles were removed from the sandwich. The proteins were transferred to the nitrocellulose membrane at 100 V for 60 min in transfer buffer (ice cooled). Prior to development, the nitrocellulose membrane was stained with Ponceau S for 5 min at room temperature to confirm transfer. The nitrocellulose membrane was destained with PBS-T for 10 min at room temperature.

Table 7.3 | Western blotting buffer recipes.

Solution	Components
Transfer buffer	Tris base (25 mM), glycine (200 mM), methanol (20% v/v)
10× PBS	Na2HPO4 (100 mM), KH2PO4 (18 mM), NaCl (1.37 M), KCl (27 mM)
PBS-T	10% v/v PBS, 0.1% Tween® 20
Blocking solution	5% w/v dried skimmed milk in PBS-T
Primary antibody solution	0.1% v/v anti-His_6 IgG1 mouse antibody in blocking solution
Secondary antibody solution	0.01% v/v anti-mouse IgG, HRP-conjugated horse antibody in blocking solution

The nitrocellulose membrane was blocked with blocking solution (2 × 30 min at room temperature) on an orbital shaker. **Note:** all stages of development were incubated on an orbital shaker unless stated. The blocking solution was decanted. The nitrocellulose membrane was incubated at 4 °C for 16 h in primary antibody solution. The primary antibody solution was decanted and the membrane washed 3× with PBS-T for 10 min at room temperature. The nitrocellulose membrane was incubated for 1 h at room temperature in secondary antibody solution. The secondary antibody solution was decanted and the membrane washed 3× in PBS-T for 10 min at room temperature. The membrane was developed in the dark for 5 min (no shaking) using ECL™ Prime Western Blotting Detection reagents according to the manufacturer's instructions. The membrane was imaged twice using a BIO-RAD ChemiDoc MP Universal Hood III: samples were visualised using the in-built "Blots-Chemi" method, and the protein

ladder using “Protein Gels-Stain Free Gel” method. Superimposition of the two images was carried out using ImageLab (6.1).

7.3.6. Bacterial Transformation and Plasmid DNA Isolation

All pipette tips, Eppendorf tubes, and MQ water used were autoclaved for 20 min at 121 °C (TouchClaveR II; LTE Scientific) prior to use. Competent cells were thawed on ice prior to use. In a sterile laminar flow hood, DNA and cells were mixed at various ratios: 1 µL DNA:15 µL cells for plasmids; 2 µL DNA:25 µL cells for nicked or ligated samples. Samples were incubated on ice for 30 min before heat shock at 42 °C for various times according to the manufacturers protocol (Table 7.4). The samples were returned to ice for 5 min, supplemented with 250 µL of SOC Outgrowth Medium, and incubated at 37 °C for 1 h on a shaker at 300 rpm. In a sterile laminar flow cabinet, 100 µL of the culture was added to 2×YT agar plates supplemented with relevant selection antibiotics for the vector and strain, streaked, and allowed to dry. Plates were incubated at 37 °C overnight. Individual colonies were transferred to 10 mL of sterile 2×YT media supplemented with the relevant selection antibiotics in a 50 mL Falcon tube and incubated at 37 °C with shaking at 200 rpm for 16 h. 5 mL of culture was centrifuged at 3,700 × *g* for 10 min at 4 °C. Plasmid DNA was isolated from the resultant pellets using a GeneJET Plasmid Miniprep kit and eluted in autoclaved MQ water. The DNA concentration and purity was assessed using a NanoDrop™. Plasmid DNA was sequenced to verify changes (TubeSeq, Eurofins Genomics) and stored at -20˚C.

Table 7.4 | Competent cell heat shock times.

Cell type	Time (s)	Additional Selection Antibiotic	Culture Temp (˚C)
XL10-Gold	30	—	37
NEB 5α	30	—	37
BL21(DE3)	10	—	37
SHuffle® T7	30	—	30
Lemo21(DE3)	10	30 µg mL^{-1} chloramphenicol	37

7.3.7. Preparation Transformed Cell Glycerol Stocks

In a sterile laminar flow hood, single colonies of transformed cells were used to inoculate 10 mL of 2×YT media supplemented with the appropriate antibiotic(s) in 50 mL Falcon tubes. Cultures were incubated at the appropriate temperature for the strain (Table 7.4), with agitation (200 rpm) for 4-6 h. 250 µL of culture was mixed with 250 µL of sterile cryoprotectant (50% v/v glycerol), added to cryovials, and stored at -80 °C.

7.4. Cloning

7.4.1. Materials

Custom oligonucleotides were synthesised by Merck or Thermo Fisher (Life Sciences) and stored as 100 µM stocks in autoclaved MQ water at -20 °C. Q5 high-fidelity polymerase, Phusion-HF polymerase, DMSO, and DNA Gel Loading Dye (6×) used for PCR were purchased from NEB. GeneJET PCR purification, GeneJET Gel Extraction, and GeneJET Plasmid Miniprep kits were purchased from Thermo Fisher Scientific. SYBR Safe DNA Gel Stain (10,000×) and GeneRuler 1 kb DNA Ladder were purchased from Invitrogen. All PCR reactions were performed using a BioRad CFX96 Real-Time PCR Detection System.

7.4.2. Ligation Independent Cloning (LIC)

7.4.2.1. Preparation of the Insert

The LIC protocol was adapted from literature[376]. The codon optimised gene (GeneArt) for His_6-ProPPE (T17-N266; Uniprot: P00772) was custom synthesised (Invitrogen). The gene for ProPPE without the His_6-tag was amplified from a pMA-RQ vector by PCR using primers with the appropriate overhangs (Table 7.5). The PCR composition is shown in Table 7.6 and the three-step reaction conditions are shown in Table 7.7. The annealing temperature was calculated using the NEB Tm Calculator[377]. The PCR products were separated by agarose gel electrophoresis (Section 7.3.3) and the insert DNA bands purified using a GeneJET Gel Extraction kit.

Table 7.5 | Primers for amplification of ProPPE LIC gene insert.

Primer	Sequence (5' – 3')
Forward	5' **TAC TTC CAA TCC** ACC CAG GAT TTT CCA GAA ACC AAT GC 3'
Reverse	5' **TAT CCA CCT TTA CTG** TTA ATT GCT GGC GAT CAC ATT ATT GAT CC 3'

Table 7.6 |PCR reaction composition for LIC gene insert amplification.

Component	Stock Concentration	Volume (µL)	Final Concentration
ProPPE in pMA-RQ	25 ng μL^{-1}	1	1 ng/µL
Forward primer	10 µM	1.25	500 nM
Reverse primer	10 µM	1.25	500 nM
Phusion reaction buffer	5×	5	1×
dNTPs	10 mM	0.5	200 µM
DMSO	100 %	1	4 %
Phusion-HF DNA polymerase	—	0.5	—
Autoclaved MQ water	—	14.5	—
Final Volume	—	25	—

Table 7.7 | PCR reaction conditions for LIC gene insert amplification.

Step	Temperature (˚C)	Time	Cycles
Initial denaturation	98	5 m	1
Denaturation	98	30 s	
Annealing	68	30 s	30
Extension	72	1 m 10 s	
Final Extension	72	10 m	1

LIC-qualified T4 DNA polymerase (Cat No. 70099-M; Merck Novagen) was used to generate cohesive ends on the purified insert DNA. The reaction composition is shown in Table 7.8. Reaction mixtures were incubated as room temperature for 30 min followed by inactivation at 75 °C for 20 min.

Table 7.8 | Insert cohesive end reaction composition.

Component	Stock Concentration	Volume (µL)	Final Concentration
Purified ProPPE insert	13.2 ng μL^{-1}	14.5	9.57 ng μL^{-1}
T4 DNA polymerase buffer	10×	2 µL	1×
dCTP	25 mM	2 µL	250 µM
DTT	100 mM	1 µL	5 mM
BSA	10 mg mL^{-1}	0.2 µL	100 µg mL^{-1}
T4 DNA polymerase	—	0.5 µL	—
Final Volume	—	20	—

7.4.2.2. Preparation of the Vector

pNIC28-Bsa4 vector (Opher Gileadi; Addgene 26103)[378] was kindly provided by Tika Malla (Schofield Group, University of Oxford). The vector was linearised by digestion with BsaI-HF (NEB) at 37 °C for 15

min using the conditions in Table 7.9. The reaction products were purified using a GeneJET PCR purification kit.

Table 7.9 | pNIC28-Bsa4 vector linearization conditions.

Component	Stock Concentration	Volume (µL)	Final Concentration
pNIC28-Bsa4 vector	95.8 ng mL^{-1}	10.4	33.2 ng μL^{-1}
BsaI-HF	—	1	—
CutSmart buffer	10×	3	1×
BSA	10 mg mL^{-1}	0.3	100 µg mL^{-1}
Autoclaved MQ water	—	15.3	—
Final Volume	—	30	—

LIC-qualified T4 DNA polymerase was used to generate cohesive ends on the purified linearised vector using the conditions in Table 7.10. The reaction mixture was incubated at room temperature for 30 min and inactivated at 75 °C for 25 min.

Table 7.10 | Vector cohesive end reaction composition.

Component	Stock Concentration	Volume (µL)	Final Concentration
Linearised pNIC28 vector	3.3 ng μL^{-1}	50	1.7 ng μL^{-1}
T4 DNA polymerase Buffer	10×	10	1×
dGTP	25 mM	10	250 µM
DTT	100 mM	5	5 mM
BSA (10 mg/mL)	10 mg mL^{-1}	1	100 µg mL^{-1}
T4 DNA polymerase	—	1	—
Autoclaved MQ water	—	23	—
Final Volume	—	100	—

7.4.2.3. Annealing and Transformation

Vector and insert DNA treated with T4 DNA polymerase was mixed at 1:2 molar ratios in 5 µL reaction volume[379]. The mixture was incubated at room temperature for 5 min, supplemented with 4 mM EDTA, and incubated at room temperature for a further 5 min. 2 µL of mixture was added to 25 µL XL10-Gold cells, and transformed, purified, and verified as described in Section 7.3.6. **Note:** cells transformed with pNIC28 vectors were selected for on 2×YT agar plates with 30 µg mL^{-1} kanamycin.

7.4.3. Gibson Cloning General Method

7.4.3.1. Amplification of Vector and Insert DNA

The His_6-ProPPE gene was cloned into a modified pCOLD-I vector (Takara Bio) with a kanamycin resistance gene instead of an ampicillin resistance gene and an additional NotI restriction site (kindly provided by Dr Patrick Rabe), using the Gibson Assembly method[380]. All *N*–terminal fusion tags (MBP, GST, SUMO, and Trx; kindly provided by Eidarus Salah) were cloned into the pCOLD-I and pNIC28 vectors containing ProPPE-His_6 gene (Section 7.4.4) using the Gibson Assembly method[380]. The details of the vectors containing the fusion tag inserts are shown in Table 7.11. Insert and vector DNA was amplified using overlapping primers (Table 7.12) designed using the NEBuilder® tool[381].

Table 7.11 | Details of fusion tag insert vectors.

Name	Source	Addgene No.	Fusion tag	Provided by
1M	Scott Gradia	29656	MBP	Eidarus Salah
pGTVL2	Pavel Savitsky (SGC)	N/A	GST	Eidarus Salah
pSUMO-LIC	SGC Vector Bank	N/A	SUMO	Eidarus Salah
pNH-TrxT	Opher Gileadi (SGC)	26106	Trx	Eidarus Salah

Table 7.12 | Gibson Assembly Primers. Overhang sequences are in bold.

Reaction	Primer		Sequence
	Template	Direction	
His_6-ProPPE into pCOLD-I	Insert	Forward	5’-CAT CAT CAC CAT CAT CAT AC-3’
		Reverse	5’-**AC CTA TCT AGA TGC GGC CGC** TTA GTT GCT GGC AAT CAC-3’
	Vector	Forward	5’-GCG GCC GCA TCT AGA TAG-3’
		Reverse	5’-**GT ATG ATG ATG GTG ATG ATG** CAT GGT GTA TTA CCT CTT AAT AAT TAA GTG-3’
MBP into ProPPE-His_6: pCOLD-I	Insert	Forward	5’-**TT AAG AGG TAA TAC ACC** ATG AAA ATC GAA GAA GGT AAA CTG GTA ATC TGG-3'
		Reverse	5’-**GT TTC TGG AAA ATC CTG GGT** GGA TTG GAA GTA CAG GTT TTC CTC G-3'
	Vector	Forward	5’- **AA AAC CTG TAC TTC CAA TCC** ACC CAG GAT TTT CCA GAA ACC AAT G-3'
		Reverse	5’-**AG TTT ACC TTC TTC GAT TTT** CAT GGT GTA TTA CCT CTT AAT AAT TAA GTG TGC C-3'
GST into ProPPE-His_6: pCOLD-I	Insert	Forward	5’-**TT AAG AGG TAA TAC ACC** ATG TCC CCT ATA CTA GGT TAT TGG AAA ATT AAG GG-3'
		Reverse	5’-**GT TTC TGG AAA ATC CTG GGT** GGA TTG GAA GTA CAG GTT TTC CTC GA-3'
	Vector	Forward	5’-**AA AAC CTG TAC TTC CAA TCC** ACC CAG GAT TTT CCA GAA ACC AAT-3'
		Reverse	5’-**CA ATA ACC TAG TAT AGG GGA** CAT GGT GTA TTA CCT CTT AAT AAT TAA GTG TGC-3'

SUMO into ProPPE-His_6: pCOLD-I	Insert	Forward	5′-**TT AAG AGG TAA TAC ACC** ATG TCG GAC TCA GAA GTC AAT CAA GAA GC-3'
		Reverse	5′-**GT TTC TGG AAA ATC CTG GGT** AGA TCC ACC AAT CTG TTC TCT GTG AGC-3'
	Vector	Forward	5′-**G AGA CAG ATT GGT GGA TCT** ACC CAG GAT TTT CCA GAA ACC-3'
		Reverse	5′-**TG ATT GAC TTC TGA GTC CGA** CAT GGT GTA TTA CCT CTT AAT AAT TAA GTG TGC C-3'
Trx into ProPPE-His_6: pCOLD-I	Insert	Forward	5′-**TT AAG AGG TAA TAC ACC** ATG AGC GAT AAA ATT ATT CAC CTG ACT GAC G-3'
		Reverse	5′-**GT TTC TGG AAA ATC CTG GGT** GGA TTG GAA GTA CAA GTT CTC GGT ACC-3'
	Vector	Forward	5′- **AG AAC TTG TAC TTC CAA TCC** ACC CAG GAT TTT CCA GAA ACC AAT GC -3'
		Reverse	5′- **AG GTG AAT AAT TTT ATC GCT** CAT GGT GTA TTA CCT CTT AAT AAT TAA GTG TGC CTT TCG GCG A -3'
MBP into ProPPE-His_6: pNIC28	Insert	Forward	5′-**AA GAA GGA GAT ATA CAT** ATG AAA ATC GAA GAA GGT AAA CTG GTA ATC TGG-3'
		Reverse	5′-**GT TTC TGG AAA ATC CTG GGT** GGA TTG GAA GTA CAG GTT TTC CTC G-3'
	Vector	Forward	5′-**AA AAC CTG TAC TTC CAA TCC** ACC CAG GAT TTT CCA GAA ACC-3'
		Reverse	5′-**AG TTT ACC TTC TTC GAT TTT** CAT ATG TAT ATC TCC TTC TTA AAG TTA AAC AAA ATT ATT TCT AGA GGG-3'
GST into ProPPE-His_6: pNIC28	Insert	Forward	5′-**AA GAA GGA GAT ATA CAT** ATG TCC CCT ATA CTA GGT TAT TGG AAA ATT AAG GG-3'
		Reverse	5′-**GT TTC TGG AAA ATC CTG GGT** GGA TTG GAA GTA CAG GTT TTC CTC G-3'
	Vector	Forward	5′-**AA AAC CTG TAC TTC CAA TCC** ACC CAG GAT TTT CCA GAA ACC-3'
		Reverse	5′-**CA ATA ACC TAG TAT AGG GGA** CAT ATG TAT ATC TCC TTC TTA AAG TTA AAC AAA ATT ATT TCT AGA GGG GA-3'
SUMO into ProPPE-His_6: pNIC28	Insert	Forward	5′-**AA GAA GGA GAT ATA CAT** ATG TCG GAC TCA GAA GTC AAT CAA GAA GC-3'
		Reverse	5′-**GT TTC TGG AAA ATC CTG GGT** AGA TCC ACC AAT CTG TTC TCT GTG AGC-3'
	Vector	Forward	5′-**GA GAA CAG ATT GGT GGA TCT** ACC CAG GAT TTT CCA GAA ACC-3'
		Reverse	5′-**TG ATT GAC TTC TGA GTC CGA** CAT ATG TAT ATC TCC TTC TTA AAG TTA AAC AAA ATT ATT TCT AGA GGG-3'
Trx into ProPPE-His_6: pNIC28	Insert	Forward	5′-**AA GAA GGA GAT ATA CAT** ATG AGC GAT AAA ATT ATT CAC CTG ACT GAC G-3'
		Reverse	5′-**GT TTC TGG AAA ATC CTG GGT** GGA TTG GAA GTA CAA GTT CTC GGT ACC-3'
	Vector	Forward	5′-**AG AAC TTG TAC TTC CAA TCC** ACC CAG GAT TTT CCA GAA ACC-3'
		Reverse	5′-**AG GTG AAT AAT TTT ATC GCT** CAT ATG TAT ATC TCC TTC TTA AAG TTA AAC AAA ATT ATT TCT AGA GGG-3'

Representative PCR compositions and reaction conditions are shown in Table 7.13 and Table 7.14, respectively. The annealing temperature for each reaction was optimised (calculated[377] and calculated –2 °C). The PCR products were separated by agarose gel electrophoresis. The desired bands were excised and purified using a GeneJET gel extraction kit according to the manufacturers protocol and the concentration determined using a NanoDrop™.

Table 7.13 | Representative PCR reaction composition for Gibson Assembly amplification with overhanging primers.

Component	Stock Concentration	Volume (µL)	Final Concentration
Template DNA	25 ng μL^{-1}	1	1 ng/µL
Forward primer	10 µM	1.25	500 nM
Reverse primer	10 µM	1.25	500 nM
Q5 reaction buffer	5×	5	1×
dNTPs	10 mM	0.5	200 µM
DMSO	100 %	0.5	2 %
Q5-HF DNA polymerase	—	0.5	—
Autoclaved MQ water	—	15	—
Final Volume	—	25	—

Table 7.14 | Representative PCR reaction conditions for Gibson Assembly amplification with overhanging primers. Annealing temperature was varied depending on the calculated Tm of the primers. Extension was varied according to the No. of base pairs of the PCR product and calculated according to the polymerase manufacturers protocol.

Step	Temperature (˚C)	Time	Cycles
Initial denaturation	98	5 m	1
Denaturation	98	30 s	
Annealing	varied	30 s	30
Extension	72	varied	
Final Extension	72	10 m	1

7.4.3.2. Assembly and Transformation

Purified insert and vector DNA was mixed in 1:1 and 1:3 molar ratios in a 5 µL reaction volume (100 ng of vector). The 5 µL mixture was added to 5 µL of 2× Gibson Assembly Master Mix (Gibson Assembly® Cloning Kit; NEB) and incubated at 50 °C for 1 h. 2 µL of the mixture was added to 25 µL of NEB® 5-α cells, and transformed, purified, and verified as described in Section 7.3.6. **Note:** cells transformed with both pNIC28 and pCOLD-1 vectors were selected for on 2×YT agar plates with 30 µg mL^{-1} kanamycin.

7.4.4. Insertion and Deletion PCRs

7.4.4.1. DNA Amplification

The pCOLD-I and pNIC28 vectors containing His_6-ProPPE were modified by deletion and insertion PCR reactions to first remove the *N*–terminal His_6-tag and subsequently to add a *C*–terminal His_6-tag. The resultant amino acid sequences are shown in Figure 4.1. *N*–terminal signal peptides were also added to the His_6-ProPPE vectors using insertion PCR reactions. The signal peptide amino acid sequences are shown in Table 4.1. The primers used for the reactions are shown in Table 7.15. Representative PCR reaction composition and conditions are shown in Table 7.16 and Table 7.17, respectively. The

annealing temperature for each reaction was optimised (calculated[377] and calculated –2 °C). A small volume of PCR product was separated by agarose gel electrophoresis to confirm target amplification (Section 7.3.3).

Table 7.15 | Insertion and deletion PCR primers. Overhangs for the insertion reactions are in bold.

Reaction	Primer Direction	Sequence
pCOLD-I:His_6-ProPPE *N*–His_6 deletion	Forward	5′-ACC CAG GAT TTT CCA GAA ACC-3′
	Reverse	5′-CAT GGT GTA TTA CCT CTT AAT AAT TAA GTG TGC C-3′
pNIC28:His_6-ProPPE *N*–His_6 deletion	Forward	5′-ACC CAG GAT TTT CCA GAA AC-3′
	Reverse	5′-CAT ATG TAT ATC TCC TTC TTA AAG TTA AAC-3′
pCOLD-I:ProPPE *C*–His_6 insertion	Forward	5′-TAA GCG GCC GCA TCT AGA TAG G-3′
	Reverse	5′-**ATG ATG ATG GTG ATG ATG ACC GCC** GTT GCT GGC AAT CAC ATT GTT AAT CC-3′
pNIC28:ProPPE *C*–His_6 insertion	Forward	5′-TAA CAG TAA AGG TGG ATA CGG ATC C-3′
	Reverse	5′-**ATG ATG ATG GTG ATG ATG ACC GCC** ATT GCT GGC GAT CAC ATT ATT GAT CC-3′
pCOLD-I:His_6-ProPPE PhoA insertion	Forward	5′-**AAA CAA TCA ACA ATA GCT TTA GCA CTA TTG CCG CTG CTG TTC ACC CCG GTG ACG AAG GCG** ACC CAG GAT TTT CCA GAA ACC-3′
	Reverse	5′-CATGGTGTATTACCTCTTAATAATTAAGTGTGCC-3′
pNIC28:His_6-ProPPE DsbA insertion	Forward	5′-**AAA AAG ATA TGG TTA GCA CTA GCT GGA TTG GTG CTG GCA TTC TCC GCT AGC GCG** ACC CAG GAT TTT CCA GAA ACC-3′
	Reverse	5′-CAT ATG TAT ATC TCC TTC TTA AAG TTA AAC-3′
pNIC28:His_6-ProPPE OmpA insertion	Forward	5′-**AAA AAG ACA GCA ATT GCT ATA GCG GTA GCG CTG GCG GGT TTC GCC ACC GTG GCT CAG GCA** ACC CAG GAT TTT CCA GAA ACC-3′
	Reverse	5′-CAT ATG TAT ATC TCC TTC TTA AAG TTA AAC-3′
pNIC28:His_6-ProPPE PhoA insertion	Forward	5′-**AAA CAA TCA ACA ATA GCT TTA GCA CTA TTG CCG CTG CTG TTC ACC CCG GTG ACG AAG GCG** ACC CAG GAT TTT CCA GAA ACC-3′
	Reverse	5′-CAT ATG TAT ATC TCC TTC TTA AAG TTA AAC-3′

Table 7.16 | Representative PCR reaction composition for the insertion and deletion PCR reaction.

Component	Stock Concentration	Volume (µL)	Final Concentration
Template DNA	25 ng μL^{-1}	1	1 ng/µL
Forward primer	10 µM	1.25	500 nM
Reverse primer	10 µM	1.25	500 nM
Q5 reaction buffer	5×	5	1×
dNTPs	10 mM	0.5	200 µM
DMSO	100 %	0.5	2 %
Q5-HF DNA polymerase	—	0.5	—
Autoclaved MQ water	—	15	—
Final Volume	—	25	—

Table 7.17 | Representative PCR reaction conditions for the insertion and deletion PCR reaction. Annealing temperature was varied depending on the calculated Tm of the primers. Extension time was varied according to the No. of base pairs of the PCR product and calculated according to the polymerase manufacturers protocol.

Step	Temperature (˚C)	Time	Cycles
Initial denaturation	98	5 m	1
Denaturation	98	30 s	
Annealing	varied	30 s	30
Extension	72	varied	
Final Extension	72	10 m	1

7.4.4.2. Blunt-End Ligation and Transformation

PCR products were ligated without further purification. Methylated template DNA was digested by DpnI, and blunt-ends were phosphorylated and ligated in the same reaction mixture using a KLD Enzyme Mix (NEB) using the composition shown in Table 7.18. Reaction mixtures were incubated at room temperature for 5 min. 2 µL of the mixture was added to 25 µL of XL10-Gold, and transformed, purified, and verified as described in Section 7.3.6. **Note:** cells transformed with both pNIC28 and pCOLD-1 vectors were selected for on 2×YT agar plates with 30 µg mL^{-1} kanamycin.

Table 7.18 | KLD Enzyme Mix composition for blunt-end ligation and circularisation of PCR products from the insertion and deletion PCR reactions.

Component	Concentration	Volume (µL)	Final Concentration
PCR product	Varied	1	Varied
KLD reaction buffer	2×	2.5	1×
KLD enzyme mix	10×	0.5	1×
Autoclaved MQ water	—	1 µL	—
Final volume	—	5 µL	—

7.5. Recombinant Protein Expression and Purification

7.5.1. Materials

BugBuster® Protein Extraction Reagent, DNase I (grade II from bovine pancreas; Cat No. 10104159001), trypsin (TPCK treated from bovine pancreas), PMSF (Cat No. P7627), Lysozyme from chicken egg white (Cat No. 62971), proline, arginine, K citrate tribasic monohydrate, trehalose dihydrate, and 10 kDa MWCO dialysis tubing were purchased from Merck. Mannitol, CTAB, and Sorbitol were purchased from Enzo Life Scientific. Betaine anhydrous, glycyl-glycine, and xylitol were purchased from Alfa Aesar. NDSB-201 was purchased from Insight Biotechnology. HisPur™ Ni-NTA Spin Columns (200 µL) and

Pierce Slide-A-Lyzer 10 kDa MWCO 3-12 mL volume dialysis cassettes were purchased from Thermo Fisher Scientific. 10 kDa MWCO centrifugal filters (3 and 15 mL) were purchased from Amicon. 0.2 µm syringe filters were purchased from Sarstedt. 5 mL HisTrap HP, 1 mL HiTrap SP cation exchange chromatography column, SP Sepharose HP resin, and Superdex® 75 resin were purchased from Cytiva (or previously GE Healthcare Life Sciences). All protein chromatography purification steps were carried out using ÄKTA Pure FPLC systems at 4 °C. All buffers for chromatography purposes were vacuum filtered through a 0.2 µm filter membrane (Merck) and degassed prior to use.

7.5.2. Cell Culture and Protein Expression

7.5.2.1. Overnight Starter Culture Preparation

Glycerol stocks of transformed cells were removed from storage and kept on dry ice. In a sterile laminar flow hood, a sterile pipette tip was used to scrape the surface of the glycerol stock and placed into 10 mL of 2×YT media supplemented with the appropriate antibiotic(s) in 50 mL Falcon tubes. Cultures were incubated at the appropriate temperature for the strain (Table 7.4), with agitation (200 rpm) for 8 h. Overnight starter cultures were prepared by inoculating 100 mL of 2×YT media supplemented with the appropriate antibiotic, with 1 mL of culture. The starter culture was incubated at the appropriate temperature for the strain (Table 7.4), with agitation (200 rpm) for 16 h.

7.5.2.2. General Cell Culture and Induction Methods

Expression conditions were optimised and thus conditions varied. The following provides a general method that is used unless explicitly stated otherwise. Cells cultures and gene expression was carried out in one of four vessels: 1 mL culture in 96 deep-well plates (covered with sterile self-adhesive breathable film (Starlab), 10 mL culture in 50 mL Falcon tubes, 100 mL culture in 250 mL Erlenmeyer flasks (covered with foam stopper and tin foil), or 600 mL culture in 2 L Erlenmeyer flasks (covered with foam stopper and tin foil). Autoclaved media supplemented with the appropriate antibiotic were inoculated with overnight starter culture (1:100 v/v) in a sterile laminar flow hood. The cultures were incubated at the appropriate temperature for the strain (Table 7.4), with agitation (180 rpm) to an

optical density 600 nm (OD_{600}) of ~0.6 (2×YT) or 0.8–1.0 (TB). Gene expression was induced by cold-shock for pCOLD-I vector (15–30 min) or addition of isopropyl-β-D-1-thiogalactopyranoside (IPTG, 1,000× stock; 0–500 µM final [C]) for pNIC28 vector in a sterile laminar flow hood. Induced cells were incubated at various temperatures for 18 h, with agitation (180 rpm). The expression temperature can be assumed to be 18 °C unless explicitly stated. Cells were harvested by centrifugation for 10 min at 7,741 g (4 °C). Cell pellets were stored at –80 °C.

7.5.3. Cell Lysis

7.5.3.1. Detergent Method for Expression Trials

Cell pellets from plate or small-scale expression trials were resuspended BugBuster® Protein Extraction Reagent supplemented with 0.01 mg mL^{-1} DNase I and 0.5 mg mL^{-1} lysozyme, in a ratio of 50 µL BugBuster to every 1 mL of cell culture. Lysis was allowed to proceed for 20 min at room temperature. Insoluble cell debris was centrifuged at 3,700 × *g* for 30 min (in plates) or 21,000 × *g* for 10 min (1 mL of lysate in 1.5 mL Eppendorf tubes), both at 4 °C. The soluble fraction (supernatant) was separated. The insoluble fraction was resuspended in 100 mM Tris pH 7.5 supplemented with 8 M urea. Soluble and insoluble samples were separated by SDS-PAGE (Section 7.3.4) and sometimes analysed by western blotting (Section 7.3.5).

7.5.3.2. High-Pressure Cell Disruption Method

Cell pellets from large-scale expressions were lysed using high-pressure cell disruption (CF1 model, Constant Cell Disruption Systems; F1000 Compact Recirculating Cooler, Julabo). The cell pellets were resuspended in relevant resuspension buffer (1:4 w/v) supplemented with 0.01 mg mL^{-1} DNase I and 0.5 mg mL^{-1} lysozyme at 4 °C. The cell disruptor was prepared by washing with 2× 100 mL MQ water, 2× 100 mL 70% v/v ethanol, 2x 100 mL MQ water, and equilibrated with 2× 100 mL of the relevant resuspension buffer, at 20,000 psi. The cells were lysed with a single pass at 20,000 psi. An additional 50 mL of resuspension buffer was passed through the cell disruptor and combined with the cell lysate. The lysate was supplemented with an additional 0.01 mg mL^{-1} DNaseI following lysis. The lysate was

centrifuged at 58,000 × *g* for 30 min at 4 °C, to pellet the insoluble cell debris. The supernatant containing the soluble fraction was passed through a 0.2 µm syringe filter prior to downstream purification steps. The cell disruptor was cleaned by washing with 2× 100 mL MQ water and 2× 70% v/v ethanol at 40,000 psi.

7.5.4. HisPur purification for Expression Trials

The soluble fractions from expression trials were sometimes purified by IMAC following cell lysis. Cell pellets were resuspended in binding buffer (100 mM Tris pH 7.5, 100 mM NaCl, 10 mM imidazole). HisPur™ Ni-NTA Spin Columns (200 µL) were prepared by washing 3× with 200 µL binding buffer, each time removing the solution by centrifuging the column at 1000 rpm in a benchtop centrifuge. 400 µL of each sample was loaded onto the column and incubated at 4 °C for 1 h on an end-over-end mixer at 10 rpm. The columns were centrifuged at 1000 rpm to remove the solution from the column. The column was washed 2× with 200 µL binding buffer, removing the solution each time by centrifugation at 1,000 rpm. Protein was eluted from the column in HisTrap elution buffer. The eluted samples were then separated by SDS-PAGE (Section 7.3.4) and sometimes analysed by western blotting (Section 7.3.5).

7.5.5. Ammonium Sulphate Precipitation

Recombinant ProPPE-His_6 was purified as a first step by ammonium sulphate precipitation. The filtered soluble fraction was supplemented with ammonium sulphate with 30% w/v causing substantial precipitation of ProPPE-His_6. The precipitate was pelleted at 38,000 × *g* in 250 mL or 58,000 × *g* in 50 mL tubes. Pellets were resuspended in the appropriate binding buffer for downstream purification steps to the original volume or more.

7.5.6. HisTrap Immobilised Metal Affinity Chromatography (IMAC)

The buffers used for IMAC are shown in Table 7.19. The HisTrap HP columns (1 or 5 mL) were stripped and recharged with nickel(II) sulphate ($NiSO_4$) using the method in Table 7.20. Samples (filtered lysate

or pre-purified) were loaded onto a 5 mL HisTrap HP column at 1–2 mL min^{-1}. The column was washed with 5 CV of binding buffer at 4 mL min^{-1}, followed by 10-15 CV of wash buffer at 4 mL min^{-1}. Protein was eluted using a gradient from 0–100% elution buffer over 20 CV at 4 mL min^{-1}. The protein content of peak fractions was separated and analysed by SDS-PAGE and western blotting.

Table 7.19 | Composition for ProPPE HisTrap HP IMAC purification buffers. † Buffers used in purification 3 were run in parallel with identical buffers containing 10 mM $CaCl_2$.

Purification	Step	Buffering agent ([C] mM)	NaCl (mM)	Imidazole (mM)	Name
1	Pre-activation	Tris pH 8.5 (100)	100	10	Binding
				20	Wash
				250	Elution
	Post-activation	Tris pH 7.5 (50)	100	10	Binding
				20	Wash
				250	Elution
2	Pre-activation	Tris pH 8.5 (100)	500	10	Binding
				20	Wash
				250	Elution
3†	Pre-activation	Tris pH 8.5 (100)	–	10	Binding
				10	Wash
				250	Elution
	On-column activation	Tris pH 7.5 (50)	–	10	Binding
				10	Wash
				250	Elution

Table 7.20 | Maintenance protocol for stripping and recharging HisTrap HP columns (1 or 5 mL). CV = column volumes. † NaOH had a 1 h contact time with the column.

Component	Composition / Concentration	Amount (CV)
Stripping buffer	Na_2HPO_4 (20 mM), NaCl (500 mM), EDTA (50 mM), pH 7.4	3
MQ water	—	10
NaCl	1 M	5
MQ water	—	10
NaOH	1 M	2†
MQ water	—	10
$NiSO_4$	100 mM	3
MQ water	—	10
Binding buffer	Varied	5

7.5.6.1. On-Column Activation HisTrap IMAC Purification

Pre-purified sample was loaded onto a 5 mL HisTrap HP column, preequilibrated with binding buffer, at 1 mL min^{-1}. The column was washed with 10 CV of wash buffer. TR (0.04–0.05 eq) was loaded onto the column at 0.5 ml min^{-1} (15 min contact time). The column was washed with 20 CV of wash buffer

at 4 mL min^{-1}. Protein was eluted from the column using a 3-step protocol: 5% elution buffer (2 CV), 50% elution buffer (10 CV), 100% elution buffer (5 CV). The protein content of UV peak fractions was separated and analysed by SDS-PAGE and western blotting.

7.5.7. Cation Exchange Chromatography (CEX)

CEX purification of recombinant protein was carried out using either a prepacked 1 mL HiTrap SP column or an 8 mL SP Sepharose HP column (packed in-house by Dr Adam Hardy, Schofield group, University of Oxford). Representative buffers for cation exchange chromatography are shown in Table 7.21. Protein samples were loaded onto the columns, pre-equilibrated with 5 CV of binding buffer, at 1 mL min^{-1}. The column was washed with 10-15 CV equilibration buffer and the protein eluted using a gradient to 100% elution buffer over 40 CV. The protein content of UV peak fractions was separated and analysed by SDS-PAGE and western blotting.

Table 7.21 | Representative cation exchange chromatography buffer compositions.

Purification	Name	Buffering Agent	[NaCl] (mM)
2	Equilibration buffer	50 mM HEPES buffer pH 7.2	0
	Elution	50 mM HEPES buffer pH 7.2	500
Low pH Purification	Equilibration buffer	100 mM acetate buffer pH 5.0	0
	Elution	100 mM acetate buffer pH 5.0	500

7.5.8. Size Exchange Chromatography (SEC)

Recombinant protein samples were purified by SEC using a 120 mL S75 column (packed in-house by Dr Adam Hardy). The samples were concentrated using centrifugal filters (10 kDa MWCO, 15 and 3 mL) at 3,700 × *g*, 4 °C, for 5-10 min cycles, prior to SEC. The samples were clarified at 21,000 × g for 5 min at 4 °C prior to loading. All samples were loaded in a volume of ≤1% the column volume, i.e., ≤1.2 mL for the 120 mL column, from an appropriately sized capillary loop preequilibrated with SEC buffer. The column was washed with 1.3 CV of SEC buffer (50 mM Tris, 10 mM imidazole, pH 7.5) prior to sample loading. Concentrated sample was loaded at 0.5 mL min-1 and eluted with 1.3 CV of SEC buffer. The protein content of UV peak fractions was separated and analysed by SDS-PAGE and western blotting. The column was washed before and after use with the protocol in Table 7.22.

Table 7.22 | S75 column maintenance protocol. Note: SEC buffer was used to prepare column, EtOH was used to store the column after use.

Component	CV	Flow rate (mL min^{-1})
MQ water	1.3	1
NaOH (500 mM)	1.3	0.5
MQ water	1.3	1
SEC buffer or EtOH (20% v/v)	1.3	1

7.5.9. Enzyme Storage

Protein samples ready for storage were pooled and concentrated to an appropriate concentration using centrifugal filters (10 kDa MWCO, 15 and 3 mL) at 3,700 × *g*, 4 °C, for 5–10 min cycles. The concentration was determined using a NanoDrop™. Protein was aliquoted and flash frozen in N_2(l), and stored at –80 °C.

7.6. Enzyme Absorbance Assays

7.6.1. Materials

Lyophilised Porcine Pancreatic Elastase (200 U/mg) was purchased from Serva Electrophoresis GmbH (Cat No. 20930) and was further purified by SEC in 50 mM acetate buffer pH 5.0 and stored at –80 °C at 5.27 mg mL^{-1}. Succinyl-Ala-Ala-Pro-Ala-pNA (AAPA) was purchased from BACHEM (Cat No. 4015680.0250) and stored at 8 mM in 100 mM Tris pH 7.5. Clear, UV-transparent, 384-well plates were purchased from Greiner Bio-One (Cat No. 781801).

7.6.2. pNA Standard Curve

25x Stock concentrations of pNA were prepared in DMSO. The stocks were diluted 25x in Tris pH 7.5 (final [DMSO] = 4% v/v and 50 µL was aliquoted in a 384-well plate (four experimental replicates). The absorbance was measured at 405 nm.

7.6.3. Kinetic Measurements

Enzyme kinetics assays were performed using a CLARIOstar Plus spectrophotometer (BMG LABTECH). The hydrolysis of Suc-AAPA-pNA substrate by PPE was measured at 405 nm, 25 °C. The assay volume

was 50 µL. The measurements were performed in 100 mM Tris pH 7.5. The concentration of PPE for the assays was 3.86 nM. A 2× dilution series of substrate was prepared manually with concentrations between 46 µM (~1/4 K_m) to 3 mM (15× K_m). The enzyme and substrate solutions were prepared as 2× stocks in 100 mM Tris pH 7.5. 25 µL of substrate was added to a 384-well plate using a Voyager adjustable electronic pipette (Integra). 25 µL of buffer was added to negative control wells using a Multidrop™ Combi Reagent Dispenser (Thermo Fisher Scientific) equipped with a small tube metal tip dispensing cassette (0.5-50 µL; Thermo Fisher Scientific, Cat No. 24073295) for background determination. The reaction was initiated by addition of 25 µL PPE to the substrate using the dispenser and dispensing cassette. Absorbance at 405 nm was measured every 30s for 2 min. Data were analysed using GraphPad Prism (9.5.0) software. Mean and standard deviation values were plotted. K_m, $V_{max,}$ and k_{cat} values were calculated using "k_{cat}" non-linear regression model.

7.6.4. Enzyme Inhibition Assay

Components for enzyme inhibition assays were prepared as in Table 7.23. The final volume for assays was 50 µL. PPE and AAPA were prepared as 5× stocks. Inhibitors were prepared at 25× stocks in a 96-well tray. A 3× inhibitor dilution series (10 concentrations) was manually prepared in DMSO. Two wells for each compound contained only DMSO for a positive (no inhibitor) and negative (no inhibitor and no enzyme) control. Buffer (100 mM Tris pH 7.5) and PPE were added to a 384-well plate using a Multidrop™ Combi Reagent Dispenser equipped with a small tube metal tip dispensing cassette. Inhibitor (or DMSO) was transferred to the 384-well plate using an Integra VIAFLO-384 electronic multipipette. PPE and inhibitor was incubated for 10 min prior to addition of substrate. AAPA was added using a Multidrop™ Combi Reagent Dispenser equipped with a small tube metal tip dispensing cassette. Separate cassettes were used for dispensing enzyme and substrate. Reactions were carried out at room-temperature for 2 min. 5 µL of 10% v/v formic acid was added to quench the reactions. Absorbance at 405 nm was measured using a CLARIOstar Plus spectrophotometer. Data were analysed using GraphPad Prism (9.5.0) software. Mean and standard deviation values were plotted. Data were

fitted using the "log(inhibitor) *vs*. normalised response - Variable slope" non-linear regression model.

Table 7.23 | PPE inhibition assay composition.

Component	Stock concentration	Final Concentration
PPE in 100 mM Tris pH 7.5	5×	3.86 nM
AAPA in 100 mM Tris pH 7.5	5×	250 µM
Inhibitor in DMSO	25×	0.01–200 µM

7.7. Protein LC-MS

7.7.1. Materials

A ProSwift RP-4H column (1 × 50 mm) column was purchased from Thermo Fisher Scientific, (Cat No. 069477). Acetonitrile (Cat No. 34851) was purchased from Sigma Aldrich. Water was purified using a Milli-Q® Direct 8 Water Purification System with a Progard® T3 Pretreatment Pack (Cat No. PR0G000T3) and a Millipak® Express 40 Filter (Cat No. MPGP02001).

7.7.2. Method

Protein LC-MS analyses were carried out using a Waters Xevo-G2-S Q-TOF mass spectrometer, equipped with an ESI lockspray™ source and Acquity Ultra Performance LC system (UPLC). 2.5 µL of each sample was loaded onto a ProSwift RP-4H column (1 × 50 mm) column. Enzyme reactions were performed at a final concentration of 2.5–10 µM. Inhibitors were added to various ratios. Where DMSO was used the final concentration was held at 4% v/v. Proteins were eluted at 0.5 µL min^{-1} using a gradient of acetonitrile in water (supplemented with 0.1% v/v formic acid). The elution method is shown in Table 7.24. Data were analysed using MassLynx (4.1) software (Waters) using the Maximum Entropy Deconvolution algorithm.

Table 7.24 | Protein LC-MS elution gradient.

Time (min)	Gradient (%)
0–10	5–95
10–12	95
12–13	95–5
13–15	5

Chapter 8: References

1. Radzicka, A. and Wolfenden, R. (1996) Rates of uncatalyzed peptide bond hydrolysis in neutral solution and the transition state affinities of proteases. *Journal of the American Chemical Society*, **118**, 6105-6109.
2. International Human Genome Sequencing, C. (2004) Finishing the euchromatic sequence of the human genome. *Nature*, **431**, 931-945.
3. Pertea, M., Shumate, A., Pertea, G. et al. (2018) CHESS: a new human gene catalog curated from thousands of large-scale RNA sequencing experiments reveals extensive transcriptional noise. *Genome Biology*, **19**, 208.
4. Pérez-Silva, J.G., Español, Y., Velasco, G. and Quesada, V. (2016) The Degradome database: expanding roles of mammalian proteases in life and disease. *Nucleic Acids Res*, **44**, D351-355.
5. Rawlings, N.D., Barrett, A.J., Thomas, P.D. et al. (2018) The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Res*, **46**, D624-d632.
6. Fujinaga, M., Cherney, M.M., Oyama, H., Oda, K. and James, M.N. (2004) The molecular structure and catalytic mechanism of a novel carboxyl peptidase from Scytalidium lignicolum. *Proc Natl Acad Sci U S A*, **101**, 3364-3369.
7. Kutzing., F. (1837) Microscopische Untersuchungen über die Hefe und Essigmutter, nebst mehreren andern dazu gehörigen vegetabilischen Gebilden. *J Prakt Chem*, **11**, 385-409.
8. Schwann, T. (1837) Vorläufige Mittheilung, betreffend Versuche über die Weingährung und Fäulniss. *Annalen der Physik*, **117**, 184-193.
9. Barnett, J.A. (2003) Beginnings of microbiology and biochemistry: the contribution of yeast research. *Microbiology*, **149**, 557-567.
10. Tajima, N., Kawai, F., Park, S.Y. and Tame, J.R. (2010) A novel intein-like autoproteolytic mechanism in autotransporter proteins. *J Mol Biol*, **402**, 645-656.
11. Rawlings, N.D., Barrett, A.J. and Bateman, A. (2011) Asparagine peptide lyases: a seventh catalytic type of proteolytic enzymes. *J Biol Chem*, **286**, 38321-38328.
12. Barrett, A.J., Rawlings, N.D. and Woessner, J.F. (2012) *Handbook of proteolytic enzymes*. San Diego: Elsevier Science & Technology, San Diego.
13. Kühne, W. (1877) Ueber das Verhalten versciedener organisirter und sog. ungeformter Fermente. *Verhandlungen Naturhistorisch-Medicinischen Vereins zu Heidelberg*, **1**.
14. Kühne, W. (1878) Erfahrungen und Bemerkungen über Enzyme und Fermente. *Untersuchungen an dem Physiologischen Institut der Universität Heidelberg*, **1**, 291-326.
15. Teich, M. (1981) Ferment or enzyme: what's in a name? *History and Philosophy of the Life Sciences*, **3**, 193-215.
16. Huber, E.M., Heinemeyer, W., Li, X. et al. (2016) A unified mechanism for proteolysis and autocatalytic activation in the 20S proteasome. *Nature Communications*, **7**, 10900.
17. Klein, T., Eckhard, U., Dufour, A., Solis, N. and Overall, C.M. (2018) Proteolytic cleavage—mechanisms, function, and "omic" approaches for a near-ubiquitous posttranslational modification. *Chemical Reviews*, **118**, 1137-1168.
18. Vernon, H.M. (1914) The Activation of Trypsinogen. *Biochem J*, **8**, 494-529.
19. Kunitz, M. and Northrop, J.H. (1935) Crystalline chymo-trypsin and chymo-trypsinogen: I. Isolation, crystallization, and general properties of a new proteoltic enzyme and its precursor. *J Gen Physiol*, **18**, 433-458.
20. Steiner, D.F., Park, S.-Y., Støy, J., Philipson, L.H. and Bell, G.I. (2009) A brief perspective on insulin production. *Diabetes, Obesity and Metabolism*, **11**, 189-196.
21. Dunkelberger, J.R. and Song, W.-C. (2010) Complement and its role in innate and adaptive immune responses. *Cell Research*, **20**, 34-50.

22. Smith, S.A., Travers, R.J. and Morrissey, J.H. (2015) How it all starts: initiation of the clotting cascade. *Crit Rev Biochem Mol Biol*, **50**, 326-336.
23. Safavi, F. and Rostami, A. (2012) Role of serine proteases in inflammation: Bowman–Birk protease inhibitor (BBI) as a potential therapy for autoimmune diseases. *Experimental and Molecular Pathology*, **93**, 428-433.
24. Chakraborti, S., Chakraborti, T. and Dhalla, N.S. (2017) *Proteases in human diseases*. Springer, Singapore.
25. Castilho, D.G., Chaves, A.F.A., Navarro, M.V. et al. (2018) Secreted aspartyl proteinase (PbSap) contributes to the virulence of Paracoccidioides brasiliensis infection. *PLOS Neglected Tropical Diseases*, **12**, e0006806.
26. Frees, D., Brøndsted, L. and Ingmer, H. (2013) In Dougan, D. A. (ed.), *Regulated Proteolysis in Microorganisms*. Springer Netherlands, Dordrecht, pp. 161-192.
27. Zephyr, J., Kurt Yilmaz, N. and Schiffer, C.A. (2021) Viral proteases: structure, mechanism and inhibition. *Enzymes*, **50**, 301-333.
28. Cannalire, R., Cerchia, C., Beccari, A.R., Di Leva, F.S. and Summa, V. (2022) Targeting SARS-CoV-2 proteases and polymerase for COVID-19 treatment: state of the art and future opportunities. *J Med Chem*, **65**, 2716-2746.
29. Fanuel, L., Thamm, I., Kostanjevecki, V. et al. (1999) Two new aminopeptidases from Ochrobactrum anthropi active on D-alanyl-p-nitroanilide. *Cellular and Molecular Life Sciences CMLS*, **55**, 812-818.
30. Buller, A.R., Labonte, J.W., Freeman, M.F. et al. (2012) Autoproteolytic activation of ThnT results in structural reorganization necessary for substrate binding and catalysis. *Journal of Molecular Biology*, **422**, 508-518.
31. Smits, S.L., Snijder, E.J. and de Groot, R.J. (2006) Characterization of a torovirus main proteinase. *J Virol*, **80**, 4157-4167.
32. Carter, P. and Wells, J.A. (1988) Dissecting the catalytic triad of a serine protease. *Nature*, **332**, 564-568.
33. Fülöp, V., Böcskei, Z. and Polgár, L. (1998) Prolyl oligopeptidase: an unusual β-propeller domain regulates proteolysis. *Cell*, **94**, 161-170.
34. Luo, Y., Pfuetzner, R.A., Mosimann, S. et al. (2001) Crystal structure of LexA: a conformational switch for regulation of self-cleavage. *Cell*, **106**, 585-594.
35. Matthews, D.A., Smith, W.W., Ferre, R.A. et al. (1994) Structure of human rhinovirus 3C protease reveals a trypsin-like polypeptide fold, RNA-binding site, and means for cleaving precursor polyprotein. *Cell*, **77**, 761-771.
36. Bazan, J.F. and Fletterick, R.J. (1988) Viral cysteine proteases are homologous to the trypsin-like family of serine proteases: structural and functional implications. *Proc Natl Acad Sci U S A*, **85**, 7872-7876.
37. Gorbalenya, A.E., Donchenko, A.P., Blinov, V.M. and Koonin, E.V. (1989) Cysteine proteases of positive strand RNA viruses and chymotrypsin-like serine proteases. *FEBS Letters*, **243**, 103-114.
38. Smith, E.L., Markland, F.S., Kasper, C.B. et al. (1966) The complete amino acid sequence of two types of subtilisin, BPN' and Carlsberg. *J Biol Chem*, **241**, 5974-5976.
39. Wright, C.S., Alden, R.A. and Kraut, J. (1969) Structure of Subtilisin BPN' at 2.5 Å Resolution. *Nature*, **221**, 235-242.
40. Hartley, B.S. (1979) Evolution of Enzyme Structure. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **205**, 443-452.
41. Ben-Shem, A., Fass, D. and Bibi, E. (2007) Structural basis for intramembrane proteolysis by rhomboid serine proteases. *Proceedings of the National Academy of Sciences*, **104**, 462-466.
42. Tsukada, H. and Blow, D.M. (1985) Structure of alpha-chymotrypsin refined at 1.68 Å resolution. *J Mol Biol*, **184**, 703-711.

43. Moult, J., Sussman, F. and James, M.N. (1985) Electron density calculations as an extension of protein structure refinement. Streptomyces griseus protease A at 1.5 A resolution. *J Mol Biol*, **182**, 555-566.
44. Mosimann, S.C., Cherney, M.M., Sia, S., Plotch, S. and James, M.N. (1997) Refined X-ray crystallographic structure of the poliovirus 3C gene product. *J Mol Biol*, **273**, 1032-1047.
45. Fisher, S.J., Blakeley, M.P., Cianci, M., McSweeney, S. and Helliwell, J.R. (2012) Protonation-state determination in proteins using high-resolution X-ray crystallography: effects of resolution and completeness. *Acta Crystallographica Section D*, **68**, 800-809.
46. Banbula, A., Yen, J., Oleksy, A. et al. (2001) Porphyromonas gingivalis DPP-7 represents a novel type of dipeptidylpeptidase. *J Biol Chem*, **276**, 6299-6305.
47. Bompard-Gilles, C., Villeret, V., Davies, G.J. et al. (2000) A new variant of the Ntn hydrolase fold revealed by the crystal structure of L-aminopeptidase D-Ala-esterase/amidase from Ochrobactrum anthropi. *Structure*, **8**, 153-162.
48. Pal, A., Kraetzner, R., Gruene, T. et al. (2009) Structure of tripeptidyl-peptidase I provides insight into the molecular basis of late infantile neuronal ceroid lipofuscinosis. *Journal of Biological Chemistry*, **284**, 3976-3984.
49. Soisson, S.M., Patel, S.B., Abeywickrema, P.D. et al. (2010) Structural definition and substrate specificity of the S28 protease family: the crystal structure of human prolylcarboxypeptidase. *BMC Structural Biology*, **10**, 16.
50. Carter, T.H. and Miller, C.G. (1984) Aspartate-specific peptidases in Salmonella typhimurium: mutants deficient in peptidase E. *J Bacteriol*, **159**, 453-459.
51. Schechter, I. and Berger, A. (1968) On the active site of proteases. III. Mapping the active site of papain; specific peptide inhibitors of papain. *Biochemical and Biophysical Research Communications*, **32**, 898-902.
52. Rawlings, N.D. (2020) Twenty-five years of nomenclature and classification of proteolytic enzymes. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, **1868**, 140345.
53. Bode, W., Meyer, E., Jr. and Powers, J.C. (1989) Human leukocyte and porcine pancreatic elastase: X-ray crystal structures, mechanism, substrate specificity, and mechanism-based inhibitors. *Biochemistry*, **28**, 1951-1963.
54. Brändén, C.-I. and Tooze, J. (2009) *Introduction to protein structure*. Second Edition. ed. Garland Publishing, New York.
55. Zheng, X.L., Kitamoto, Y. and Sadler, J.E. (2009) Enteropeptidase, a type II transmembrane serine protease. *Front Biosci (Elite Ed)*, **1**, 242-249.
56. Lu, D., Fütterer, K., Korolev, S. et al. (1999) Crystal structure of enteropeptidase light chain complexed with an analog of the trypsinogen activation peptide. *Journal of Molecular Biology*, **292**, 361-373.
57. Goettig, P., Brandstetter, H. and Magdolen, V. (2019) Surface loops of trypsin-like serine proteases as determinants of function. *Biochimie*, **166**, 52-76.
58. Guo, S., Skala, W., Magdolen, V. et al. (2016) A single glycan at the 99-loop of human kallikrein-related peptidase 2 regulates activation and enzymatic activity. *J Biol Chem*, **291**, 593-604.
59. Ohlsson, K. and Olsson, I. (1974) The neutral proteases of human granulocytes. Isolation and partial characterization of granulocyte elastases. *Eur J Biochem*, **42**, 519-527.
60. Dollery, C.M., Owen, C.A., Sukhova, G.K. et al. (2003) Neutrophil elastase in human atherosclerotic plaques. *Circulation*, **107**, 2829-2836.
61. Dewald, B., Rindler-Ludwig, R., Bretz, U. and Baggiolini, M. (1975) Subcellular localization and heterogeneity of neutral proteases in neutrophilic polymorphonuclear leukocytes. *Journal of Experimental Medicine*, **141**, 709-723.
62. Turk, B., Turk, D. and Turk, V. (2012) Protease signalling: the cutting edge. *Embo j*, **31**, 1630-1643.
63. van Gent, D., Sharp, P., Morgan, K. and Kalsheker, N. (2003) Serpins: structure, function and molecular evolution. *The International Journal of Biochemistry & Cell Biology*, **35**, 1536-1547.

64. Lucas, A., Yaron, J.R., Zhang, L. and Ambadapadi, S. (2018) In Lucas, A. (ed.), *Serpins: Methods and Protocols*. Springer New York, New York, NY, pp. 1-7.
65. Wang, K., Meng, X. and Guo, Z. (2021) Elastin structure, synthesis, regulatory mechanism and relationship with cardiovascular diseases. *Frontiers in Cell and Developmental Biology*, **9**.
66. Vrhovski, B. and Weiss, A.S. (1998) Biochemistry of tropoelastin. *European Journal of Biochemistry*, **258**, 1-18.
67. Shapiro, S.D., Endicott, S.K., Province, M.A., Pierce, J.A. and Campbell, E.J. (1991) Marked longevity of human lung parenchymal elastic fibers deduced from prevalence of D-aspartate and nuclear weapons-related radiocarbon. *The Journal of Clinical Investigation*, **87**, 1828-1834.
68. Heinz, A. (2020) Elastases and elastokines: elastin degradation and its significance in health and disease. *Critical Reviews in Biochemistry and Molecular Biology*, **55**, 252-273.
69. Baló, J. and Banga, I. (1949) Elastase and Elastase-Inhibitor. *Nature*, **164**, 491-491.
70. Baló, J. and Banga, I. (1950) The elastolytic activity of pancreatic extracts. *Biochem J*, **46**, 384-387.
71. Talas, U., Dunlop, J., Khalaf, S., Leigh, I.M. and Kelsell, D.P. (2000) Human elastase 1: evidence for expression in the skin and theidentification of a frequent frameshift polymorphism. *Journal of Investigative Dermatology*, **114**, 165-170.
72. Largman, C., Brodrick, J.W. and Geokas, M.C. (1976) Purification and characterization of two human pancreatic elastases. *Biochemistry*, **15**, 2491-2500.
73. Kawashima, I., Tani, T., Shimoda, K. and Takiguchi, Y. (1987) Characterization of pancreatic elastase II cDNAs: two elastase II mRNAs are expressed in human pancreas. *DNA*, **6**, 163-172.
74. Szepessy, E. and Sahin-Tóth, M. (2006) Inactivity of recombinant ELA2B provides a new example of evolutionary elastase silencing in humans. *Pancreatology*, **6**, 117-122.
75. Mallory, P.A. and Travis, J. (1975) Human pancreatic enzymes: purification and characterization of a nonelastolytic enzyme, protease E. resembling elastase. *Biochemistry*, **14**, 722-730.
76. Tani, T., Ohsumi, J., Mita, K. and Takiguchi, Y. (1988) Identification of a novel class of elastase isozyme, human pancreatic elastase III, by cDNA and genomic gene cloning. *Journal of Biological Chemistry*, **263**, 1231-1239.
77. Gertler, A. and Birk, Y. (1970) Isolation and characterization of porcine proelastase. *Eur J Biochem*, **12**, 170-176.
78. Shotton, D.M. and Hartley, B.S. (1970) Amino-acid sequence of porcine pancreatic elastase and its homologies with other serine proteinases. *Nature*, **225**, 802-806.
79. Adkison, A.M., Raptis, S.Z., Kelley, D.G. and Pham, C.T. (2002) Dipeptidyl peptidase I activates neutrophil-derived serine proteases and regulates the development of acute experimental arthritis. *J Clin Invest*, **109**, 363-371.
80. Sinha, S., Watorek, W., Karr, S. et al. (1987) Primary structure of human neutrophil elastase. *Proc Natl Acad Sci U S A*, **84**, 2228-2232.
81. Shotton, D.M. (1970), *Methods in Enzymology*. Academic Press, Vol. 19, pp. 113-140.
82. Korkmaz, B., Horwitz, M.S., Jenne, D.E. and Gauthier, F. (2010) Neutrophil elastase, proteinase 3, and cathepsin G as therapeutic targets in human diseases. *Pharmacological Reviews*, **62**, 726.
83. Brink, N.G., Lewis, U.J. and Williams, D.E. (1956) Pancreatic elastase: purification, properties, and function. *J Biol Chem*, **222**, 705-720.
84. Korkmaz, B., Moreau, T. and Gauthier, F. (2008) Neutrophil elastase, proteinase 3 and cathepsin G: Physicochemical properties, activity and physiopathological functions. *Biochimie*, **90**, 227-242.
85. Gertler, A. and Hofmann, T. (1967) The involvement of the amino-terminal amino acid in the activity of pancreatic proteases: I. The effects of nitrous acid on elastase. *Journal of Biological Chemistry*, **242**, 2522-2527.

86. Shotton, D.M., Hartley, B.S., Camerman, N. et al. (1968) Crystalline porcine pancreatic elastase. *Journal of Molecular Biology*, **32**, 155-IN158.
87. Shotton, D.M. and Watson, H.C. (1970) Three-dimensional structure of tosyl-elastase. *Nature*, **225**, 811-816.
88. Meyer, E., Cole, G., Radhakrishnan, R. and Epp, O. (1988) Structure of native porcine pancreatic elastase at 1.65 A resolutions. *Acta Crystallogr B*, **44 (Pt 1)**, 26-38.
89. Würtele, M., Hahn, M., Hilpert, K. and Höhne, W. (2000) Atomic resolution structure of native porcine pancreatic elastase at 1.1 A. *Acta Crystallogr D Biol Crystallogr*, **56**, 520-523.
90. Tamada, T., Kinoshita, T., Kurihara, K. et al. (2009) Combined high-resolution neutron and X-ray analysis of inhibited elastase confirms the active-site oxyanion hole but rules against a low-barrier hydrogen bond. *Journal of the American Chemical Society*, **131**, 11033-11040.
91. Bode, W., Wei, A.Z., Huber, R. et al. (1986) X-ray crystal structure of the complex of human leukocyte elastase (PMN elastase) and the third domain of the turkey ovomucoid inhibitor. *Embo j*, **5**, 2453-2458.
92. Hansen, G., Gielen-Haertwig, H., Reinemer, P. et al. (2011) Unexpected active-site flexibility in the structure of human neutrophil elastase in complex with a new dihydropyrimidone inhibitor. *Journal of Molecular Biology*, **409**, 681-691.
93. Bode, W. and Schwager, P. (1975) The single calcium-binding site of crystallin bovin beta-trypsin. *FEBS Lett*, **56**, 139-143.
94. Freer, S.T., Kraut, J., Robertus, J.D., Wright, H.T. and Xuong, N.H. (1970) Chymotrypsinogen: 2.5-angstrom crystal structure, comparison with alpha-chymotrypsin, and implications for zymogen activation. *Biochemistry*, **9**, 1997-2009.
95. Sawyer, L., Shotton, D.M., Campbell, J.W. et al. (1978) The atomic structure of crystalline porcine pancreatic elastase at 2.5 Å resolution: Comparisons with the structure of α-chymotrypsin. *Journal of Molecular Biology*, **118**, 137-208.
96. Meyer, E.F., Jr., Clore, G.M., Gronenborn, A.M. and Hansen, H.A. (1988) Analysis of an enzyme-substrate complex by X-ray crystallography and transferred nuclear Overhauser enhancement measurements: porcine pancreatic elastase and a hexapeptide. *Biochemistry*, **27**, 725-730.
97. Narayanan, A.S. and Anwar, R.A. (1969) The specificity of purified porcine pancreatic elastase. *Biochem J*, **114**, 11-17.
98. Kasafírek, E., Fric, P., Slabý, J. and Malis, F. (1976) p-Nitroanilides of 3-carboxypropionyl-peptides. Their cleavage by elastase, trypsin, and chymotrypsin. *Eur J Biochem*, **69**, 1-13.
99. Powers, J.C., Gupton, B.F., Harley, A.D., Nishino, N. and Whitley, R.J. (1977) Specificity of porcine pancreatic elastase, human leukocyte elastase and cathepsin G Inhibition with peptide chloromethyl ketones. *Biochimica et Biophysica Acta (BBA) - Enzymology*, **485**, 156-166.
100. Thompson, R.C. and Blout, E.R. (1970) Evidence for an extended active center in elastase. *Proc Natl Acad Sci U S A*, **67**, 1734-1740.
101. Atlas, D. (1975) The active site of porcine elastase. *J Mol Biol*, **93**, 39-53.
102. Thompson, R.C. and Blout, E.R. (1973) Restrictions on the binding of proline-containing peptides to elastase. *Biochemistry*, **12**, 51-57.
103. McRae, B., Nakajima, K., Travis, J. and Powers, J.C. (1980) Studies on reactivity of human leukocyte elastase, cathepsin G, and porcine pancreatic elastase toward peptides including sequences related to the reactive site of .alpha.1-protease inhibitor (.alpha.1-antitrypsin). *Biochemistry*, **19**, 3973-3978.
104. Stein, R.L., Strimpler, A.M., Hori, H. and Powers, J.C. (1987) Catalysis by human leukocyte elastase: mechanistic insights into specificity requirements. *Biochemistry*, **26**, 1301-1305.
105. Del Mar, E.G., Largman, C., Brodrick, J.W., Fassett, M. and Geokas, M.C. (1980) Substrate specificity of human pancreatic elastase 2. *Biochemistry*, **19**, 468-472.
106. Wilmouth, R.C., Clifton, I.J., Robinson, C.V. et al. (1997) Structure of a specific acyl-enzyme complex formed between beta-casomorphin-7 and porcine pancreatic elastase. *Nat Struct Biol*, **4**, 456-462.

107. Wright, P.A., Wilmouth, R.C., Clifton, I.J. and Schofield, C.J. (2001) Kinetic and crystallographic analysis of complexes formed between elastase and peptides from beta-casein. *Eur J Biochem*, **268**, 2969-2974.
108. Renaud, A., Lestienne, P., Hughes, D.L., Bieth, J.G. and Dimicoli, J.L. (1983) Mapping of the S' subsites of porcine pancreatic and human leucocyte elastases. *J Biol Chem*, **258**, 8312-8316.
109. Polgár, L. (2005) The catalytic triad of serine peptidases. *Cellular and Molecular Life Sciences CMLS*, **62**, 2161-2172.
110. Jansen, E.F., Nutting, F. and Balls, A.K. (1949) Mode of inhibition of chymotrypsin by diisopropyl fluorophosphate; introduction of phosphorus. *J Biol Chem*, **179**, 201-204.
111. Schaffer, N.K., May, S.C., Jr. and Summerson, W.H. (1953) Serine phosphoric acid from diisopropylphosphoryl chymotrypsin. *J Biol Chem*, **202**, 67-76.
112. Gutfreund, H. (1955) The characterization of the catalytic site of trypsin. *Transactions of the Faraday Society*, **51**, 441-446.
113. Cunningham, L.W. and Brown, C.S. (1956) The influence of pH on the kinetic constants of alpha-chymotrypsin-catalyzed esterolysis. *J Biol Chem*, **221**, 287-299.
114. Schoellmann, G. and Shaw, E. (1963) Direct evidence for the presence of histidine in the active center of chymotrypsin. *Biochemistry*, **2**, 252-255.
115. Matthews, B.W., Sigler, P.B., Henderson, R. and Blow, D.M. (1967) Three-dimensional Structure of Tosyl-α-chymotrypsin. *Nature*, **214**, 652-656.
116. Sigler, P.B., Blow, D.M., Matthews, B.W. and Henderson, R. (1968) Structure of crystalline α-chymotrypsin: II. A preliminary report including a hypothesis for the activation mechanism. *Journal of Molecular Biology*, **35**, 143-164.
117. Blow, D.M., Birktoft, J.J. and Hartley, B.S. (1969) Role of a buried acid group in the mechanism of action of chymotrypsin. *Nature*, **221**, 337-340.
118. Hedstrom, L. (2002) Serine protease mechanism and specificity. *Chem Rev*, **102**, 4501-4524.
119. Bachovchin, W.W. and Roberts, J.D. (1978) Nitrogen-15 nuclear magnetic resonance spectroscopy. The state of histidine in the catalytic triad of .alpha.-lytic protease. Implications for the charge-relay mechanism of peptide-bond cleavage by serine proteases. *Journal of the American Chemical Society*, **100**, 8041-8047.
120. Liang, T.C. and Abeles, R.H. (1987) Complex of α-chymotrypsin and N-acetyl-L-leucyl-L-phenylalanyl trifluoromethyl ketone: structural studies with NMR spectroscopy. *Biochemistry*, **26**, 7603-7608.
121. Zhong, S., Haghjoo, K., Kettner, C. and Jordan, F. (1995) Proton magnetic resonance studies of the active center histidine of chymotrypsin complexed to peptideboronic acids: Solvent accessibility to the N.delta. and N.epsilon. sites can differentiate slow-binding and rapidly reversible inhibitors. *Journal of the American Chemical Society*, **117**, 7048-7055.
122. Spink, E., Cosgrove, S., Rogers, L., Hewage, C. and Malthouse, J.P.G. (2007) 13C and 1H NMR Studies of Ionizations and Hydrogen Bonding in Chymotrypsin-Glyoxal Inhibitor Complexes*. *Journal of Biological Chemistry*, **282**, 7852-7861.
123. Howe, N., Rogers, L., Hewage, C. and Malthouse, J.P. (2009) Oxyanion and tetrahedral intermediate stabilisation by subtilisin: detection of a new tetrahedral adduct. *Biochim Biophys Acta*, **1794**, 1251-1258.
124. Petrillo, T., O'Donohoe, C.A., Howe, N. and Malthouse, J.P.G. (2012) Importance of tetrahedral intermediate formation in the catalytic mechanism of the serine proteases chymotrypsin and subtilisin. *Biochemistry*, **51**, 6164-6170.
125. Henderson, R. (1970) Structure of crystalline alpha-chymotrypsin. IV. The structure of indoleacryloyl-alpha-chyotrypsin and its relevance to the hydrolytic mechanism of the enzyme. *J Mol Biol*, **54**, 341-354.
126. Weiner, H., White, W.N., Hoare, D.G. and Koshland, D.E., Jr. (1966) The formation of anhydrochymotrypsin by removing the elements of water from the serine at the active site. *Journal of the American Chemical Society*, **88**, 3851-3859.

127. Nakagawa, Y. and Bender, M.L. (1970) Methylation of histidine-57 in α-chymotrypsin by methyl ρ-nitrobenzenesulfonate. New approach to enzyme modification. *Biochemistry*, **9**, 259-267.
128. Henderson, R. (1971) Catalytic activity of -chymotrypsin in which histidine-57 has been methylated. *Biochem J*, **124**, 13-18.
129. Corey, D.R. and Craik, C.S. (1992) An investigation into the minimum requirements for peptide hydrolysis by mutation of the catalytic triad of trypsin. *Journal of the American Chemical Society*, **114**, 1784-1790.
130. Carter, P. and Wells, J.A. (1990) Functional interaction among catalytic residues in subtilisin BPN'. *Proteins*, **7**, 335-342.
131. Carter, P. and Wells, J.A. (1987) Engineering enzyme specificity by "substrate-assisted catalysis". *Science*, **237**, 394-399.
132. Carter, P., Abrahmsen, L. and Wells, J.A. (1991) Probing the mechanism and improving the rate of substrate-assisted catalysis in subtilisin BPN'. *Biochemistry*, **30**, 6142-6148.
133. Corey, D., Willett, W., Coombs, G. and Craik, C. (1995) Trypsin specificity increased through substrate-assisted catalysis. *Biochemistry*, **34**, 11521-11527.
134. Craik, C.S., Roczniak, S., Largman, C. and Rutter, W.J. (1987) The catalytic role of the active site aspartic acid in serine proteases. *Science*, **237**, 909-913.
135. Sprang, S., Standing, T., Fletterick, R.J. et al. (1987) The three-dimensional structure of Asn102 mutant of trypsin: role of Asp102 in serine protease catalysis. *Science*, **237**, 905-909.
136. Corey, D.R., McGrath, M.E., Vasquez, J.R., Fletterick, R.J. and Craik, C.S. (1992) An alternate geometry for the catalytic triad of serine proteases. *Journal of the American Chemical Society*, **114**, 4905-4907.
137. Rao, S.N., Singh, U.C., Bash, P.A. and Kollman, P.A. (1987) Free energy perturbation calculations on binding and catalysis after mutating Asn 155 in subtilisin. *Nature*, **328**, 551-554.
138. Braxton, S. and Wells, J.A. (1991) The importance of a distal hydrogen bonding group in stabilizing the transition state in subtilisin BPN'. *Journal of Biological Chemistry*, **266**, 11797-11800.
139. Bryan, P., Pantoliano, M.W., Quill, S.G., Hsiao, H.Y. and Poulos, T. (1986) Site-directed mutagenesis and the role of the oxyanion hole in subtilisin. *Proceedings of the National Academy of Sciences*, **83**, 3743-3745.
140. Wells, J.A., Cunningham, B.C., Graycar, T.P. et al. (1986) Importance of hydrogen-bond formation in stabilizing the transition state of subtilisin. *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, **317**, 415-423.
141. Robillard, G. and Shulman, R.G. (1972) High resolution nuclear magnetic resonance study of the histidine--aspartate hydrogen bond in chymotrypsin and chymotrypsinogen. *J Mol Biol*, **71**, 507-511.
142. Bachovchin, W.W. (1985) Confirmation of the assignment of the low-field proton resonance of serine proteases by using specifically nitrogen-15 labeled enzyme. *Proc Natl Acad Sci U S A*, **82**, 7948-7951.
143. Hunkapiller, M.W., Smallcombe, S.H., Whitaker, D.R. and Richards, J.H. (1973) Ionization behavior of the histidine residue in the catalytic triad of serine proteases: mechanistic implications. *Journal of Biological Chemistry*, **248**, 8306-8308.
144. Hunkapiller, M.W., Smallcombe, S.H., Whitaker, D.R. and Richards, J.H. (1973) Carbon nuclear magnetic resonance studies of the histidine residue in α-lytic protease. Implications for the catalytic mechanism of serine proteases. *Biochemistry*, **12**, 4732-4743.
145. Kossiakoff, A.A. and Spencer, S.A. (1980) Neutron diffraction identifies His 57 as the catalytic base in trypsin. *Nature*, **288**, 414-416.
146. Frey, P.A., Whitt, S.A. and Tobin, J.B. (1994) A low-barrier hydrogen bond in the catalytic triad of serine proteases. *Science*, **264**, 1927-1930.

147. Everill, P., Sudmeier, J.L. and Bachovchin, W.W. (2012) Direct NMR observation and pKa determination of the Asp102 side chain in a serine protease. *J Am Chem Soc*, **134**, 2348-2354.
148. Ash, E.L., Sudmeier, J.L., De Fabo, E.C. and Bachovchin, W.W. (1997) A low-barrier hydrogen bond in the catalytic triad of serine proteases? Theory versus experiment. *Science*, **278**, 1128-1132.
149. Stratton, J.R., Pelton, J.G. and Kirsch, J.F. (2001) A novel engineered subtilisin BPN' lacking a low-barrier hydrogen bond in the catalytic triad. *Biochemistry*, **40**, 10411-10416.
150. Warshel, A., Naray-Szabo, G., Sussman, F. and Hwang, J.K. (1989) How do serine proteases really work? *Biochemistry*, **28**, 3629-3637.
151. Kemp, M.T., Lewandowski, E.M. and Chen, Y. (2021) Low barrier hydrogen bonds in protein structure and function. *Biochim Biophys Acta Proteins Proteom*, **1869**, 140557.
152. Kuhn, P., Knapp, M., Soltis, S.M. et al. (1998) The 0.78 Å structure of a serine protease: Bacillus lentus subtilisin. *Biochemistry*, **37**, 13446-13452.
153. Bachovchin, W.W. (2001) Review: contributions of NMR spectroscopy to the study of hydrogen bonds in serine protease active sites. *Magnetic Resonance in Chemistry*, **39**, S199-S213.
154. Robillard, G. and Shulman, R.G. (1974) High resolution nuclear magnetic resonance studies of the active site of chymotrypsin: I. The hydrogen bonded protons of the "charge relay" system. *Journal of Molecular Biology*, **86**, 519-540.
155. Bachovchin, W.W. (1986) 15N NMR spectroscopy of hydrogen-bonding interactions in the active site of serine proteases: evidence for a moving histidine mechanism. *Biochemistry*, **25**, 7751-7759.
156. Brayer, G.D., Delbaere, L.T. and James, M.N. (1979) Molecular structure of the alpha-lytic protease from Myxobacter 495 at 2.8 Angstroms resolution. *J Mol Biol*, **131**, 743-775.
157. Fujinaga, M., Delbaere, L.T.J., Brayer, G.D. and James, M.N.G. (1985) Refined structure of α-lytic protease at 1.7 Å resolution analysis of hydrogen bonding and solvent structure. *Journal of Molecular Biology*, **184**, 479-502.
158. Smith, S.O., Farr-Jones, S., Griffin, R.G. and Bachovchin, W.W. (1989) Crystal versus solution structures of enzymes: NMR spectroscopy of a crystalline serine protease. *Science*, **244**, 961-964.
159. James, M.N., Sielecki, A.R., Brayer, G.D., Delbaere, L.T. and Bauer, C.A. (1980) Structures of product and inhibitor complexes of Streptomyces griseus protease A at 1.8 A resolution. A model for serine protease catalysis. *J Mol Biol*, **144**, 43-88.
160. Schmidt, A., Jelsch, C., Østergaard, P., Rypniewski, W. and Lamzin, V.S. (2003) Trypsin revisited: crystallography at (sub) atomic resolution and quantum chemistry revealing details of catalysis. *Journal of Biological Chemistry*, **278**, 43357-43362.
161. Hartley, B.S. and Kilby, B.A. (1954) The reaction of p-nitrophenyl esters with chymotrypsin and insulin. *Biochem J*, **56**, 288-297.
162. Fastrez, J. and Fersht, A.R. (1973) Demonstration of the acyl-enzyme mechanism for the hydrolysis of peptides and anilides by chymotrypsin. *Biochemistry*, **12**, 2025-2034.
163. Dixon, M. and Matthews, B.W. (1989) Is γ-chymotrypsin a tetrapeptide acyl-enzyme adduct of α-chymotrypsin? *Biochemistry*, **28**, 7033-7038.
164. Burgi, H.B., Dunitz, J.D. and Shefter, E. (1973) Geometrical reaction coordinates. II. Nucleophilic addition to a carbonyl group. *Journal of the American Chemical Society*, **95**, 5065-5067.
165. Dixon, M.M., Brennan, R.G. and Matthews, B.W. (1991) Structure of γ -chymotrypsin in the range pH 2.0 to pH 10.5 suggests that gamma-chymotrypsin is a covalent acyl-enzyme adduct at low pH. *Int J Biol Macromol*, **13**, 89-96.
166. Brayer, G.D., Delbaere, L.T.J. and James, M.N.G. (1978) Molecular structure of crystalline Streptomyces griseus protease A at 2.8 Å resolution: I. Crystallization, data collection and structural analysis. *Journal of Molecular Biology*, **124**, 243-259.

167. Sielecki, A.R., Hendrickson, W.A., Broughton, C.G. et al. (1979) Protein structure refinement: Streptomyces griseus serine protease A at 1.8 Å resolution. *Journal of Molecular Biology*, **134**, 781-804.
168. Avbelj, F., Moult, J., Kitson, D.H., James, M.N.G. and Hagler, A.T. (1990) Molecular dynamics study of the structure and dynamics of a protein molecule in a crystalline ionic environment, Streptomyces griseus protease A. *Biochemistry*, **29**, 8658-8676.
169. Blanchard, H. and James, M.N.G. (1994) A crystallographic re-investigation into the structure of Streptomyces griseus proteinase A reveals an acyl-enzyme intermediate. *Journal of Molecular Biology*, **241**, 574-587.
170. Wilmouth, R.C., Edman, K., Neutze, R. et al. (2001) X-ray snapshots of serine protease catalysis reveal a tetrahedral intermediate. *Nat Struct Biol*, **8**, 689-694.
171. Radisky, E.S., Lee, J.M., Lu, C.-J.K. and Koshland, D.E. (2006) Insights into the serine protease mechanism from atomic resolution structures of trypsin reaction intermediates. *Proceedings of the National Academy of Sciences*, **103**, 6835-6840.
172. Tonge, P.J. and Carey, P.R. (1989) Direct observation of the titration of substrate carbonyl groups in the active site of. alpha.-chymotrypsin by resonance Raman spectroscopy. *Biochemistry*, **28**, 6701-6709.
173. Tonge, P.J. and Carey, P.R. (1990) Length of the acyl carbonyl bond in acyl-serine proteases correlates with reactivity. *Biochemistry*, **29**, 10723-10727.
174. Pauling, L. (1946) Molecular architecture and biological reactions. *Chemical & Engineering News Archive*, **24**, 1375-1377.
175. Pauling, L. (1948) Chemical achievement and hope for the future. *American Scientist*, **36**, 50-58.
176. Bender, M.L. (1960) Mechanisms of Catalysis of Nucleophilic Reactions of Carboxylic Acid Derivatives. *Chemical Reviews*, **60**, 53-113.
177. Kettner, C.A. and Shenvi, A.B. (1984) Inhibition of the serine proteases leukocyte elastase, pancreatic elastase, cathepsin G, and chymotrypsin by peptide boronic acids. *J Biol Chem*, **259**, 15106-15114.
178. Yennawar, N.H., Yennawar, H.P. and Farber, G.K. (1994) X-ray crystal structure of γ-chymotrypsin in hexane. *Biochemistry*, **33**, 7326-7336.
179. Zaks, A. and Klibanov, A.M. (1986) Substrate specificity of enzymes in organic solvents vs. water is reversed. *Journal of the American Chemical Society*, **108**, 2767-2768.
180. Gololobov, M., Voyushina, T.L., Stepanov, V.M. and Adlercreutz, P. (1992) Organic solvent changes the chymotrypsin specificity with respect to nucleophiles. *FEBS Lett*, **307**, 309-312.
181. Homandberg, G.A., Mattis, J.A. and Laskowski, M., Jr. (1978) Synthesis of peptide bonds by proteinases. Addition of organic cosolvents shifts peptide bond equilibria toward synthesis. *Biochemistry*, **17**, 5220-5227.
182. Lee, T.-W. and James, M.N.G. (2008) 1.2Å-resolution crystal structures reveal the second tetrahedral intermediates of streptogrisin B (SGPB). *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, **1784**, 319-334.
183. Kendrew, J.C., Bodo, G., Dintzis, H.M. et al. (1958) A three-dimensional model of the myoglobin molecule obtained by X-ray analysis. *Nature*, **181**, 662-666.
184. Holton, J. (2009) A beginner's guide to radiation damage. *Journal of Synchrotron Radiation*, **16**, 133-142.
185. Garman, E.F. and Schneider, T.R. (1997) Macromolecular cryocrystallography. *Journal of Applied Crystallography*, **30**, 211-237.
186. Juers, D.H. and Matthews, B.W. (2004) Cryo-cooling in macromolecular crystallography: advantages, disadvantages and optimization. *Q Rev Biophys*, **37**, 105-119.
187. Barends, T.R.M., Stauch, B., Cherezov, V. and Schlichting, I. (2022) Serial femtosecond crystallography. *Nat Rev Methods Primers*, **2**.

188. Hwu, Y. and Margaritondo, G. (2021) Synchrotron radiation and X-ray free-electron lasers (X-FELs) explained to all users, active and potential. *Journal of Synchrotron Radiation*, **28**, 1014-1029.
189. Chapman, H.N., Caleman, C. and Timneanu, N. (2014) Diffraction before destruction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **369**, 20130313.
190. Chapman, H.N., Fromme, P., Barty, A. et al. (2011) Femtosecond X-ray protein nanocrystallography. *Nature*, **470**, 73-77.
191. Lomb, L., Barends, T.R., Kassemeyer, S. et al. (2011) Radiation damage in protein serial femtosecond crystallography using an x-ray free-electron laser. *Phys Rev B Condens Matter Mater Phys*, **84**, 214111.
192. Orville, A.M. (2020) Recent results in time resolved serial femtosecond crystallography at XFELs. *Curr Opin Struct Biol*, **65**, 193-208.
193. Nogly, P., Weinert, T., James, D. et al. (2018) Retinal isomerization in bacteriorhodopsin captured by a femtosecond x-ray laser. *Science*, **361**, eaat0094.
194. Skopintsev, P., Ehrenberg, D., Weinert, T. et al. (2020) Femtosecond-to-millisecond structural changes in a light-driven sodium pump. *Nature*, **583**, 314-318.
195. Nass Kovacs, G., Colletier, J.-P., Grünbein, M.L. et al. (2019) Three-dimensional view of ultrafast dynamics in photoexcited bacteriorhodopsin. *Nature Communications*, **10**, 3177.
196. Oda, K., Nomura, T., Nakane, T. et al. (2021) Time-resolved serial femtosecond crystallography reveals early structural changes in channelrhodopsin. *eLife*, **10**, e62389.
197. Coquelle, N., Sliwa, M., Woodhouse, J. et al. (2018) Chromophore twisting in the excited state of a photoswitchable fluorescent protein captured by time-resolved serial femtosecond crystallography. *Nature Chemistry*, **10**, 31-37.
198. Woodhouse, J., Nass Kovacs, G., Coquelle, N. et al. (2020) Photoswitching mechanism of a fluorescent protein revealed by time-resolved crystallography and transient absorption spectroscopy. *Nature Communications*, **11**, 741.
199. Kern, J., Chatterjee, R., Young, I.D. et al. (2018) Structures of the intermediates of Kok's photosynthetic water oxidation clock. *Nature*, **563**, 421-425.
200. Suga, M., Akita, F., Yamashita, K. et al. (2019) An oxyl/oxo mechanism for oxygen-oxygen coupling in PSII revealed by an x-ray free-electron laser. *Science*, **366**, 334-338.
201. Ibrahim, M., Fransson, T., Chatterjee, R. et al. (2020) Untangling the sequence of events during the S₂ → S₃ transition in photosystem II and implications for the water oxidation mechanism. *Proceedings of the National Academy of Sciences*, **117**, 12624-12635.
202. Olmos, J.L., Pandey, S., Martin-Garcia, J.M. et al. (2018) Enzyme intermediates captured “on the fly” by mix-and-inject serial crystallography. *BMC Biology*, **16**, 59.
203. Ishigami, I., Lewis-Ballester, A., Echelmeier, A. et al. (2019) Snapshot of an oxygen intermediate in the catalytic reaction of cytochrome c oxidase. *Proceedings of the National Academy of Sciences*, **116**, 3572-3577.
204. Fuller, F.D., Gul, S., Chatterjee, R. et al. (2017) Drop-on-demand sample delivery for studying biocatalysts in action at X-ray free-electron lasers. *Nat Methods*, **14**, 443-449.
205. Rabe, P., Kamps, J.J.A.G., Sutherlin, K.D. et al. (2021) X-ray free-electron laser studies reveal correlated motion during isopenicillin. *Sci Adv*, **7**.
206. Kern, J., Alonso-Mori, R., Tran, R. et al. (2013) Simultaneous femtosecond X-ray spectroscopy and diffraction of photosystem II at room temperature. *Science*, **340**, 491-495.
207. Fransson, T., Chatterjee, R., Fuller, F.D. et al. (2018) X-ray emission spectroscopy as an in situ diagnostic tool for X-ray crystallography of metalloproteins using an X-ray free-electron laser. *Biochemistry*, **57**, 4629-4637.
208. Ohmer, C.J., Dasgupta, M., Patwardhan, A. et al. (2022) XFEL serial crystallography reveals the room temperature structure of methyl-coenzyme M reductase. *Journal of Inorganic Biochemistry*, **230**, 111768.

209. Hoffman, J.R. and Falvo, M.J. (2004) Protein - which is best? *J Sports Sci Med*, **3**, 118-130.
210. Brantl, V., Pfeiffer, A., Herz, A., Henschen, A. and Lottspeich, F. (1982) Antinociceptive potencies of beta-casomorphin analogs as compared to their affinities towards mu and delta opiate receptor sites in brain and periphery. *Peptides*, **3**, 793-797.
211. Yoshdcawa, M., Yoshimura, T. and Chiba, H. (1984) Opioid Peptides from Human β-Casein. *Agricultural and Biological Chemistry*, **48**, 3185-3187.
212. Koch, G., Wiedemann, K. and Teschemacher, H. (1985) Opioid activities of human beta-casomorphins. *Naunyn Schmiedebergs Arch Pharmacol*, **331**, 351-354.
213. Cade, R., Privette, M., Fregly, M. et al. (2000) Autism and schizophrenia: intestinal disorders. *Nutr Neurosci*, **3**, 57-72.
214. Sun, Z., Zhang, Z., Wang, X. et al. (2003) Relation of beta-casomorphin to apnea in sudden infant death syndrome. *Peptides*, **24**, 937-943.
215. Beales, P.E., Elliott, R.B., Flohé, S. et al. (2002) A multi-centre, blinded international trial of the effect of A(1) and A(2) beta-casein variants on diabetes incidence in two rodent models of spontaneous Type I diabetes. *Diabetologia*, **45**, 1240-1246.
216. McLachlan, C.N. (2001) beta-casein A1, ischaemic heart disease mortality, and other illnesses. *Med Hypotheses*, **56**, 262-272.
217. EFSA. (2009) Review of the potential health impact of β-casomorphins and related peptides. *EFSA Scientific Report*, **231**, 1-107.
218. Hunter, L.C., O'Hare, A., Herron, W.J., Fisher, L.A. and Jones, G.E. (2003) Opioid peptides and dipeptidyl peptidase in autism. *Dev Med Child Neurol*, **45**, 121-128.
219. Elwood, P.C., Pickering, J.E., Fehily, A.M., Hughes, J. and Ness, A.R. (2004) Milk drinking, ischaemic heart disease and ischaemic stroke I. Evidence from the Caerphilly cohort. *Eur J Clin Nutr*, **58**, 711-717.
220. Elwood, P.C., Pickering, J.E., Hughes, J., Fehily, A.M. and Ness, A.R. (2004) Milk drinking, ischaemic heart disease and ischaemic stroke II. Evidence from cohort studies. *Eur J Clin Nutr*, **58**, 718-724.
221. Teschemacher, H., Koch, G. and Brantl, V. (1997) Milk protein-derived opioid receptor ligands. *Biopolymers*, **43**, 99-117.
222. Chang, K.J., Lillian, A., Hazum, E., Cuatrecasas, P. and Chang, J.K. (1981) Morphiceptin (NH4-tyr-pro-phe-pro-COHN2): a potent and specific agonist for morphine (mu) receptors. *Science*, **212**, 75-77.
223. Brantl, V. (1984) Novel opioid peptides derived from human beta-casein: human beta-casomorphins. *Eur J Pharmacol*, **106**, 213-214.
224. Jinsmaa, Y. and Yoshikawa, M. (1999) Enzymatic release of neocasomorphin and beta-casomorphin from bovine beta-casein. *Peptides*, **20**, 957-962.
225. Nyberg, F., Carlsson, A. and Hallberg, M. (2013), *Handbook of biologically active peptides*. Second edition. ed. Elsevier Academic Press, Amsterdam, Vol. 211, pp. 1550-1555.
226. Petrilli, P., Addeo, F. and Chianese, L. (1983) Primary structure of water buffalo beta-casein tryptic and CNBr peptides. *Ital J Biochem*, **32**, 336-344.
227. Clare, D.A. and Swaisgood, H.E. (2000) Bioactive milk peptides: a prospectus. *J Dairy Sci*, **83**, 1187-1195.
228. Summer, A., Di Frangia, F., Ajmone Marsan, P., De Noni, I. and Malacarne, M. (2020) Occurrence, biological properties and potential effects on human health of β-casomorphin 7: Current knowledge and concerns. *Crit Rev Food Sci Nutr*, **60**, 3705-3723.
229. Ruivo, E.F.P., Gonçalves, L.M., Carvalho, L.A.R. et al. (2016) Clickable 4-Oxo-β-lactam-Based Selective Probing for Human Neutrophil Elastase Related Proteomes. *ChemMedChem*, **11**, 2037-2042.
230. Carvalho, L.A.R., Almeida, V.T., Brito, J.A. et al. (2020) 3-Oxo-β-sultam as a Sulfonylating Chemotype for Inhibition of Serine Hydrolases and Activity-Based Protein Profiling. *ACS Chemical Biology*, **15**, 878-883.

231. Mueller, C., Marx, A., Epp, S.W. et al. (2015) Fixed target matrix for femtosecond time-resolved and in situ serial micro-crystallography. *Struct Dyn*, **2**, 054302.
232. Owen, R.L., Axford, D., Sherrell, D.A. et al. (2017) Low-dose fixed-target serial synchrotron crystallography. *Acta Crystallogr D Struct Biol*, **73**, 373-378.
233. Horrell, S., Axford, D., Devenish, N.E. et al. (2021) Fixed target serial data collection at Diamond Light Source. *J Vis Exp*.
234. Mehrabi, P., Schulz, E.C., Agthe, M. et al. (2019) Liquid application method for time-resolved analyses by serial synchrotron crystallography. *Nat Methods*, **16**, 979-982.
235. Glerup, J. (2022) Development of drop-on-chip time-resolved serial crystallography. MChem Thesis, University of Edinburgh.
236. Wilmouth, R.C., Kassamally, S., Westwood, N.J. et al. (1999) Mechanistic insights into the inhibition of serine proteases by monocyclic lactams. *Biochemistry*, **38**, 7989-7998.
237. Cleland, W.W. and Hengge, A.C. (1995) Mechanisms of phosphoryl and acyl transfer. *The FASEB Journal*, **9**, 1585-1594.
238. Markley, J.L., Travers, F. and Balny, C. (1981) Lack of evidence for a tetrahedral intermediate in the hydrolysis of nitroanilide substrates by serine proteinases. *European Journal of Biochemistry*, **120**, 477-485.
239. Fastrez, J. and Fersht, A.R. (1973) Mechanism of chymotrypsin. Structure, reactivity, and nonproductive binding relations. *Biochemistry*, **12**, 1067-1074.
240. Fersht, A.R. and Renard, M. (1974) pH Dependence of chymotrypsin catalysis. Appendix. Substrate binding to dimeric α-chymotrypsin studied by X-ray diffraction and the equilibrium method. *Biochemistry*, **13**, 1416-1426.
241. Ishida, T. and Kato, S. (2003) Theoretical Perspectives on the Reaction Mechanism of Serine Proteases: The Reaction Free Energy Profiles of the Acylation Process. *Journal of the American Chemical Society*, **125**, 12035-12048.
242. Topf, M. and Richards, W.G. (2004) Theoretical Studies on the Deacylation Step of Serine Protease Catalysis in the Gas Phase, in Solution, and in Elastase. *Journal of the American Chemical Society*, **126**, 14631-14641.
243. Bauer, C.A. (1976) The active centers of Streptomyces griseus protease 3 and alpha-chymotrypsin. Enzyme-substrate interactions beyond subsite S'1. *Biochim Biophys Acta*, **438**, 495-502.
244. Coombs, G.S., Dang, A.T., Madison, E.L. and Corey, D.R. (1996) Distinct mechanisms contribute to stringent substrate specificity of tissue-type plasminogen activator. *J Biol Chem*, **271**, 4461-4467.
245. Tosha, T., Nomura, T., Nishida, T. et al. (2017) Capturing an initial intermediate during the P450nor enzymatic reaction using time-resolved XFEL crystallography and caged-substrate. *Nature Communications*, **8**, 1585.
246. Mangubat-Medina, A.E. and Ball, Z.T. (2021) Triggering biological processes: methods and applications of photocaged peptides and proteins. *Chemical Society Reviews*, **50**, 10403-10421.
247. Mangubat-Medina, A.E., Martin, S.C., Hanaya, K. and Ball, Z.T. (2018) A Vinylogous Photocleavage Strategy Allows Direct Photocaging of Backbone Amide Structure. *Journal of the American Chemical Society*, **140**, 8401-8404.
248. Bender, M.L. and Hamilton, G.A. (1962) Kinetic Isotope Effects of Deuterium Oxide on Several α-Chymotrypsin-catalyzed Reactions. *Journal of the American Chemical Society*, **84**, 2570-2576.
249. Matern, U., Schleberger, C., Jelakovic, S., Weckesser, J. and Schulz, G.E. (2003) Binding structure of elastase inhibitor scyptolin A. *Chem Biol*, **10**, 997-1001.
250. Kinoshita, T., Tamada, T., Imai, K. et al. (2007) Crystallization of porcine pancreatic elastase and a preliminary neutron diffraction experiment. *Acta Crystallogr Sect F Struct Biol Cryst Commun*, **63**, 315-317.
251. Gilliland, G.L. and Teplyakov, A. (2011), *Encyclopedia of inorganic and bioinorganic chemistry*.

252. Katona, G., Wilmouth, R.C., Wright, P.A. et al. (2002) X-ray structure of a serine protease acyl-enzyme complex at 0.95-A resolution. *J Biol Chem*, **277**, 21962-21970.
253. Wright, P.A. (1999) Mechanistic studies on the catalysis and inhibition of serine proteases. D. Phil Thesis, University of Oxford.
254. Butryn, A., Simon, P.S., Aller, P. et al. (2021) An on-demand, drop-on-drop method for studying enzyme catalysis by serial crystallography. *Nat Commun*, **12**, 4461.
255. Pandey, S., Calvey, G., Katz, A.M. et al. (2021) Observation of substrate diffusion and ligand binding in enzyme crystals using high-repetition-rate mix-and-inject serial crystallography. *IUCrJ*, **8**, 878-895.
256. Imaki, K., Arai, Y. and Okegawa, T. (1989) EP 0347168 A1.
257. Kawabata, K., Suzuki, M., Sugitani, M. et al. (1991) ONO-5046, a novel inhibitor of human neutrophil elastase. *Biochemical and Biophysical Research Communications*, **177**, 814-820.
258. Crocetti, L., Giovannoni, M.P., Cantini, N. et al. (2020) Novel sulfonamide analogs of sivelestat as potent human neutrophil elastase inhibitors. *Frontiers in Chemistry*, **8**.
259. Wewers, M.D., Casolaro, M.A., Sellers, S.E. et al. (1987) Replacement therapy for alpha 1-antitrypsin deficiency associated with emphysema. *N Engl J Med*, **316**, 1055-1062.
260. Zeiher, B.G., Artigas, A., Vincent, J.-L. et al. (2004) Neutrophil elastase inhibition in acute lung injury: Results of the STRIVE study. *Critical Care Medicine*, **32**, 1695-1702.
261. Aikawa, N. and Kawasaki, Y. (2014) Clinical utility of the neutrophil elastase inhibitor sivelestat for the treatment of acute respiratory distress syndrome. *Ther Clin Risk Manag*, **10**, 621-629.
262. Ono Pharmaceutical Co., L. (2002, June 13) Announcement of the launch of Elaspol® 100 for injection, the world's first drug for acute lung injury [press release]. https://www.ono-pharma.com/sites/default/files/en/news/press/enews20020613.pdf
263. Imaki, K., Okada, T., Nakayama, Y. et al. (1996) Non-peptidic inhibitors of human neutrophil elastase: the design and synthesis of sulfonanilide-containing inhibitors. *Bioorg Med Chem*, **4**, 2115-2134.
264. Nakayama, Y., Odagaki, Y., Fujita, S. et al. (2002) Clarification of mechanism of human sputum elastase inhibition by a new inhibitor, ONO-5046, using electrospray ionization mass spectrometry. *Bioorg Med Chem Lett*, **12**, 2349-2353.
265. Hwang, T.-L., Wang, W.-H., Wang, T.-Y., Yu, H.-P. and Hsieh, P.-W. (2015) Synthesis and pharmacological characterization of 2-aminobenzaldehyde oxime analogs as dual inhibitors of neutrophil elastase and proteinase 3. *Bioorganic & Medicinal Chemistry*, **23**, 1123-1134.
266. Giovannoni, M.P., Schepetkin, I.A., Quinn, M.T. et al. (2018) Synthesis, biological evaluation, and molecular modelling studies of potent human neutrophil elastase (HNE) inhibitors. *Journal of Enzyme Inhibition and Medicinal Chemistry*, **33**, 1108-1124.
267. Bone, R., Shenvi, A.B., Kettner, C.A. and Agard, D.A. (1987) Serine protease mechanism: structure of an inhibitory complex of .alpha.-lytic protease and a tightly bound peptide boronic acid. *Biochemistry*, **26**, 7609-7614.
268. Katz, B.A., Finer-Moore, J., Mortezaei, R., Rich, D.H. and Stroud, R.M. (1995) Episelection: novel Ki .apprx. nanomolar inhibitors of serine proteases selected by binding or chemistry on an enzyme surface. *Biochemistry*, **34**, 8264-8280.
269. Gráf, L., Jancsó, A., Szilágyi, L. et al. (1988) Electrostatic complementarity within the substrate-binding pocket of trypsin. *Proc Natl Acad Sci U S A*, **85**, 4961-4965.
270. Hedstrom, L., Szilagyi, L. and Rutter, W.J. (1992) Converting trypsin to chymotrypsin: the role of surface loops. *Science*, **255**, 1249-1253.
271. Hedstrom, L., Perona, J.J. and Rutter, W.J. (1994) Converting Trypsin to Chymotrypsin: Residue 172 Is a Substrate Specificity Determinant. *Biochemistry*, **33**, 8757-8763.
272. Venekei, I., Szilágyi, L., Gráf, L. and Rutter, W.J. (1996) Attempts to convert chymotrypsin to trypsin. *FEBS Letters*, **379**, 143-147.

273. Hung, S.H. and Hedstrom, L. (1998) Converting trypsin to elastase: substitution of the S1 site and adjacent loops reconstitutes esterase specificity but not amidase activity. *Protein Engineering, Design and Selection*, **11**, 669-673.
274. Lobstein, J., Emrich, C.A., Jeans, C. et al. (2012) SHuffle, a novel Escherichia coli protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm. *Microb Cell Fact*, **11**, 56.
275. Baneyx, F. and Mujacic, M. (2004) Recombinant protein folding and misfolding in Escherichia coli. *Nat Biotechnol*, **22**, 1399-1408.
276. Bardwell, J.C., McGovern, K. and Beckwith, J. (1991) Identification of a protein required for disulfide bond formation in vivo. *Cell*, **67**, 581-589.
277. Landeta, C., Boyd, D. and Beckwith, J. (2018) Disulfide bond formation in prokaryotes. *Nat Microbiol*, **3**, 270-280.
278. Karyolaimos, A., Ampah-Korsah, H., Hillenaar, T. et al. (2019) Enhancing Recombinant Protein Yields in the E. coli Periplasm by Combining Signal Peptide and Production Rate Screening. *Frontiers in Microbiology*, **10**.
279. Hand, K., Wilkinson, M.C. and Madine, J. (2018) Isolation and purification of recombinant immunoglobulin light chain variable domains from the periplasmic space of Escherichia coli. *PLoS One*, **13**, e0206167.
280. Costa, S., Almeida, A., Castro, A. and Domingues, L. (2014) Fusion tags for protein solubility, purification and immunogenicity in Escherichia coli: the novel Fh8 system. *Front Microbiol*, **5**, 63.
281. Magnusdottir, A., Johansson, I., Dahlgren, L.G., Nordlund, P. and Berglund, H. (2009) Enabling IMAC purification of low abundance recombinant proteins from E. coli lysates. *Nat Methods*, **6**, 477-478.
282. Hanson, A.D., McCarty, D.R., Henry, C.S. et al. (2021) The number of catalytic cycles in an enzyme's lifetime and why it matters to metabolic engineering. *Proc Natl Acad Sci U S A*, **118**.
283. Leibly, D.J., Nguyen, T.N., Kao, L.T. et al. (2012) Stabilizing additives added during cell lysis aid in the solubilization of recombinant proteins. *PLoS One*, **7**, e52482.
284. Schlegel, S., Löfblom, J., Lee, C. et al. (2012) Optimizing membrane protein overexpression in the Escherichia coli strain Lemo21(DE3). *J Mol Biol*, **423**, 648-659.
285. Campbell, A.K., Naseem, R., Holland, I.B., Matthews, S.B. and Wann, K.T. (2007) Methylglyoxal and other carbohydrate metabolites induce lanthanum-sensitive Ca2+ transients and inhibit growth in E. coli. *Arch Biochem Biophys*, **468**, 107-113.
286. Graf, L., Craik, C.S., Patthy, A. et al. (1987) Selective alteration of substrate specificity by replacement of aspartic acid-189 with lysine in the binding pocket of trypsin. *Biochemistry*, **26**, 2616-2623.
287. Gangola, P. and Rosen, B.P. (1987) Maintenance of intracellular calcium in Escherichia coli. *J Biol Chem*, **262**, 12570-12574.
288. Naseem, R., Wann, K.T., Holland, I.B. and Campbell, A.K. (2009) ATP regulates calcium efflux and growth in E. coli. *J Mol Biol*, **391**, 42-56.
289. Bolanos-Garcia, V.M. and Davies, O.R. (2006) Structural analysis and classification of native proteins from E. coli commonly co-purified by immobilised metal affinity chromatography. *Biochim Biophys Acta*, **1760**, 1304-1313.
290. Higaki, J.N., Evnin, L.B. and Craik, C.S. (1989) Introduction of a cysteine protease active site into trypsin. *Biochemistry*, **28**, 9256-9263.
291. Baghban, R., Farajnia, S., Rajabibazl, M. et al. (2019) Yeast Expression Systems: Overview and Recent Advances. *Molecular Biotechnology*, **61**, 365-384.
292. Hohenblum, H., Vorauer-Uhl, K., Katinger, H. and Mattanovich, D. (2004) Bacterial expression and refolding of human trypsinogen. *Journal of Biotechnology*, **109**, 3-11.

293. Kim, Y.-M., Haghighat, L., Spiekerkoetter, E. et al. (2011) Neutrophil Elastase Is Produced by Pulmonary Artery Smooth Muscle Cells and Is Linked to Neointimal Lesions. *The American Journal of Pathology*, **179**, 1560-1572.
294. Bronze-da-Rocha, E. and Santos-Silva, A. (2018) Neutrophil elastase inhibitors and chronic kidney disease. *Int J Biol Sci*, **14**, 1343-1360.
295. Belaaouaj, A., McCarthy, R., Baumann, M. et al. (1998) Mice lacking neutrophil elastase reveal impaired host defense against gram negative bacterial sepsis. *Nat Med*, **4**, 615-618.
296. Belaaouaj, A.a., Kim, K.S. and Shapiro, S.D. (2000) Degradation of outer membrane protein A in Escherichia coli killing by neutrophil elastase. *Science*, **289**, 1185-1187.
297. Standish, A.J. and Weiser, J.N. (2009) Human neutrophils kill Streptococcus pneumoniae via serine proteases. *The Journal of Immunology*, **183**, 2602-2609.
298. Hahn, I., Klaus, A., Janze, A.K. et al. (2011) Cathepsin G and neutrophil elastase play critical and nonredundant roles in lung-protective immunity against Streptococcus pneumoniae in mice. *Infect Immun*, **79**, 4893-4901.
299. Tkalcevic, J., Novelli, M., Phylactides, M. et al. (2000) Impaired immunity and enhanced resistance to endotoxin in the absence of neutrophil elastase and cathepsin G. *Immunity*, **12**, 201-210.
300. Reeves, E.P., Lu, H., Jacobs, H.L. et al. (2002) Killing activity of neutrophils is mediated through activation of proteases by K+ flux. *Nature*, **416**, 291-297.
301. Nordahl, E.A., Rydengård, V., Nyberg, P. et al. (2004) Activation of the complement system generates antibacterial peptides. *Proceedings of the National Academy of Sciences*, **101**, 16879-16884.
302. Weinrauch, Y., Drujan, D., Shapiro, S.D., Weiss, J. and Zychlinsky, A. (2002) Neutrophil elastase targets virulence factors of enterobacteria. *Nature*, **417**, 91-94.
303. Jorch, S.K. and Kubes, P. (2017) An emerging role for neutrophil extracellular traps in noninfectious disease. *Nature Medicine*, **23**, 279-287.
304. Brinkmann, V., Reichard, U., Goosmann, C. et al. (2004) Neutrophil extracellular traps kill bacteria. *Science*, **303**, 1532-1535.
305. Kessenbrock, K., Fröhlich, L., Sixt, M. et al. (2008) Proteinase 3 and neutrophil elastase enhance inflammation in mice by inactivating antiinflammatory progranulin. *J Clin Invest*, **118**, 2438-2447.
306. Henry, C.M., Sullivan, G.P., Clancy, D.M. et al. (2016) Neutrophil-derived proteases escalate inflammation through activation of IL-36 family cytokines. *Cell Rep*, **14**, 708-722.
307. Brostjan, C. and Oehler, R. (2020) The role of neutrophil death in chronic inflammation and cancer. *Cell Death Discov*, **6**, 26.
308. Lucas, S.D., Costa, E., Guedes, R.C. and Moreira, R. (2013) Targeting COPD: advances on low-molecular-weight inhibitors of human neutrophil elastase. *Medicinal Research Reviews*, **33**, E73-E101.
309. Voynow, J.A. and Shinbashi, M. (2021) Neutrophil elastase and chronic lung disease. *Biomolecules*, **11**, 1065.
310. Taylor, S., Dirir, O., Zamanian, R.T., Rabinovitch, M. and Thompson, A.A.R. (2018) The role of neutrophils and neutrophil elastase in pulmonary arterial hypertension. *Frontiers in Medicine*, **5**.
311. Lerman, I. and Hammes, S.R. (2018) Neutrophil elastase in the tumor microenvironment. *Steroids*, **133**, 96-101.
312. Huang, H., Zhang, H., Onuma, A.E. and Tsung, A. (2020) In Birbrair, A. (ed.), *Tumor Microenvironment: State of the Science*. Springer International Publishing, Cham, pp. 13-23.
313. Fleming, A. (1929) On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of B. influenzæ. *Br J Exp Pathol.*, **10**, 226-236.
314. Bush, K. and Bradford, P.A. (2016) β-Lactams and β-lactamase inhibitors: an overview. *Cold Spring Harb Perspect Med*, **6**.

315. Dolk, F.C.K., Pouwels, K.B., Smith, D.R.M., Robotham, J.V. and Smieszek, T. (2018) Antibiotics in primary care in England: which antibiotics are prescribed and for which conditions? *Journal of Antimicrobial Chemotherapy*, **73**, ii2-ii10.
316. Ventola, C.L. (2015) The antibiotic resistance crisis: part 1: causes and threats. *P t*, **40**, 277-283.
317. Wasserman, H.H., Precopio, F.M. and Liu, T.-C. (1952) Studies on the mucohalic acids. II. The synthesis of fused α-lactam-thiazolidines related to Penicillin1. *Journal of the American Chemical Society*, **74**, 4093-4095.
318. Baldwin, J.E., Chan, M.F., Gallacher, G., Monk, P. and Prout, K. (1983) γ-Lactam analogues of penicillanic and carbapenicillanic acids. *Journal of the Chemical Society, Chemical Communications*, 250-252.
319. Baldwin, J.E., Lynch, G.P. and Pitlik, J. (1991) Gamma-lactam analogues of beta-lactam antibiotics. *J Antibiot (Tokyo)*, **44**, 1-24.
320. Baldwin, J.E., Adlington, R.M., Jones, R.H. et al. (1986) γ-Lactam analogues of carbapenicillanic acids. *Tetrahedron*, **42**, 4879-4888.
321. Boyd, D.B., Elzey, T.K., Hatfield, L.D., Kinnick, M.D. and Morin, J.M. (1986) γ-Lactam analogues of the penems. *Tetrahedron Letters*, **27**, 3453-3456.
322. Boyd, D.B., Foster, B.J., Hatfield, L.D. et al. (1986) γ-Lactam analogues of carbapenems. *Tetrahedron Letters*, **27**, 3457-3460.
323. Osterhage, C., Kaminsky, R., König, G.M. and Wright, A.D. (2000) Ascosalipyrrolidinone A, an antimicrobial alkaloid, from the obligate marine fungus Ascochyta salicorniae. *J Org Chem*, **65**, 6412-6417.
324. Jang, D.S., Lee, G.Y., Lee, Y.M. et al. (2009) Flavan-3-ols having a gamma-lactam from the roots of Actinidia arguta inhibit the formation of advanced glycation end products in vitro. *Chem Pharm Bull (Tokyo)*, **57**, 397-400.
325. Feling, R.H., Buchanan, G.O., Mincer, T.J. et al. (2003) Salinosporamide A: a highly cytotoxic proteasome inhibitor from a novel microbial source, a marine bacterium of the new genus salinospora. *Angew Chem Int Ed Engl*, **42**, 355-357.
326. Janecka, A., Wyrębska, A., Gach, K., Fichna, J. and Janecki, T. (2012) Natural and synthetic α-methylenelactones and α-methylenelactams with anticancer potential. *Drug Discov Today*, **17**, 561-572.
327. Caruano, J., Muccioli, G.G. and Robiette, R. (2016) Biologically active gamma-lactams: synthesis and natural sources. *Org Biomol Chem*, **14**, 10134-10156.
328. Georgopapadakou, N., Hammarström, S. and Strominger, J.L. (1977) Isolation of the penicillin-binding peptide from D-alanine carboxypeptidase of Bacillus subtilis. *Proc Natl Acad Sci U S A*, **74**, 1009-1012.
329. A Firestone, R., arker, P.L., isano, J.M., she, B.M. and ahlgren, M.E. (1990) Monocyclic β-lactam inhibitors of human leukocyte elastase. *Tetrahedron*, **46**, 2255-2262.
330. Knight, W.B., Green, B.G., Chabin, R.M. et al. (1992) Specificity, stability, and potency of monocyclic beta-lactam inhibitors of human leukocyte elastase. *Biochemistry*, **31**, 8160-8170.
331. Migaud, M.E., Wilmouth, R.C., Mills, G.I. et al. (2002) 5,5-Fused thiophene γ-lactams as templates for serine protease inhibition. *Chem Commun (Camb)*, 1274-1275.
332. Wilmouth, R.C., Westwood, N.J., Anderson, K. et al. (1998) Inhibition of elastase by N-sulfonylaryl β-lactams: anatomy of a stable acyl−enzyme complex. *Biochemistry*, **37**, 17506-17513.
333. Westwood, N.J., Claridge, T.D.W., Edwards, P.N. and Schofield, C.J. (1997) Reversible acylation of elastase by γ-lactam analogues of β-lactam inhibitors. *Bioorganic & Medicinal Chemistry Letters*, **7**, 2973-2978.
334. Wilmouth, R.C., Kassamally, S., Westwood, N.J. et al. (1999) Mechanistic Insights into the Inhibition of Serine Proteases by Monocyclic Lactams. *Biochemistry*, **38**, 7989-7998.

335. Macdonald, S.J.F., Dowle, M.D., Harrison, L.A. et al. (2001) The discovery of a potent, intracellular, orally bioavailable, long duration inhibitor of human neutrophil elastase—GW311616A a development candidate. *Bioorganic & Medicinal Chemistry Letters*, **11**, 895-898.
336. Macdonald, S.J.F., Belton, D.J., Buckley, D.M. et al. (1998) Syntheses of trans-5- Oxo-hexahydro-pyrrolo[3,2-b]pyrroles and trans-5-Oxo-hexahydro-furo[3,2-b]pyrroles (Pyrrolidine trans-Lactams and trans-Lactones): New Pharmacophores for Elastase Inhibition. *Journal of Medicinal Chemistry*, **41**, 3919-3922.
337. Wright, P.A., Wilmouth, R.C., Clifton, I.J. and Schofield, C.J. (2000) 'pH-jump' crystallographic analyses of γ-lactam–porcine pancreatic elastase complexes. *Biochemical Journal*, **351**, 335-340.
338. Kupitz, C., OlmosJr., J.L., Holl, M. et al. (2017) Structural enzymology using X-ray free electron lasers. *Structural Dynamics*, **4**, 044003.
339. Wilamowski, M., Sherrell, D.A., Kim, Y. et al. (2022) Time-resolved β-lactam cleavage by L1 metallo-β-lactamase. *Nature Communications*, **13**, 7379.
340. Gao, S., Zhou, Y., Zhang, W. et al. (2017) Structural insights into the γ-lactamase activity and substrate enantioselectivity of an isochorismatase-like hydrolase from Microbacterium hydrocarbonoxydans. *Scientific Reports*, **7**, 44542.
341. Wimouth, R.C. (1998) Structural studies on the mechanism and inhibition of elastase. D. Phil. Thesis, University of Oxford.
342. Mishra, R., Sachan, N., Kumar, N., Mishra, I. and Chand, P. (2018) Thiophene Scaffold as Prospective Antimicrobial Agent: A Review. *Journal of Heterocyclic Chemistry*, **55**, 2019-2034.
343. Wright, P.A., Wilmouth, R.C., Clifton, I.J. and Schofield, C.J. (2000) 'pH-jump' crystallographic analyses of gamma-lactam-porcine pancreatic elastase complexes. *Biochem J*, **351 Pt 2**, 335-340.
344. Takahashi, L.H., Radhakrishnan, R., Rosenfield, R.E., Jr. and Meyer, E.F., Jr. (1989) Crystallographic analysis of the inhibition of porcine pancreatic elastase by a peptidyl boronic acid: structure of a reaction intermediate. *Biochemistry*, **28**, 7610-7617.
345. Smoum, R., Rubinstein, A., Dembitsky, V.M. and Srebnik, M. (2012) Boron containing compounds as protease inhibitors. *Chemical Reviews*, **112**, 4156-4220.
346. Adamczyk-Woźniak, A., Borys, K.M. and Sporzyński, A. (2015) Recent developments in the chemistry and biological applications of benzoxaboroles. *Chemical Reviews*, **115**, 5224-5247.
347. Baker, S.J., Zhang, Y.-K., Akama, T. et al. (2006) Discovery of a new boron-containing antifungal agent, 5-fluoro-1,3-dihydro-1-hydroxy-2,1- benzoxaborole (AN2690), for the potential treatment of onychomycosis. *Journal of Medicinal Chemistry*, **49**, 4447-4450.
348. Rock, F.L., Mao, W., Yaremchuk, A. et al. (2007) An antifungal agent inhibits an aminoacyl-tRNA synthetase by trapping tRNA in the editing Site. *Science*, **316**, 1759-1761.
349. Xia, Y., Cao, K., Zhou, Y. et al. (2011) Synthesis and SAR of novel benzoxaboroles as a new class of β-lactamase inhibitors. *Bioorganic & Medicinal Chemistry Letters*, **21**, 2533-2536.
350. McKinney, D.C., Zhou, F., Eyermann, C.J. et al. (2015) 4,5-Disubstituted 6-aryloxy-1,3-dihydrobenzo[c][1,2]oxaboroles are broad-spectrum serine β-lactamase inhibitors. *ACS Infectious Diseases*, **1**, 310-316.
351. Newman, H., Krajnc, A., Bellini, D. et al. (2021) High-throughput crystallography reveals boron-containing inhibitors of a penicillin-binding protein with di- and tricovalent binding modes. *J Med Chem*, **64**, 11379-11394.
352. Xiao, Y.-C., Chen, X.-P., Deng, J. et al. (2021) Design and enantioselective synthesis of 3-(α-acrylic acid) benzoxaboroles to combat carbapenemase resistance. *Chemical Communications*, **57**, 7709-7712.
353. Li, X., Zhang, Y.K., Liu, Y. et al. (2010) Synthesis of new acylsulfamoyl benzoxaboroles as potent inhibitors of HCV NS3 protease. *Bioorg Med Chem Lett*, **20**, 7493-7497.
354. Li, X., Zhang, S., Zhang, Y.K. et al. (2011) Synthesis and SAR of acyclic HCV NS3 protease inhibitors with novel P4-benzoxaborole moieties. *Bioorg Med Chem Lett*, **21**, 2048-2054.

355. Akama, T., Dong, C., Virtucio, C. et al. (2013) Linking phenotype to kinase: identification of a novel benzoxaborole hinge-binding motif for kinase inhibition and development of high-potency Rho kinase inhibitors. *Journal of Pharmacology and Experimental Therapeutics*, **347**, 615.
356. Jacobs, R.T., Nare, B., Wring, S.A. et al. (2011) SCYX-7158, an orally-active benzoxaborole for the treatment of stage 2 human African trypanosomiasis. *PLoS Negl Trop Dis*, **5**, e1151.
357. Mäser, P., Wittlin, S., Rottmann, M. et al. (2012) Antiparasitic agents: new drugs on the horizon. *Curr Opin Pharmacol*, **12**, 562-566.
358. Kühl, N., Lang, J., Leuthold, M.M. and Klein, C.D. (2022) Discovery of potent benzoxaborole inhibitors against SARS-CoV-2 main and dengue virus proteases. *Eur J Med Chem*, **240**, 114585.
359. António, J.P.M., Gonçalves, L.M., Guedes, R.C., Moreira, R. and Gois, P.M.P. (2018) Diazaborines as new inhibitors of human neutrophil elastase. *ACS Omega*, **3**, 7418-7423.
360. Inglis, S., Zervosen, A., Woon, E. et al. (2009) Synthesis and evaluation of 3-(dihydroxyboryl)benzoic acids as D,D-carboxypeptidase R39 inhibitors. *Journal of medicinal chemistry*, **52**, 6097-6106.
361. Graczyk, K. (2011) Synthesis of functionalized benzoboroxoles for the construction of boronolectins. *Synthesis*, 4059.
362. Kinoshita, T., Yamaguchi, A. and Tada, T. (2006) Tris(hydroxymethyl)aminomethane induces conformational change and crystal-packing contraction of porcine pancreatic elastase. *Acta Crystallogr Sect F Struct Biol Cryst Commun*, **62**, 623-626.
363. Smith, E.T., Kruppa, M., Johnson, D.A. et al. (2023) High yield expression in Pichia pastoris of human neutrophil elastase fused to cytochrome B5. *Protein Expression and Purification*, **206**, 106255.
364. The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC.
365. Winter, G., Lobley, C.M. and Prince, S.M. (2013) Decision making in xia2. *Acta Crystallogr D Biol Crystallogr*, **69**, 1260-1273.
366. Lee, D., Baek, S., Park, J. et al. (2019) Nylon mesh-based sample holder for fixed-target serial femtosecond crystallography. *Sci Rep*, **9**, 6971.
367. Brewster, A.S., Young, I.D., Lyubimov, A., Bhowmick, A. and Sauter, N.K. (2019) Processing serial crystallographic data from XFELs or synchrotrons using the cctbx.xfel GUI. *Computational Crystallography Newsletter*, **10**, 22–39.
368. Sauter, N.K., Hattne, J., Grosse-Kunstleve, R.W. and Echols, N. (2013) New Python-based methods for data processing. *Acta Crystallogr D Biol Crystallogr*, **69**, 1274-1282.
369. Hattne, J., Echols, N., Tran, R. et al. (2014) Accurate macromolecular structures using minimal measurements from X-ray free-electron lasers. *Nat Methods*, **11**, 545-548.
370. Sauter, N.K., Hattne, J., Brewster, A.S. et al. (2014) Improved crystal orientation and physical properties from single-shot XFEL stills. *Acta Crystallogr D Biol Crystallogr*, **70**, 3299-3309.
371. Sauter, N.K. (2015) XFEL diffraction: developing processing methods to optimize data quality. *J Synchrotron Radiat*, **22**, 239-248.
372. Adams, P.D., Grosse-Kunstleve, R.W., Hung, L.W. et al. (2002) PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr*, **58**, 1948-1954.
373. Emsley, P., Lohkamp, B., Scott, W.G. and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr*, **66**, 486-501.
374. Murshudov, G.N., Vagin, A.A. and Dodson, E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr*, **53**, 240-255.
375. Grade2 version 1.3.2. Cambridge, United Kingdom: Global Phasing Ltd. https://grade.globalphasing.org/cgi-bin/grade2_server.cgi

376. Gileadi, O., Burgess-Brown, N.A., Colebrook, S.M. et al. (2008) In Kobe, B., Guss, M. and Huber, T. (eds.), *Structural Proteomics: High-Throughput Methods*. Humana Press, Totowa, NJ, pp. 221-246.
377. NEB Tm Calculator. https://tmcalculator.neb.com/#!/main
378. Savitsky, P., Bray, J., Cooper, C.D. et al. (2010) High-throughput production of human proteins for crystallization: the SGC experience. *J Struct Biol*, **172**, 3-13.
379. NEBio Calculator. https://nebiocalculator.neb.com/#!/ligation
380. Gibson, D.G., Young, L., Chuang, R.-Y. et al. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*, **6**, 343-345.
381. NEBuilder. https://nebuilder.neb.com/#!/
382. Expasy ProtParam. https://web.expasy.org/protparam/

Appendix A: Zymogen Sequences for PPE, HNE and CTR A

PPE numbering systems. Residues are numbered using the chymotrypsinogen A system. Orange: propeptide residues that are not present in the active enzyme after cleavage, blue: propeptide residues that remain attached after cleavage. Note: chymotrypsin is additionally cleaved in certain forms which results in the loss of residues, 14–15 (α, π, δ, and γ), and 147–148 (neochymotrypsinogen, α, and γ).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
PPE	–	–	–	–	–	Thr	Gln	Asp	Phe	Pro	Glu	Thr	Asn	Ala	Arg
HNE	–	–	–	–	–	–	–	–	–	–	–	–	–	Ser	Glu
CTR	Cys	Gly	Val	Pro	Ala	Ile	Gln	Pro	Val	Leu	Ser	Gly	Leu	Ser	Arg

	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
PPE	Val	Val	Gly	Gly	Thr	Glu	Ala	Gln	Arg	Asn	Ser	Trp	Pro	Ser	Gln
HNE	Ile	Val	Gly	Gly	Arg	Arg	Ala	Arg	Pro	His	Ala	Trp	Pro	Phe	Met
CTR	Ile	Val	Asn	Gly	Glu	Glu	Ala	Val	Pro	Gly	Ser	Trp	Pro	Trp	Gln

	31	32	33	34	35	36	36A	36B	36C	37	38	39	40	41	42
PPE	Ile	Ser	Leu	Gln	Tyr	Arg	Ser	Gly	Ser	Ser	Trp	Ala	His	Thr	Cys
HNE	Val	Ser	Leu	Gln	Leu	Arg	–	–	–	–	Gly	Gly	His	Phe	Cys
CTR	Val	Ser	Leu	Gln	Asp	Lys	–	–	–	Try	Gly	Phe	His	Phe	Cys

	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57
PPE	Gly	Gly	Thr	Leu	Ile	Arg	Gln	Asn	Trp	Val	Met	Thr	Ala	Ala	His
HNE	Gly	Ala	Thr	Leu	Ile	Ala	Pro	Asn	Phe	Val	Met	Ser	Ala	Ala	His
CTR	Gly	Gly	Ser	Leu	Ile	Asn	Glu	Asn	Trp	Vak	Val	Thr	Ala	Ala	His

	58	59	60	61	62	63	63A	63B	63C	64	65	65A	66	67	68
PPE	Cys	Val	Asp	Arg	Glu	Leu	–	–	–	Thr	Phe	Arg	Val	Val	Val
HNE	Cys	Val	Ala	Asn	Val	Asn	Val	Arg	Ala	Val	Arg	–	Val	Val	Leu
CTR	Cys	Gly	Val	Thr	Thr	Ser	–	–	–	Asp	Val	–	Val	Val	Ala

	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83
PPE	Gly	Glu	His	Asn	Leu	Asn	Gln	Asn	Asp	Gly	Thr	Gly	Gln	Tyr	Val
HNE	Gly	Ala	His	Asn	Leu	Ser	Arg	Arg	Glu	Pro	Thr	Arg	Gln	Val	Phe
CTR	Gly	Glu	Phe	Asp	Gln	Gly	Ser	Ser	Ser	Glu	Lys	Ile	Gln	Lys	Leu

	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98
PPE	Gly	Val	Gln	Lys	Ile	Val	Val	His	Pro	Tyr	Trp	Asn	Thr	Asp	Asp
HNE	Ala	Val	Gln	Arg	Ile	Phe	Glu	–	Asn	Gly	Tyr	Asp	Pro	Val	Asn
CTR	Lys	Ile	Ala	Lys	Val	Phe	Lys	Asn	Ser	Lys	Tyr	Asn	Ser	Leu	Thr

	99	99A	99B	100	101	102	103	104	105	106	107	108	109	110	111
PPE	Val	Ala	Ala	Gly	Tyr	Asp	Ile	Ala	Leu	Leu	Arg	Leu	Ala	Gln	Ser
HNE	Leu	–	–	Leu	Asn	Asp	Ile	Val	Ile	Leu	Gln	Leu	Asn	Gly	Ser
CTR	Ile	–	–	Asn	Asn	Asp	Ile	Thr	Leu	Leu	Lys	Leu	Ser	Thr	Ala

	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126
PPE	Val	Thr	Leu	Asn	Ser	Tyr	Val	Gln	Leu	Gly	Val	Leu	Pro	Arg	Ala
HNE	Ala	Thr	Ile	Asn	Ala	Asn	Val	Gln	Val	Ala	Gln	Leu	Pro	Ala	Gln
CTR	Ala	Ser	Phe	Ser	Gln	Thr	Val	Ser	Ala	Val	Cys	Leu	Pro	Ser	Ala

	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141
PPE	Gly	Thr	Ile	Leu	Ala	Asn	Asn	Ser	Pro	Cys	Tyr	Ile	Thr	Gly	Trp
HNE	Gly	Arg	Arg	Leu	Gly	Asn	Gly	Val	Gln	Cys	Leu	Ala	Met	Gly	Trp
CTR	Ser	Asp	Asp	Phe	Ala	Ala	Gly	Thr	Thr	Cys	Val	Thr	Thr	Gly	Trp

	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156
PPE	Gly	Leu	Thr	Arg	–	Thr	Asn	Gly	Gln	Leu	Ala	Gln	Thr	Leu	Gln
HNE	Gly	Leu	Leu	Gly	Arg	Asn	Arg	–	Gly	Ile	Ala	Ser	Val	Leu	Gln
CTR	Gly	Leu	Thr	Arg	Tyr	Thr	Asn	Ala	Asn	Thr	Pro	Asp	Arg	Leu	Gln

	157	158	159	160	161	162	163	164	165	166	167	168	169	170	170A
PPE	Gln	Ala	Tyr	Leu	Pro	Thr	Val	Asp	Tyr	Ala	Ile	Cys	Ser	Ser	Ser
HNE	Glu	Leu	Asn	Val	Thr	Val	Val	Thr	Ser	Leu	–	Cys	–	–	–
CTR	Gln	Ala	Ser	Leu	Pro	Leu	Leu	Ser	Asn	Thr	Asn	Cys	Lys	Lys	–

	170B	171	172	173	174	175	176	177	178	179	180	181	182	183	184
PPE	Ser	Tyr	Trp	Gly	Ser	Thr	Val	Lys	Asn	Ser	Met	Val	Cys	Ala	Gly
HNE	–	–	–	–	–	–	–	Arg	Arg	Ser	Asn	Val	Cys	Thr	Leu
CTR	–	Tyr	Trp	Gly	Thr	Lys	Ile	Lys	Asp	Ala	Met	Ile	Cys	Ala	Gly

	185	186	186A	186B	187	188	188A	189	190	191	192	193	194	195	196
PPE	Gly	Asp	–	–	Gly	Val	Arg	Ser	Gly	Cys	Gln	Gly	Asp	Ser	Gly
HNE	Val	Arg	Gly	Arg	Gln	Ala	–	Gly	Val	Cys	Phe	Gly	Asp	Ser	Gly
CTR	Ala	Ser	–	–	Gly	Val	–	Ser	Ser	Cys	Met	Gly	Asp	Ser	Gly

	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211
PPE	Gly	Pro	Leu	His	Cys	Leu	Val	Asn	Gly	Gln	Tyr	Ala	Val	His	Gly
HNE	Ser	Pro	Leu	Val	Cys	Asn	—	—	—	—	Gly	Leu	Ile	His	Gly
CTR	Gly	Pro	Leu	Val	Cys	Lys	Lys	Asn	Gly	Ala	Trp	Thr	Leu	Val	Gly

	212	213	214	215	216	217	217A	218	219	220	221	221A	222	223	224
PPE	Val	Thr	Ser	Phe	Val	Ser	Arg	Leu	Gly	Cys	Asn	Val	Thr	Arg	Lys
HNE	Ile	Ala	Ser	Phe	Val	Arg	—	Gly	Gly	Cys	Ala	Ser	Gly	Leu	Tyr
CTR	Ile	Val	Ser	Trp	Gly	Ser	—	Ser	Thr	Cys	Ser	—	Thr	Ser	Thr

	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239
PPE	Pro	Thr	Val	Phe	Thr	Arg	Val	Ser	Ala	Tyr	Ile	Ser	Trp	Ile	Asn
HNE	Pro	Asp	Ala	Phe	Ala	Pro	Val	Ala	Gln	Phe	Val	Asn	Trp	Ile	Asp
CTR	Pro	Gly	Val	Tyr	Ala	Arg	Val	Thr	Ala	Leu	Val	Asn	Trp	Val	Gln

	240	241	242	243	244	245
PPE	Asn	Val	Ile	Ala	Ser	Asn
HNE	Ser	Ile	Ile	Gln	—	—
CTR	Gln	Thr	Leu	Ala	Ala	Asn

Appendix B: Comparison of PPE numbering systems

PPE numbering systems. The chymotrypsinogen A numbering system is shown above the amino acid sequence, the standard numbering using the initiator methionine as residue number 1 is below the sequence. Orange: propeptide residues, yellow: signal peptide sequence.

PPE	Met	Leu	Arg	Leu	Leu	Val	Val	Ala	Ser	Leu	Val	Leu	Tyr	Gly	His
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
		6	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**
	Ser	Thr	Gln	Asp	Phe	Pro	Glu	Thr	Asn	Ala	Arg	Val	Val	Gly	Gly
	16	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**
	20	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**	**31**	**32**	**33**	**34**
PPE	Thr	Glu	Ala	Gln	Arg	Asn	Ser	Trp	Pro	Ser	Gln	Ile	Ser	Leu	Gln
	31	**32**	**33**	**34**	**35**	**36**	**37**	**38**	**39**	**40**	**41**	**42**	**43**	**44**	**45**
	35	**36**	**36A**	**36B**	**36C**	**37**	**38**	**39**	**40**	**41**	**42**	**43**	**44**	**45**	**46**
PPE	Tyr	Arg	Ser	Gly	Ser	Ser	Trp	Ala	His	Thr	Cys	Gly	Gly	Thr	Leu
	46	**47**	**48**	**49**	**50**	**51**	**52**	**53**	**54**	**55**	**56**	**57**	**58**	**59**	**60**
	47	**48**	**49**	**50**	**51**	**52**	**53**	**54**	**55**	**56**	**57**	**58**	**59**	**60**	**61**
PPE	Ile	Arg	Gln	Asn	Trp	Val	Met	Thr	Ala	Ala	His	Cys	Val	Asp	Arg
	61	**62**	**63**	**64**	**65**	**66**	**67**	**68**	**69**	**70**	**71**	**72**	**73**	**74**	**75**
	62	**63**	**64**	**65**	**65A**	**66**	**67**	**68**	**69**	**70**	**71**	**72**	**73**	**74**	**75**
PPE	Glu	Leu	Thr	Phe	Arg	Val	Val	Val	Gly	Glu	His	Asn	Leu	Asn	Gln
	76	**77**	**78**	**79**	**80**	**81**	**82**	**83**	**84**	**85**	**86**	87	**88**	**89**	**90**
	76	**77**	**78**	**79**	**80**	**81**	**82**	**83**	**84**	**85**	**86**	**87**	**88**	**89**	**90**
PPE	Asn	Asp	Gly	Thr	Gly	Gln	Tyr	Val	Gly	Val	Gln	Lys	Ile	Val	Val
	91	**92**	**93**	**94**	**95**	**96**	**97**	**98**	**99**	**100**	**101**	**102**	**103**	**104**	**105**
	91	**92**	**93**	**94**	**95**	**96**	97	**98**	**99**	**99A**	**99B**	**100**	**101**	**102**	**103**
PPE	His	Pro	Tyr	Trp	Asn	Thr	Asp	Asp	Val	Ala	Ala	Gly	Tyr	Asp	Ile
	16	**107**	**108**	**109**	**110**	**111**	**112**	**113**	**114**	**115**	**116**	**117**	**118**	**119**	**120**
	104	**105**	**106**	**107**	**108**	**109**	**110**	**111**	**112**	**113**	**114**	**115**	**116**	**117**	**118**
PPE	Ala	Leu	Leu	Arg	Leu	Ala	Gln	Ser	Val	Thr	Leu	Asn	Ser	Tyr	Val
	121	**122**	**123**	**124**	**125**	**126**	**127**	**128**	**129**	**130**	**131**	**132**	**133**	**134**	**135**
	119	**120**	**121**	**122**	**123**	**124**	**125**	**126**	**127**	**128**	**129**	**130**	**131**	**132**	**133**
PPE	Gln	Leu	Gly	Val	Leu	Pro	Arg	Ala	Gly	Thr	Ile	Leu	Ala	Asn	Asn
	136	**137**	**138**	**139**	**140**	**141**	**142**	**143**	**144**	**145**	**146**	**147**	**148**	**149**	**150**
	134	**135**	**136**	**137**	**138**	**139**	**140**	**141**	**142**	**143**	**144**	**145**	**147**	**148**	**149**
PPE	Ser	Pro	Cys	Tyr	Ile	Thr	Gly	Trp	Gly	Leu	Thr	Arg	Thr	Asn	Gly
	151	**152**	**153**	**154**	**155**	**156**	**157**	**158**	**159**	**160**	**161**	**162**	**163**	**164**	**165**
	150	**151**	**152**	**153**	**154**	**155**	**156**	**157**	**158**	**159**	**160**	**161**	**162**	**163**	**164**
PPE	Gln	Leu	Ala	Gln	Thr	Leu	Gln	Gln	Ala	Tyr	Leu	Pro	Thr	Val	Asp
	166	**167**	**168**	**169**	**170**	**171**	**172**	**173**	**174**	**175**	**176**	**177**	**178**	**179**	**180**
	165	**166**	**167**	**168**	**169**	**170**	**170A**	**170B**	**171**	**172**	173	**174**	**175**	**176**	**177**
PPE	Tyr	Ala	Ile	Cys	Ser	Ser	Ser	Ser	Tyr	Trp	Gly	Ser	Thr	Val	Lys
	181	**182**	**183**	**184**	**185**	**186**	**187**	**188**	**189**	**190**	**191**	**192**	**193**	**194**	**195**
	178	**179**	**180**	**181**	**182**	**183**	**184**	**185**	**186**	**187**	**188**	**188A**	**189**	**190**	**191**
PPE	Asn	Ser	Met	Val	Cys	Ala	Gly	Gly	Asp	Gly	Val	Arg	Ser	Gly	Cys
	196	**197**	**198**	**199**	**200**	**201**	**202**	**203**	**204**	**205**	**206**	**207**	**208**	**209**	**210**
	192	**193**	**194**	**195**	**196**	**197**	**198**	**199**	**200**	**201**	**202**	**203**	**204**	**205**	**206**
PPE	Gln	Gly	Asp	Ser	Gly	Gly	Pro	Leu	His	Cys	Leu	Val	Asn	Gly	Gln
	211	**212**	**213**	**214**	**215**	**216**	**217**	**218**	**219**	**220**	**221**	**222**	**223**	**224**	**225**
	207	**208**	**209**	**210**	**211**	**212**	**213**	**214**	**215**	**216**	**217**	**217A**	**218**	**219**	**220**
PPE	Tyr	Ala	Val	His	Gly	Val	Thr	Ser	Phe	Val	Ser	Arg	Leu	Gly	Cys
	226	**227**	**228**	**229**	**230**	**231**	**232**	**233**	**234**	**235**	**236**	**237**	**238**	**239**	**240**
	221	**221A**	**222**	**223**	**224**	**225**	**226**	**227**	**228**	**229**	**230**	**231**	**232**	**233**	**234**
PPE	Asn	Val	Thr	Arg	Lys	Pro	Thr	Val	Phe	Thr	Arg	Val	Ser	Ala	Tyr
	241	**242**	**243**	**244**	**245**	**246**	**247**	**248**	**249**	**250**	**251**	**252**	**253**	**254**	**255**
	235	**236**	**237**	**238**	**239**	**240**	**241**	**242**	**243**	**244**	**245**				
PPE	Ile	Ser	Trp	Ile	Asn	Asn	Val	Ile	Ala	Ser	Asn				
	256	**257**	**258**	**259**	**260**	**261**	**262**	**263**	**264**	**265**	**266**				

Appendix C: Crystal and Refinement Statistics

Crystallisation, data collection and refinement statistics.

Dataset	**1. Apo:PPE screening**	**2. Apo:PPE screening**	**3. Apo:PPE screening**
Crystallisation Conditions	33.5 mg mL^{-1} PPE 0.3 M Na phosphate, 0.3 M K phosphate, 0.033 M Na acetate pH 4.6, +20% v/v glycerol	33.5 mg mL^{-1} PPE 5% v/v PEG 3350, 0.033 M Na succinic acid, +20% v/v glycerol	25 mg mL^{-1} PPE 10% v/v PEG 3000, 0.05 M Na Citrate pH 5.5, +20% v/v glycerol
Figure	Figure 2.2, A	Figure 2.2, B	Figure 2.1, C
Data collection (T in K)	Cryo-MX (100 K)	Cryo-MX (100 K)	Cryo-MX (100 K)
Beamline (wavelength Å)	DLS I03 (0.9763 Å)	DLS I03 (0.9763 Å)	DLS I03 (0.9763 Å)
Detector	Dectris Eiger2 XE 16M	Dectris Eiger2 XE 16M	Dectris Eiger2 XE 16M
Data processing	Dials	Dials	Dials
Merging Statistics			
Space group	*P* 2_1 2_1 2_1	*P* 2_1 2_1 2_1	*P* 2_1 2_1 2_1
Cell dimensions			
- *a, b, c* (Å)	49.82, 57.72, 74.23	49.83, 57.44, 73.85	49.87, 57.62, 73.81
- α, β, γ (˚)	90, 90, 90	90, 90, 90	90, 90, 90
No. of molecules/ASU	1	1	1
Unique reflections	42859 (2098)*	42501 (2091)*	53090 (2590)*
Resolution (Å)	45.57-1.40 (1.42-1.40)*	45.34-1.40 (1.42-1.40)*	45.42-1.30 (1.32-1.30)*
R_{split}/R_{meas}	0.041 (0.213)*	0.058 (0.314)*	0.072 (0.718)*
I/σI	39.1 (8.0)*	27.3 (5.0)*	17.0 (1.0)*
$CC_{1/2}$	1.0 (0.998)*	0.999 (0.975)*	1.0 (0.843)*
Completeness (%)	100 (99.76)*	100 (99.15)*	100 (98.14)*
Multiplicity	13.26 (13.56)*	13.23 (13.41)*	12.65 (7.62)*
Wilson B value ($Å^2$)	12.20	11.91	14.7
Refinement	PHENIX	PHENIX	PHENIX
PHASER	3EST	3EST	3EST
R_{work}/R_{free}	0.1424/0.1579	0.1434/0.1523	0.1553/0.1733
No. atoms	2143	2137	2092
- Enzyme	1816	1816	1818
- Ligand	25	27	37
- Water	310	306	247
Average B-factors	16.32	15.48	18.63
- Enzyme	14.32	13.59	16.90
- Ligand	31.18	28.91	32.14
- Water	27.25	26.07	29.85
R.m.s deviations			
- Bond lengths (Å)	0.008	0.009	0.005
- Bond angles (Å)	1.04	1.08	0.88
Ramachandran Plot (%)			
Favoured	97.9	97.06	97.48
Allowed	2.1	2.94	2.52
Outliers	0	0	0

* values in parentheses correspond to the highest resolution shell.

Crystallisation, data collection and refinement statistics.

Dataset	**4. Apo:PPE screening**	**5. Apo:PPE screening**	**6. Apo:PPE screening**
Crystallisation Conditions	33.5 mg mL^{-1} PPE 8.25% v/v PEG 1500, 0.033 M MIB buffer pH 4.0, +20% v/v glycerol	25 mg mL^{-1} PPE 12.5% v/v PEG 1500, 0.05 M MMT buffer pH 4.0, +20% v/v glycerol	33.5 mg mL^{-1} PPE 0.15 M Na F, 0.025 M acetate buffer pH 4.6, +20% v/v glycerol
Figure	Figure 2.2, D	Figure 2.2, E	Figure 2.2, F
Data collection (T in K)	Cryo-MX (100 K)	Cryo-MX (100 K)	Cryo-MX (100 K)
Beamline (wavelength Å)	DLS I03 (0.9763 Å)	DLS I03 (0.9763 Å)	DLS I03 (0.9763 Å)
Detector	Dectris Eiger2 XE 16M	Dectris Eiger2 XE 16M	Dectris Eiger2 XE 16M
Data processing	Dials	3dii	Dials
Merging Statistics			
Space group	*P* 2_1 2_1 2_1	*P* 2_1 2_1 2_1	*P* 2_1 2_1 2_1
Cell dimensions			
- *a, b, c* (Å)	50.02, 57.62, 74.18	50.12, 57.60, 74.40	49.38, 57.64, 73.68
- α, β, γ (˚)	90, 90, 90	90, 90, 90	90, 90, 90
No. of molecules/ASU	1	1	1
Unique reflections	42338 (1991)*	43159 (2086)*	42159 (2043)*
Resolution (Å)	45.51-1.40 (1.42-1.40)*	41.57-1.40 (1.42-1.40)*	45.40-1.40 (1.42-1.40)*
R_{split}/R_{meas}	0.076 (0.422)*	0.048 (0.392)*	0.038 (0.314)*
I/σI	20.2 (2.9)*	29.3 (5.8)*	18.3 (2.2)*
$CC_{1/2}$	0.999 (0.957)*	1.0 (0.954)*	0.999 (0.949)*
Completeness (%)	98.52 (95.40)*	100 (98.63)*	100 (98.79)*
Multiplicity	13.37 (13.76)*	13.03 (11.55)*	13.26 (13.41)*
Wilson B value ($Å^2$)	12.51	14.88	13.32
Refinement	PHENIX	PHENIX	PHENIX
PHASER	3EST	3EST	3EST
R_{work}/R_{free}	0.1596/0.1753	0.1409/0.1573	0.1530/0.1681
No. atoms	2128	2142	2056
- Enzyme	1822	1817	1820
- Ligand	11	14	26
- Water	300	315	224
Average B-factors	16.58	18.73	17.88
- Enzyme	14.81	16.81	16.57
- Ligand	22.22	28.32	31.09
- Water	27.20	29.53	27.85
R.m.s deviations			
- Bond lengths (Å)	0.009	0.006	0.007
- Bond angles (Å)	1.15	0.86	0.95
Ramachandran Plot (%)			
Favoured	97.06	97.9	97.06
Allowed	2.94	2.1	2.94
Outliers	0	0	0

* values in parentheses correspond to the highest resolution shell.

Crystallisation, data collection and refinement statistics.

Dataset	**7. Apo:PPE halides**	**8. Apo:PPE halides**	**9. Apo:PPE halides**
Crystallisation Conditions	25 mg mL^{-1} 50 mM NaCl, 25 mM Na acetate pH 5.0	25 mg mL^{-1}, 50 mM NaCl, 25 mM Na acetate pH 5.0, +20% v/v glycerol	25 mg mL^{-1}, 25 mM NaBr, 25 mM Na acetate pH 5.0, +20% v/v glycerol
Figure	Figure 2.3, B	Figure 2.3, C	Figure 2.3, D
Data collection (T in K)	Cryo-MX (100 K)	Cryo-MX (100 K)	Cryo-MX (100 K)
Beamline (wavelength Å)	DLS I03 (0.9763 Å)	DLS I03 (0.9763 Å)	DLS I03 (0.9763 Å)
Detector	Dectris Eiger2 XE 16M	Dectris Eiger2 XE 16M	Dectris Eiger2 XE 16M
Data processing	3dii	3dii	3dii
Merging Statistics			
Space group	*P* 2_1 2_1 2_1	*P* 2_1 2_1 2_1	*P* 2_1 2_1 2_1
Cell dimensions			
- *a, b, c* (Å)	50.27, 58.07, 74.69	50.11, 58.04, 74.38	50.21, 57.96, 74.59
- α, β, γ (˚)	90, 90, 90	90, 90, 90	90, 90, 90
No. of molecules/ASU	1	1	1
Unique reflections	48768 (2377)*	54106 (2653)*	54304 (2698)*
Resolution (Å)	45.84-1.35 (1.37-1.35)*	41.56-1.30 (1.32-1.30)*	45.77-1.30 (1.32-1.30)*
R_{split}/R_{meas}	0.064 (1.755)*	0.088 (1.500)*	0.079 (2.208)*
I/σI	19.1 (1.4)*	15.4 (1.7)*	16.5 (1.1)*
$CC_{1/2}$	1.0 (0.659)*	0.999 (0.671)*	1.0 (0.572)*
Completeness (%)	100 (100)*	100 (99.70)*	100 (99.56)*
Multiplicity	12.95 (12.95)*	13.19 (12.71)*	13.12 (12.53)*
Wilson B value ($Å^2$)	17.62	14.33	15.96
Refinement	PHENIX	PHENIX	PHENIX
PHASER	**6**	**6**	**6**
R_{work}/R_{free}	0.1623/0.1843	0.1571/0.1834	0.1601/0.1800
No. atoms	2176	2179	2152
- Enzyme	1825	1823	1816
- Ligand	8	36	37
- Water	346	339	318
Average B-factors	21.87	17.93	21.03
- Enzyme	20.11	16.09	19.25
- Ligand	21.94	24.93	35.45
- Water	31.18	27.47	30.36
R.m.s deviations			
- Bond lengths (Å)	0.007	0.006	0.007
- Bond angles (Å)	0.83	0.83	0.87
Ramachandran Plot (%)			
Favoured	97.9	97.06	97.48
Allowed	2.1	2.94	2.52
Outliers	0	0	0

* values in parentheses correspond to the highest resolution shell.

Crystallisation, data collection and refinement statistics.

Dataset	**10. Apo:PPE halides**	**11. PPE:BCM7 (NaF)**	**12. PPE:BCM7 (NaCl)**
Crystallisation Conditions	25 mg mL^{-1}, 37.5 mM NaI, 25 mM Na acetate pH 5.0, +20% v/v glycerol	16.7 mg mL^{-1}, 30 eq BCM7, 200 mM NaF, 25 mM Na acetate pH 5.0,	16.7 mg mL^{-1}, 30 eq BCM7, 125 mM NaCl, 50 mM Na acetate pH 5.0, +20% v/v glycerol
Figure	Figure 2.3, E	Figure 2.4, C	Figure 2.4, D
Data collection (T in K)	Cryo-MX (100 K)	Cryo-MX (100 K)	Cryo-MX (100 K)
Beamline (wavelength Å)	DLS I03 (0.9763 Å)	DLS I03 (0.9763 Å)	DLS I03 (0.9763 Å)
Detector	Dectris Eiger2 XE 16M	Dectris Eiger2 XE 16M	Dectris Eiger2 XE 16M
Data processing	Dials	Dials	Dials
Merging Statistics			
Space group	*P* 2_1 2_1 2_1	*P* 2_1 2_1 2_1	*P* 2_1 2_1 2_1
Cell dimensions			
- *a, b, c* (Å)	50.17, 58.02, 74.74	50.12, 58.00, 74.80	50.05, 58.33, 74.66
- α, β, γ (˚)	90, 90, 90	90, 90, 90	90, 90, 90
No. of molecules/ASU	1	1	1
Unique reflections	43776 (2184)*	43451 (2131)*	54438 (2642)*
Resolution (Å)	45.83-1.40 (1.42-1.40)*	45.83-1.40 (1.42-1.40)*	41.57-1.30 (1.32-1.30)*
R_{split}/R_{meas}	0.140 (1.506)*	0.133 (1.057)*	0.047 (0.369)*
I/σI	10.8 (0.9)*	10.9 (1.0)*	19.2 (2.6)*
$CC_{1/2}$	0.999 (0.750)*	0.996 (0.828)*	0.999 (0.910)*
Completeness (%)	100 (100)*	99.51 (98.38)*	99.96 (98.14)*
Multiplicity	13.24 (13.04)*	13.26 (12.88)*	12.86 (8.20)*
Wilson B value ($Å^2$)	15.07	13.78	10.60
Refinement	PHENIX	PHENIX & REFMAC	PHENIX & REFMAC
PHASER	**6**	**37**	**37**
R_{work}/R_{free}	0.1853/0.2050	0.1794/0.1983	0.1599/0.1797
No. atoms	2099	2206	2201
- Enzyme	1822	1816	1823
- Ligand	24	32	39
- Water	264	358	339
Average B-factors	19.10	23.62	20.65
- Enzyme	17.84	15.29	12.01
- Ligand	37.18	19.49	15.93
- Water	26.91	29.29	26.95
R.m.s deviations			
- Bond lengths (Å)	0.003	0.010	0.012
- Bond angles (Å)	0.62	1.62	1.77
Ramachandran Plot (%)			
Favoured	97.90	97.5	97.5
Allowed	2.1	2.5	2.5
Outliers	0	0	0

* values in parentheses correspond to the highest resolution shell.

Crystallisation, data collection and refinement statistics.

Dataset	**13.1 PPE:BCM7 pH jump (tetrahedral)**	**13.2 PPE:BCM7 pH jump (acetate and glycerol)**	**14. PPE:BCM7 pH jump (deacylated with glycerol)**
Crystallisation Conditions	16.7 mg mL^{-1} PPE, 30 eq BCM7, 200 mM NaCl, 50 mM acetate buffer pH 5.0. The crystal was soaked in NaCl at the same concentration as the mother liquor mixed 1:1 v/v with pH jump buffer (final [C] = 100 mM CHES pH 9.0, 20% v/v glycerol) for 1 min 30 s		16.7 mg mL^{-1}, 30 eq BCM7, 162.5 mM NaCl, 50 mM acetate buffer pH 5.0. The crystal was soaked in mother liquor mixed 1:1 v/v with pH jump buffer (final [C] = 100 mM Tris pH 9.0, 20% v/v glycerol) for 45 s
Figure	Figure 2.6, A-D	Figure 2.6, G-J	Figure 2.7, A
Data collection (T in K)	Cryo-MX (100 K)		Cryo-MX (100 K)
Beamline (wavelength Å)	DLS I03 (0.9762 Å)		DLS I03 (0.9762 Å)
Detector	Dectris Eiger2 XE 16M		Dectris Eiger2 XE 16M
Data processing	Dials		Dials
Merging Statistics			
Space group	*P* 2_1 2_1 2_1		*P* 2_1 2_1 2_1
Cell dimensions			
- *a, b, c* (Å)	50.68, 57.87, 74.68		51.39, 58.03, 75.02
- α, β, γ (˚)	90, 90, 90		90, 90, 90
No. of molecules/ASU	1		1
Unique reflections	39688 (1890)*		21432 (1027)*
Resolution (Å)	45.74-1.45 (1.47-1.45)*		45.90-1.80 (1.83-1.80)*
R_{split}/R_{meas}	0.103 (1.320)*		0.141 (1.088)*
I/σI	15.1 (1.0)*		12.7 (1.6)*
$CC_{1/2}$	0.999 (0.750)*		0.998 (0.822)*
Completeness (%)	100 (98.59)*		100 (97.07)*
Multiplicity	13.34 (13.61)*		13.23 (13.41)*
Wilson B value ($Å^2$)	16.35		21.17
Refinement	PHENIX	PHENIX	PHENIX
PHASER	1HAZ		**12**
R_{work}/R_{free}	0.1668/0.1960	0.1618/0.1927	0.1661/0.2020
No. atoms	2099	2100	1988
- Enzyme	1826	1817	1820
- Ligand	15	36	15
- Water	266	266	161
Average B-factors	20.96	21.74	26.51
- Enzyme	19.46	20.16	26.09
- Ligand	33.16	30.74	31.36
- Water	30.91	31.96	31.06
R.m.s deviations			
- Bond lengths (Å)	0.005	0.005	0.003
- Bond angles (Å)	0.84	0.78	0.63
Ramachandran Plot (%)			
Favoured	98.32	98.32	97.48
Allowed	1.68	1.68	2.52
Outliers	0	0	0

* values in parentheses correspond to the highest resolution shell.

Crystallisation, data collection and refinement statistics.

Dataset	**15. Apo-PPE soaked in CHES buffer pH 9.0**	**16. PPE:BCM7 conformational change**
Crystallisation Conditions	25 mg mL^{-1}, 50 mM NaCl, 50 mM acetate buffer pH 5.0. The crystal was soaked in NaCl at the same concentration as the mother liquor mixed 1:1 with pH jump solution (final [C] = 100 mM CHES pH 9.0, 20% v/v glycerol) for 30 s	16.7 mg mL^{-1} PPE, 30 eq BCM7, 200 mM NaF, 50 mM acetate buffer pH 5.0. The crystal was soaked in NaF at the same concentration as the mother liquor mixed 1:1 with pH jump solution (final [C] = 100 mM CHES pH 9.0, 20% v/v glycerol) for 1 min.
Figure	Figure 2.7, B	Figure 2.8
Data collection (T in K)	Cryo-MX (100 K)	Cryo-MX (100 K)
Beamline (wavelength Å)	DLS I03 (0.9763 Å)	DLS I03 (0.9763 Å)
Detector	Dectris Eiger2 XE 16M	Dectris Eiger2 XE 16M
Data processing	Dials	Dials
Merging Statistics		
Space group	*P* 2_1 2_1 2_1	*P* 2_1 2_1 2_1
Cell dimensions		
- *a, b, c* (Å)	50.12, 57.90, 74.68	52.86, 57.30, 75.06
- α, β, γ (˚)	90, 90, 90	90, 90, 90
No. of molecules/ASU	1	1
Unique reflections	53564 (2599)*	56751 (5574)*
Resolution (Å)	45.76-1.30 (1.32-1.30)*	43.22-1.30 (1.346-1.30)*
R_{split}/R_{meas}	0.094 (0.919)*	0.08069 (0.9321)*
I/σI	15.9 (1.7)*	19.06 (1.56)*
$CC_{1/2}$	0.998 (0.846)*	1.0 (0.880)*
Completeness (%)	98.94 (97.27)*	99.96 (99.93)*
Multiplicity	13.37 (13.34)*	13.2 (12.4)*
Wilson B value ($Å^2$)	12.17	12.84
Refinement	PHENIX	PHENIX
PHASER	**6**	**37**
R_{work}/R_{free}	0.1514/0.1684	0.1524/0.1666
No. atoms	2145	2224
- Enzyme	1820	1836
- Ligand	49	61
- Water	303	361
Average B-factors	16.90	17.30
- Enzyme	15.07	15.11
- Ligand	26.11	25.66
- Water	27.19	27.82
R.m.s deviations		
- Bond lengths (Å)	0.013	0.148
- Bond angles (Å)	1.19	3.05
Ramachandran Plot (%)		
Favoured	97.48	97.9
Allowed	2.52	2.1
Outliers	0	0

* values in parentheses correspond to the highest resolution shell.

Crystallisation, data collection and refinement statistics.

Dataset	**17. Apo-PPE Na_2SO_4 SFX**	**18. PPE:BCM7 (NaF) SFX**	**19. PPE:BCM7 (NaF) SSX: ground**
Crystallisation Conditions	10 mg mL^{-1} PPE, 50 mM Na_2SO_4, 50 mM acetate buffer pH 5.0.	10 mg mL^{-1} PPE, 30 eq BCM7, 400 mM NaF, 50 mM acetate buffer pH 5.0 + microseeds.	10 mg mL^{-1} PPE, 30 eq BCM7, 400 mM NaF, 50 mM acetate buffer pH 5.0 + microseeds.
Figure	Figure 2.12, A	Figure 2.13	Figure 2.14, A
Data collection (T in K)	Fixed target-SFX (298)	Fixed target-SFX (298)	Fixed target-SSX (298)
Beamline (wavelength Å)	NCI SFX (0.998213 Å)	NCI SFX (0.999179 Å)	DLS I24 (0.999870 Å)
Detector	Jungfrau 4M	Jungfrau 4M	Dectris Pilatus3 6M
Data processing	cctbx.xfel	cctbx.xfel	cctbx.xfel
Merging Statistics			
Space group	*P* 2_1 2_1 2_1	*P* 2_1 2_1 2_1	*P* 2_1 2_1 2_1
Cell dimensions			
- *a, b, c* (Å)	51.21, 58.05, 75.31	51.31, 58.38, 75.64	51.70, 58.07, 75.40
- α, β, γ (˚)	90, 90, 90	90, 90, 90	90, 90, 90
No. of molecules/ASU	1	1	1
Unique reflections	50009 (4922)*	33665 (1660)	23072 (1086)
Resolution (Å)	42.35-1.35 (1.3733-1.35)*	46.21-1.55 (1.5767-1.55)*	46.0-1.76 (1.7903-1.76)*
R_{split}/R_{meas}	0.157 (0.597)*	0.140 (0.128)*	0.120 (0.471)*
I/σI	2.053 (0.269)*	5.021 (3.869)*	4.966 (1.376)*
$CC_{1/2}$	0.968 (0.199)*	0.963 (0.341)*	92.4 (67.2)*
Completeness (%)	99.99 (99.96)*	100 (100)*	99.78 (97.49)*
Multiplicity	89.05 (17.23)*	136.78 (87.73)*	62.91 (9.02)*
Wilson B value ($Å^2$)	15.29	1.14	19.64
Serial data			
- No. of lattices	17011	34044	11864
Refinement	PHENIX	PHENIX	PHENIX & REFMAC
PHASER	3EST	**12**	**37**
R_{work}/R_{free}	0.1779/0.2105	0.2261/0.2700	0.1661/0.1914
No. atoms	2000	2053	1986
- Enzyme	1812	1823	1825
- Ligand	11	32	32
- Water	177	198	129
Average B-factors	21.96	15.83	27.76
- Enzyme	20.82	10.79	23.86
- Ligand	48.74	11.15	26.69
- Water	31.99	22.10	35.21
R.m.s deviations			
- Bond lengths (Å)	0.005	0.013	0.009
- Bond angles (Å)	0.77	1.20	1.55
Ramachandran Plot (%)			
Favoured	98.32	97.49	97.5
Allowed	1.68	2.51	2.5
Outliers	0	0	0

* values in parentheses correspond to the highest resolution shell.

Crystallisation, data collection and refinement statistics.

Dataset	**20. PPE:BCM7 (NaF) SSX: CHES (200 mM) 1.12s**	**21. PPE:BCM7 (NaF) SSX: CHES (200 mM) 5.60 s**	**22. PPE:BCM7 (NaF) SSX: ground**
Crystallisation Conditions	10 mg mL^{-1} PPE, 30 eq BCM7, 400 mM NaF, 50 mM acetate buffer pH 5.0 + microseeds.	10 mg mL^{-1} PPE, 30 eq BCM7, 400 mM NaF, 50 mM acetate buffer pH 5.0 + microseeds.	10 mg mL^{-1} PPE, 30 eq BCM7, 400 mM NaF, 50 mM acetate buffer pH 5.0 + microseeds.
Figure	Figure 2.14, B	Figure 2.14, C	Figure 2.15, A
Data collection (T in K)	DoC-SSX (298 K)	DoC-SSX (298 K)	Fixed target-SSX (298 K)
Beamline (wavelength Å)	DLS I24 (0.999870 Å)	DLS I24 (0.999870 Å)	DLS I24 (0.999870 Å)
Detector	Dectris Pilatus3 6M	Dectris Pilatus3 6M	Dectris Pilatus3 6M
Data processing	cctbx.xfel	cctbx.xfel	cctbx.xfel
Merging Statistics			
Space group	*P* 2_1 2_1 2_1	*P* 2_1 2_1 2_1	*P* 2_1 2_1 2_1
Cell dimensions			
- *a, b, c* (Å)	52.07, 57.97, 75.44	51.83, 57.97, 75.70	51.08, 58.19, 75.38
- α, β, γ (˚)	90, 90, 90	90, 90, 90	90, 90, 90
No. of molecules/ASU	1	1	1
Unique reflections	19128 (925)*	21676 (1040)*	27690 (1342)*
Resolution (Å)	45.97-1.88 (1.9124-1.88)*	42.73-1.80 (1.831-1.80)*	46.06-1.65 (1.6785-1.65)*
R_{split}/R_{meas}	0.131 (0.486)*	0.129 (0.446)*	0.133(0.375)*
I/σI	4.539 (1.419)*	4.954 (1.746)*	4.8 (1.8)*
$CC_{1/2}$	0.973 (0.676)*	0.976 (0.658)*	0.976 (0.679)*
Completeness (%)	99.69 (98.30)*	99.87 (98.86)*	99.97 (99.63)*
Multiplicity	48.3 (9.42)*	51.01 (10.76)*	47.64 (11.36)*
Wilson B value ($Å^2$)	23.47	20.33	16.19
Serial data			
No. of lattices	7503	11598	12621
Refinement	PHENIX & REFMAC	PHENIX & REFMAC	PHENIX & REFMAC
PHASER	**37**	**37**	**37**
R_{work}/R_{free}	0.1698/0.2143	0.1757/0.2085	0.1529/0.1810
No. atoms	1970	2002	2035
- Enzyme	1824	1822	1826
- Ligand	32	32	32
- Water	114	148	177
Average B-factors	32.08	28.62	24.54
- Enzyme	28.84	24.18	18.56
- Ligand	31.73	27.15	21.08
- Water	39.07	36.00	32.88
R.m.s deviations			
- Bond lengths (Å)	0.009	0.010	0.015
- Bond angles (Å)	1.62	1.62	2.03
Ramachandran Plot (%)			
Favoured	97.5	97.5	97.5
Allowed	2.5	2.5	2.5
Outliers	0	0	0

* values in parentheses correspond to the highest resolution shell.

Crystallisation, data collection and refinement statistics.

Dataset	**23. PPE:BCM7 (NaF) SSX: Tris (200 mM) 1.12s**	**24. PPE:BCM7 (NaF) SSX: Tris (200 mM) 5.60 s**	**25. PPE:BCM7 (NaCl) SSX: ground**
Crystallisation Conditions	10 mg mL^{-1} PPE, 30 eq BCM7, 400 mM NaF, 50 mM acetate buffer pH 5.0 + microseeds.	10 mg mL^{-1} PPE, 30 eq BCM7, 400 mM NaF, 50 mM acetate buffer pH 5.0 + microseeds.	16.7 mg mL^{-1} PPE, 30 eq BCM7, 100 mM NaCl, 50 mM acetate buffer pH 5.0 + microseeds
Figure	Figure 2.15, B	Figure 2.15, C	Figure 2.16, A
Data collection (T in K)	DoC-SSX (298 K)	DoC-SSX (298 K)	Fixed target-SSX (298 K)
Beamline (wavelength Å)	DLS I24 (0.999870 Å)	DLS I24 (0.999870 Å)	DLS I24 (0.999870 Å)
Detector	Dectris Pilatus3 6M	Dectris Pilatus3 6M	Dectris Pilatus3 6M
Data processing	cctbx.xfel	cctbx.xfel	cctbx.xfel
Merging Statistics			
Space group	*P* 2_1 2_1 2_1	*P* 2_1 2_1 2_1	*P* 2_1 2_1 2_1
Cell dimensions			
- *a, b, c* (Å)	51.25, 58.15, 75.45	51.25, 58.13, 75.45	51.26, 58.14, 75.42
- α, β, γ (˚)	90, 90, 90	90, 90, 90	90, 90, 90
No. of molecules/ASU	1	1	1
Unique reflections	27785 (1347)*	25451 (1251)	27778 (1355)
Resolution (Å)	46.06-1.65 (1.6785-1.65)*	46.05-1.70 (1.7293-1.70)*	46.05-1.65 (1.6785-1.65)*
R_{split}/R_{meas}	0.101 (0.624)*	0.128 (0.399)*	0.121 (0.273)*
I/σI	6.0 (1.9)*	4.5 (1.6)*	5.8 (2.9)*
$CC_{1/2}$	0.966 (0.014)*	0.979 (0.717)*	0.972 (0.425)*
Completeness (%)	99.99 (99.70)*	99.96 (99.52)*	100 (99.93)*
Multiplicity	78.04 (12.71)*	42.5 (10.71)*	84.43 (30.22)*
Wilson B value ($Å^2$)	16.71	17.98	15.12
Serial data			
No. of lattices	19820	11363	16108
Refinement	PHENIX & REFMAC	PHENIX & REFMAC	PHENIX & REFMAC
PHASER	**37**	**37**	**37**
R_{work}/R_{free}	0.1436/0.1824	0.1557/0.1872	0.1533/0.1784
No. atoms	2010	2027	2041
- Enzyme	1819	1823	1829
- Ligand	32	32	32
- Water	159	172	180
Average B-factors	24.17	26.04	23.75
- Enzyme	19.11	20.36	17.68
- Ligand	20.76	24.80	18.80
- Water	32.00	34.11	32.15
R.m.s deviations			
- Bond lengths (Å)	0.015	0.015	0.015
- Bond angles (Å)	2.01	2.01	2.12
Ramachandran Plot (%)			
Favoured	97.92	98.33	97.08
Allowed	2.08	1.67	2.92
Outliers	0	0	0

* values in parentheses correspond to the highest resolution shell.

Crystallisation, data collection and refinement statistics.

Dataset	**26. PPE:BCM7 (NaCl) SSX: Tris (200 mM) 1.12s**	**27. PPE:BCM7 (NaCl) SSX: Tris (1 M) 1.12 s**	**28. PPE:BCM7 (NaCl) SSX: Tris (500 mM) 1.12 s**
Crystallisation Conditions	16.7 mg mL^{-1} PPE, 30 eq BCM7, 100 mM NaCl, 50 mM acetate buffer pH 5.0 + microseeds	16.7 mg mL^{-1} PPE, 30 eq BCM7, 100 mM NaCl, 50 mM acetate buffer pH 5.0 + microseeds	16.7 mg mL^{-1} PPE, 30 eq BCM7, 100 mM NaCl, 50 mM acetate buffer pH 5.0 + microseeds
Figure	Figure 2.16, B	Figure 2.17, A	Figure 2.17, C
Data collection (T in K)	DoC-SSX (298 K)	DoC-SSX (298 K)	DoC-SSX (298 K)
Beamline (wavelength Å)	DLS I24 (0.999870 Å)	DLS I24 (0.999870 Å)	DLS I24 (0.999870 Å)
Detector	Dectris Pilatus3 6M	Dectris Pilatus3 6M	Dectris Pilatus3 6M
Data processing	cctbx.xfel	cctbx.xfel	cctbx.xfel
Merging Statistics			
Space group	*P* 2_1 2_1 2_1	*P* 2_1 2_1 2_1	*P* 2_1 2_1 2_1
Cell dimensions			
- *a, b, c* (Å)	51.43, 58.07, 75.43	51.74, 57.88, 75.36	51.01, 58.01, 75.31
- α, β, γ (˚)	90, 90, 90	90, 90, 90	90, 90, 90
No. of molecules/ASU	1	1	1
Unique reflections	29389 (1434)*	30462 (1455)*	27395 (1273)*
Resolution (Å)	46.02-1.62 (1.6479-1.62)*	31.58-1.60 (1.6276-1.60)*	34.14-1.65 (1.6785-1.65)*
R_{split}/R_{meas}	0.120 (0.523)*	0.081 (0.353)*	0.115 (0.786)*
I/σI	5.8 (2.4)*	8.063 (2.447)*	5.553 (1.915)*
$CC_{1/2}$	0.977 (0.239)*	0.988 (0.299)*	0.982 (0.110)*
Completeness (%)	99.98 (99.93)*	99.97 (99.87)*	99.95 (99.48)*
Multiplicity	73.75 (18.57)*	122.75 (20.83)*	78.67 (15.70)*
Wilson B value ($Å^2$)	16.02	16.13	17.45
Serial data			
No. of lattices	13146	19618	9449
Refinement	PHENIX & REFMAC	PHENIX & REFMAC	PHENIX & REFMAC
PHASER	**37**	**12**	**12**
R_{work}/R_{free}	0.1570/0.1918	0.1514/0.1746	0.1768/0.2063
No. atoms	2014	2007	1995
- Enzyme	1820	1819	1817
- Ligand	32	32	32
- Water	162	156	146
Average B-factors	24.94	24.91	25.63
- Enzyme	19.18	19.61	21.04
- Ligand	20.34	24.86	24.12
- Water	33.70	33.16	33.35
R.m.s deviations			
- Bond lengths (Å)	0.015	0.015	0.015
- Bond angles (Å)	2.07	2.1	2.06
Ramachandran Plot (%)			
Favoured	98.33	97.5	98.33
Allowed	1.67	2.5	1.67
Outliers	0	0	0

* values in parentheses correspond to the highest resolution shell.

Crystallisation, data collection and refinement statistics.

Dataset	**29. PPE:BCM7 (NaCl) SSX: Tris (500 mM) 5.60 s**	**30. PPE:Sivelestat**	**31. PPE:4**
Crystallisation Conditions	16.7 mg mL^{-1} PPE, 30 eq BCM7, 100 mM NaCl, 50 mM acetate buffer pH 5.0 + microseeds	20 mg mL^{-1} PPE, 15 mM NaCl, 50 mM acetate buffer pH 5.0. The crystal was soaked in mother liquor with 20 mM Sivelestat and 20% v/v DMSO for 5 min.	25 mg mL^{-1} PPE, 37.5 mM NaCl, 25 mM acetate buffer pH 5.0. The crystal was soaked in mother liquor with 5 mM **4** and 20% v/v DMSO for 2 min 35 s.
Figure	Figure 2.17, D	Figure 3.4, C	Figure 5.12, A
Data collection (T in K)	DoC-SSX (298 K)	Cryo-MX (100 K)	Cryo-MX (100 K)
Beamline (wavelength Å)	DLS I24(0.999870 Å)	DLS I04 (0.8739 Å)	DLS I03 (0.9762 Å)
Detector	Dectris Pilatus3 6M	Dectris Eiger2 XE 16M	Dectris Eiger2 XE 16M
Data processing	cctbx.xfel	Dials	Dials
Merging Statistics			
Space group	*P* 2_1 2_1 2_1	*P* 2_1 2_1 2_1	*P* 2_1 2_1 2_1
Cell dimensions			
- *a, b, c* (Å)	51.65, 57.90, 75.36	51.01, 58.84, 74.76	50.52, 58.25, 75.00
- α, β, γ (˚)	90, 90, 90	90, 90, 90	90, 90, 90
No. of molecules/ASU	1	1	1
Unique reflections	30207 (1318)*	36274 (1723)	27337 (1354)*
Resolution (Å)	31.58-1.60 (1.6276-1.60)*	42.13-1.50 (1.53-1.50)*	46.01-1.65 (1.68-1.60)*
R_{split}/R_{meas}	0.129 (0.412)*	0.116 (0.837)*	0.111 (0.939)*
I/σI	4.920 (1.785)*	13.9 (1.9)*	14.8 (1.5)*
$CC_{1/2}$	0.972 (0.422)*	0.999 (0.881)*	0.999 (0.895)*
Completeness (%)	99.93 (99.20)*	98.75 (97.13)*	100 (99.85)*
Multiplicity	59.03 (11.36)*	13.69 (14.13)*	13.32 (13.66)*
Wilson B value ($Å^2$)	16.55	15.46	18.99
Serial data			
No. of lattices	13462	—	—
Refinement	PHENIX & REFMAC	PHENIX	PHENIX
PHASER	**12**	**7**	**7**
R_{work}/R_{free}	0.1569/0.1887	0.1698/0.1849	0.1856/0.2139
No. atoms	2017	1993	2009
- Enzyme	1823	1804	1814
- Ligand	32	116	120
- Water	162	142	141
Average B-factors	25.07	20.85	25.35
- Enzyme	19.49	19.74	24.73
- Ligand	26.78	33.60	31.74
- Water	33.40	30.69	30.98
R.m.s deviations			
- Bond lengths (Å)	0.015	0.008	0.005
- Bond angles (Å)	2.08	0.89	0.81
Ramachandran Plot (%)			
Favoured	97.08	97.9	97.48
Allowed	2.92	2.1	2.52
Outliers	0	0	0

* values in parentheses correspond to the highest resolution shell.

Crystallisation, data collection and refinement statistics.

Dataset	**32. PPE:10**	**33. PPE:12**	**34. PPE:26**
Crystallisation Conditions	25 mg mL^{-1} PPE, 62.5 mM NaCl, 25 mM acetate buffer pH 5.0. The crystal was soaked in mother liquor with 5 mM **10** and 20% v/v DMSO for 3 min.	25 mg mL^{-1} PPE, 62.5 mM NaCl, 25 mM acetate buffer pH 5.0. The crystal was soaked in mother liquor with 10 mM **12** and 20% v/v DMSO for 1 min 20 s.	25 mg mL^{-1} PPE, 62.5 mM NaCl, 50 mM acetate buffer pH 5.0. The crystal was soaked in mother liquor with 10 mM **26** and 20% v/v DMSO for 2 min.
Figure	Figure 5.12, B	Figure 5.12, C	Figure 5.25, A
Data collection (T in K)	Cryo-MX (100 K)	Cryo-MX (100 K)	Cryo-MX (100 K)
Beamline (wavelength Å)	DLS I03 (0.9763 Å)	DLS I03 (0.9763 Å)	DLS I03 (0.9762 Å)
Detector	Dectris Eiger2 XE 16M	Dectris Eiger2 XE 16M	Dectris Eiger2 XE 16M
Data processing	Dials	Dials	3dii
Merging Statistics			
Space group	*P* 2_1 2_1 2_1	*P* 2_1 2_1 2_1	*P* 2_1 2_1 2_1
Cell dimensions			
- *a, b, c* (Å)	51.31, 57.56, 74.66	51.72, 56.93, 75.69	51.61, 57.68, 74.45
- α, β, γ (˚)	90, 90, 90	90, 90, 90	90, 90, 90
No. of molecules/ASU	1	1	1
Unique reflections	43391 (2091)*	44792 (2178)*	36235 (1781)*
Resolution (Å)	45.58-1.40 (1.42-1.40)*	38.28-1.40 (1.42-1.40)*	38.46-1.50 (1.53-1.50)*
R_{split}/R_{meas}	0.063 (0.914)*	0.088 (0.837)*	0.108 (2.334)*
I/σI	20.3 (1.4)*	16.1 (1.2)*	13.7 (1.1)*
$CC_{1/2}$	1.0 (0.909)*	0.999 (0.824)*	1.0 (0.636)*
Completeness (%)	98.03 (96.32)*	100 (98.91)*	99.78 (97.80)*
Multiplicity	13.54 (13.64)*	13.20 (12.95)*	13.21 (12.30)*
Wilson B value ($Å^2$)	16.77	14.22	18.64
Refinement	PHENIX	PHENIX	PHENIX
PHASER	**7**	**7**	**7**
R_{work}/R_{free}	0.1774/0.1900	0.1691/0.1922	0.2028/0.2384
No. atoms	2099	2118	2088
- Enzyme	1823	1823	1817
- Ligand	91	29	45
- Water	231	277	246
Average B-factors	24.71	18.50	29.57
- Enzyme	23.68	17.02	28.66
- Ligand	25.18	15.82	29.11
- Water	32.73	28.44	36.37
R.m.s deviations			
- Bond lengths (Å)	0.004	0.004	0.029
- Bond angles (Å)	0.72	0.78	2.94
Ramachandran Plot (%)			
Favoured	97.9	97.48	97.06
Allowed	2.1	2.52	2.94
Outliers	0	0	0

* values in parentheses correspond to the highest resolution shell.

Crystallisation, data collection and refinement statistics.

Dataset	**35. PPE:29**	**36. PPE:23**	**37. PPE:BCM7 in NaF**
Crystallisation Conditions	25 mg mL^{-1} PPE, 62.5 mM NaCl, 50 mM acetate buffer pH 5.0. The crystal was soaked in mother liquor with 10 mM **29** and 20% v/v DMSO for 3 min.	25 mg mL^{-1} PPE, 62.5 mM NaCl, 50 mM acetate buffer pH 5.0. The crystal was soaked in mother liquor with 20 mM **23** and 20% v/v DMSO for 2 min.	25 mg mL^{-1}, 225 mM NaF, 50 mM acetate buffer, pH 5.0.
Figure	Figure 5.25, D	Figure 5.25, F	N/A
Data collection (T in K)	Cryo-MX (100 K)	Cryo-MX (100 K)	Cryo-MX (100 K)
Beamline (wavelength Å)	DLS I03 (0.9762 Å)	DLS I03 (0.9762 Å)	DLS I03 (0.9762 Å)
Detector	Dectris Eiger2 XE 16M	Dectris Eiger2 XE 16M	Dectris Eiger2 XE 16M
Data processing	Dials	Dials	Dials
Merging Statistics			
Space group	*P* 2_1 2_1 2_1	*P* 2_1 2_1 2_1	*P* 2_1 2_1 2_1
Cell dimensions			
- *a, b, c* (Å)	51.52, 57.46, 74.53	51.65, 57.25, 74.58	49.91, 58.28, 74.55
- α, β, γ (˚)	90, 90, 90	90, 90, 90	90, 90, 90
No. of molecules/ASU	1	1	1
Unique reflections	36182 (1763)*	39943 (1962)*	68999 (3354)*
Resolution (Å)	42.38-1.50 (1.53-1.50)*	45.41-1.45 (1.48-1.45)*	45.91-1.20 (1.22-1.20)*
R_{split}/R_{meas}	0.089 (1.146)*	0.107 (1.141)*	0.087 (0.716)*
I/σI	15.0 (1.0)*	13.7 (1.1)*	16.4 (1.6)*
$CC_{1/2}$	0.999 (0.893)*	0.999 (0.878)*	0.999 (0.823)*
Completeness (%)	100 (99.10)*	100 (99.04)*	100 (99.53)*
Multiplicity	13.32 (13.59)*	13.36 (13.69)*	12.62 (7.08)*
Wilson B value ($Å^2$)	17.82	15.23	10.77
Refinement	PHENIX	PHENIX	PHENIX
PHASER	**7**	**7**	**6**
R_{work}/R_{free}	0.1847/0.2104	0.1799/0.1995	0.1512/0.1622
No. atoms	2088	2131	2158
- Enzyme	1817	1817	1826
- Ligand	30	42	31
- Water	253	292	300
Average B-factors	26.64	20.64	20.88
- Enzyme	25.52	19.31	13.69
- Ligand	32.72	23.71	14.49
- Water	32.24	28.69	26.83
R.m.s deviations			
- Bond lengths (Å)	0.006	0.003	0.006
- Bond angles (Å)	0.82	0.63	0.97
Ramachandran Plot (%)			
Favoured	96.64	97.48	97.91
Allowed	3.36	2.52	2.09
Outliers	0	0	0

* values in parentheses correspond to the highest resolution shell.

Appendix D: Protein Sequences

General Information: The protein sequences are shown with the propeptide sequence of PPE shown in bold and the trypsin cleavage site shown by the arrow (▼). Theoretical protein parameters were calculated using Expasy ProtParam[382]. The values in parentheses are for the cleaved protein product.

Construct: His_6-ProPPE (pNIC)
MHHHHHHSSGVDLGTENLYFQS**TQDFPETNAR**▼VVGGTEAQRNSWPSQISLQYRSGSSWAHTCGGTLIRQNW
VMTAAHCVDRELTFRVVVGEHNLNQNDGTEQYVGVQKIVVHPYWNTDDVAAGYDIALLRLAQSVTLNSYVQLGV
LPRAGTILANNSPCYITGWGLTRTNGQLAQTLQQAYLPTVDYAICSSSSYWGSTVKNSMVCAGGDGVRSGCQGDS
GGPLHCLVNGQYAVHGVTSFVSRLGCNVTRKPTVFTRVSAYISWINNVIASN*

Number of amino acids: 272 (240)
Calculated molecular weight: 29621 (25908)
Ext. coefficient (assuming all Cys pairs form cystines): 56880 (55390)
Calculated pI: 7.17 (8.44)

Construct: His_6-ProPPE (pCOLD)
MHHHHHH**TQDFPETNAR**▼VVGGTEAQRNSWPSQISLQYRSGSSWAHTCGGTLIRQNWVMTAAHCVDRELTFR
VVVGEHNLNQNDGTEQYVGVQKIVVHPYWNTDDVAAGYDIALLRLAQSVTLNSYVQLGVLPRAGTILANNSPCYI
TGWGLTRTNGQLAQTLQQAYLPTVDYAICSSSSYWGSTVKNSMVCAGGDGVRSGCQGDSGGPLHCLVNGQYAV
HGVTSFVSRLGCNVTRKPTVFTRVSAYISWINNVIASN*

Number of amino acids: 257 (240)
Calculated molecular weight: 28022 (25908)
Ext. coefficient (assuming all Cys pairs form cystines): 55390 (55390)
Calculated pI: 8.14 (8.44)

Construct: ProPPE-His_6 (pNIC/pCOLD)
M**TQDFPETNAR**▼VVGGTEAQRNSWPSQISLQYRSGSSWAHTCGGTLIRQNWVMTAAHCVDRELTFRVVVGEH
NLNQNDGTEQYVGVQKIVVHPYWNTDDVAAGYDIALLRLAQSVTLNSYVQLGVLPRAGTILANNSPCYITGWGLT
RTNGQLAQTLQQAYLPTVDYAICSSSSYWGSTVKNSMVCAGGDGVRSGCQGDSGGPLHCLVNGQYAVHGVTSF
VSRLGCNVTRKPTVFTRVSAYISWINNVIASNGGHHHHHH*

Number of amino acids: 259 (248)
Calculated molecular weight: 28136 (26845)
Ext. coefficient (assuming all Cys pairs form cystines): 55390 (55390)
Calculated pI: 8.14 (8.45)

Construct: GST-ProPPE-His_6 (pNIC/pCOLD)
MSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFPNLPYYIDGDVKLTQSMAIIRYIADKH
NMLGGCPKERAEISMLEGAVLDIRYGVSRIAYSKDFETLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTHPDFMLY
DALDVVLYMDPMCLDAFPKLVCFKKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPPKGIEENLYFQS**TQDF**
PETNAR▼VVGGTEAQRNSWPSQISLQYRSGSSWAHTCGGTLIRQNWVMTAAHCVDRELTFRVVVGEHNLNQN
DGTEQYVGVQKIVVHPYWNTDDVAAGYDIALLRLAQSVTLNSYVQLGVLPRAGTILANNSPCYITGWGLTRTNGQ
LAQTLQQAYLPTVDYAICSSSSYWGSTVKNSMVCAGGDGVRSGCQGDSGGPLHCLVNGQYAVHGVTSFVSRLGC
NVTRKPTVFTRVSAYISWINNVIASNGGHHHHHH*

Number of amino acids: 486 (248)
Calculated molecular weight: 54667 (26845)
Ext. coefficient (assuming all Cys pairs form cystines): 99990 (55390)
Calculated pI: 6.65 (8.45)

Construct: MBP-ProPPE-His_6 (pNIC/pCOLD)
MKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHDRFGGYAQSGL
LAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIPALDKELKAKGKSALMFNLQEP
YFTWPLIAADGGYAFKYENGKYDIKDVGVDNAGAKAGLTFLVDLIKNKHMNADTDYSIAEAAFNKGETAMTINGP
WAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFLENYLLTDEGLEAVNKDKPLGAVALKS
YEEELAKDPRIAATMENAQKGEIMPNIPQMSAFWYAVRTAVINAASGRQTVDEALKDAQTNGIEENLYFQS**TQDF**
PETNAR▼VVGGTEAQRNSWPSQISLQYRSGSSWAHTCGGTLIRQNWVMTAAHCVDRELTFRVVVGEHNLNQN
DGTEQYVGVQKIVVHPYWNTDDVAAGYDIALLRLAQSVTLNSYVQLGVLPRAGTILANNSPCYITGWGLTRTNGQ
LAQTLQQAYLPTVDYAICSSSSYWGSTVKNSMVCAGGDGVRSGCQGDSGGPLHCLVNGQYAVHGVTSFVSRLGC
NVTRKPTVFTRVSAYISWINNVIASNGGHHHHHH*

Number of amino acids: 636 (248)
Calculated molecular weight: 69623 (26845)
Ext. coefficient (assuming all Cys pairs form cystines): 123230 (55390)
Calculated pI: 5.88 (8.45)

Construct: Trx-ProPPE-His_6 (pNIC/pCOLD)
MSDKIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQGKLTVAKLNIDQNPGTAPKYGIRGIP
TLLLFKNGEVAATKVGALSKGQLKEFLDANLAGTENLYFQS**TQDFPETNAR▼**VVGGTEAQRNSWPSQISLQYRSGS
SWAHTCGGTLIRQNWVMTAAHCVDRELTFRVVVGEHNLNQNDGTEQYVGVQKIVVHPYWNTDDVAAGYDIAL
LRLAQSVTLNSYVQLGVLPRAGTILANNSPCYITGWGLTRTNGQLAQTLQQAYLPTVDYAICSSSSYWGSTVKNSM
VCAGGDGVRSGCQGDSGGPLHCLVNGQYAVHGVTSFVSRLGCNVTRKPTVFTRVSAYISWINNVIASNGGHHHH
HH*

Number of amino acids: 376 (248)
Calculated molecular weight: 40834 (26845)
Ext. coefficient (assuming all Cys pairs form cystines): 70985 (55390)
Calculated pI: 6.34 (8.45)

Construct: SUMO-ProPPE-His_6 (pNIC/pCOLD)
MSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLMEAFAKRQGKEMDSLRFLYDGIRIQADQTP
EDLDMEDNDIIEAHREQIGGS**TQDFPETNAR**▼VVGGTEAQRNSWPSQISLQYRSGSSWAHTCGGTLIRQNWVM
TAAHCVDRELTFRVVVGEHNLNQNDGTEQYVGVQKIVVHPYWNTDDVAAGYDIALLRLAQSVTLNSYVQLGVLPR
AGTILANNSPCYITGWGLTRTNGQLAQTLQQAYLPTVDYAICSSSSYWGSTVKNSMVCAGGDGVRSGCQGDSGG
PLHCLVNGQYAVHGVTSFVSRLGCNVTRKPTVFTRVSAYISWINNVIASNGGHHHHHH*

Number of amino acids: 357 (248)
Calculated molecular weight: 39336 (26845)
Ext. coefficient (assuming all Cys pairs form cystines): 56880 (55390)
Calculated pI: 6.39 (8.45)