Utilising anhydride reactivity to target endogenous human proteins



Sheryl Ying Ting Lim

Department of Biochemistry
Linacre College
University of Oxford

Hilary 2022

This thesis is submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy.

Word count: 46,548

Abstract

Covalent conjugation between proteins has been used to develop novel drug delivery methods, enhance therapeutic efficacy, and improve diagnostic sensitivity. Proximity-directed ligation methods enable coupling to unmodified endogenous proteins. A diverse range of proteins contain autoproteolytic domains that cleave at an Asp-Pro bond via an aspartic anhydride intermediate. In this thesis, the anhydride is harnessed as an electrophilic warhead for targeted conjugation. I developed NeissLock, a technology that uses the self-processing module (SPM) of the FrpC protein from *Neisseria meningitidis*. SPM is fused to the C-terminus of the binder protein so that the anhydride generated upon calcium-induced activation reacts with proximal nucleophiles on the target. NeissLock, using the standard genetic code, enables rapid and specific coupling under mild conditions.

I discovered that the FrpA SPM variant has an enhanced activation rate compared to the first-generation NeissLock. I developed a bacterial surface display system for directed evolution of SPM, towards improving the cleavage yield and stability of the domain. Finally, I established the split SPM system by splitting the domain into 2 fragments to provide another level of control over SPM activation. The fusion of SpyTag003/SpyCatcher003 to the split SPM fragments increased the coupling efficiency. Split SPM facilitates the mammalian expression of binder-SPM fusions, broadening the applicability of NeissLock. I fused various human red blood cell-binding peptides to SPM to test for coupling to human red blood cells. Covalent conjugation of therapeutic proteins to red blood cells using NeissLock would increase the stability of coupling compared to non-covalent methods, potentially enhancing drug delivery using red blood cell carriers.

Authorship Declaration

I declare that this thesis is entirely my own work, except for the following exceptions:

Chapter 4

Section 4.1 SEC-MALS experiment was run by Dr David Staunton

Figure 4.17 Coupling experiment was run by Arne Scheu

Section 4.5 DogTag-MBP protein was made by Dr Vikash Yadav

Section 4.9.2 TGFα-GSY-SPM was made by Arne Scheu

This thesis has not been submitted for any other degree or to any other institution.

Publications over the course of my graduate studies:

Scheu, A.H.A.*, Lim, S.Y.T.*, Metzner, F.J., Mohammed, S., and Howarth M. NeissLock provides an inducible protein anhydride for covalent targeting of endogenous proteins. Nature Communications 12, 717 (2021). https://doi.org/10.1038/s41467-021-20963-5
*Joint first author

Scheu, A.H.A., Lim, S.Y.T. and Howarth, M. System for covalently linking proteins. UK patent application 2003683.6 (2020).

Acknowledgements

I would like to thank my primary supervisor, Professor Mark Howarth, for his guidance throughout the last 3 years. His mentorship has taught me to think more critically and creatively.

I am grateful to my funding body, the Agency for Science, Technology and Research Singapore for providing me with the opportunity to develop my research skills at the University of Oxford.

I would also like to thank everyone in the Howarth laboratory, past and present. I am grateful for all your help with troubleshooting experiments, brainstorming for new ideas, and being a source of support and encouragement. I would especially like to thank Arne Scheu, with whom I have had the privilege of working closely on the NeissLock project.

Many thanks to all my friends in Oxford who have made my life outside of the laboratory immensely enjoyable. Lastly, I'd like to express my gratitude to my family for believing in me and their unending support throughout my studies.

Abbreviations

AF	AlexaFluor
AZI	Antizyme Inhibitor
AP	Acceptor peptide/ AviTag
BLAST	Basic Local Alignment Search Tool
BSA	Bovine serum albumin
cfu	Colony forming unit
CHO	Chinese hamster ovary
DMEM	Dulbecco's modified eagle's medium
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EGTA	Ethylene glycol-bis(β-aminoethyl ether)-N,N,N′,N′-tetraacetic acid
EGFR	Epidermal Growth Factor receptor
FACS	Fluorescence-activated cell sorting
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GST	Glutathione-S-transferase
HBS	HEPES-buffered saline
HEK	Human embryonic kidney
HEPES	N-2-hydroxyethylpiperazine-N-2-ethane sulfonic acid
HRP	Horseradish peroxidase
IPTG	Isopropyl β-D-1-thiogalactopyranoside
LB	Luria Bertani broth
MBP	Maltose binding protein
Ni-NTA	Nickel-nitrilotriacetic acid
NMR	Nuclear magnetic resonance
OAZ	Ornithine decarboxylase antizyme
OD	Optical density
ODC	Ornithine decarboxylase
PAGE	Polyacrylamide gel electrophoresis
PE	Phycoerythrin
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PEG	Polyethylene glycol

PLP	Pyridoxal phosphate
PMSF	Phenylmethylsulfonylfluoride
PTM	Post-translational modification
RGM	Repulsive guidance molecule
SDS	Sodium dodecyl sulfate
SEC-MALS	Size exclusion chromatography with Multi Angle Light Scattering
sfGFP	Superfolder green fluorescent protein
SPM	Self-processing module
STAT	Signal transducer and activator of transcription
SUMO	Small ubiquitin modifier
TBS	Tris-buffered saline
TCEP	Tris(2-carboxyethyl)phosphine
TGFα	Transforming growth factor alpha
TMB	3,3',5,5'-Tetramethylbenzidine
tPA	Tissue-type plasminogen activator
Ubq	Ubiquitin

Contents

Abstract **2**

Authorship Declaration **3**

Acknowledgements **4**

Abbreviations **5**

List of Figures **10**

Chapter 1: Introduction **12**

1.1. Anhydride-forming domains 12
1.1.1. Mechanism for autocatalytic cleavage 12
1.1.2. RGMs and other autoproteolytic domains 14
1.1.3. FrpC 18

1.2. Methods for covalent conjugation to endogenous proteins 22
1.2.1. Photo-induced crosslinking 23
1.2.2. Proximity-based labelling 25

1.3. Aims of the Thesis 35

Chapter 2: Methods **37**

2.1. Molecular Cloning 37
2.1.1. Polymerase Chain Reaction (PCR) 37
2.1.2. Agarose Gel Electrophoresis, Gel extraction and DNA purification 37
2.1.3. Gibson Assembly 38
2.1.4. E. coli Transformation 38
2.1.5. Plasmid purification 39
2.1.6. DNA sequencing 39
2.1.7. Description of plasmids generated 39

2.2. Protein Expression and Purification 42
2.2.1. Expi293F and ExpiCHO growth conditions 42
2.2.2. Transfection of Expi293F and ExpiCHO 42
2.2.3. Media and E. coli growth conditions 43
2.2.4. Cell lysis 44
2.2.5. IMAC purification of His_6-tagged proteins 44
2.2.6. Purification of C-tagged proteins 45
2.2.7. Size-Exclusion Chromatography 46
2.2.8. Quantification of protein concentration 47

2.3. Protein Biotinylation 47

2.4. SDS-PAGE 48

2.5. Western Blotting 48
2.6. Protein Characterisation 50
2.6.1. Intact protein Mass Spectrometry 50
2.6.2. SEC-MALS 51
2.7. *In Vitro* Protein Cleavage and/or Conjugation Assays 51
2.7.1. Analysing RGMC cleavage process 51
2.7.2. Analysing anhydride reactivity with alkoxyamine-PEG$_4$-biotin 52
2.7.3. General set-up for OAZ/ODC conjugation 52
2.7.4. Quantification of cleavage and crosslinking rates 53
2.7.5. OAZ coupling in cell lysate 54
2.7.6. General set-up for split SPM cleavage and conjugation 55
2.8. Plate-based assay of anhydride reactivity 55
2.9. TGFα-SPM coupling to EGFR on live cells 56
2.10. Surface Plasmon Resonance 57
2.11. Lyophilisation 58
2.12. Bacterial display and flow cytometry 58
2.12.1. Library generation 58
2.12.2. Grow-up and induction of intimin expression 59
2.12.3. Flow cytometry and fluorescence-activated cell sorting (FACS) 60
2.13. Bioinformatics and Computational Analyses 61
2.13.1. Homology searches using BLAST and secondary structure prediction 61
2.13.2. Structure visualisation 62

Chapter 3: RGMs 63

3.1. Expression of RGMB and RGMC 64
3.2. Treatment with additives to isolate pre-cleavage VWD 68
3.3. Use of a different cell-line to isolate pre-cleavage VWD 70
3.4. Targeting for ER retention or protein body formation to isolate pre-cleavage VWD 72
3.5. Prokaryotic expression to isolate pre-cleavage VWD 74
3.6. Analysing RGMC VWD cleavage 78
3.7. Covalent coupling of RGMC to alkoxyamine-PEG$_4$-biotin 79
3.8. Discussion 81

Chapter 4: Understanding anhydride reactivity of FrpC SPM 85

4.1. NeissLock: inducible protein anhydride for covalent targeting of endogenous proteins 86
4.2. Biochemical characterisation of FrpC SPM coupling 90
4.2.1. pH-dependence of FrpC SPM coupling 90
4.2.2. Calcium-dependence of FrpC SPM coupling 93
4.2.3. Buffer-dependence of FrpC SPM coupling 94

4.2.4. Effect of lyophilisation on FrpC SPM coupling 95
4.3. Effect of C-terminal truncation on FrpC SPM cleavage rate 95
4.4. Improving OAZ-GSY-SPM stability 98
4.5. Specificity of FrpC SPM coupling 99
4.6. Effect of binding affinity on FrpC SPM coupling 103
4.7. Identifying crosslinking sites on ODC 107
4.8. Crosslinking to other nucleophiles 112
4.9. Other complexes 114
4.9.1. Crosslinking to Antizyme Inhibitor (AZI) 114
4.9.2. Effect of TGFα crosslinking to EGFR on downstream signalling 115
4.10. Discussion 119

Chapter 5: Optimising SPM for an enhanced NeissLock 124

5.1. Cleavage and coupling activity of SPM homologues 124
5.2. Establishing bacterial display system for directed evolution of SPM 131
5.2.1. Validation of display by flow cytometry 134
5.2.2. Calcium-induced cleavage of intimin-HA-FrpA-myc (WT) 135
5.3. Generation of a library of SPM variants 137
5.4. Sorting of the library and evaluation of clones 140
5.5. Discussion 144

Chapter 6: Extending NeissLock for mammalian binding proteins and red blood cell targets 151

6.1. Engineering split SPM for mammalian expression 151
6.1.1. Mammalian expression of SPM 151
6.1.2. Optimising the SPM split site 155
6.1.3. Fusion of SpyTag003/SpyCatcher003 to increase rate of activation 158
6.2. Application of split SPM: Conjugation of tPA-TGFα-SPM$_N$-SpyTag003 162
6.3. Towards a modular system of utilising binding scaffold fused to SPM 167
6.3.1. Validation of covalent coupling of UbqK0-HA-ERY-GSY-SPM 169
6.4. Expression of red blood cell bait antigens 172
6.5. Discussion 174

Chapter 7: Summary and Discussion 180

Bibliography 194

List of Figures

Chapter 1: Introduction

Figure 1.1 Suggested mechanism for cleavage at Asp-Pro bond 14
Figure 1.2 Structure of RGMB VWD domain 16
Figure 1.3 FrpC cleavage 19
Figure 1.4 Structure of FrpC SPM (residues 415-591) post-cleavage 20
Figure 1.5 Applications of FrpC SPM in purification 21
Figure 1.6 Photo-induced crosslinking 23
Figure 1.7 Proximity-based affinity labelling 27
Figure 1.8 Traceless Affinity Labelling 28
Figure 1.9 Electrophilic affibody 31
Figure 1.10 Disulfide trapping 33

Chapter 3: RGMs

Figure 3.1 Use of autocatalytic VWD domain for crosslinking to endogenous proteins 64
Figure 3.2 Schematic map of RGMs 65
Figure 3.3 Expression of RGMB and RGMC in Expi293F cells 68
Figure 3.4 Treatment with additives 70
Figure 3.5 Use of ExpiCHO cell line to inhibit cleavage 72
Figure 3.6 Targeting RGMB for ER retention or protein body formation 74
Figure 3.7 Prokaryotic expression of RGMB and RGMC VWD 77
Figure 3.8 pH-dependence of RGMC cleavage *in vitro* 79
Figure 3.9 Covalent coupling of alkoxyamine-PEG4-biotin to RGMC N-terminal cleavage product 81

Chapter 4: Understanding anhydride reactivity of FrpC SPM

Figure 4.1 FrpC calcium-induced autoproteolysis 87
Figure 4.2 NeissLock vision 88
Figure 4.3 Coupling of OAZ to ODC using NeissLock 89
Figure 4.4 SEC-MALS characterisation of OAZ-GSY-SPM 90
Figure 4.5 pH-dependence of OAZ-GSY-SPM crosslinking to ODC 92
Figure 4.6 Calcium-dependence of OAZ-GSY-SPM crosslinking to ODC 93
Figure 4.7 Buffer-dependence of OAZ-GSY-SPM crosslinking to ODC 94
Figure 4.8 Effect of lyophilisation 95
Figure 4.9 Effect of SPM truncation on cleavage activity 97
Figure 4.10 C175A mutation to improved OAZ-GSY-SPM stability 99
Figure 4.11 Specificity of NeissLock coupling 100
Figure 4.12 Specificity of NeissLock coupling in cell lysate 102
Figure 4.13 Dependence of coupling on cleavage of the D-P bond 102
Figure 4.14 Generation of OAZ binding mutants 105
Figure 4.15 Effect of K_d on coupling extent of OAZ to ODC 106
Figure 4.16 Potential crosslinking sites on ODC 108
Figure 4.17 Presence of at least two other crosslinking sites apart from K92 108
Figure 4.18 Identifying crosslinking sites on ODC 111

Figure 4.19 Crosslinking to other nucleophiles 113
Figure 4.20 Crosslinking to Antizyme Inhibitor (AZI) 115
Figure 4.21 TGFα crosslinking to EGFR on A431 cells 118

Chapter 5: Optimising SPM for an enhanced NeissLock

Figure 5.1: Cleavage and coupling activity of SPM homologues 127
Figure 5.2: FrpA SPM shows enhanced reactivity compared to FrpC SPM 130
Figure 5.3 Intimin-based bacterial surface display of SPM 133
Figure 5.4 Validation of intimin-HA-SPM-myc expression by flow cytometry 134
Figure 5.5 Validation of intimin-HA-SPM-myc cleavage activity by flow cytometry 137
Figure 5.6 ClustalO alignment of amino acid sequences of clones from the naïve library 139
Figure 5.7 Anti-myc fluorescence of cells expressing wild-type FrpA SPM using the optimised staining procedure 140
Figure 5.8 Staining procedure for FACS to minimise cross-reactivity 141
Figure 5.9 FACS analysis of dual-stained cells expressing intimin-HA-SPM-myc 142
Figure 5.10 FACS analysis of dual-stained cells expressing intimin-HA-SPM-myc during sequential sorting 144
Figure 5.11 Intimin-based bacterial display for selection of covalent binders against arbitrary protein targets 150

Chapter 6: Extending NeissLock for mammalian binding proteins and red blood cell targets

Figure 6.1 Split SPM system for mammalian expression of binder-SPM fusions 154
Figure 6.2 Optimising SPM split site for enhanced conjugation rate 157
Figure 6.3 Fusion of SpyTag003/SpyCatcher003 to split SPM increase cleavage rate 161
Figure 6.4 Mammalian expression of tPA-TGFα-SPM$_N$-SpyTag003 164
Figure 6.5 Conjugation of mammalian-expressed tPA-TGFα-SPM$_N$-SpyTag003 to sEGFR 166
Figure 6.6 Cleavage and conjugation of UbqK0-HA-ERY-GSY-SPM for use in red blood cell coupling 171
Figure 6.7 Selection of covalent peptide binders against red blood cell bait antigens 172
Figure 6.8 Mammalian expression and purification of Basigin 174
Figure 6.9 Using split-SPM for purification of ‘tag-less’ mammalian protein 175

Chapter 7: Summary and Discussion

Figure 7.1 Structure of *Ecballium elaterium* trypsin inhibitor II (EETI-II) (PDB 2IT7) 188
Figure 7.2 Schematic for (a) yeast surface display, or (b) mammalian surface display of knottin-SPM 190

Chapter 1: Introduction

1.1. Anhydride-forming domains

1.1.1. Mechanism for autocatalytic cleavage

Covalent coupling to endogenous proteins is useful for research as well as various other applications such as enhancing the efficacy of vaccines/therapeutics (Krall et al., 2016), increasing diagnostic sensitivity (Jain et al., 2013) and developing novel drug delivery (Shi et al., 2014) and imaging tools (Xue et al., 2015). In the same way that nature inspired some of the existing tools for covalent targeting (Zakeri *et al.*, 2012; Shah and Muir, 2014), nature can likely provide lessons for developing new tools for covalent targeting. It was previously suggested that autoproteolytic domains may be used in nature for covalent coupling between proteins by virtue of the reactive intermediate formed during cleavage (Lidell et al., 2003).

The first example of an autoproteolytic domain that generates an anhydride intermediate is the heavy chain precursor of the pre-α-inhibitor (pre-H3) (Thuveson and Fries, 2000). Following the first discovery, similar autoproteolysis in further mammalian proteins were characterised, specifically, mucin 2 (MUC2) (Lidell et al., 2003) and mucin 5AC (MUC5AC) (Lidell and Hansson, 2006). The autoproteolysis mechanism is not just limited to eukaryotic proteins. A similar autocatalytic cleavage at an Asp-Pro peptide bond was observed around the same time in the repeats-in-toxin (RTX) protein, FrpC, from *Neisseria meningitidis*, a Gram-negative pathogen (Osicka et al., 2004).

Given that it was proposed that the anhydride intermediate could be used natively for covalent coupling of H3 to chondroitin sulfate (Kaczmarczyk et al., 2002) and the adherence of FrpC to endogenous host cell proteins (Prochazkova et al., 2005), the key question in this thesis is to understand the reactivity of anhydrides

generated from autoproteolytic domains and investigate if the anhydride can be harnessed and redirected towards other target proteins.

In the proposed mechanism, under acidic conditions, the aspartate carboxylate initiates a nucleophilic attack on the carbonyl carbon of the peptide bond with proline, leading to a cleavage in the peptide bond (Thuveson and Fries, 2000) (Figure 1.1). As a tertiary amine, proline's main-chain N-H has a higher degree of protonation under acidic conditions than for other amino acid residues, making it a good leaving group upon cleavage of the peptide bond (Piszkiewicz et al., 1970). The cleavage of the peptide bond gives rise to an anhydride intermediate. The generation of a reactive aspartic anhydride after Asp-Pro cleavage in MUC2 was supported by labelling of the N-terminal cleavage fragment with biotin conjugated with a primary amine (Lidell et al., 2003). The anhydride is subsequently hydrolysed by water or susceptible to nucleophilic attack by neighbouring nucleophiles (Figure 1.1).

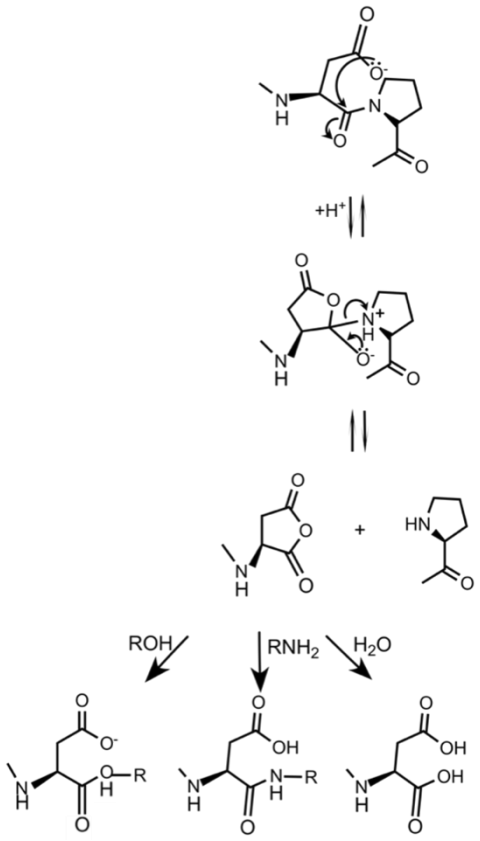


Figure 1.1 Suggested mechanism for cleavage at Asp-Pro bond. The β-carboxylate of Asp initiates a nucleophilic attack on the α-carbon of the Asp-Pro bond, leading to cleavage of the peptide bond. The main chain N-H of proline has a higher degree of protonation, making it easier for the tetrahedral intermediate to collapse. An anhydride is formed at the N-terminal cleavage fragment. The anhydride may react with water or other neighbouring nucleophiles.

The cleavage of Asp-Pro peptide bonds, especially under acidic conditions, has been observed outside of these autoproteolytic domains (Piszkiewicz et al., 1970). Apart from the nucleophilic displacement of the protonated amide by the carboxylate of Asp like in Figure 1.1, Piszkiewics *et al.* suggested that the cleavage could involve isomerisation of the α-amide bond to a ß-amide bond via a cyclic imide intermediate before anhydride formation and hydrolysis (Piszkiewicz et al., 1970). The kinetic rates of the Asp-Pro cleavage reactions described by Piszkiewicz *et al.* and in smaller peptides are however, much slower than that the cleavage reactions of autoproteolytic domains described in this thesis. The rate constant of cleavage of an Asp-Pro bond in a decapeptide at 40 °C and pH 4.5 was determined to be approximately 0.0038 h^{-1}, with a half-life of 7.6 days (Li et al., 2009). The anomalous cleavage of Asp-Pro bonds during chemical and enzymatic digest of proteins under low pH conditions described by Piszkiewicz *et al.* required incubation for 4 to 24 h (Schultz, 1967) while cleavage of the Asp-Pro bond in FrpC SPM is mostly complete within 10 - 20 min (Osicka et al., 2004).

1.1.2. RGMs and other autoproteolytic domains

The Repulsive Guidance Molecule proteins (RGMs) are involved in a range of functions including axon guidance, cell motility and differentiation, immune regulation

and iron homeostasis (Siebold et al., 2017). The RGMs are expressed in most vertebrates, while invertebrates like *Caenorhabditis elegans* have a single RGM gene (Siebold et al., 2017). The RGM family includes RGMA, RGMB/DRAGON, RGMC/hemojuvelin and RGMD (Siebold et al., 2017). RGMA and RGMB are primarily expressed in the central nervous system (CNS) while RGMC is predominantly expressed in the liver and skeletal muscle (Siebold et al., 2017). The RGMs consists of an N-terminal signal peptide, a conserved Arg-Gly-Asp (RGD) motif for binding to integrins, a partial von Willebrand factor type D (VWD) domain and a C-terminal glycosylphosphatidylinositol (GPI)-anchor for association with the cell membrane (Siebold et al., 2017). All RGMs contain a Gly-Asp-Pro-His (GDPH) sequence in the VWD domain (Siebold et al., 2017).

Zhang *et al.* was first to propose that RGMC undergoes autocatalytic cleavage at the GDPH sequence during its processing (Zhang et al., 2005). It was observed that cleavage of purified recombinant RGMC was enhanced after incubation at pH 5.5 compared to pH 8 (Zhang et al., 2005). Under reducing conditions that cleaved the intramolecular disulfide bonds, the observed 33 kDa and 14 kDa bands matched the predicted molecular weights of fragments that would be formed from cleavage of the Asp-Pro bond in the GDPH sequence of the VWD domain (Zhang et al., 2005). Crystallisation of the ectodomain of RGMB with the fifth and sixth fibronectin type III domains of its neogenin receptor provided structural insights into the VWD domain and confirmed the presence of the autocatalytic cleavage site at the Asp-Pro bond in the GDPH sequence (Bell et al., 2013).

The crystal structure of the VWD domain in RGMB revealed a novel protein fold consisting of a β-sandwich followed by four C-terminal α-helices (Bell et al., 2013) (Figure 1.2). According to the crystal structure, the cleavage site is located in the loop

between β-strands 1 and 2 (Bell et al., 2013) (Figure 1.2). The autocatalytic cleavage may have been enhanced by burial of Pro into the hydrophobic core of the protein domain post-cleavage (Bell et al., 2013) (Figure 1.2).

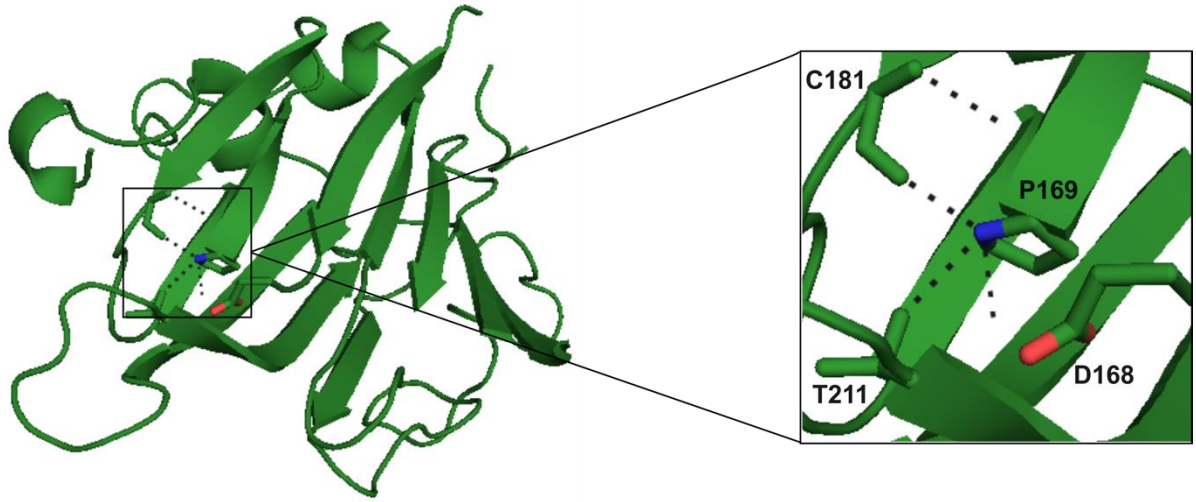


Figure 1.2 Structure of RGMB VWD domain. Autocatalytic cleavage site at D168 and P169 (stick residues, coloured by element), as well as other residues in the vicinity (stick residues, green) that form polar contacts (dashed lines) with the side-chains of D168 and P169. (PDB file: 4BQ6).

During the normal processing of RGMC, the aspartic anhydride of RGMC is hydrolysed after cleavage of the Asp-Pro bond (Bell et al., 2013). The autoproteolytic cleavage does not lead to secretion of the N-terminal fragment alone with the leader sequence because the cleaved fragments remain held together by intramolecular disulfide bonds (Siebold et al., 2017). The autoproteolytic processing was found to be essential for the plasma membrane localisation of RGMC (Bell et al., 2013). Cleavage is likely required for the proper folding of the VWD domain as well (Bell et al., 2013). The importance of the cleavage in the normal physiological functions of RGMC and its role in iron homeostasis was further suggested by the clustering of a large number of juvenile hemochromatosis disease-associated mutations in the vicinity of the cleavage site (Papanikolaou et al., 2004).

Autoproteolysis in other VWD domains has been previously characterised. Autoproteolysis was observed to take place in the VWD domains of mucin 2 (MUC2) (Lidell et al., 2003), mucin 5AC (MUC5AC) (Lidell and Hansson, 2006), zonadhesin (Bi et al., 2003) and crossveinless 2 (CV2) (Serpe et al., 2008). Some evidence for possible autoproteolysis in the VWD domains of IgG Fc-binding protein (FCGP) (Johansson et al., 2009), hemolectin (Goto *et al.*, 2001) and sialomucin complex (SMC) (Sheng et al., 1992) were reported as well. The proteins containing autoproteolytic VWD domains have diverse functions. For example, mucins are involved in maintaining the mucosal barriers for defence, zonadhesin is involved in fertilization while hemolectin is involved in clotting and immunity (Lang et al., 2007).

In all the autoproteolytic VWD domains, cleavage was reported or hypothesised to take place at the Asp-Pro bond. The function and biological relevance of autoproteolysis in proteins like RGMs, zonadhesin and CV2 are better understood. For instance, the autoproteolysis of zonadhesin to its mature forms was found to be essential for mediating species-specific binding of the sperm to the zona pellucida of the ovum (Bi et al., 2003). In other proteins like MUC2, it was proposed that the anhydride intermediate could react with neighbouring amines or hydroxyl groups, possibly in the O-glycans of other mucins (Lidell et al., 2003). Crosslinking between mucins could affect the viscosity of mucus and be of physiological importance (Lidell et al., 2003). However, there was no direct experimental evidence supporting the proposed autoproteolysis-dependent crosslinking between mucins (Lidell et al., 2003). Crosslinking between MUC2 and FCGBP post-cleavage is in line with the observation of a reduction-insensitive bond between MUC2 and FCGBP but there was no biochemical evidence that the crosslink is due to autoproteolysis of the VWD domain in either protein (Johansson et al., 2009). It is therefore unclear whether

autoproteolysis and anhydride formation is used by native proteins containing the VWD domain to crosslink with specific interacting partners.

Autoproteolysis does not take place exclusively at Asp-Pro bonds or VWD domains. The *Salmonella typhimurium* FlhB membrane protein and *Yersinia pseudotuberculosis* YscU protein undergo autocleavage at an Asn-Pro bond (Björnfot et al., 2009; Ferris et al., 2005). Unlike cleavage at Asp-Pro bonds, cleavage at Asn-Pro bonds would generate a succinimide rather than anhydride intermediate (Ferris et al., 2005).

A subset of proteins containing the Sea urchin sperm protein, Enterokinase, and Agrin (SEA) domain undergo autoproteolysis at a Gly-Ser bond that was accelerated by conformational strain of the scissile peptide bond (Johansson et al., 2008). During activation of cephalosporin acylase, cephalosporin acylase also undergoes primary autocleavage at a Gly-Ser bond (Kim et al., 2006). The cleavage at Gly-Ser bonds was postulated to involve an N→O acyl shift and an ester intermediate (Kim et al., 2006). The aforementioned non-exhaustive list of autoproteolytic domains highlights the wide range of distinct mechanisms for autocatalytic cleavage, each of which is mediated by different residues. Autoproteolysis of the Asp-Pro bond is of most interest given the electrophilicity of the aspartic anhydride intermediate that can be harnessed for covalent targeting.

1.1.3. FrpC

FrpC is a repeats-in-toxin (RTX) protein secreted by *Neisseria meningitidis* (Thompson and Sparling, 1993). Osicka *et al.* first reported the observation of calcium-induced autocatalytic cleavage at the Asp-Pro bond in the self-processing module (SPM) of FrpC (Osicka et al., 2004) (Figure 1.3). It was proposed that the reactivity of

the anhydride intermediate formed at the N-terminal cleavage fragment was responsible for the observed crosslinking to other FrpC molecules upon calcium addition. A minimal scar of a single aspartate residue is expected after crosslinking. The single residue scar is ideal when harnessing the SPM domain for crosslinking to other target proteins so as not to impact the function of the target protein post-conjugation.

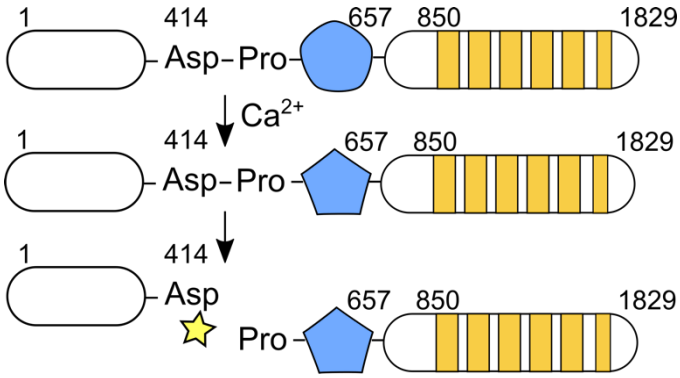


Figure 1.3 FrpC cleavage. FrpC consists of a self-processing module (SPM) (blue) between residues 414-657 and repeats-in-toxin (RTX) domains (yellow), each composed of a nonapeptide repeat, in the C-terminus. Calcium addition induces conformational change in the SPM. The Asp-Pro bond between D414 and P415 is cleaved, and an aspartic anhydride is generated (yellow star).

It was proposed that the autocatalytic cleavage of SPM in the calcium-rich extracellular environment enables crosslinking of the FrpC N-terminal cleavage fragment to host cell proteins, whilst the same fragment remains bound to FrpD, another outer membrane lipoprotein of *Neisseria meningitidis* (Sviridova et al., 2017). The covalent bond formation between FrpC N-terminus and host cell proteins could aid in adhesion of *Neisseria meningitidis* to the host cell surface.

FrpC SPM undergoes a conformational change in the presence of high calcium concentrations ($\geq 10^{-4}$ M) (Osicka et al., 2004). The solution structure of the SPM, specifically residues 415-591 of FrpC, in the presence of calcium was elucidated by nuclear magnetic resonance (NMR) (Kuban *et al.*, 2020). Post-cleavage SPM was

found to possess a unique protein fold that does not resemble any other known motif. Other features in post-cleavage SPM include four calcium-binding sites and π-π interaction between two tryptophan residues, W519 and W451, arranged in a T-shape (Figure 1.4) (Kuban *et al.*, 2020). *In silico* modelling of the pre-cleavage SPM structure and molecular dynamic (MD) simulations suggested a twisted amide leading to conformational strain at the scissile Asp-Pro bond could have increased susceptibility to autocleavage (Kuban *et al.*, 2020).

In the same paper, it was suggested that residues T430-G450 likely do not participate in the cleavage of the SPM since it was observed that deletion of T430-G450 did not affect the calcium-dependence of cleavage (Kuban *et al.*, 2020). However, it was not shown if the deletion affected the cleavage rate. Therefore, the involvement of residues T430-G450 in cleavage is inconclusive. Other residues like D478 and D579, initially hypothesised to catalyse the cleavage process, were found to only play a structural role given that the double D478N D579N mutant still exhibited cleavage activity (Kuban *et al.*, 2020). It is therefore still unclear which other residues catalyse the Asp-Pro bond cleavage.

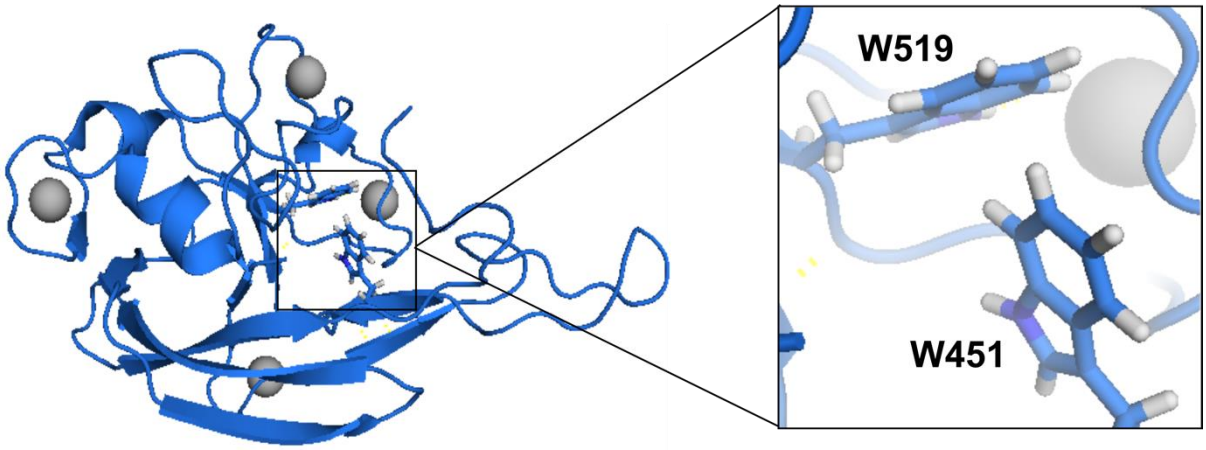


Figure 1.4 Structure of FrpC SPM (residues 415-591) post-cleavage. There are 4 bound calcium ions (grey spheres) and π-π interaction between W519 and W451 (stick residues) arranged in a T-shape. (PDB 6SJW)

The autocleaving property of FrpC SPM has been exploited for protein purification applications. In the first example of using SPM for purifying recombinant proteins, SPM with a C-terminal affinity tag was expressed as a genetic fusion to the target protein (Sadilkova et al., 2008). After standard binding and wash steps of affinity purification, cleavage of SPM at the Asp-Pro bond was induced by addition of calcium, leading to release of the target protein with a single aspartate at the C-terminus (Figure 1.5a). The rest of the SPM remains bound to the affinity matrix (Sadilkova et al., 2008). To inhibit the crosslinking activity of the anhydride intermediate, thiol compounds were added during the cleavage step. The resulting thioester bond in the protein-of-interest was found to be efficiently hydrolysed (Sadilkova et al., 2008). Various proteins including glutathione-S-transferase (GST), maltose-binding protein (MBP) and β-galactosidase were purified using SPM as a self-cleaving purification tag and yields were comparable to intein-based purification (Sadilkova et al., 2008).

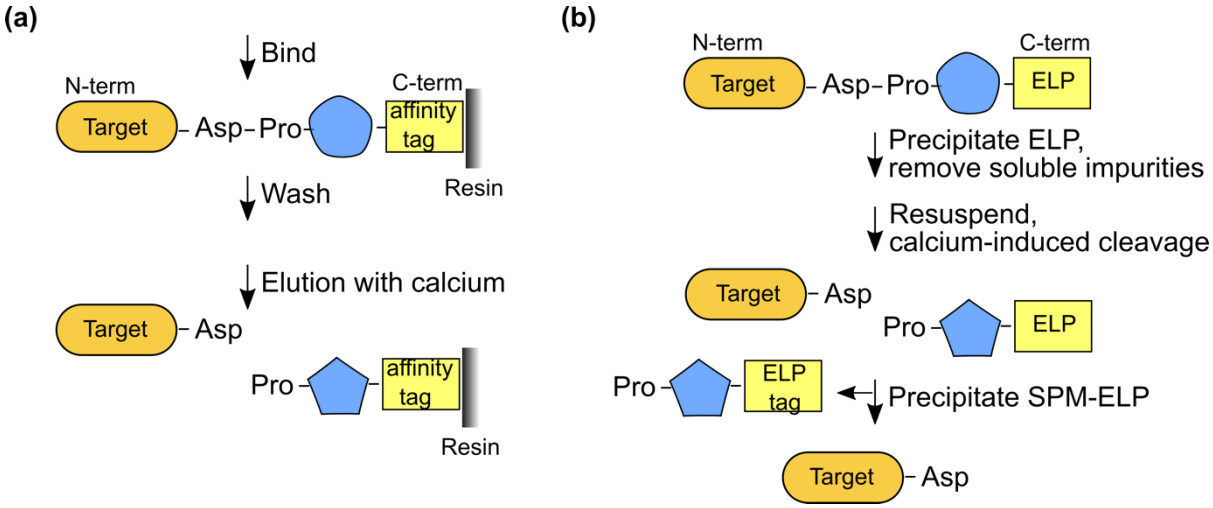


Figure 1.5 Applications of FrpC SPM in purification (a) Scheme for use of FrpC SPM (blue) together with a C-terminal affinity tag (yellow) as a self-cleaving tag, to purify the target protein (orange). A single Asp residue "scar" is left on the protein after elution. (b) Scheme for use of FrpC SPM (blue) paired with a C-terminal elastin-like polypeptide (ELP) tag (yellow) as a self-cleaving tag, to purify the target protein (orange) by Inverse Transition Cycling.

In the second example of utilising SPM as a self-cleaving purification tag, SPM was fused to an elastin-like polypeptide (ELP) for purification of the recombinant protein by Inverse Transition Cycling (ITC) (Figure 1.5b) (Liu et al., 2014). The ELP was used to replace the affinity tag described in the previous scheme, so that the recombinant protein has the following organisation: target protein-SPM-ELP. The ELP tag undergoes reversible phase transition when heated. Repeated cycling through the insoluble and soluble phases of the fusion protein leads to isolation of the protein (Figure 1.5b) (Liu et al., 2014). The self-cleaving activity of SPM enables the removal of the entire ELP tag upon calcium addition. The SPM-ELP purification method was demonstrated for green fluorescent protein (GFP), the Fc region of porcine IgG (pFc) and human β defensin 3 (HBD3). For GFP, pFc and HBD3, the SPM-ELP purification method led to 90 - 100 % purity and yields of 1.1 to 36 mg/ L culture (Liu et al., 2014).

1.2. Methods for covalent conjugation to endogenous proteins

There are diverse methods for covalent conjugation to proteins, encompassing genetic, chemical, and enzymatic means. Genetically engineered fusion proteins are highly stable. However, depending on the specific fusion site, genetic fusions may affect the structure and function of the protein-of-interest (Snapp, 2005). In some cases, a peptide-tag is genetically encoded with the protein-of-interest, enabling downstream post-translational conjugation. The peptide-tag can serve as a recognition handle for transglutaminase (Steffen et al., 2017), or sortase (Popp et al., 2007), or a split-intein partner (Stevens et al., 2017), or SpyCatcher (Zakeri et al., 2012a). Genetic and peptide-based conjugation methods are not within the scope of this thesis since they are not applicable to targeting unmodified endogenous proteins.

Instead, the focus of this thesis is chemical conjugation techniques that target the reactive side-chains of natural amino acids, such as Cys, Lys and Tyr.

1.2.1. Photo-induced crosslinking

Photo-induced crosslinking is a method that relies on ultraviolet (UV) light irradiation of photoreactive groups, e.g., diazirine, aryl azides or benzophenone, to form a covalent bond with nearby molecules (Pham et al., 2013) (Figure 1.6). Due to the high reactivity that is induced only under specific conditions and short half-life of the reactive intermediate, photo-induced crosslinking is more specific than traditional chemical crosslinking that utilises moderately reactive bifunctional reagents that react with any available protein functional groups (Suchanek et al., 2005). Photo-induced crosslinking can not only be used to target proteins (He et al., 2017; Lopez et al., 2017), but also DNA (Fujimoto et al., 2018) and RNA (Fujimoto et al., 2019; Mayo and Cole, 2017). Carbohydrate photo-induced crosslinking probes have been developed to map sugar-lectin interactions and modify glycoproteins (Wu and Kohler, 2019).

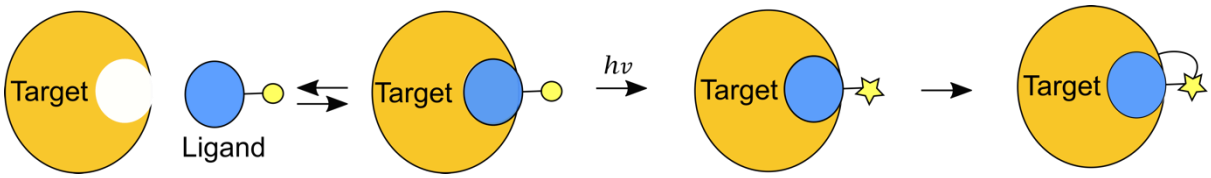


Figure 1.6 Photo-induced crosslinking. Ligand (blue) docks onto the target protein (orange). Upon irradiation by light, the photoreactive group (yellow) on the ligand is activated (yellow star) and crosslinks to proximal sites on the target protein.

With the development of the expanded genetic code, unnatural photoactivatable amino acid analogues such as L-photo-leucine and L-photo-methionine have been incorporated into proteins for photo-induced crosslinking (Suchanek et al., 2005). L-photo-methionine, for example, was incorporated into the

spider toxin Huwentoxin-IV around its binding interface with voltage-gated sodium channel subtype 1.7 ($Na_v1.7$) to gain insights into the structure of $Na_v1.7$ (Tzakoniati et al., 2020). Activation by UV led to the formation of a reactive carbene group that efficiently crosslinked to $Na_v1.7$ and specific crosslinking sites in the complex were identified using tandem mass spectrometry (tandem MS) (Tzakoniati et al., 2020).

Photo-inducibility enables temporal and spatial control of crosslinking in live cells. Improvements to the method allowed photo-induced crosslinking to be used to capture low-affinity and/or transient intracellular protein-protein interactions, for example, enzyme-substrate interactions. Lopez et al. identified specific substrates of histone deacetylase 8 (HDAC 8) by incorporating p-benzoyl-L-phenylalanine at various positions of HDAC's active site (Lopez et al., 2017). He et al. devised a quantitative proteomics method to identify novel substrates of the *E. coli* protease-chaperone, DegP (He et al., 2017). Changes in the interactome in response to different stimuli can also be elucidated (Zhang et al., 2011).

However, it has been observed that the use of different photoreactive probes can lead to biases in labelling (Chen and Topp, 2019). In a study that compared the use of different benzophenone and diazirine photoaffinity probes, different photoreactive probes labelled different sets of target proteins (Park et al., 2016). The difference in labelling could be due to different non-specific non-covalent interactions between the probe and other proteins present in cell lysate (Park et al., 2016). Testing multiple photoreactive probes may therefore be necessary to eliminate non-specific interactions (Smith and Collins, 2015). Compounding this problem, the synthesis and incorporation of unnatural photoactivatable amino acid analogues in proteins may be challenging. Growth in media containing the unnatural amino acid analogues rather than their structurally-related essential amino acids may lead to cell toxicity (Suchanek

et al., 2005). In other cases, the incorporation of the photoactivatable unnatural amino acid analogues depends on the simultaneous expression of an engineered suppressor tRNA-tRNA synthetase pair as well as the protein-of-interest with an amber codon (Hino et al., 2005). The final frequency of incorporation of the unnatural amino acids into proteins may vary (Suchanek et al., 2005). Finally, possible cytotoxicity and unintended effects on pathways-on-interest due to irradiation by UV light is another concern when performing *in vivo* photo-induced crosslinking (Chen and Topp, 2019).

1.2.2. Proximity-based labelling

Proximity-based labelling is a method for mapping protein-protein interactions (PPI) or the proteomes of cellular compartments by fusing a probe to a bait-protein so that the probe covalently labels neighbouring proteins (Chen and Perrimon, 2017). Proximity-based labelling provides high spatial and temporal resolution of protein-protein interactions in their native environment. Proximity-based labelling include common enzyme-catalysed methods like BioID (Roux et al., 2018) and APEX (Rhee et al., 2013), as well as affinity-guided methods like Ligand-directed (LD) chemistry (Tsukiji et al., 2009) and ligand-tethered DMAP (Hayashi et al., 2015).

In Biotin Identification (BioID), the bait, which can be a protein or small molecular ligand, is genetically fused to a promiscuous *E. coli* biotin ligase (Roux et al., 2012). When the bait binds to its prey, BirA catalyses biotinylation at proximal lysine residues on the prey within 10 to 20 nm radius (Roux et al., 2012). Biotin-affinity capture is then used to isolate biotinylated target proteins and the isolated proteins can be identified by mass spectrometry (Roux et al., 2012). Spence *et al.* recently reported the use of BioID in mapping the proteome of the excitatory synapse in the developing spines of mice, leading to a better understanding of synaptic regulation

and development (Spence et al., 2019). However, limitations of the BioID method include the possible obstruction of binding between the bait and prey due to the size of BirA (Varnaitė and MacNeill, 2016). The relatively slow labelling also meant that long incubation times were required, impairing temporal resolution and possibly leading to increased toxicity due to the labelling reagents (Branon et al., 2018). Some of these limitations have been partially overcome through improved variants of the BioID method such as BASU (Ramanathan et al., 2018) and TurboID (Branon et al., 2018).

In the APEX labelling method, ascorbic peroxidase (APEX) is fused to the bait and upon binding of the bait to prey, electron-rich amino acids e.g., Tyr, Trp, Cys and His on the prey are modified with biotin phenol tags (Martell et al., 2012). Compared to BioID, labelling with APEX is more rapid and can be completed in minutes (Martell et al., 2012). APEX was recently used to track the sub-minute time-resolved downstream signalling network after G-protein receptor (GPCR) activation (Lobingier et al., 2017). However, the need for hydrogen peroxide to generate the radical intermediate translates to relatively harsh conditions for *in vivo* labelling.

Unlike enzymatic methods, affinity labelling is based on the concept of introducing a chemically or photochemically reactive handle to a ligand (Hayashi and Hamachi, 2012). After ligand binding, the target protein is covalently modified at residues proximal to the ligand-binding site (Hayashi and Hamachi, 2012) (Figure 1.7). The labelling is dependent on the initial non-covalent but specific binding between the ligand and target, increasing the local concentration of the reactive group proximal to the binding site, thus promoting covalent bond formation (Wofsy et al., 1962). The chemically modified residues on the target protein are usually proximal nucleophilic residues such as Cys, Lys and Tyr whose side chains with heightened nucleophilicity

are susceptible to reaction with the handle (Tamura and Hamachi, 2019). The inherent moderate reactivity of the handle enables labelling of selected proximal residues near the binding interface rather than random solvent-exposed residues on the target (Ruan et al., 2018).

Affinity-based methods ensure high selectivity in labelling and capture even weak/transient interactions that otherwise would not be identified using classical affinity purification (Kim and Roux, 2016). With improved mass spectrometry methods, affinity labelling allows further characterisation of the target protein, including identification of the binding site residues (Tamura and Hamachi, 2019). In the case of enzymes, activity-based protein profiling (ABPP) utilises covalent probes with an electrophilic reactive group to characterise the structure and reactivity of active sites of enzymes such as proteases and kinases *in vivo* (Evans and Cravatt, 2006).

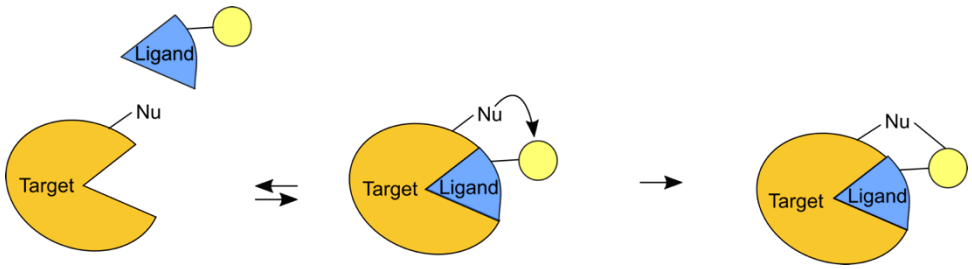


Figure 1.7 Proximity-based affinity labelling. Ligand (blue) with attached reactive handle (yellow) binds the target protein (orange) reversibly. The reactive handle is brought into close proximity to nucleophile ("Nu") on the target protein and forms a covalent bond to the nucleophile.

1.2.2.1 Traceless Affinity Labelling

Hamachi *et al.* pioneered traceless affinity labelling that differentiates itself from other affinity-labelling methods by the removal of the non-covalently bound ligand post-labelling. The retention of ligand post-labelling could impact the function of the target protein. In the exchange/cleavage reaction type of traceless affinity labelling,

the target is selectively labelled by the cleavage and subsequent transfer of probe between the ligand and target protein (Figure 1.8) (Matsuo and Hamachi, 2017).

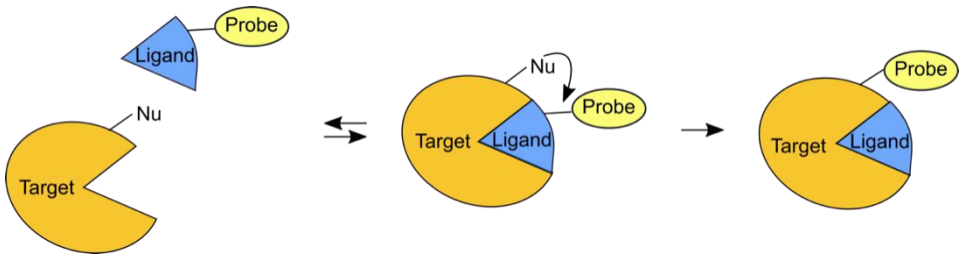


Figure 1.8 Traceless Affinity Labelling. Ligand (blue) with attached probe (yellow) docks to the target protein (orange). The reactive group is brought into close proximity to a nucleophile ("Nu") on the target protein. The probe is cleaved from the ligand and only the probe becomes covalently conjugated to the target.

Ligand-directed tosyl (LDT) chemistry utilises a phenyl sulfonate (tosylate) linker in the labelling reagent that is both a reactive electrophile and also a good leaving group for ligand cleavage (Tsukiji et al., 2009). Upon ligand-binding to the target, the tosyl ester group is brought into proximity to and reacts with nucleophilic residues e.g., His at the ligand-binding site (Tsukiji et al., 2009). The nucleophilic substitution reaction leads to dissociation of the ligand and transfer of the probe to the target protein. It was observed that LDT-labelling led to higher selectivity. Carbonic anhydrase in intact erythrocytes was labelled using the 7-dimethylaminocoumarin (Dc)-type labelling reagent after 48 h (Tsukiji et al., 2009). Alternative, a biotin-tag type reagent was used to label endogenous carbonic anhydrase in living mice (Tsukiji et al., 2009). The advantages of LDT chemistry-labelling are its sufficient stability *in vivo*, easy labelling process and high selectivity in whole proteomes (Hayashi and Hamachi, 2012). Other endogenous proteins like the molecular chaperone, FK506-binding protein 12 (FKBP12) have been labelled using LDT chemistry (Tamura et al., 2012). However, one disadvantage of LDT chemistry is its slow reaction kinetics (Tamura et

al., 2012). Slow kinetics mandated long incubations times and resulted in low yields, especially for labelling of native intracellular target proteins.

The latest example of traceless affinity labelling utilises *N*-acyl-*N*-alkyl sulfonamide (LDNASA) chemistry that has exceptionally rapid kinetics compared to previous generations of LD chemistry (Tamura et al., 2018). The rate constant of the LNASA reaction was found to be in the range of self-labelling (Benink and Urh, 2015; Cole, 2013) and bio-orthogonal reactions (Lang and Chin, 2014). Similar to LDBB, LDNASA showed enhanced stability under physiological conditions (Tamura et al., 2018). LDNASA-labelling was demonstrated on endogenous FKBP12, Hsp90 and folate receptors in live cells (Tamura et al., 2018). A high yield of labelled FKBP12 was observed within 1 h with minimal off-target labelling (Tamura et al., 2018).

Some limitations of LD chemistry remain. For example, LD chemistry requires the use of affinity ligands without reactive nucleophiles that would otherwise self-react with the cleavable electrophile (Tamura and Hamachi, 2019). Although LD chemistry-labelling does not require pre-identification of the target protein, it is limited to target proteins with small-molecule compatible binding pockets (Hayashi and Hamachi, 2012). Furthermore, the design and synthesis of the synthetic labelling reagents can be complex (Tamura et al., 2018).

Affinity-guided DMAP (4-dimethylaminopyridine) (AGD) is an example of affinity-based labelling that utilises catalyst tethering to selectively label native proteins. Unlike LD chemistry, the ligand in the labelling reagent can be a small molecule (Hayashi et al., 2013), peptide or protein (Hayashi et al., 2015), expanding the range of targetable proteins. For instance, antibody-DMAP was used to label endogenous cell surface receptors, human epidermal growth factor receptor 2 (HER2)

and epidermal growth factor receptor (EGFR) on live cells (Hayashi et al., 2015). HER2 and EGFR are otherwise not targetable by LD chemistry.

In the first demonstration of AGD, Koshi *et al*. tethered 4-dimethylaminopyridine (DMAP) to a carbohydrate-ligand for labelling of lectins (Koshi et al., 2008). In the presence of an acyl donor e.g., fluorescein, an acyl transfer reaction was catalysed by DMAP, leading to transfer of the acyl ester to a proximal nucleophilic residue (Lys, Tyr or Ser) on the target (Koshi et al., 2008). DMAP tethered to lactose was used for specific labelling of congerin II, a lactose-binding lectin (Koshi et al., 2008). The labelling of congerin II was found to be site-specific and occurred exclusively at Tyr 51 (Koshi et al., 2008). However, labelling was relatively slow. There was only a 35% yield in 3 h (Koshi et al., 2008). AGD chemistry also required basic pH conditions (pH > 8) that is less compatible with physiological conditions (Koshi et al., 2008). Furthermore, the use of highly reactive thioester acyl donors meant that AGD often led to non-specific labelling of other proteins in complex samples (Tamura et al., 2017). An improved AGD chemistry, affinity-guided oxime (AGOX) chemistry was developed to improve the efficiency and specificity of catalyst-tethered labelling (Tamura et al., 2017). AGOX was used to selectively label the endogenous AMPA neurotransmitter receptor in mouse brain tissue (Tamura et al., 2017). Even with AGOX, the labelling of intracellular proteins using AGD remains a challenge.

1.2.2.2 Electrophilic affibodies

Another method for proximity-induced labelling is the introduction of an electrophilic group into a binding scaffold, such as an affibody or antibody. Binding of the affibody to the target protein will bring the unnatural electrophile into close proximity with a nucleophile e.g., Cys, Lys, or His on the target, driving covalent bond

formation (Holm et al., 2009) (Figure 1.9). The engineering of complementary reactive groups between the antibody and ligand to promote a covalent reaction and enhance the antibody's binding affinity was explored by Chmura *et. al* (Chmura et al., 2001). Extending from this idea, the use of electrophilic affibodies was first demonstrated by Holm *et al.* (Holm et al., 2009).

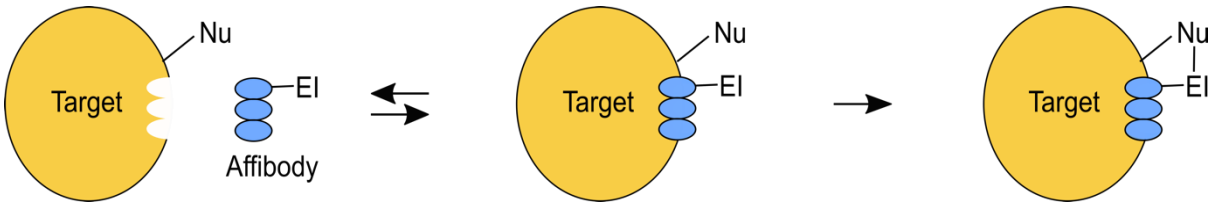


Figure 1.9 Electrophilic affibody. Affibody (blue) with attached electrophile ("El") docks to the target protein (orange). The electrophile is brought into close proximity with a nucleophile ("Nu") on the target protein. A covalent bond forms between the electrophile and nucleophile.

Acrylamide, a weak electrophile was introduced to an affibody at a site near the binding interface with its target (Holm et al., 2009) (Figure 1.9). The affibody-target covalent conjugate was formed and the coupling reaction was observed to be affinity-dependent and specific (Holm et al., 2009). Minimal cross-reactivity with other proteins in complex protein mixtures like cell lysate was observed (Holm et al., 2009). Electrophilic affibodies were applied in live cell imaging and ELISA assays. Apart from improved specificity and stability in imaging, it was observed that electrophilic affibodies enhanced the sensitivity of ELISA assays (Holm et al., 2009).

With bio-orthogonal chemistry, electrophilic unnatural amino acids can be incorporated into specific sites of binding scaffolds more easily (Chen et al., 2014). More recently, electrophilic unnatural amino acids were introduced into affibodies to covalently bind proximal Cys residues in Z protein (Xiang et al., 2013), or His and Lys residues in HER2 receptors on live cells (Chen et al., 2014). To target Cys residues,

weakly electrophilic phenylalanine-derived unnatural amino acids like p-2′-fluoroacetylphenylalanine were used (Xiang et al., 2013). To target His or Lys residues, an alkyl bromide-derived unnatural amino acid, (S)-2-amino-6-(6-bromohexanamido)hexanoic acid (BrC6K) was used instead (Chen et al., 2014). The reactivity of the unnatural amino acids was tuned to reduce non-specific reactivity with non-target amino acids or proteins (Xiang et al., 2013). The non-specific reactivity could otherwise inhibit genetic incorporation of the unnatural amino acid (Xiang et al., 2013). For example, passage of the protein containing the unnatural amino acid through the translation machinery may be hindered (Xiang et al., 2013). The weak electrophilicity of the unnatural amino acid ensured that reactivity was induced only when the unnatural amino acid and the target residue were in proximity (Chen et al., 2014; Xiang et al., 2013).

However, the use of electrophilic affibodies for conjugation to native proteins is limited to known and structurally well-characterised affibody/target pairs. This limits the range of targetable proteins in the proteome. There must be known nucleophilic residues in the vicinity on the target for efficient coupling, especially since the electrophile is a weak one and reactivity is proximity-induced. If a wider range of electrophiles is tested, a balance will have to be found between reactivity and specificity. Crosslinking to cysteines using electrophilic affibodies was rapid and high yielding but crosslinking to other nucleophilic residues (Lys and His) was slow and required hours of incubation (Chen et al., 2014; Holm et al., 2009). Labelling of Lys and His residues also required basic pH conditions (pH 8) that was non-toxic to cells but less favourable (Chen et al., 2014; Holm et al., 2009).

1.2.2.3 Disulfide trapping

Disulfide trapping is a method for site-specific labelling of cysteines on target proteins. In disulfide trapping, ligands that can be cysteine-containing peptides (Buck and Wells, 2005) or thiol/ disulfide-containing small molecules (Erlanson et al., 2000) bind to the target protein non-covalently initially. Proximity between the thiol group on the ligand and a cysteine on the target protein upon binding promotes disulfide bond formation (Figure 1.10). The cysteine residue on the target protein can be either natural or engineered by site-directed mutagenesis. The use of partially reducing conditions during the reaction promoted reversible thiol-disulfide exchange (Buck and Wells, 2005; Erlanson et al., 2000). Although the ligands may have low to moderate affinity (0.1–5 mM) for the target initially, the formation of a covalent tether leads to the formation of a stable complex and effectively “traps” the ligand (Erlanson et al., 2000). On the other hand, if the ligand does not dock to the target protein, a disulfide bond does not form.

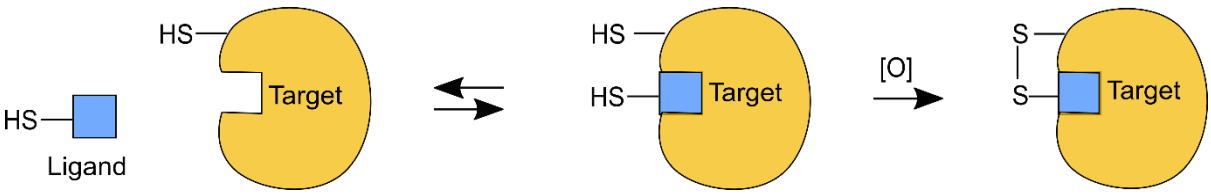


Figure 1.10 Disulfide trapping. Ligand (blue) with attached thiol (“-SH”) binds to the target protein (orange) non-covalently. The thiol group on the ligand is brought into close proximity to the cysteine residue (“SH”) on the target protein. A disulfide bond forms between the ligand and target.

The conjugation of the ligand to specific cysteines on the target protein is usually verified by mass spectrometry and characterised by X-ray crystallography. Disulfide trapping has been used for various applications including identifying binding sites on the target protein (Buck et al., 2005). For example, disulfide trapping was used

to elucidate the agonist and antagonistic molecular mechanisms of various cysteine-containing peptide analogues. The importance of the transmembrane helical triad (α-helices III, VI and VII) in the activation of the chemotactic cytokine complement factor 5a receptor (C5aR) was revealed in the same study (Buck and Wells, 2005).

Additionally, disulfide trapping can be used for developing and validating site-directed small molecule or peptide inhibitors. Small molecule inhibitors with enhanced affinities (up to 3,000-fold) against thymidylate synthase were identified after screening a library of thiol-containing analogues (Erlanson et al., 2000). Covalent small molecule inhibitors against oncogenic mutant KRAS (G12C) (Ostrem et al., 2013) and KRAS (M72C) (Gentile et al., 2017) were also identified after first screening a library of disulfide-containing small molecules. The binding of the inhibitors was verified to be dependent on the mutant cysteines introduced to KRAS and do not affect wild-type KRAS (Gentile et al., 2017; Ostrem et al., 2013). X-ray crystallography was used to identify the allosteric binding pocket of the inhibitors and characterisation of the molecular features aided structure-based design and optimisation of the inhibitors (Ostrem et al., 2013). In order to achieve irreversible binding, Ostrem *et al*. switched from the disulfide-containing small molecules to analogues containing electrophiles like acrylamides and vinyl sulphonamides (Ostrem et al., 2013).

A key limitation of disulfide trapping is that cysteine is a relatively rare amino acid in proteins, especially at the solvent-accessible surface (Lins et al., 2003; Miseta and Csutora, 2000). It may not always be possible to artificially introduce cysteines into the target protein without altering the protein structure. In any case, if artificial introduction of cysteine is required, this labelling method will not be applicable to unmodified endogenous proteins without native cysteines. The use of highly reducing

conditions for thiol-disulfide exchange also renders disulfide-trapping unsuitable for targeting proteins with pre-existing disulfides.

1.3. Aims of the Thesis

The overall objective of this thesis was to develop a general platform for covalent labelling of unmodified endogenous proteins. NeissLock is an inducible covalent labelling method where a binder protein that is fused to a SPM binds to its target such that in the presence of Ca^{2+}, an anhydride forms and reacts with a nearby nucleophile on the target protein, forming an irreversible binder:target complex. Key criteria for effective coupling are: 1) a rapid reaction rate (reaction completion within 1 h), 2) high conjugation yield (>50% yield), and 3) high specificity in complex mixture like cell lysate 4) mild reaction conditions for *in vivo* labelling (pH 7.4, ≤37 °C, aqueous).

In the first part of this thesis, the aim was to explore the potential of using an alternative autoproteolytic domain, the VWD domain for anhydride generation and covalent targeting. A better understanding of the *in vivo* regulation of the autocatalytic cleavage process as well as the pH-dependence of reaction aid evaluation of its potential as an inducible covalent probe.

In the second part of this thesis, the goal was to better understand the first generation NeissLock technology that utilised FrpC SPM. The aim was to characterise the anhydride reactivity. This includes addressing the key questions of whether anhydride reactivity is compatible with mild reaction conditions (pH 7.4) and whether the coupling is specific. Apart from minimising off-target coupling to irrelevant proteins that do not bind to the binder protein, the question arises regarding the affinity-dependence of the coupling and whether there is a correlation between binder-target

affinity and the extent of coupling observed. To optimise the design of NeissLock probes, I investigated whether C-terminal truncations of FrpC SPM affect the anhydride generation and reactivity. In the last part of this section, I addressed the question regarding the selectivity of crosslinking sites. By identifying the amino acid residues that the anhydride reacts with in the case of the ornithine decarboxylase antizyme (OAZ)/ ornithine decarboxylase (ODC), it helps us to understand which types of residues can be targeted and what the distance requirement is between the anhydride and target residue.

The third part of this thesis involved developing a second generation NeissLock with enhanced reaction rate and yield. A search amongst SPM homologues led to the identification of a faster-reacting variant, FrpA SPM. I performed initial experiments in bacterial surface display of SPM for the directed evolution of an SPM variant that outperforms FrpA SPM. To enable the wider applicability of NeissLock, such as expression of the binder-SPM fusions in eukaryotic systems, I established a modified NeissLock system based on splitting SPM, whilst retaining high reactivity. I found that fusing SpyTag003/SpyCatcher003 to the split SPM fragments enhanced the rate of activation. The split SPM system was demonstrated in the coupling of mammalian-expressed human tissue plasminogen activator-transforming growth factor alpha (tPA-TGFα) to epidermal growth factor receptor (EGFR). Working towards demonstrating the utility of NeissLock in the decoration of red blood cells, I generated several constructs comprising human red blood cell binding peptides fused to SPM to test their covalent reactivity to human red blood cells, in collaboration with Ashley Toye's laboratory in the University of Bristol.

Chapter 2: Methods

2.1. Molecular Cloning

2.1.1. Polymerase Chain Reaction (PCR)

Single-stranded DNA oligonucleotides were ordered from Integrated DNA technologies (IDT). The dry oligonucleotides were resuspended in nuclease-free water (Promega) to a final concentration of 100 µM. Long-term storage of DNA was at -20 °C. Desiccated gBlock gene fragments were ordered from IDT and resuspended in nuclease-free water to a final concentration of 10 ng/µL. Q5 High-Fidelity Polymerase (NEB) was used for Polymerase Chain Reaction (PCR). 25 µL of the final reaction mix consisted of 5 ng template DNA, 1.25 µL each of 10 µM forward and reverse primers and 12.5 µL of 2× Q5 high-fidelity polymerase master mix (NEB). The reaction components were assembled on ice before transferring to a C100 Thermal Cycler (Bio-Rad) that was preheated to 95 °C. The thermocycling conditions used were as follows: initial denaturation at 95 °C for 30 s, 30 cycles of denaturation at 95 °C for 20 s, primer annealing at 60 °C for 15 s, and elongation at 72 °C for 30 s per kb, followed by a final extension at 72 °C for 5 min. The primer annealing temperatures were altered according to the NEB T_m Calculator (NEB) and optimised for each primer pair.

2.1.2. Agarose Gel Electrophoresis, Gel extraction and DNA purification

After thermocycling, PCR samples at 25 µL each were mixed with 5 µL of 6× loading dye [10 mM Tris-HCl pH 7.4, 0.15% (w/v) orange G, 60% (v/v) glycerol, 60 mM EDTA] before loading on 1 % (w/v) agarose gels containing 0.005 % (v/v) SYBR Safe (Thermo Fisher). Samples were run at 130 V for 30-45 min in 1× TAE buffer (40 mM Tris base, 20 mM acetic acid, 1 mM EDTA, pH 8.0). Agarose gels were visualised

through a XcitaBlue Conversion Screen and Filter (Bio-Rad) and imaged using a ChemiDoc XRS imager (Bio-Rad). Bands-of-interest were excised from the gel and DNA was extracted and purified using the Wizard SV Gel and PCR Clean-Up Kit (Promega), following the manufacturer's protocol. DNA concentrations were estimated using the Nanodrop ND-1000 Spectrophotometer (Nanodrop).

2.1.3. Gibson Assembly

Gibson Assembly of DNA fragments was carried out as previously described (Gibson et al., 2009) using an in-house 2× Gibson assembly master mix. The reaction mix contained Phusion-like DNA polymerase (Wang et al., 2004), Taq DNA ligase and T5 exonuclease (NEB). The DNA polymerase and ligase were expressed and purified by Karl Brune, a previous member of the Howarth laboratory. 10 µL of the gel-purified PCR products were mixed on ice with 10 µL of the 2× Gibson assembly master mix before incubation for 2 – 4 h at 50 °C in a pre-heated C1000 Thermal Cycler (Bio-Rad). DNA fragments were added either in equimolar ratios for fragments of similar lengths or a 2 - 5-fold excess of insert(s) was added to 100 ng of vector.

2.1.4. *E. coli* Transformation

50 µL NEB Turbo Competent *E. coli* cells (NEB) or BL21-CodonPlus (DE3)-RIPL Competent Cells (Agilent Technologies) were thawed on ice. 8 µL of the Gibson assembly reaction mix or 1 µL of purified plasmid DNA was added to the cells and gently mixed on ice. Heat shock was performed by incubating the cells in a 42 °C water bath for 45 s before transferring it back on ice. 100 µL of SOC media was added for cells to recover and then incubated in a 37°C shaking incubator for 1 h. After 1 h, the

transformation was plated onto LB agar plate containing 100 µg/mL ampicillin or 50 µg/mL kanamycin and incubated at 37 °C for 16-18 h.

2.1.5. Plasmid purification

Colonies of transformed bacteria from the LB agar plate were used to inoculate 5-10 mL of LB (High Salt) media described below (Section 2.2.3), with 100 µg/mL ampicillin or 50 µg/mL kanamycin. Turbo Competent *E. coli* cells (NEB) were grown for 4-6 h at 37 °C in a shaking incubator before plasmid isolation. The GeneJET Plasmid Miniprep Kit (Thermo Fisher) was used to purify plasmids, following the manufacturer's protocol.

2.1.6. DNA sequencing

For all constructs, the gene sequences were validated by Sanger sequencing using services provided by Source BioScience (Nottingham and Cambridge, UK).

2.1.7. Description of plasmids generated

The plasmids used for expressing the ectodomains of *Homo sapiens* RGMB (GenBank ID AK074887, residues 53-412) and *Homo sapiens* RGMC (GenBank ID AY372521, residues 36-400) have been previously described (Bell et al., 2013). The -KDEL endoplasmic reticulum (ER) retention sequence and Zera sequence to induce protein bodies formation (Torrent et al., 2009) were separately cloned in the RGMB and RGMC constructs by Gibson assembly.

The SPM sequence was based on residues 414-657 of FrpC (UniProt Q9JYV5) (Osicka et al., 2004); the SPM635 variant was based on residues 414-635; the

SPM613 variant was based on residues 414-613 ; the SPM591 variant was based on residues 414-591. Numbering of OAZ and ODC were based on OAZ:ODC crystal structure (PDB 4ZGY) (Wu et al., 2015). Residues 95-219 of human OAZ (UniProt P54368) was used. pET28a-His_6-OAZ-SPM-Ctag has the following organization: N-terminal (M)GSS-linker, His_6–tag, OAZ, SPM, GSG-linker, C-tag. The mutations K153A, K153E, V198A, A215R and R188E were introduced into OAZ by site-directed mutagenesis using Gibson assembly. Human ODC1 (UniProt P11926) was cloned into pET28a-His_6-ODC-Ctag that has following organization: N-terminal (M)GSS-linker, His_6–tag, SSG-linker, ODC1, GSG-linker, C-tag. The lysine to arginine mutations (K92R, K121R, K74R and K78R) as well as N-terminal truncation (ΔN) were introduced into ODC by site-directed mutagenesis using Gibson assembly. pet28a-SpyTag003-sfGFP (Addgene 133454) has been described previously (Keeble et al., 2019). pET28a-DogTag-MBP (Addgene 171773) has been described previously (Keeble et al., 2021). Residues 1-448 of human AZIN1 (UniProt O14977) was used. pET28a-His_6-AZI-Ctag has the organization: N-terminal (M)GSS-linker, His_6–tag, SSG-linker, AZI1, GSG-linker, C-tag. Residues 40-89 of human protransforming growth factor alpha (UniProt P01135) was used in pET28a-TGFα-GSY-SPM that has the organisation: N-terminal (M), TGFα, GSY-linker, SPM, GSS-linker, His_6–tag, SSG-linker, C-tag. Residues 25-125 of human epidermal growth factor receptor (EGFR) (UniProt P00533) that encodes the soluble extracellular domain was used in pENTR4-sEGFR-His_6 that has the organisation: tPA leader sequence, sEGFR, GSGESG-linker, His_6-tag.

The sequences of the SPM homologues used included position 298-543 of FrpA from *Neisseria meningitidis* serotype C (Uniprot: P55126), position 231-476 of haemolysin-type calcium-binding protein related domain-containing protein from

Alysiella filiformis (GenBank:SOD68839.1 position 190-434 of the bifunctional hemolysin/adenylate cyclase precursor from *Kingella negevensis* (GenBank: SNB83338.1) and position 244-491 of the calcium-binding hemolysin protein from *Pseudomonas caricapapayae* (GenBank: RMV74036.1). The SPM homologues were cloned into pET28a-AviTag-His_6-OAZ-SPM-Ctag.

pet28a-intimin-HA-SPM-myc used for bacterial surface display has the organisation: Eae leader sequence, intimin, GGGSGSS linker, DYDIPTT spacer, HA-tag, TEV site, GSY-linker, FrpA SPM, GSS$(GSG)_2$KL-linker, myc-tag. Residues 1-659 encoding the LysM, ß-barrel, D0 and D00 domains of the Eae intimin from *Escherichia coli* O157:H7 (UniProt P43261) was used.

For split FrpA, pet28a-AP-His_6-OAZ-SPM_{N318}-SpyTag003 has the organisation: N-terminal (M)GS-linker, AviTag, GSS-linker, His_6–tag, OAZ (residues 95-219), G-linker, FrpA SPM (residues 298-318), GSG-linker, SpyTag003. pet28a-His_6-SUMO-SPM_{C322}-SpyCatcher003 has the organisation: N-terminal (M)GSWS-linker, His_6–tag, -G linker, SUMO-tag, FrpA SPM (residues 322-543), GSG-linker, SpyCatcher003, GSG-linker, C-tag. The SUMO-tag has been described previously (Buldun et al., 2018).

Residues 211-562 of human tissue-type plasminogen activator (tPA) (UniProt P00750) was used to encode truncated tPA in pcDNA3.1-tPA-TGFa-SpyTag003 that has the following organisation: tPA leader sequence, tPA, GSS-linker, TGFa, GSY-linker, FrpA SPM (residues 300-318), GSG-linker, SpyTag003. Site-directed mutagenesis and Gibson assembly were used to introduce the tetra-alanine mutations at residues 296-299 of tPA.

2.2. Protein Expression and Purification

2.2.1. Expi293F and ExpiCHO growth conditions

Expi293F cells (Thermo Fisher) were cultured in Expi293F Expression media (Thermo Fisher) and ExpiCHO cells (Thermo Fisher) were cultured in ExpiCHO Expression Media (Thermo Fisher). Both media were supplemented with 50 U/mL penicillin/streptomycin (Gibco). For routine cell culture maintenance, cells were grown up to a density of 3×10^6 cells/mL in a 125 mL flask and split 1:10 every 3-4 days. Cells were incubated at 37 °C and 8% CO_2 in a shaking incubator at 125 rpm. During treatment of transfected Expi293F with additives, 10% ammonium chloride (Sigma-Aldrich) was added to Expi293F media to a final concentration of 25 mM. Alternatively, 20 mM chloroquine diphosphate (Sigma-Aldrich) dissolved in water was added to the media to a final concentration of 50 μM. Lastly, 0.16 mM bafilomycin A1 (Sigma-Aldrich) was added to the media to a final concentration of 50 nM. Additives were added on the day of transfection.

2.2.2. Transfection of Expi293F and ExpiCHO

The RGM proteins, SPM-fusion proteins, tPA and sEGFR and were expressed by transient transfection in Expi293F cells (Thermo Fisher) using the Expifectamine293 Transfection kit (Thermo Fisher), following the standard manufacturer's protocol. RGM proteins and TGFα-containing fusion proteins were transfected in ExpiCHO cells (Thermo Fisher) using ExpifectamineCHO Transfection kit (Thermo Fisher), following the manufacturer's protocol. The plasmids used for expression of the RGMB and RGMC were kindly provided by Christian Siebold (University of Oxford). The plasmid DNA used for transfection were amplified in *E. coli*

Turbo Competent cells (NEB) and isolated using the PureLink HiPure Plasmid Maxiprep Kit (Invitrogen). When expressing sEGFR, kifunensine (Sigma-Aldrich) was added to the culture concurrent to transfection, to a final concentration of 5 μM. Proteins were harvested 4-8 days post-transfection.

2.2.3. Media and *E. coli* growth conditions

All LB media were prepared by the Media and Glassware Sterilization Laboratory at the Department of Biochemistry. LB (high salt) media contained 10 g/L tryptone, 5 g/L yeast extract and 10 g/L NaCl in MilliQ water. Media sterilisation was carried out by autoclaving at 121 °C for 20 min. Media were stored at room temperature prior to usage. Proteins were mostly expressed in *E. coli* BL21 (DE3) RIPL (Agilent). For RGMB and RGMC, expression in other *E. coli* strains including Rosetta-gami 2 (DE3) (NEB), SHuffle T7 Express (NEB) and C41(DE3) (Miroux and Walker, 1996), obtained from Anthony Watts (University of Oxford), were tested. For TGFα-SPM, T7Express (NEB) was used as well. *E. coli* cultures were grown at 37 °C in an Infors Multitron shaking incubator (Infors UK) at 200 rpm. Single colonies were picked from antibiotic-containing LB agar plate and used to inoculate 11 mL of LB (high salt) media containing 100 μg/mL ampicillin or 50 μg/mL kanamycin. The culture was left to grow for 16-18 h before 10 mL of overnight culture was added to 1 L of LB (High Salt) media containing 100 μg/mL ampicillin or 50 μg/mL kanamycin and 34 μg/mL chloramphenicol. Cells were grown at 37 °C until OD_{600} of 0.6 and protein expression was induced by addition of 0.42 mM IPTG. OAZ-SPM and ODC were expressed for 16-18 h at 25 °C. Cultures were centrifuged at 5,000 g at 4 °C for 20 min to harvest the cell pellet.

2.2.4. Cell lysis

To purify RGMB and RGMC, the Expi293F cultures were centrifuged at 1,000 g for 10 min at 4 °C. The supernatant after filtration through a 0.22 µm syringe filter (Starlab) was then used for Immobilised Metal Affinity Chromatography (IMAC) purification (Section 2.2.5). To purify intracellular RGMB and RGMC, the cell pellet was resuspended in ice-cold modified RIPA buffer [50 mM Tris-HCl, 150 mM NaCl, 1% (v/v) Triton-X-100, pH 7.5] supplemented with 1 mM PMSF and cOmplete protease inhibitor mix (Roche) and incubated on ice for 20 min. Cell lysates were centrifuged at 4°C for 20 min at 17,000 g and the clear supernatant was used for IMAC purification (Section 2.2.5).

For bacterial expression, the cell pellet from 1 L of bacterial culture was resuspended in 10 mL ice-cold lysis buffer [30 mM Tris-HCl, 200 mM NaCl, 5% (v/v) glycerol, 15 mM imidazole, pH 7.5] supplemented with cOmplete mini EDTA-free mixed protease inhibitors (Roche), 1 mM phenylmethylsulfonyl fluoride (PMSF) (Sigma), 1mg/mg lysozyme (Sigma) and 2 U/mL benzonase (Sigma). For purification of OAZ and ODC, 5 mM 2-Mercaptoethanol (Sigma) was added to the lysis buffer. The cell lysate was sonicated on ice three times for 1 min each time at 50% duty cycle with 1 min rest in between each cycle using a Sonics Vibra-cell sonicator (Ultrasonic Processor). After sonication, the cell lysates were centrifuged using a JA 25-50 rotor (Beckman) at 4 °C for 20 min at 30,000 g and the supernatant was used for further purification steps (Section 2.2.5, 2.2.6).

2.2.5. IMAC purification of His_6-tagged proteins

The clear supernatant was incubated with 0.5 mL Ni-NTA resin (Qiagen) per L of bacterial culture or per 30 mL of transfected Expi293F/ ExpiCHO culture for 30 min

at 4 °C on a rolling incubator. Samples were added to polyprep columns (Bio-Rad) and allowed to flow through by gravity. Flow-through was collected for SDS-PAGE analysis. The Ni-NTA resin was washed twice with 10× resin volume of Ni-NTA binding buffer (50 mM Tris-HCl, 300 mM NaCl, pH 7.8) with 10 mM imidazole, followed by another two washes with 10× resin volume of Ni-NTA binding buffer with 30 mM imidazole. The protein was eluted from the column after 5 min incubation with 1× resin volume of Ni-NTA binding buffer with 200 mM imidazole for 5 min. The elution step was repeated thrice. When purifying OAZ and ODC, 5 mM 2-Mercaptoethanol was added to the wash and elution buffers. All wash and elution steps were done at 4 °C. After IMAC purification, the eluted fractions of OAZ and ODC were added to a Vivaspin centrifugal concentrator with 10 or 30 kDa cut-off (GE Healthcare) and centrifuged at 4,000 g at 4 °C. RGMB and RGMC were dialysed thrice using Spectra/Por 3 3,500 Da cut-off membrane (Spectrum Labs) for at least 3 h each time at 4 °C. RGMB and RGMC were dialysed in 1,000-fold volume excess of buffer containing 10 mM HEPES and 150mM NaCl, pH 7.5. For cleavage of the N-linked glycans for mammalian-expressed proteins like sEGFR, PNGaseF (NEB) digestion was carried out under either native or denaturing conditions, following the standard manufacturer's protocol.

2.2.6. Purification of C-tagged proteins

To purify ODC variants and SPM_C fusion proteins without a His_6-tag, the clear bacterial lysate was incubated with CaptureSelect™ C-tagXL Affinity Matrix (ThermoFisher) for 30 min at 4 °C on a rolling incubator. Samples were added to polyprep columns (Bio-Rad) and allowed to flow through by gravity. The C-tagXL resin was washed four times with 10 × resin volumes of wash buffer containing 20 mM Tris-

HCl, pH 7.4. The protein was eluted from the column after 5 min incubation with 1× resin volume of elution buffer containing 50 mM HEPES, 2 M $MgCl_2$, pH 7.4, repeating the elution step thrice. 5 mM 2-Mercaptoethanol was added to the wash and elution buffers for ODC variants. All wash and elution steps were done at 4 °C. The eluted fractions of ODC were added to a Vivaspin centrifugal concentrator with 30 kDa cut-off (GE Healthcare) and centrifuged at 4,000 g at 4 °C.

2.2.7. Size-Exclusion Chromatography

Proteins samples were loaded at 1 mL/min flow rate onto a HiLoad 16/600 Superdex 200 pg column (GE Healthcare) connected to the ÄKTA pure 25 fast protein liquid chromatography (FPLC) system (GE Healthcare) for size-exclusion chromatography (SEC). The Superdex 200 pg column was pre-equilibrated with the running buffer (50 mM HEPES, 150 mM NaCl, 2 mM TCEP, pH 7.4) used for gel filtration of OAZ and ODC. When purifying ODC variants, the running buffer was supplemented with an additional 0.02 mM pyridoxal phosphate (PLP), a co-factor. SEC was performed at 4 °C. Based on the A_{280} peak, fractions-of-interest were collected, and the purity of the fractions were verified by SDS-PAGE. Fractions containing the protein-of-interest were concentrated to a final volume of ~500 µL using a Vivaspin centrifugal concentrator with 10 or 30 kDa cut-off (GE Healthcare). Long-term storage of the proteins was at -80 °C.

2.2.8. Quantification of protein concentration

Protein concentrations were estimated using the A_{280} measured by the Nanodrop ND-1000 Spectrophotometer (Nanodrop). The extinction coefficient of proteins were predicted by ExPASy ProtParam (Gasteiger et al., 2003).

2.3. Protein Biotinylation

An AviTag (also referred to as AP) was genetically-fused to OAZ-SPM and the proteins were biotinylated using GST-BirA, as described previously (Fairhead and Howarth, 2015). The 1 mL biotinylation reaction mixture consisted of 100 µM AP-tagged protein, 6.7 µM GST-BirA, 5 mM $MgCl_2$, 1.5 mM D-biotin and 1 mM ATP in phosphate buffered saline (PBS), pH 7.4. The reaction components were incubated at 25 °C for 1 h before addition of another 6.7 µM GST-BirA and 1.5 mM D-biotin. The reaction components were incubated at 25 °C for another 1 h. For removal of GST-BirA, the sample was incubated with 200 µL of glutathione-HiCap resin (Qiagen) at 4 °C for 1 h while rotating. The resin was removed from the sample by centrifuging at 17,000 g for 1 min at 4 °C. To remove excess biotin, the supernatant was added to Spectra/Por 3 3,500 Da cut-off membrane (Spectrum Labs). Proteins were dialysed thrice for at least 3 h each time in 1,000-fold volume excess of buffer (50 mM HEPES, 150 mM NaCl, 2 mM TCEP, pH 7.4) at 4 °C. A streptavidin shift assay was used to assay the extent of biotinylation. A molar excess of monovalent streptavidin was added to the biotinylated protein and a complete upward shift in the band corresponding to the biotinylated protein verifies that the AP-tagged protein was completely biotinylated (Fairhead and Howarth, 2015).

2.4. SDS-PAGE

Protein analysis was performed using Tris-glycine sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). The 12, 16 or 18% Tris-glycine gels were run at 180 or 200 V for 60-100 min at 25 °C using XCell SureLock Mini-cell gel chambers (Thermo Fisher). The running buffer consisted of 25 mM Tris–HCl, 192 mM glycine and 0.1% (w/v) SDS. Protein samples were denatured by addition of the 6× SDS loading buffer [0.23 M Tris–HCl, 24 % (v/v) glycerol, 0.23 M SDS, 120 μM bromophenol blue, pH 6.8] and boiled at 95 °C for 3-6 min. 2 mM DTT was added to the 6× SDS loading buffer for samples that required reducing. PageRuler Plus Prestained Protein Ladder (ThermoFisher Scientific) was used as molecular weight marker. After resolving the samples by SDS-PAGE, the gels were stained for 16-18 h at 25 °C using InstantBlue (Expedeon) and destained using MilliQ water for 16-18 h. SDS-PAGE gels were imaged using the ChemiDoc XRS+ Imager (Bio-Rad). Image Lab version 5.2.1 (Bio-Rad) was used for densitometry analysis.

2.5. Western Blotting

With the RGM proteins, samples were transferred from Tris-glycine gels onto a nitrocellulose membrane (Bio-Rad) using transfer buffer [25 mM Tris base, 192 mM glycine and 20% (v/v) methanol]. Protein transfer was carried out at 35 V for 90 min at 25 °C in the XCell II Blot Module (Thermo Fisher). After transfer, the nitrocellulose membrane was blocked for 16-18 h at 4°C using 5% (w/v) skimmed milk (Thermo Fisher) in PBS-T [0.05% (v/v) Tween-20 in PBS, pH 7.4] or 5% (w/v) bovine serum albumin (BSA) (Sigma) in PBS-T. For detection of RGMB, the membrane was probed

with primary anti-RGMB antibody (SAB2700346, Sigma-Aldrich) diluted 1:1,000 in 5% (w/v) skimmed milk in PBS-T at 25 °C for 1 h. The membrane was washed 6× 5 min in PBS-T to remove excess primary antibody. The membrane was then probed with secondary goat anti-rabbit HRP (Thermo Fisher) diluted 1:5,000 in 5% (w/v) skimmed milk in PBS-T for 1 h at 25 °C. The membrane was washed 6× 5 min in PBS-T to remove excess secondary antibody. For detection of alkoxyamine-PEG_4-biotin, streptavidin-HRP (RABHRP3, Sigma-Aldrich) diluted 1:1,000 in 5% (w/v) bovine serum albumin (BSA) (Sigma) in PBS-T was used and this was followed by 6× 5 min washes in PBS-T.

For Western blotting with anti-TGFα and anti-EGFR, transfer was carried out at 30 V for 16 h at 4 °C for wet transfer or 20 V for 10 min when using the iBlot 2 Dry Blotting System (ThermoFisher). Subsequently, the membrane was incubated in 5 % (w/v) milk (Sigma) in PBS-T blocking solution for 1 h at 25 °C. The anti-TGFα antibody (MF9, Novus Biologicals) was diluted to 0.2 µg/mL in the blocking solution. The anti-EGFR antibody (LA22, Sigma-Aldrich) was diluted 1:1,1000 in the blocking solution. The membranes were incubated with the primary antibodies for 20 h at 4 °C. 6× 5 min washes with PBS-T were carried out before incubation with the secondary antibody, goat anti-mouse horseradish peroxidase (HRP) (A4416, Merck) diluted 1:1,000 in the blocking solution. The membranes were incubated with the secondary antibodies for 1 h at 25 °C before another 6× 5 min washes with PBS-T.

For Western blotting with anti-pSTAT1 and anti-GAPDH, protein transfer was carried out at 30 V for 3 h at 25 °C. For blocking, the membranes were incubated in 5 % (w/v) BSA (Sigma) in PBS-T for 1 h at 25 °C. Both the rabbit anti-pSTAT1(Y701) antibody (58D6, Cell Signaling Technology) and mouse anti-GAPDH antibody (GA1R, Thermo Fisher) were diluted 1:1,1000 in the blocking solution. The membranes were

incubated with the primary antibodies for 20 h at 4 °C. 6× 5 min washes with PBS-T were carried out before incubation with the secondary antibodies, goat anti-rabbit horseradish peroxidase (HRP) (65–6120, Thermo Fisher) for anti-pSTAT1 or goat anti-mouse horseradish peroxidase (HRP) (A4416, Merck) for anti-GAPDH, both diluted 1:1,000 in the blocking solution. The membranes were incubated with the secondary antibodies for 1 h at 25 °C before another 6× 5 min washes with PBS-T and addition of the HRP substrate.

For all the Western blots, the Supersignal West Pico chemiluminescent substrate (Thermo Fisher) was used to develop the membranes. The membranes were imaged using the ChemiDoc XRS+ Imager (Bio-Rad) and analysed using Image Lab version 5.2.1 (Bio-Rad).

2.6. Protein Characterisation

2.6.1. Intact protein Mass Spectrometry

For intact protein mass spectrometry, a RapidFire 365 platform (Agilent) was used. The RapidFire platform comprised an electrospray ionization (EI) source and 6550 Accurate-Mass Quadrupole Time-of-Flight (Q-TOF) detector (Agilent). 10 µM protein samples were injected and acidified to 0.9 % (v/v) formic acid. Samples were aspirated for 0.3 s before they were adsorbed onto a C4 solid-phase cartridge for extraction. Washes using 0.1% (v/v) formic acid were carried out for 5.5 s before elution onto the detector for 5.5 s. Mass spectrometry data was analysed with MassHunter Qualitative Analysis B.07.00 (Agilent). The expected molecular weights of proteins were determined by ExPASy ProtParam (Gasteiger et al., 2003).

2.6.2. SEC-MALS

Dr David Staunton from the Department of Biochemistry Biophysical Instrument Facility performed the size exclusion chromatography with multi-angle light scattering (SEC-MALS) experiment for OAZ-GSY-SPM. A Superdex 200 HR 10/30 column (GE Healthcare) connected to a Shimadzu HPLC system was used. The HPLC system was coupled to a Dawn HELEOS-II 8-angle light scattering detector (Wyatt technology) and Optilab rEX refractive index monitor (Wyatt technology). The running buffer used for SEC consisted of 50 mM HEPES, 150 mM NaCl, 2 mM TCEP, 0.02 mM PLP, pH 7.4. 2 mg/mL OAZ-GSY-SPM diluted in 100 µL of running buffer was injected into the column. SEC-MALS was carried out at 25 °C. Data analysis was performed using the Astra version 6.1 software (Wyatt technology).

2.7. *In Vitro* Protein Cleavage and/or Conjugation Assays

2.7.1. Analysing RGMC cleavage process

0.02 mg RGMC, initially kept on ice, was diluted 1:10 in citrate-phosphate buffers spanning pH 2.2- 7.4. The buffers were prepared from 0.2 M Na_2HPO_4 and 0.1 M citric acid stock solutions, as described previously (C. McIlvaine, 1921). RGMC samples were incubated at 37 °C after dilution in the buffers. After 2 h, a 20 µL aliquot was taken and cleavage was stopped by the addition of 6× SDS loading buffer [0.23 M Tris–HCl, 24 % (v/v) glycerol, 0.23 M SDS, 120 µM bromophenol blue, pH 6.8]. For samples that required reducing, the SDS loading buffer was supplemented with 50 mM DTT. All samples were boiled at 95 °C for 5 min before resolution by 16% SDS-PAGE. Tris-glycine gels were stained using InstantBlue Coomassie stain (Expedeon).

Image Lab version 5.2.1 (Bio-Rad) was used for densitometry analysis. The following formula was used to calculate the cleavage extent of RGMC:

$$\%\ cleavage = \frac{(Total\ RGMC) - (Pre\text{-}cleavage\ RGMC)}{Total\ RGMC}$$

Total RGMC was approximated by the band intensity of full-length RGMC under non-reducing conditions and pre-cleavage RGMC was approximated by the band intensity of full length RGMC under reducing conditions

2.7.2. Analysing anhydride reactivity with alkoxyamine-PEG$_4$-biotin

1.25 mM, 0.31 mM, or 0.08 mM EZ-link alkoxyamine-PEG$_4$-biotin (Thermo Fisher) dissolved in DMSO was added to RGMC that was diluted in 0.1 M 2-(N-morpholino) ethanesulfonic acid (MES) buffer, pH 4.5 or pH 7.4. Samples were incubated at 37 °C for 1 h at 300 rpm in a thermomixer. The reaction was stopped by the addition of 6× SDS loading buffer. For samples that required reducing, the SDS loading buffer was supplemented with 50 mM DTT. All samples were boiled at 95 °C for 5 min before resolution by 16% SDS-PAGE (Section 2.4). Western blotting was performed for the detection of EZ-link alkoxyamine-PEG$_4$-biotin (Section 2.5).

2.7.3. General set-up for OAZ/ODC conjugation

To investigate OAZ cleavage and/or conjugation to ODC, 10 μM OAZ (wild-type or variant) was incubated with 10 μM ODC (wildtype or variant) in reaction buffer (50 mM HEPES, 150 mM NaCl, 2 mM TCEP, pH 7.4). When probing the pH-dependence, the reaction buffer was supplemented with additional 50 mM 2-(N-morpholino) ethanesulfonic acid (MES) to buffer over the entire pH range. Reactions were carried

out at 37 °C. At the 0 h time point, a 15 µL aliquot was added to 5 µL of stop solution that consisted of 5× SDS-loading buffer [0.19 M Tris-HCl, 20% (v/v) glycerol, 100 µM bromophenol blue, 0.19 M SDS, pH 6.8] with additional 75 mM EDTA, before addition of the start solution (40 mM $CaCl_2$ diluted in the reaction buffer). To induce cleavage, the start solution, pre-equilibrated to 37 °C, was added to the sample to give a final concentration of 10 mM $CaCl_2$ in the reaction mix. At the indicated time-points, a 20 µL aliquot was taken from each sample and the reaction was stopped by addition of 5 µL of stop solution. Samples were boiled at 95 °C for 3 min.

2.7.4. Quantification of cleavage and crosslinking rates

Cleavage of OAZ and conjugation to ODC or AZI were analysed by Coomassie-stained SDS-PAGE (Section 2.4). The assays were performed in triplicates for densitometry analysis. The following formulas were used for calculating cleavage and conjugation rates:

$$\%\ Cleavage = \frac{OAZ\text{-}SPM_{time\ 0} - OAZ\text{-}SPM_{time\ x}}{OAZ\text{-}SPM_{time\ 0}} \times 100\%$$

$$\%\ Conjugated = \frac{AZI_{time\ 0} - AZI_{time\ x}}{AZI_{time\ 0}} \times 100\%$$

To calculate relative conjugation in the pH/calcium/buffer/lyophilisation-dependence experiments, the band intensity of the OAZ:ODC conjugate was divided by the band intensity of OAZ:ODC at 120 min for the reaction with highest conjugation yield i.e.:

$$\text{Relative}\ \%\ Conjugated = \frac{OAZ{:}ODC_{time\ x}}{OAZ{:}ODC_{time\ 120\ min}} \times 100\%$$

To calculate relative conjugation to different ODC mutants for mapping the crosslinking sites on ODC, the band intensity of the OAZ:ODC(mut) conjugate was

divided by the band intensity of OAZ:ODC(wt) with adjustments for the difference in M_w of the conjugates i.e.:

$$\text{Relative \% } Conjugated = \frac{OAZ{:}ODC\ (mut)_{time\ x}}{OAZ{:}ODC(wt)_{time\ x}} \times \frac{Mw_{OAZ:ODC(wt)}}{Mw_{OAZ:ODC(mut)}} \times 100\%$$

The rate constants of FrpC and FrpA SPM cleavage were determined by fitting the cleavage time-course to a non-linear regression model, specifically, a two-phase exponential model using GraphPad Prism 9.

2.7.5. OAZ coupling in cell lysate

A431 cells was washed once with PBS before approximately 2 X 10^7 cells was resuspended in 1 mL of ice-cold cell lysis buffer consisting of 20 mM Tris, 150 NaCl, 1% (v/v) Triton-X-100, supplemented with additional mixed protease inhibitors and 1 mM PMSF, pH 7.4. The cells were incubated in lysis buffer on ice with occasional agitation for 20 min before centrifugation at 12,000 g at 4 °C for 10 min to remove insoluble cell debris. The supernatant was used to replace the reaction buffer for OAZ-GSY-SPM coupling as described in 2.7.3. A final concentration of 2 mM Ca^{2+} was used instead of 10 mM Ca^{2+} and the coupling reaction was carried out at 25 °C instead of 37 °C. The denatured samples post-reaction were run on SDS-PAGE and transferred onto a nitrocellulose membrane using the iBlot2 Dry Blotting system (Thermo Fisher). The membrane was blocked with 5 % (w/v) bovine serum albumin (BSA) in PBS with 0.05% Tween-20 (PBS-T) at 25 °C for 1 h before incubation with mouse anti-His (Invitrogen, HISH8) at 1:1,000 dilution in PBS-T at 4 °C for 20 h. The membrane was washed 6 times with PBS-T before incubation with goat anti-mouse HRP (Merck, A4416) at 1:1,000 dilution in 5% (w/v) BSA in PBS-T at 25 °C for 1 h.

The membrane was washed 6 times with PBS-T before addition of the HRP substrate, SuperSignal West Pico chemiluminescent substrate (Thermo Fisher) and imaged with a ChemiDoc XRS imager using Image Lab Software version 5.2.1 (Bio-Rad).

2.7.6. General set-up for split SPM cleavage and conjugation

To investigate cleavage and/or conjugation with the split SPM system, the SPM_N-fusion protein was incubated with the SPM_C-fusion protein in HBS reaction buffer (50 mM HEPES, 150 mM NaCl, with or without 2 mM TCEP, pH 7.4) at 37 °C. The split-SPM fusion proteins were pre-incubated for 1 h at 37 °C. before addition of the target protein. At the 0 h time point, a 15 µL aliquot was added to 5 µL of stop solution consisting of 5× SDS-loading buffer [0.19 M Tris-HCl, 20% (v/v) glycerol, 100 µM bromophenol blue, 0.19 M SDS, pH 6.8] with additional 75 mM EDTA, before addition of the start solution (8 mM $CaCl_2$ diluted in the HEPES buffer). To induce cleavage, the start solution, pre-equilibrated to 37 °C, was added to the sample to give a final concentration of 2 mM $CaCl_2$ in the reaction mix. At the indicated time-points, a 20 µL aliquot was taken from each sample and the reaction was stopped by addition of 5 µL of stop solution. Samples were boiled at 95 °C for 3 min.

2.8. Plate-based assay of anhydride reactivity

For immobilisation, SpyCatcher003-sfGFP at a concentration of 2 µg/mL in 100 µL PBS pH 7.4 was added to each well of a Nunc MaxiSorp 96-well flat bottom microtiter plate (Thermo Fisher) for 16-18 h at 4 °C. Excess unbound SpyCatcher003-sfGFP was washed off the plate by 6 X 400 µL PBS-T washes. 200 µL of 3 % (w/v) BSA in PBS-T was added to each well and incubated for 1 h at 25 °C to block non-

specific binding. After removing the blocking solution and 1 X 400 µL PBS-T wash, 100 µL of SpyTag-Y-SPM at 2 µg/mL in HBS pH 7.4 was added to each well and incubated for 30 min at 25 °C for the SpyTag:SpyCatcher conjugation. Excess unreacted SpyTag-Y-SPM was removed by 6 X 400 µL PBS-T washes. 100 uL of the biotinylated nucleophiles i.e., biotin pentyldiamine (Thermo Fisher, 10404034), biocytinamide (Merck), biotin-PEG-alkoxyamine (Thermo Fisher,26137), biotin tyramide (Merck) and biotin-PEG-SH (Iris Biotech, PEG1213) were added to each well to a final concentration of 0.1, 1 or 10 mM in 10% (v/v) DMSO and 10 mM $CaCl_2$ in HBS pH 7.4 for 30 min at 37 °C and 300 rpm. Excess unreacted biotinylated nucleophiles were removed by 6 X 400 µL PBS-T washes. Following that, 100 µL of 1-Step Ultra 3,3',5,5'-Tetramethylbenzidine (TMB) substrate (Thermo Fisher) was added to each well and incubated for 1 min at 25 °C before 100 µL of 1 M HCl was added to each well to stop the reaction. Absorbance was measured using the FLUOstar Omega microplate plate reader (BMG Labtech). The assay was performed in triplicates.

2.9. TGFα-SPM coupling to EGFR on live cells

For cell surface coupling, A431 cells (CRUK, Lincoln's Inn Fields) were seeded at 5 X 10^5 cells/well in a 6-well plate and incubated at 37 °C with 5% (v/v) CO_2 in complete media [high glucose Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% (v/v) fetal bovine serum (FBS), 1 U/ mL penicillin/streptomycin (Gibco), and 1× GlutaMAX (Gibco)] for 24 h. After 2 X 4 mL washes with FBS-free complete media, the A431 cells were serum-starved for 24 h at 37 °C with 5% (v/v) CO_2. After 24 h, the cells were washed with 2 X 2 mL HBS-Mg (50 mM HEPES, 150 mM NaCl, 5 mM $MgCl_2$, pH 7.4, sterile-filtered). 150 µL of TGFα-SPM at 2 µM in HBS

was added to each well, with or without hydroxylamine (Sigma) dissolved to a final concentration of 5 mM in 2% (v/v) DMSO in HBS-Mg. To activate coupling, 150 µL of 4 mM $CaCl_2$ in HBS-Mg, with or without hydroxylamine (Sigma) dissolved to final concentration of 5 mM in 2% (v/v) DMSO in HBS-Mg, was added to give a final concentration of 2 mM $CaCl_2$. The cells were incubated with TGFα-SPM and $CaCl_2$ for 10 min at 37 °C with 5% (v/v) CO_2. After removal of the protein solution, the cells were further incubated in FBS-free complete media containing final concentration 2% (v/v) DMSO, with or without 5 mM hydroxylamine for the indicated time. For cell lysis, the A431 cells were scrapped from the plates and incubated on ice in 150 µL modified RIPA buffer [20 mM Tris-HCl, 150 mM NaCl, 5 mM NaF, 1% (v/v) Triton-X-100, 0.1% (w/v) SDS, 0.5% (w/v) sodium deoxycholate, pH 7.4], supplemented with cOmplete protease inhibitor mix, 1 mM PMSF and 1 mM sodium orthovanadate for 20 min. The clear supernatant was removed after centrifuging at 12,000g for 20 min at 4 °C. The proteins were denatured by addition of 6 X SDS loading buffer, supplemented with 120 mM DTT, to the clear lysate before boiling for 6 min at 95 °C.

2.10. Surface Plasmon Resonance

Surface plasmon resonance (SPR) for measuring OAZ:ODC binding affinity was performed using a Biacore T200 (GE Healthcare). Measurements were made at 25 °C. The running buffer used consisted of 50 mM HEPES, 150 mM NaCl, 2 mM TCEP and 0.02 mM PLP, pH 7.4. The Biotin CAPture Kit Series S (GE Healthcare) was utilised, following the manufacturer's protocol. In each run, the streptavidin-based CAPture reagent was first added, before 0.5 µM biotinylated AP-OAZ-GSY-SPM ligand was immobilized on the chip. Finally, the ODC analyte was flowed over the chip to detect k_{on} and washed off by addition of more running buffer. Serial dilutions of ODC

were tested. 1.25 µM ODC was diluted two-fold down to 78.1 nM for wild-type OAZ-GSY-SPM binding to wild-type ODC and wild-type OAZ-GSY-SPM binding to ODC(ΔN 4KR). For binding mutants of OAZ-GSY-SPM, 97 µM of ODC was diluted two-fold down to 1.51 µM. Following each run, the chip was regenerated using the commercial regeneration solutions. Each experiment was carried out in duplicate. Kinetic data analysis was carried out using the Biacore T200 software (GE Healthcare).

2.11. Lyophilisation

100 µl of OAZ-GSY-SPM in HBS reaction buffer [50 mM HEPES, 150 mM NaCl, 2 mM TCEP, pH 7.4] was incubated at -80 °C for 16 h. After the sample was frozen, lyophilisation was carried out by incubating the sample in a benchtop 2 K freeze-dryer (VisTis) at -72.5 °C and 0.14 mbar for 20 h. The freeze-dried sample was reconstituted by resuspending in 100 µl of MilliQ water followed by centrifuging at 16,900 g, 4 °C for 10 min to remove aggregates. The clear supernatant containing soluble protein was used for the coupling experiment subsequently.

2.12. Bacterial display and flow cytometry

2.12.1. Library generation

To generate the error-prone PCR library of SPM variants, the GeneMorph II random mutagenesis kit (Agilent) was used with 50 ng of input template encoding SPM and 30 PCR cycles. The PCR products were digested using Dpn1 (NEB) at 37 °C for 2 h before heat inactivation at 80 °C for 20 min. The PCR products were gel-extracted and purified using the Wizard SV Gel and PCR Clean-Up Kit (Promega), following the manufacturer's protocol. The 2 × HiFi DNA Assembly mix (NEB) was used to insert

the SPM fragments into the pet28a vector backbone, and the reaction mix was incubated at 50 °C for 2 h. After heat inactivation at 80 °C for 20 min, the DNA was purified using the Wizard SV Gel and PCR Clean-Up Kit (Promega) and eluted into nuclease-free water. Approximately 600 ng of the purified DNA was added into each 60 µl aliquot of commercial TG1 Duo cells (Lucigen). Electroporation at 2.5 kV per pulse was caried out using 0.2 mm cuvettes (Bio-Rad) with a MicroPulser (Bio-Rad). 1 mL of pre-warmed rescue media (Lucigen) was immediately added per aliquot of cells after electroporation. Transformed bacteria were incubated at 37 °C for 1 h for outgrowth. Serial dilutions of the library were plated on LB agar containing 50 µg/mL kanamycin to determine the library size. The TG1 library size was approximately 3.7 $\times$ 10^5 cfu. The rest of the TG1 culture was plated onto bioassay plates containing LB agar with 50 µg/mL kanamycin. After incubation at 37 °C for 16 h, the TG1 colonies were scrapped and resuspended in 40 mL of LB media. Plasmids from the pooled colonies were isolated using the GeneJET Plasmid Miniprep Kit (Thermo Fisher), following the manufacturer's protocol, and used for heat shock transformation of commercial chemically competent BL21(DE3) cells (Lucigen). 300 ng of DNA was added to each 40 µl aliquot of BL21 cells. The BL21 library size was determined by plating serial dilutions on LB agar containing 50 µg/mL kanamycin and found to be approximately 3.8 $\times$ 10^5 cfu.

2.12.2. Grow-up and induction of intimin expression

A single colony of chemically-competent BL21(DE3) cells containing the intimin-HA-SPM(wild-type)-myc plasmid was picked from the LB agar containing 50 µg/mL kanamycin and used to inoculate 10 mL of LB (high salt) media with 50 µg/mL kanamycin. In the case of the library encoding the SPM variants, the BL21 colonies

were pooled and resuspended in 50 mL LB media with 50 µg/mL kanamycin. An overnight culture of the library was set up by a 1:1,000 dilution of the stock culture. After overnight growth at 37 °C for 16-18 h, the overnight cultures were diluted 1:100 or 1:200 into 10 mL of fresh LB media containing 50 µg/mL kanamycin and 2 mM EGTA. After incubation at 37°C for 2.5 h, the OD_{600} reaches 0.3 – 0.5 and protein expression was induced by adding 0.42 mM IPTG and incubating the cultures at 30 °C for another 2.5 h or 4 h, as indicated. The cells were harvested by centrifugation at 3,000 g for 3 min at 4 °C. The cells were washed twice with HBS and approximately 10^7 cells were used for antibody staining of each sample.

2.12.3. Flow cytometry and fluorescence-activated cell sorting (FACS)

The cells were incubated with mouse anti-myc antibody (9B11, Cell Signaling Technology) diluted 1:20 in 50 µL HBS at 25 °C for 20 min. For the appropriate samples, 2 mM Ca^{2+} was added to the cells subsequently and incubated at 37 °C for another 10 minutes. Cells were washed twice with PBS before incubating at 25 °C for 20 min in the dark with goat anti-mouse IgG, Phycoerythrin (PE) antibody (Invitrogen) diluted 1:30 in 50 µl PBS. Cells were washed another 3 times with PBS.

When testing the anhydride reactivity of the intimin-SPM fusion, 1 mM alkoxyamine-PEG_4-biotin (from a 20 mM stock solution in DMSO) was added to the sample, together with the calcium solution. Subsequently, the cells were washed 4 times with PBS before incubating in the dark at 25 °C for 20 min with 33 µg/mL Streptavidin, R-Phycoerythrin Conjugate (SAPE) (Invitrogen) diluted in PBS. Cells were then washed another 3 times with PBS before flow cytometry.

For dual-stained cells, after the anti-myc staining, the cells were incubated at 25 °C for 20 min with biotin-rat anti-HA (3F10, Roche) diluted 1:50 in 50 µL PBS. Cells

were washed twice with PBS before incubating in the dark with 10 µg/mL Streptavidin, Alexa Fluor™ 488 conjugate (Invitrogen) diluted in 50 µl PBS at 25 °C for 20 min. Cells were washed another 3 times with PBS before the pellet was resuspended in 1 mL of PBS. For experiments involving sequential sorting, calcium was not added after incubation with the anti-myc antibody but the sorted cells from the first round (high anti-HA and high anti-myc) were centrifuged at 3,000 g for 3 min at 4 °C before the cell pellet was resuspended in 50 µL HBS containing 2 mM Ca^{2+} and incubated at 37 °C for another 10 minutes. The cells were washed once with PBS before the next round of sorting for cells that are high in anti-HA but low in anti-myc fluorescence. For all the samples, cells were thoroughly vortexed before flow cytometry using the S3e Cell Sorter (Bio-Rad). 100,000 events were recorded for flow cytometry. Gating by size was performed during analysis to identify *E. coli* cells and exclude doublets.

2.13. Bioinformatics and Computational Analyses

2.13.1. Homology searches using BLAST and secondary structure prediction

Homology searches for the VWD domains of RGMB, RGMC or SPM of FrpC were performed using standard protein Basic Local Alignment Search Tool (BLAST) (Altschul et al., 1990). The query sequences were searched against the non-redundant protein sequences (nr) database using the blastp algorithm for RGMB and RGMC. Parameters of the searches include an expect threshold of “10” and word size of “6”. The “BLOSUM62” matrix was used for scoring, factoring in gap costs of “Existence:11 and Extension:1”. Conditional compositional score matrix adjustment was also used. The PSI-BLAST algorithm was used to identify FrpC SPM homologues,

with 5 iterations and a statistical significance threshold of 0.005. Parameters of the search include an expect threshold of “10”, word size of “3”. The same scoring parameters as RGMB and RGMC were used for FrpC SPM. ClustalO (Madeira et al., 2019) was used for multiple sequence alignment of homologous sequences identified from the BLAST searches. Jpred (ver 4) (Drozdetskiy et al., 2015) was used for secondary structure prediction of FrpC SPM.

2.13.2. Structure visualisation

The structure of RGMB was obtained from PDB 4BQ6 (Bell et al., 2013). The structure of OAZ/ODC was obtained from PDB 4ZGY (Wu et al., 2015) and the structure of ODC dimer was obtained from PDB 1D7K (Almrud et al., 2000). The structure of OAZ/AZI was obtained from PDB 4ZGZ (Wu et al., 2015). FrpC SPM structure was obtained from PDB 6SJW (Kuban et al., 2020). PyMOL (version 2.0) (Schrödinger) was used to visualise crystal and NMR structures. Figures were prepared using the open-source graphics editor tool Inkscape (inkscape.org).

Chapter 3: RGMs

Repulsive Guidance Molecules (RGMs) are involved in functions ranging from cell motility and adhesion to iron metabolism (Bell et al., 2013). RGMB and RGMC contain a partial von Willebrand type D (VWD) domain (Monnier et al., 2002). Multiple studies have reported autocatalytic cleavage in the Asp-Pro bond of the VWD domain of RGMC (Bell et al., 2013; Zhang et al., 2005). Similar to other reported autocatalytic cleavage at Asp-Pro bonds, cleavage likely occurs via an anhydride intermediate. In RGMC, the anhydride formed after cleavage is rapidly hydrolysed but the cleaved fragments remain held together by disulfide bonds (Zhang et al., 2005).

Crystallisation of an ectodomain fragment of RGMB containing the VWD domain after cleavage revealed a novel protein fold (PDB 4BQ6) (Bell et al., 2013). The Asp-Pro cleavage site was found in the loop between β-strands 1 and 2. After cleavage, the newly formed N-terminal Pro was found to be buried in the hydrophobic core of the domain (Bell et al., 2013). However, it is not known which other residues are important for cleavage.

Bell *et al.* observed that the cleavage of RGMB and RGMC was essential for localisation and secretion of the proteins (Bell et al., 2013). It was hypothesised that cleavage of the VWD domains is promoted by acidic pH and occurs in the late secretory pathway like in the cleavage of MUC2 (Bell et al., 2013; Lidell et al., 2003). The biological relevance of autocatalytic cleavage and its function in iron metabolism was further suggested by the clustering of a large number disease-causing mutations proximal to the cleavage site (Papanikolaou et al., 2004).

In this chapter, I describe experiments I have performed to understand the cleavage mechanism of the VWD domain and the amino acid environment necessary for anhydride formation. This analysis was done with a view to potentially engineering

the VWD domain and exploiting the anhydride for conjugation to endogenous proteins, as illustrated in Figure 3.1.

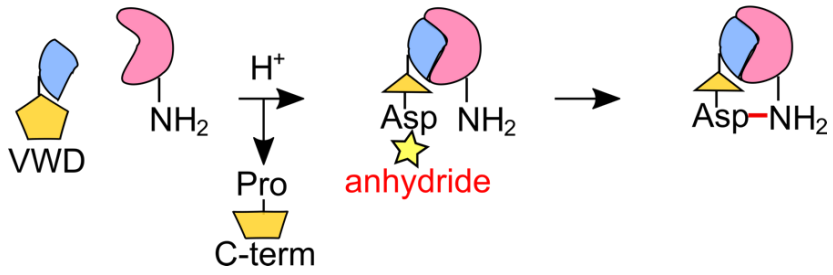


Figure 3.1 Use of autocatalytic VWD domain for crosslinking to endogenous proteins. Binder protein (blue) is genetically fused to VWD domain (yellow) that contains the scissile Asp-Pro bond. The binder protein docks to the target protein (pink). Under acidic conditions, VWD domain undergoes cleavage at the Asp-Pro bond, releasing the C-terminal fragment of VWD. An aspartic anhydride is generated at the terminus of the N-terminal domain of VWD and reacts with a proximal nucleophile on the target protein.

Experiments to isolate the VWD domains of RGMB and RGMC before cleavage are described in the first half of this chapter. Results from the experiments and study of the *in vitro* cleavage of RGMC provided insights into control over the cleavage process. I also show that the N-terminal cleavage fragment of RGMC is capable of covalently reacting with alkoxyamine. This conjugation confirms the presence of an anhydride intermediate that can be used for coupling to endogenous proteins.

3.1. Expression of RGMB and RGMC

Fragments of the extracellular regions of RGMB and RGMC containing the VWD-like domain (Figure 3.2) were expressed by transient transfection of Expi293F cells. These ectodomain fragments were previously shown to be stably expressed in

HEK293T cells (Bell et al., 2013). The RGMB fragment, in particular, was shown to retain binding to neogenin and was capable of initiating downstream signalling and inhibiting neurite growth (Bell et al., 2013). The GPI anchor was excluded so that the ectodomain fragment would be secreted rather than membrane-associated. The N-terminal domain was excluded as well since the VWD domain-of-interest is found in the C-terminal neogenin-binding region.

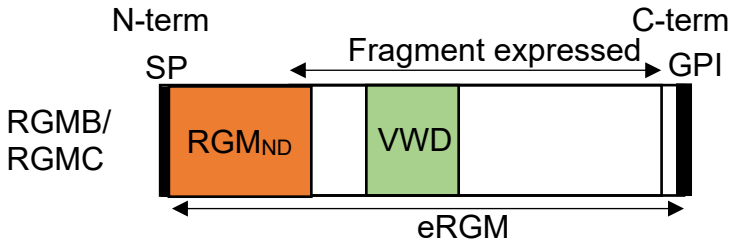


Figure 3.2 Schematic map of RGMs. SP: Signal peptide; RGM$_{ND}$ (orange): RGM N-terminal domain containing the RGD motif. RGM$_{ND}$ binds bone morphogenetic protein (BMP); eRGM: ectodomain of RGM; VWD (green): von Willebrand Factor type D-like domain; GPI: glycosylphosphatidylinositol anchor. Adapted from Bell *et al.*, 2013.

Like HEK293T, Expi293F is derived from the 293 human cell line. Since it is a suspension rather than adherent culture, Expi293F cells grow to much higher cell densities than HEK293T and generally lead to higher protein yields. The presence of an N-terminal secretory signal peptide in the construct enabled the secretion of the RGMs into the extracellular media.

4-days post-transfection, RGMB and RGMC were harvested from the expression media. Utilising the C-terminal His$_6$-tag on both proteins, RGMB and RGMC were purified by immobilised metal-affinity chromatography (IMAC). The eluted fractions were run on SDS-PAGE with Coomassie staining for analysis of their purity (Figure 3.3a).

9 mg of RGMB was purified from 30 mL of Expi293F culture. A lower yield of RGMC was obtained: 1 mg of RGMC was purified from 30 mL of Expi293F culture.

Both proteins were purified with reasonable purities (Figure 3.3a). The identity of purified RGMB was verified by Western blot using an anti-RGMB antibody (Figure 3.3c). RGMC was used as a negative control in the Western blot.

The expected molecular weight of RGMB was 41 kDa and the expected molecular weight of RGMC was 40 kDa. The observed deviations from the expected molecular weights on SDS-PAGE were due to glycosylation (Figure 3.3a). Cleavage of the N-linked oligosaccharides from RGMB and RGMC after overnight Peptide:N-glycosidase F (PNGase F) treatment led to downward shifts in the bands so that they correspond to the expected molecular weights (Figure 3.3a). The initial double bands observed for RGMC may correspond to different RGMC isoforms dependent on furin cleavage, as has been reported previously (Kuninger et al., 2008; Silvestri et al., 2008). A furin cleavage site is present in the C-terminus of RGMC, N-terminal of the GPI anchor (Kuninger et al., 2008). The furin cleavage site is not present in RGMA or RGMB (Siebold et al., 2017). Previously, RGMC was found as a mixture of 50 kDa and 40 kDa species in extracellular fluid, depending on cleavage by furin (Kuninger et al., 2008). The cleaved 40 kDa species is a non-membrane bound soluble protein that acts as a decoy receptor for BMP ligands (Silvestri et al., 2008). In the RGMC ectodomain fragment used for expression, furin cleavage will lead to formation of an isoform that is 7 kDa smaller than the full-length protein that is 40 kDa.

After cleavage of the Asp-Pro bond in the VWD domain of the RGMs, the resulting fragments are held together by disulfide bonds (Bell et al., 2013). The addition of a reducing agent is required for cleavage of the disulfide bonds and separation of the cleavage fragments. The extent of cleavage of purified RGMB and RGMC were visualised by Coomassie-stained SDS-PAGE (Figure 3.3b). It was observed that purified RGMB was completely cleaved. Little to no full-length RGMB

was observed under reducing conditions (Figure 3.3b). The disappearance of the full-length band was accompanied by the appearance of two new lower molecular weight bands (~32 kDa and ~17 kDa) that correspond to the C- and N-terminal cleavage fragments (Figure 3.3b).

On the other hand, purified RGMC was a mixture of cleaved and pre-cleavage forms. The bands corresponding to full-length RGMC were still observed in the presence of DTT (Figure 3.3b). Similar to RGMB, lower molecular weight bands corresponding to the cleavage fragments were only observed under reducing conditions. The complete cleavage of secreted RGMB and cleavage of only a proportion of secreted RGMC is in agreement with previous studies (Bell et al., 2013; Zhang et al., 2005).

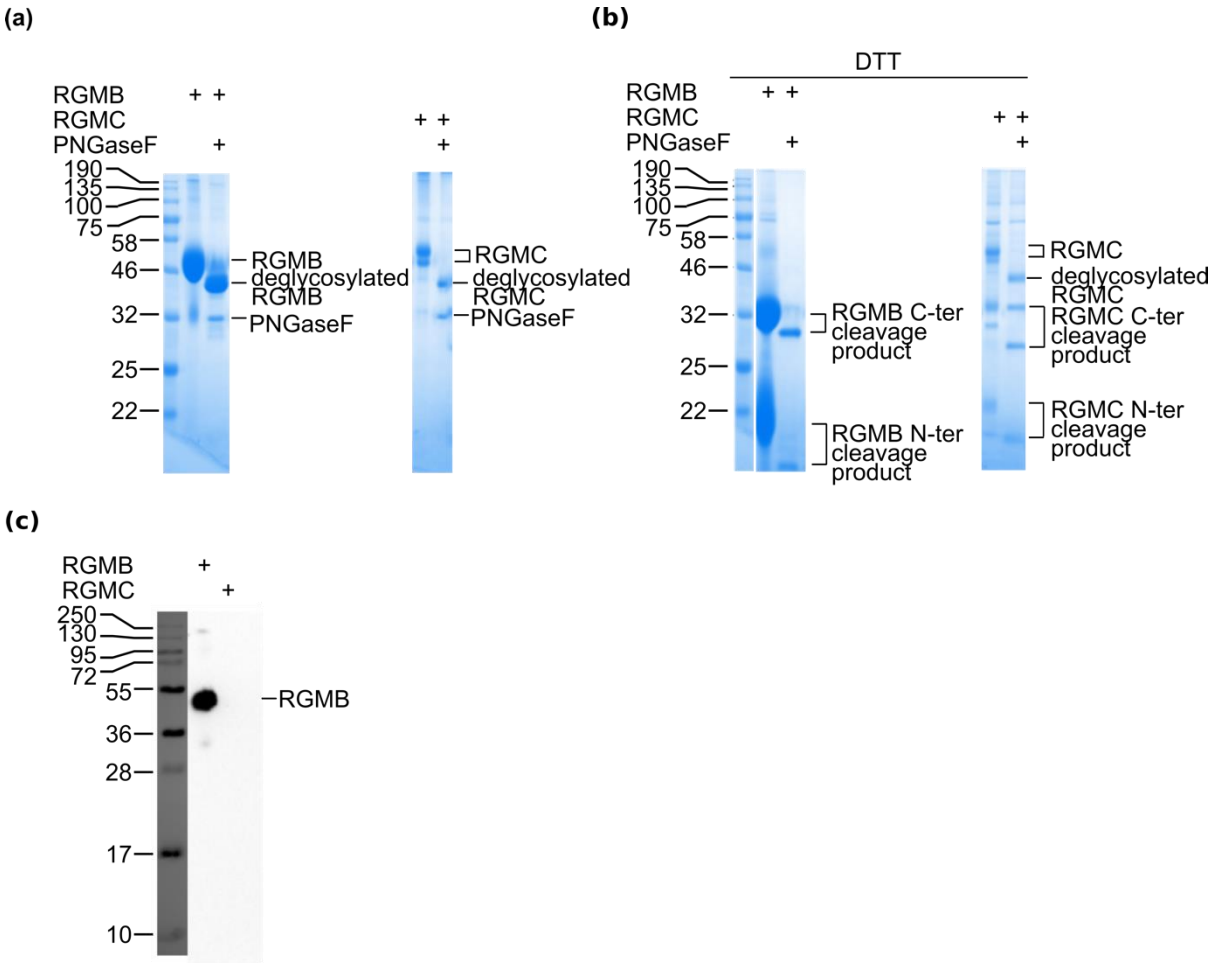


Figure 3.3 Expression of RGMB and RGMC in Expi293F cells. (a) Coomassie-stained SDS-PAGE of IMAC-purified RGMB and RGMC. For deglycosylation, RGMB and RGMC were treated with PNGase F. (b) Coomassie-stained SDS-PAGE of purified RGMB and RGMC under reducing conditions. 50 mM DTT was added to the SDS loading buffer. (c) Western blot using anti-RGMB antibody on IMAC-purified RGMB and RGMC under non-reducing conditions. The epitope for the anti-RMB antibody is located at the C-terminal cleavage fragment.

3.2. Treatment with additives to isolate pre-cleavage VWD

To study the cleavage process, I tried different approaches to isolate RGMB and RGMC VWD domains before cleavage. The first approach was to treat the transiently-transfected Expi293F cells with additives. I hypothesised that additives like ammonium chloride, bafilomycin A1 and chloroquine would inhibit the *in vivo* cleavage of RGMB and RGMC.

Treatment with ammonium chloride leads to an influx of ammonia into cells and was found to increase the pH in the trans-Golgi from approximately 6.2 to 6.8 (Demaurex et al., 1998). On the other hand, bafilomycin A1 is an inhibitor of the vacuolar ATPase proton pump and was found to increase the pH in the *trans*-Golgi to approximately 6.5 (Axelsson et al., 2001). When studying the autocatalytic cleavage of the MUC2 and MUC5AC VWD domains, Lidell *et al.* treated transiently transfected CHO-K1 cells with 25 mM ammonium chloride or 300 nM bafilomycin A1 to neutralise the secretory pathway (Lidell and Hansson, 2006; Lidell et al., 2003). The ammonium chloride and bafilomycin A1 treatments were found to inhibit MUC2 cleavage (Lidell et al., 2003). Chloroquine is another weak base that neutralises lysosomes and the secretory pathway (Back and Soinila, 1996). Addition of chloroquine could alter the cleavage of the RGMs if cleavage occurs in the late secretory pathway.

Transiently-transfected Expi293F cells were treated with the additives from the day of transfection until 4 days post-transfection when the proteins were harvested. Initially, shorter durations (72 h with ammonium chloride and 24 h with chloroquine) of treatments were tested but since there was no observed change in the cleavage extent of RGMB, I decided to increase the duration of treatment and maintain the treatment throughout protein expression. A range of concentrations of the additives were tested: 25 mM, 50 mM, or 100 mM ammonium chloride; 50 nM and 100 nM of bafilomycin A1; 50 μM, 100 μM or 200 μM chloroquine.

I observed that treatment with the additives lowered the yields of the purified proteins. Treatment with 25 mM ammonium chloride over 93 h led to a decrease in yield of RGMB from 1 mg to 0.5 mg per 30 mL of culture (Figure 3.4a). With addition of 50 nM bafilomycin A1, the yield of purified RGMB was much lower than the untreated control and the band corresponding to the purified protein was barely visible in Coomassie-stained SDS-PAGE (Figure 3.4a). 50 μM chloroquine treatment also led to a decrease in yield of RGMB (Figure 3.4a). The observed reduced yields were likely due to the cytotoxic nature of the drugs (Mains and May, 1988; Yan et al., 2016).

In all cases, the addition of additives did not inhibit cleavage of RGMB. In all the treated cells, the band corresponding to full-length RGMB disappeared under reducing conditions (Figure 3.4a). This suggests that secreted RGMB is still completely cleaved.

Similarly, a reduced yield of RGMC was observed after treatment with the additives. RGMC expression was almost completely abolished when treated with 25 mM ammonium chloride and 50 nM bafilomycin (Figure 3.4b). Purified RGMC after treatment with 50 μM chloroquine was still in a mixture of cleaved and uncleaved

forms. Under reducing conditions, bands corresponding to both full length RGMC and the cleavage products were observed, similar to the untreated control (Figure 3.4b).

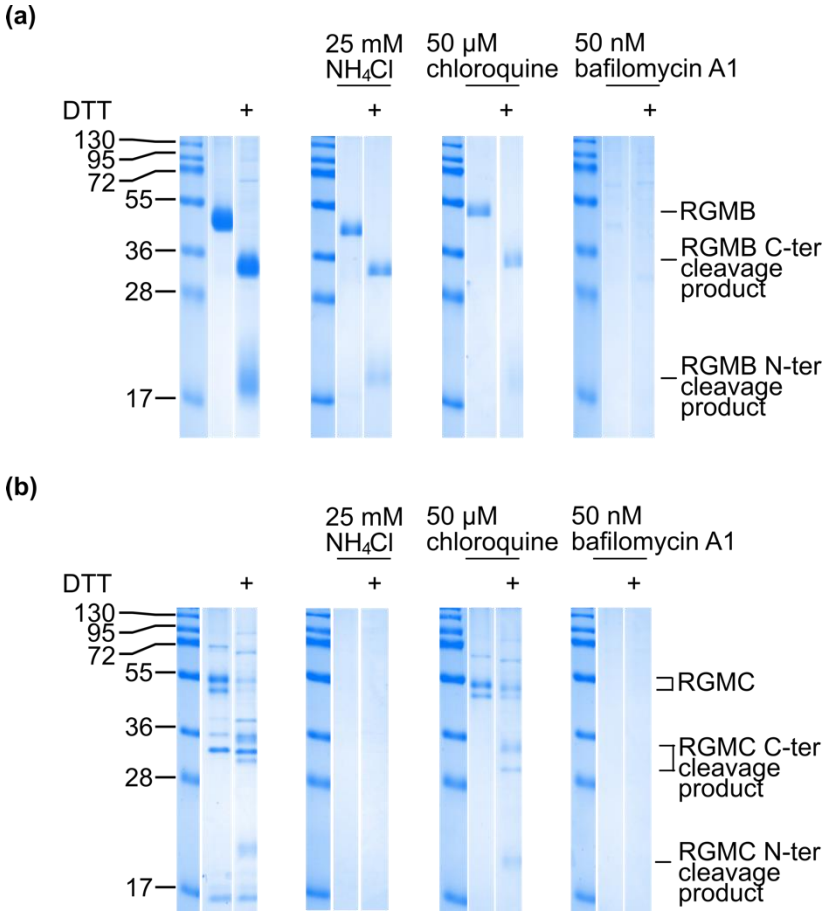


Figure 3.4 Treatment with additives. (a) Transiently transfected Expi293F cells were treated with 25 mM ammonium chloride, 50 µM chloroquine or 50 nM bafilomycin for 93 h, from the day of transfection. IMAC-purified RGMB were analysed under non-reducing or reducing conditions by Coomassie-stained SDS-PAGE. (b) Same as (a) but for IMAC-purified RGMC.

3.3. Use of a different cell-line to isolate pre-cleavage VWD

Another approach I attempted to isolate the VWD domains of RGMB and RGMC before cleavage was to express the ectodomain fragments in a different cell line. Apart from Expi293F, I expressed RGMB and RGMC in ExpiCHO. Like Expi293F, ExpiCHO is a mammalian cell-line that grows in suspension culture and gives high recombinant protein yields.

Differences between the secretory pathways of mammalian CHO and human HEK293 cells are known. For example, there is a difference in the glycosylation patterns of recombinant proteins expressed in CHO compared to HEK293 (Bosques et al., 2010). Additional Galactose-*alpha*-1,3-galactose (α-gal) and *N*-Glycolylneuraminic acid (NGNA) were observed in glycans of CHO-expressed proteins (Bosques et al., 2010). MUC2 VWD domain was shown to be cleaved to a lesser extent when expressed in CHO-K1 compared to the LS174T cell line (Lidell et al., 2003). Lidell *et al.* hypothesised that the reduced cleavage of MUC2 in CHO-K1 was due to a shorter time spent in the later, more acidic parts of the secretory pathway (Lidell et al., 2003).

Comparing purified RGMB and RGMC from ExpiCHO and Expi293, I did not observe any difference in the cleavage extents of RGMB and RGMC expressed in the two cell lines. All the purified RGMB from ExpiCHO was completely cleaved, similar to RGMB purified from Expi293F (Figure 3.5a). The complete cleavage of RGMB can be inferred from the disappearance of the band for full length RGMB under reducing conditions. There was little difference in yields of RGMB purified from ExpiCHO and Expi293. A majority of RGMC purified from ExpiCHO was likewise in its cleaved form. A faint band corresponding to full length RGMC remained under reducing conditions (Figure 3.5b). A new band was apparent at 55 kDa under reducing conditions for ExpiCHO-expressed RGMB and a new band was apparent at 60 kDa under reducing conditions for ExpiCHO-expressed RGMC.

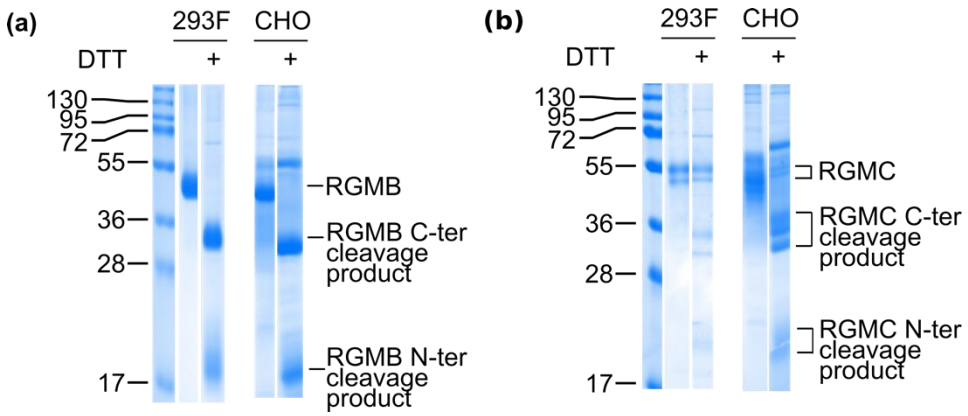


Figure 3.5 Use of ExpiCHO cell line to inhibit cleavage. Coomassie-stained SDS-PAGE gel of (a) RGMB and (b) RGMC purified using IMAC from the media of transiently-transfected ExpiCHO compared to Expi293F.

3.4. Targeting for ER retention or protein body formation to isolate pre-cleavage VWD

The third approach I tried to isolate pre-cleavage VWD of RGMB and RGMC was to target the proteins for retention in the endoplasmic reticulum (ER) or ER-derived protein bodies. If the autocatalytic cleavage of the VWD occurs in the acidic compartments of the Golgi apparatus, redirecting the proteins from the Golgi apparatus should reduce the cleavage extents of RGMB and RGMC.

ER-resident proteins in the lumen of the ER usually have an additional C-terminal -KDEL signal sequence (Stornaiuolo et al., 2003). The -KDEL motif is recognised by ER retention receptors that enable retrieval from the cis-Golgi (Stornaiuolo et al., 2003). Recombinant proteins have been targeted into the ER by appending a C-terminal -KDEL sequence (Petruccelli et al., 2006). An alternative ER pathway is the accumulation in ER or vacuole-derived protein bodies. The proline-rich Zera sequence was originally discovered in the N-terminus of the maize storage protein, γ zein that accumulates in large amounts in seed tissues (Torrent et al., 2009).

It was shown that fusing the Zera sequence to the N-terminus of other recombinant proteins induced stable accumulation of the recombinant proteins in protein bodies of even eukaryotic cells, including mammalian cells (Torrent et al., 2009).

I transfected Expi293F cells with new constructs of RGMB and RGMC, with either an additional C-terminal -KDEL sequence or N-terminal Zera sequence attached. 4 days post transfection, I isolated the intracellular protein fraction by cell lysis and subjected the lysate to immobilised metal affinity chromatography (IMAC) for further purification.

The intracellular fraction had a lot more impurities as compared to the expression media, and it was unclear from Coomassie-stained SDS-PAGE if RGM-KDEL and Zera-RGM were expressed. Hence, I used Western blotting with an anti-RGMB antibody to detect RGMB-KDEL and Zera-RGMB in the intracellular fraction (Figure 3.6). Secreted RGMB previously purified from the expression media was used as a positive control and RGMC as a negative control. The double bands observed for secreted RGMB in the absence of DTT may have been due to the leakage of reducing agent from the neighbouring right lanes (Figure 3.6).

I observed a faint band corresponding to the C-terminal cleavage product of RGMB-GKDEL under reducing conditions. This suggests that RGMB-GKDEL was expressed but was still cleaved, like the secreted RGMB control. Zera-RGMB was not detectable in the intracellular fraction of transfected Expi293F (Figure 3.6). RGMC-GKDEL and Zera-RGMC were not detected by Coomassie-stained SDS-PAGE or Western blotting with an anti-His antibody (data not shown). Anti-RGMC is commercially available but the antibody that was tested did not detect the expressed ectodomain fragment.

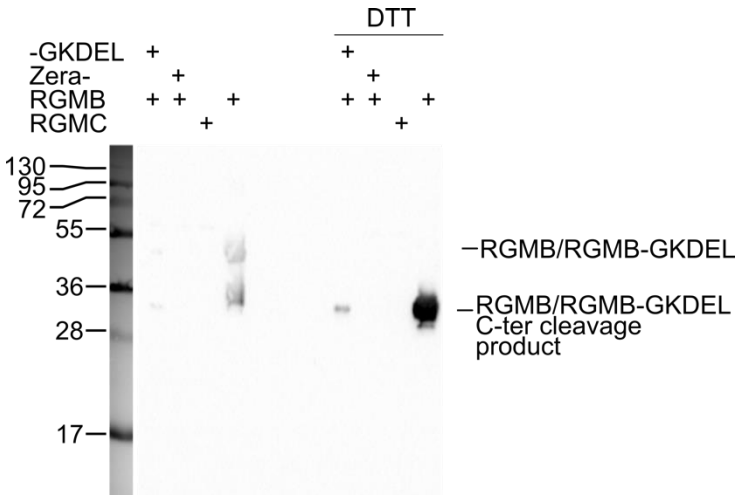


Figure 3.6 Targeting RGMB for ER retention or protein body formation. Western blot using anti-RGMB antibody on IMAC-purified RGMB-GKDEL, Zera-RGMB, secreted RGMB and secreted RGMC. RGMB-GKDEL and Zera-RGMB were purified by IMAC from the intracellular fraction of transiently transfected Expi293F.The anti-RGMB antibody does not recognise the N-terminal cleavage fragment.

3.5. Prokaryotic expression to isolate pre-cleavage VWD

Unlike eukaryotes, prokaryotes do not contain membrane-bound organelles including the ER and Golgi. The intracellular pH of *E. coli* is maintained at about 7.6 (Zilberstein et al., 1984). If the autocatalytic cleavage of RGMB and RGMC takes place under acidic conditions, I hypothesised that prokaryotic expression of the protein may inhibit the cleavage.

Given the presence of multiple disulfides in the RGM ectodomain, I chose to test expression in different *E. coli* strains that promote disulfide bond formation. The *E. coli* strains included Rosettagami2 (DE3), SHuffle T7 Express, C41(DE3) and BL21 (DE3) Erv1p DsbC. Rosettagami2 is a double mutant ($trxB^-$, gor^-) expression strain that has an oxidising cytoplasmic environment (de Marco, 2009). SHuffle T7 expression strain has an oxidising cytoplasmic environment and also overexpresses DsbC isomerase (de Marco, 2009). C41(DE3) was superior to BL21(DE3) in

overexpression of some proteins which were otherwise toxic to BL21(DE3) (Miroux and Walker, 1996). BL21 (DE3) Erv1p DsbC expresses both DsbC isomerase and also Erv1p, a sulfhydryl oxidase for disulfide bond formation (Nguyen et al., 2011). Finally, I also tested directing RGMB and RGMC to the bacterial periplasm by the addition of a N-terminal PelB secretion leader peptide. The bacterial periplasm provides an oxidising environment for disulfide bond formation and also contains a variety of disulfide isomerases, chaperones and foldases (de Marco, 2009).

The coding sequences for ectodomain fragments of RGMB and RGMC were cloned into a pET28a expression vector for *E. coli* expression. Multiple conditions were tested to optimise expression, including varying the protein induction temperature and time. The *E. coli* cells were lysed, and the intracellular fraction was purified by IMAC. For the PelB-fusion proteins, the periplasm was extracted, and the soluble periplasmic fraction was subjected to IMAC. The eluted fractions from IMAC were then analysed by Coomassie-stained SDS-PAGE gel.

There was no detectable expression of RGMB or RGMC in the soluble lysate of Rosettagami2 (DE3), SHuffle T7 Express, C41(DE3) or BL21 (DE3) Erv1p DsbC (Figure 3.7a). There were no observable bands at the expected molecular weight (~41 kDa for RGMB and ~40 kDa for RGMC) (Figure 3.7a). Similarly, there was no detectable PelB-RGMB or PelB-RGMC in the periplasm (Figure 3.7a).

The observation of a strong band at ~40 kDa in the total lysate fraction of BL21 (DE3) Erv1p DsbC but not the soluble cell lysate fraction (gel not shown) suggested that RGMB and RGMC were found in inclusion bodies. Therefore, I decided to isolate the inclusion bodies in BL21 (DE3) Erv1p, DsbC and refold the proteins. The refolding buffer contained oxidised and reduced glutathione to promote disulfide shuffling. Refolded RGMB and RGMC were then analysed by Coomassie-stained SDS-PAGE

(Figure 3.7b). The formation of disulfide bonds in refolded RGMB and RGMC was verified by the shift in the bands upon addition of DTT (Figure 3.7b).

To verify if refolded RGMB was in its native structure, I investigated if RGMB retained its autocatalytic cleavage activity. I incubated the refolded RGMB in citrate-phosphate buffer, pH 5.0 for 1 h. Previous experiments showed that cleavage of RGMC was enhanced at pH <7.4 and occurred within 2 h (Section 3.6). However, after 1 h incubation in acidic pH, all of the RGMB remained in its pre-cleavage form, appearing as a full-length band under reducing conditions (Figure 3.7c). This suggests that the refolded RGMB was inactive and may not have been properly refolded to its native structure.

Additionally, higher molecular weight soluble aggregates were also observed in the Coomassie-stained SDS-PAGE under non-reducing conditions (Figure 3.7c). The disappearance of the aggregates under reducing conditions suggests that aggregation was due to improper disulfide bond formation (Figure 3.7c).

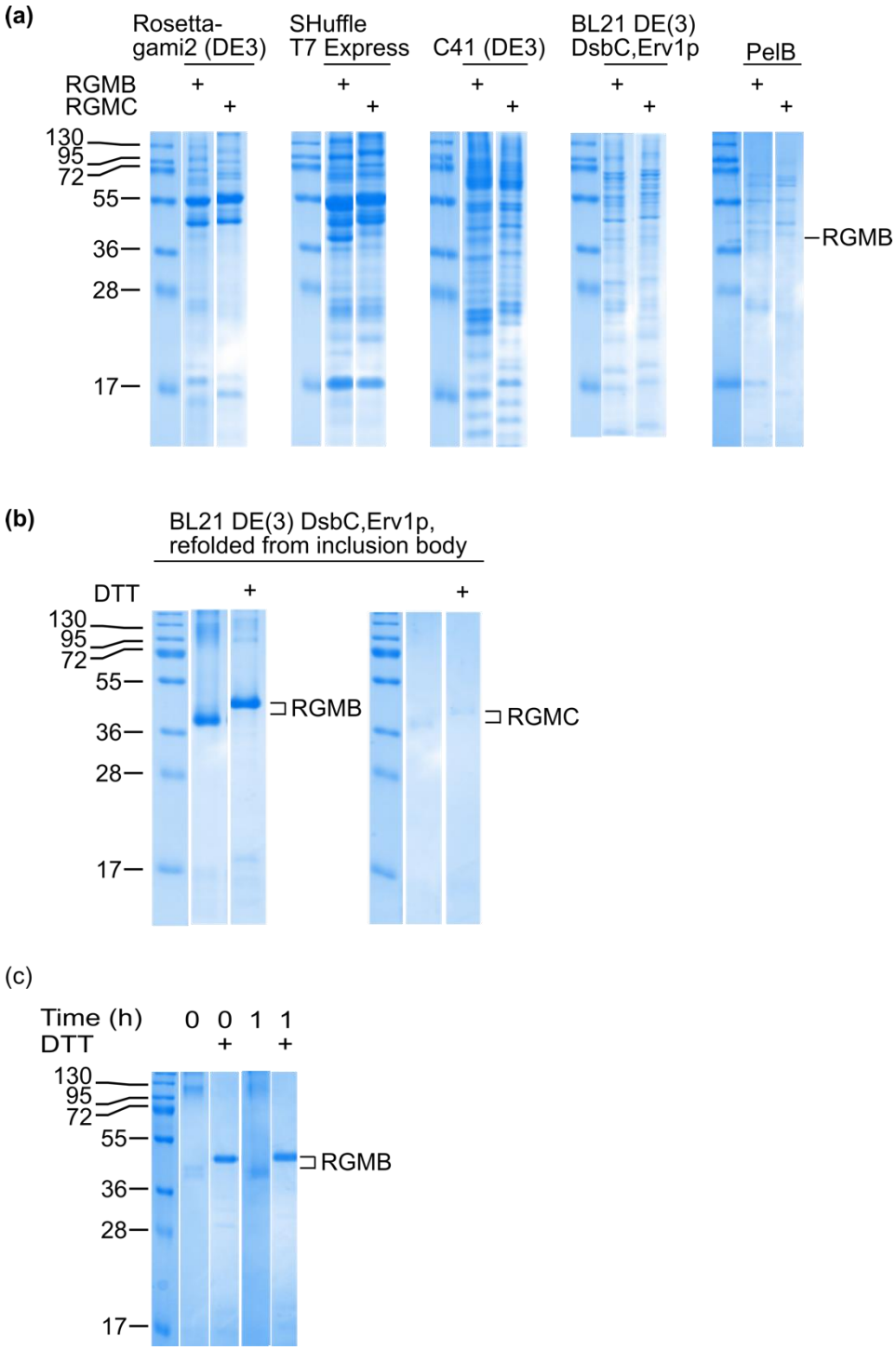


Figure 3.7 Prokaryotic expression of RGMB and RGMC VWD. (a) Coomassie-stained SDS-PAGE of RGMB and RGMC IMAC-purified from the soluble lysate of Rosettagami2 (DE3), SHuffle T7 Express, C41(DE3) and BL21 (DE3) Erv1p DsbC, as well as PelB-RGMs purified from the periplasm of BL21 (DE3). (b) Coomassie-stained SDS-PAGE of RGMB and RGMC refolded from inclusion bodies of BL21 (DE3) Erv1p DsbC. (c) Refolded RGMB was incubated in citrate-phosphate buffer, pH 5.0 for 1 h, at 37 °C to assay for autocatalytic cleavage activity.

3.6. Analysing RGMC VWD cleavage

To study the cleavage process of the VWD domain, I incubated RGMC in buffers spanning pH 2.2 - 7.4 at 37 °C. A citric acid/disodium phosphate (Na_2HPO_4) buffer was used to ensure proper buffering over the entire pH range. Secreted RGMC from Expi293F cells was initially already in a mixture of pre-cleavage and cleaved forms (Section 3.1). I hypothesised that incubation in acidic conditions would induce further cleavage, similar to MUC2 (Lidell et al., 2003).

The cleavage extent of RGMC was monitored by Coomassie-stained SDS-PAGE and densitometry. The band corresponding to full-length RGMC under reducing conditions represents pre-cleavage RGMC while the band for full-length RGMC under non-reducing conditions represents the total amount of RGMC (Figure 3.8a). The fraction of total RGMC that was cleaved after 2 h under each pH condition was calculated. There was no clear pH-dependence, although the highest cleavage extents were observed at pH 2.2 - 3.4. Cleavage at pH 4.2 – 5.8 were slightly enhanced compared to pH 7.4 (Figure 3.8b).

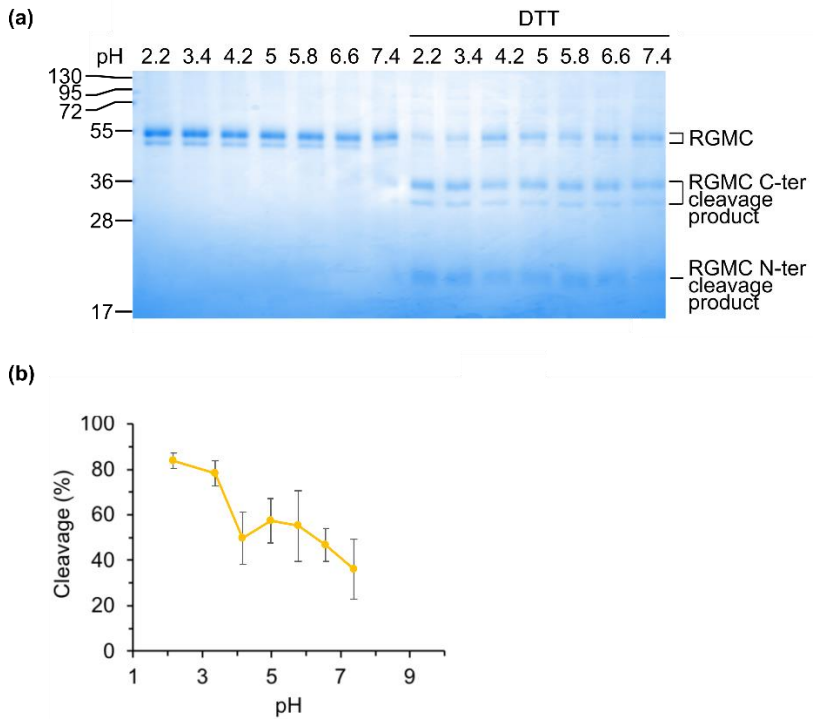


Figure 3.8 pH-dependence of RGMC cleavage *in vitro*. Secreted RGMC was incubated at 37 °C for 2 h in a citrate-phosphate buffer spanning pH 2.2-7.4. The reaction was quenched by adding SDS loading buffer and boiling at 95 °C for 3 min. a) Coomassie-stained SDS-PAGE to monitor cleavage of RGMC. b) pH-dependence of RGMC cleavage determined by densitometry analysis of Coomassie-stained SDS-PAGE as in (a) (mean ± 1 S.D., n=3).

3.7. Covalent coupling of RGMC to alkoxyamine-PEG$_4$-biotin

To confirm the formation of an anhydride and ensure that the lower molecular weight bands observed in Figure 3.8a were not due to non-specific proteolysis by contaminating proteases, I next tested the reactivity of the anhydride formed during autocatalytic cleavage of the VWD domain in RGMC. I incubated RGMC with an excess of alkoxyamine-PEG$_4$-biotin (Figure 3.9a). I hypothesised that the anhydride electrophile will react with the nucleophilic alkoxyamine and drive covalent bond formation upon calcium addition.

To verify the pH-dependence of conjugation, I incubated RGMC with alkoxyamine-PEG_4-biotin at pH 4.5 and pH 7.4. Based on the previous experiment, cleavage of RGMC should be enhanced at pH 4.5 compared to pH 7.4 (Section 3.6). The pK_a of alkoxyamine is about 4.7. Therefore, at pH 7.4, a majority of alkoxyamine-PEG_4-biotin should be in its deprotonated form that acts as a good nucleophile. At pH 4.5, approximately 50% of the available alkoxyamine-PEG_4-biotin would be deprotonated, which is likely more than sufficient for reaction as well. With increased cleavage at pH 4.5, I expected that conjugation of RGMC to alkoxyamine-PEG_4-biotin will be enhanced at pH 4.5 compared to pH 7.4.

The presence of alkoxyamine-PEG_4-biotin bound to RGMC was detected by Western blotting using streptavidin-HRP. As was hypothesised, this analysis indicated that alkoxyamine-PEG_4-biotin was covalently bound to RGMC, predominantly at the N-terminal cleavage fragment at pH 4.5 (Figure 3.9b, 3.9c). The N-terminal cleavage fragment is where an anhydride is postulated to form. RGMC was conjugated to alkoxyamine-PEG_4-biotin when incubated at 0.08 mM or higher concentrations of alkoxyamine-PEG_4-biotin. Higher concentrations of alkoxyamine-PEG_4-biotin led to increased conjugation to RGMC (Figure 3.9b, 3.9c). Some conjugation to the C-terminal fragment was also observed (Figure 3.9c). This could be due non-specific binding of the antibody or reaction of alkoxyamine-PEG_4-biotin with oxidised sialic acid residues in the glycans attached to RGMC. It was observed that alkoxyamine-PEG_4-biotin coupled to the N-terminal cleavage fragment even at pH 7.4, when 1.25 mM alkoxyamine-PEG_4-biotin was added (Figure 3.9b, 3.9c).

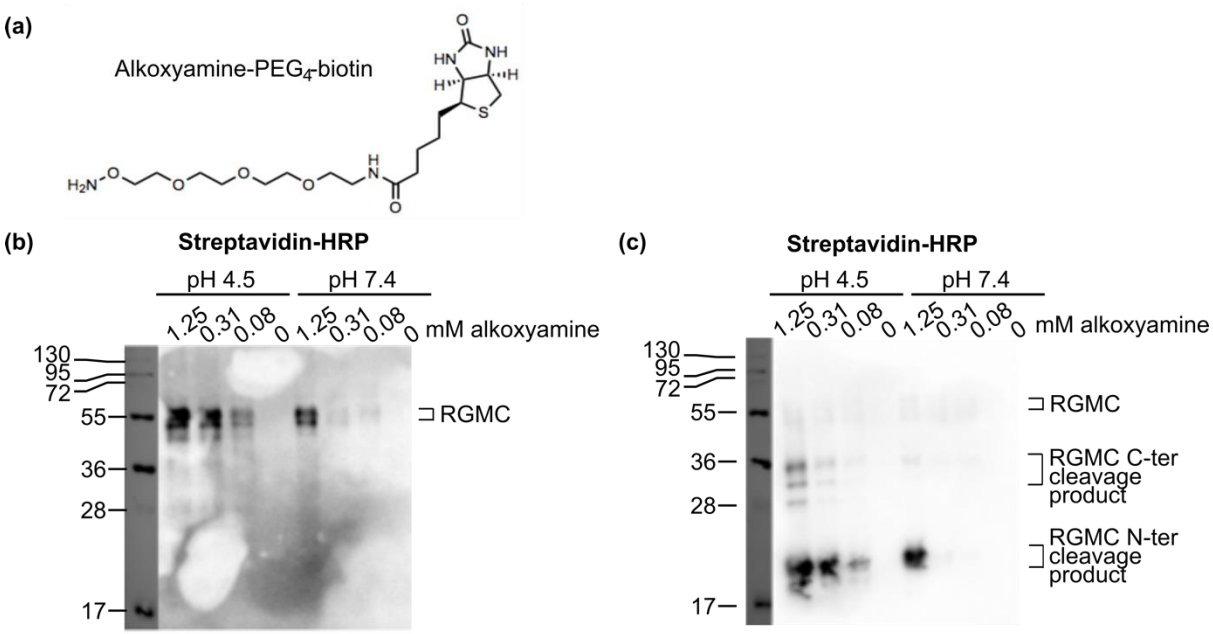


Figure 3.9 Covalent coupling of alkoxyamine-PEG4-biotin to RGMC N-terminal cleavage product. RGMC was incubated with varying concentrations of alkoxyamine-PEG$_4$-biotin in MES buffer pH 4.5 or pH 7.4 for 1 h at 37 °C. a) Chemical structure of alkoxyamine-PEG$_4$-biotin. (b) Western blot using streptavidin-HRP to detect alkoxyamine-PEG$_4$-biotin under non-reducing conditions. c) Western blot using streptavidin-HRP to detect alkoxyamine-PEG$_4$-biotin under reducing conditions.

3.8. Discussion

The aim of this chapter was to evaluate the potential of using the VWD domains of RGMB and RGMC as an anhydride-generating domain for targeting endogenous proteins. To develop the VWD domains as a covalent probe, I first sought to understand the cleavage mechanism of the Asp-Pro bond in the VWD domain and the amino acid environment necessary for anhydride formation.

In the first part of the chapter, I attempted to isolate the VWD domains of RGMB and RGMC before cleavage. The use of additives and expression in a different mammalian cell line, however, did not alter the cleavage extent. Previously, it was observed that additives inhibited cleavage of the VWD domain in MUC2 and the use of CHO-K1 for expression altered the cleavage extent of MUC2 (Lidell et al., 2003).

The difference in the sensitivity to the use of additives and different cell lines suggests possible differences in the cleavage mechanisms of MUC2 and RGMs. The observation that redirecting the VWD domain of RGMB to the ER did not inhibit cleavage suggests it is possible that cleavage of RGMB occurs in the ER rather than only later in the secretory pathway like MUC2. Alternatively, the short time RGMB spends in the cis-Gogi before ER retrieval was sufficient to induce cleavage of the protein. Cleavage in the nearly neutral environment of the ER was also postulated for MUC5B (Lidell and Hansson, 2006). The presence of multiple disulfides and differences in prokaryotic and mammalian pathways for post-translational modifications made prokaryotic expression and refolding of the VWD domains of RGMB and RGMC to their native structures difficult.

Given the low sequence homologies between the partial VWD domains of RGMs and other VWD domains, it was unclear if the same protein fold observed for RGMB (Bell et al., 2013) is conserved across different VWD domains. The crystal structures of the VWD domains of ANGTPL4 (PDB 6EUB) and the von Willebrand Factor (vWF) VWD'D3 (PDB 6N29) domains were recently published (Biterova et al., 2018; Dong et al., 2019). The ANGTPL4 VWD domain does not contain the same tightly-packed β-sandwich like in RGMB and vWF VWD domains. Like RGMB, the FGDPH sequence of ANGTPL4 is located in a short turn between two β-strands. However, there has been no reported cleavages at the Asp-Pro bonds in ANGTPL4 (Biterova et al., 2018) or vWF VWD domain (Dong et al., 2019). Cleavage of the Asp-Pro bond in some but not all VWD domains suggests the importance of the protein's tertiary structure in the cleavage process. The crystal structure of cleaved RGMB provided us with some clues for factors that contribute to cleavage. For instance, the stabilising π-stacking interaction between the Phe and His residues in the FGDPH

sequence post-cleavage may promote cleavage. However, without isolating the pre-cleavage VWD domains of the RGMs, we are unable to compare the pre- and post-cleavage structures.

The study of the *in vitro* cleavage of RGMC in this chapter provided us with some insights into the pH-dependence of the reaction. At all the tested pH conditions, cleavage of RGMC did not reach 100 % after 2 h. The enhanced cleavage of RGMC at pH 2.2 – 3.4 may be due to the general acid-lability of Asp-Pro bonds. The acid lability of Asp-Pro has been reported in previous studies involving mass spectrometry of peptides and proteins with Asp-Pro bonds (Piszkiewicz et al., 1970). The strongly acidic condition is unlikely representative of the *in vivo* cleavage of RGMB and RGMC during expression. On the other hand. the observed higher cleavage extent between pH 4.2 – 5.8 compared to pH 7.4 may reflect cleavage of RGMC in the *trans*-Golgi network or secretory vesicles. Previously, Zhang *et al.* also observed that cleavage of purified recombinant RGMC was increased after incubation for 6 h at pH 5.5 compared to pH 8 (Zhang et al., 2005). The *trans*-Golgi network has a reported pH of 6.0 and secretory vesicles have a pH around 5.2 (Paroutis et al., 2004). Although the enhanced cleavage of RGMC under mildly acidic conditions concurs with previous studies on autoproteolysis of VWD domains like in MUC2 (Lidell et al., 2003), it is important to note that the citrate-phosphate buffer used to test RGMC cleavage *in vitro* may not be representative of *in vivo* conditions in the secretory pathway and the extracellular media. More studies will be required to understand the *in vivo* autocatalytic cleavage of RGMC, including its mechanism, where it occurs and time dependence of the reaction.

In the last part of this chapter, I showed that the N-terminal cleavage fragment of RGMC can covalently react with alkoxyamine-PEG_4-biotin in the presence of excess

alkoxyamine-PEG_4-biotin. Apart from confirming the role of autoproteolysis in generation of the anhydride, the reactivity with alkoxyamine-PEG_4-biotin verified that hydrolysis can be outcompeted and supports the use of the anhydride for cross-linking to other nucleophiles including primary amines. However, upon conjugation to a target protein, the 14 kDa N-terminal fragment of RGMC will be retained and this translates into a much larger “scar” post-conjugation (Figure 3.1). It remains to be investigated if the N-terminus of the VWD domains from RGMs can be further truncated whilst retaining cleavage activity. For this reason, the use of the VWD domain is less favourable compared to alternative anhydride-generating domains like FrpC self-processing module (SPM) that leaves only a single aspartate residue “scar” post-conjugation. Moreover, the two cleavage fragments of the VWD domains in the RGMs are held together by disulfide bonds and separation of the fragments would require the addition of reducing agents. The addition of reducing agents may be incompatible with target proteins containing pre-existing disulfide bonds.

It was an interesting observation that alkoxyamine-PEG_4-biotin coupled to the N-terminal cleavage fragment of RGMC even at pH 7.4 when 1.25 mM alkoxyamine-PEG_4-biotin was added. It is possible that further cleavage of RGMC occurs even at pH 7.4, albeit at a slower rate. Conjugation of alkoxyamine-PEG_4-biotin to the N-terminal cleavage fragment may have driven the cleavage reaction.

All in all, the difficulty in expressing and isolating the pre-cleavage forms, as well as the unclear pH-dependence of the cleavage reaction suggest that the VWD domains of the RGMs are less promising to engineer as a “pH-inducible” covalent probes.

Chapter 4: Understanding anhydride reactivity of FrpC SPM

Covalent coupling to endogenous proteins is useful for various applications including enhancing therapeutic efficacy, diagnostic sensitivity and for developing novel drug delivery systems (Han et al., 2018; Jain et al., 2013; Oh et al., 2006). Most current approaches utilise either tags on both protein partners e.g., SpyTag/SpyCatcher (Banerjee and Howarth, 2018) and split inteins (Wood and Camarero, 2014), bio-orthogonal click chemistry (Shieh and Bertozzi, 2014) or proximity-directed ligation (Chmura et al., 2001). However, the utility of proximity-directed ligation for coupling to unmodified proteins is often limited to targets with deep "druggable" pockets. The deep pocket allows ligand binding to bring the reactive group close to nucleophiles on the target (Zhang et al., 2019). Alternatively, an electrophile can be introduced to a specific site of a protein by chemical attachment or genetically-encoded unnatural amino acids (Holm et al., 2009; Xiang et al., 2013). The electrophile reacts with known nucleophilic sites on binding partners. Currently, there is no general method for coupling to endogenous proteins using the standard genetic code and without requiring complex synthesis of labelling reagents.

The concept of using the autoproteolytic FrpC self-processing module (SPM) for covalent coupling first began with the work of Felix Metzner, previously working in the Howarth Laboratory. He verified the retention of autocleavage activity when SPM was genetically fused to the C-terminus of various proteins and showed that SPM could be used for intramolecular ligation.

The project was then taken up by Arne Scheu in the Howarth Laboratory who first showed intermolecular coupling using SPM. He tested intermolecular ligation using two protein complexes isolated from a database search. The database search was performed to filter protein complexes from the protein database bank (PDB) based

on 3 predicted requirements for efficient coupling: 1) the C-terminus of the binder protein must be near the binding interface 2) a lysine residue on the target protein must be accessible and near to the cleavage site 3) there must be no other nucleophiles in close proximity on the binder protein itself to compete for reaction (Scheu et al., 2021). This led to identification of the first protein complex, Ornithine Decarboxylase Antizyme (OAZ)/ Ornithine Decarboxylase (ODC) that was used for proof-of-concept of 'NeissLock'. 'NeissLock' is defined as the approach for covalent coupling to endogenous proteins by directing a recombinant protein with an anhydride generated by autoproteolysis to nucleophiles on the target.

In this chapter, I describe experiments that I have undertaken to further understand and develop NeissLock. I characterise the anhydride reactivity and illustrate lessons that can be derived from the OAZ/ODC model system. I show that the system is specific in coupling to the target protein and potentially more broadly applicable beyond complexes isolated from the initial database search. A better understanding of the NeissLock technology using OAZ/ODC as a model system aids future design of NeissLock probes and optimisation of the reaction. Finally, I demonstrate specific coupling of transforming growth factor alpha (TGFα) to epidermal growth factor receptor (EGFR) on live cells and I show that covalent coupling potentially induces changes in downstream signalling.

4.1. NeissLock: inducible protein anhydride for covalent targeting of endogenous proteins

FrpC is a protein secreted by *Neisseria meningitidis* (Thompson and Sparling, 1993) with a self-processing module (SPM) (Osicka et al., 2004). In the presence of calcium, the SPM undergoes conformational change, inducing cleavage at an Asp-

Pro bond (Osicka et al., 2004) (Figure 4.1). The cleavage mechanism involves formation of an aspartic anhydride intermediate at the C-terminus of the N-terminal cleavage fragment i.e., residues 1-414 (Osicka et al., 2004; Sadilkova et al., 2008) (Figure 4.1).

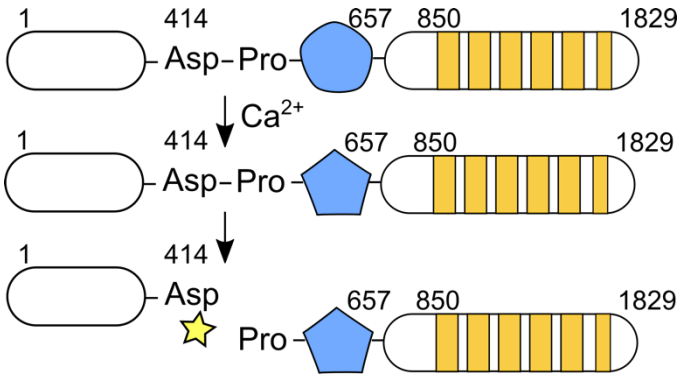


Figure 4.1 FrpC calcium-induced autoproteolysis. FrpC consists of a self-processing module (SPM) (blue) between residues 414-657 and repeats-in-toxin (RTX) domains (orange), each composed of a nonapeptide repeat, in the C-terminus. Calcium addition induces conformational change in the SPM. Asp-Pro bond between D414 and P415 is cleaved, and an aspartic anhydride is generated (yellow star).

Previous studies found that the SPM could be genetically fused to the C-terminus of different proteins whilst retaining its autoproteolytic activity (Liu et al., 2014; Sadilkova et al., 2008). The anhydride intermediate, as a strong electrophile, can participate in intermolecular covalent conjugation. In *Neisseria* infection, the anhydride was proposed to be utilised for conjugation of the FrpC proximal cleavage fragment to host cell proteins (Sviridova et al., 2017). The N-terminus of FrpC was also found to bind FrpD, another *Neisseria* outer-membrane lipoprotein (Sviridova et al., 2017). Some non-specific crosslinking to other FrpC molecules due to anhydride reactivity were previously observed (Osicka et al., 2004; Sadilkova et al., 2008).

The NeissLock approach was devised based on the idea of redirecting the anhydride reactivity to target endogenous proteins. In NeissLock, the binder protein is

a recombinant protein with a C-terminal SPM that undergoes autoproteolysis in the presence of calcium. The anhydride is then directed to available nucleophiles on the protein to which the binder protein docks to ("target protein") (Figure 4.2). This method of coupling has the advantage of leaving a minimal scar of a single aspartate residue in the final conjugated product (Figure 4.2).

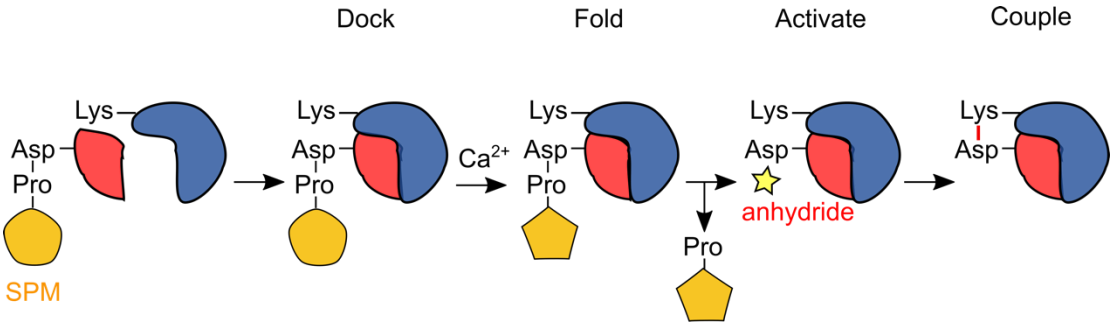


Figure 4.2 NeissLock vision. The recombinant binder protein (red) with a genetically-fused C-terminal SPM docks with the target protein (blue). Calcium addition leads to conformational change of SPM and formation of the anhydride. The anhydride reacts with proximal nucleophiles on the target protein to form a covalent bond.

The NeissLock approach for intermolecular ligation was first demonstrated using the OAZ/ODC where SPM was fused to the C-terminus of OAZ. OAZ was truncated to E219, the last resolved residue of OAZ in the crystal structure (PDB 4ZGY). Truncations of OAZ were previously reported to have little effect on the binding affinity of OAZ to ODC (Hsieh et al., 2011). Arne Scheu from the Howarth laboratory had promising preliminary data on effective coupling. I first verified that I could achieve the same effective coupling in OAZ/ODC.

When I incubated OAZ-Y-SPM in the presence of high calcium concentration, there was resulting formation of the cleavage product "SPM" (Figure 4.3). Additionally, the OAZ N-terminal cleavage fragment that contained the anhydride either underwent hydrolysis to give "OAZ linear" or self-reaction with a nucleophile on itself to give "OAZ cyclized" (Figure 4.3). The formation of the higher molecular weight crosslinked

product “ODC:OAZ-Y” was observed when OAZ-Y-SPM was incubated with ODC in a 1:1 ratio and in the presence of high calcium concentration (Figure 4.3). Previously, Arne Scheu in the Howarth laboratory verified by intact mass spectrometry the identities of the various species post-reaction, including the “ODC:OAZ-Y” conjugate, “OAZ linear” and “OAZ cyclized” (Scheu et al., 2021).

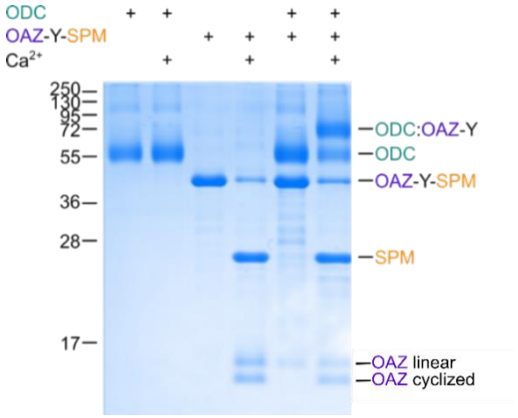


Figure 4.3 Coupling of OAZ to ODC using NeissLock. ODC and OAZ-Y-SPM were incubated at 10 µM each in reaction buffer consisting of 50 mM HEPES,150 mM sodium chloride, 2 mM TCEP with or without 10 mM CaCl2, pH 7.4 at 37 °C for 16 h, boiled in 1×SDS loading buffer with 15 mM EDTA, and analysed by Coomassie-stained SDS-PAGE.

For further characterisation of the NeissLock probe used in the OAZ/ODC model system, I investigated the solution behaviour of OAZ-GSY-SPM using size exclusion chromatography with multi-angle light scattering (SEC-MALS). OAZ-GSY-SPM was chosen over OAZ-Y-SPM as the addition of a -GS linker showed an enhanced crosslinking rate (Scheu et al., 2021). With the help of Dr David Staunton from the Department of Biochemistry Biophysical Instrument Facility who performed the SEC-MALS, I found little deviation between the predicted (42.2 kDa) and observed molecular weight (42.63 $\pm$ 0.4 kDa) for a monomeric OAZ-GSY-SPM, suggesting that the observed crosslinking was not due to dimerization of OAZ-GSY-SPM (Figure 4.4).

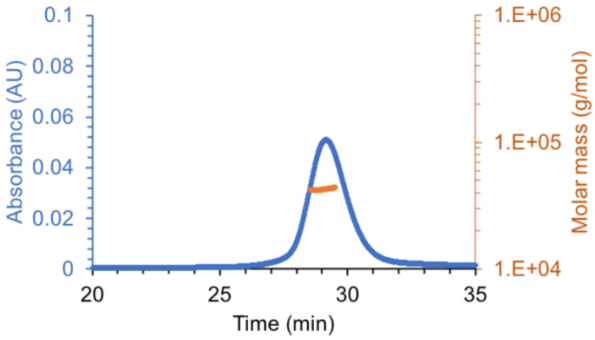


Figure 4.4 SEC-MALS characterisation of OAZ-GSY-SPM. Graph showing elution time of OAZ-GSY-SPM plotted against Absorbance (blue) and MALS signal (orange). OAZ-GSY-SPM was monomeric in solution and had observed M_w of 42.63 $\pm$ 0.4 kDa (Expected M_w: 42.2 kDa).

4.2. Biochemical characterisation of FrpC SPM coupling

4.2.1. pH-dependence of FrpC SPM coupling

To better understand the anhydride reactivity in NeissLock using the OAZ/ODC model system, I first investigated the pH-dependence of the crosslinking reaction. The ε-amino side-chain of lysine has an approximate pK_a of 10.5 in polypeptides (Zhang and Vogel, 1993). This would mean that at physiological pH of about 7.4, most ε-amines would be protonated, rendering them a poor nucleophile.

Nevertheless, it is known that specific protein microenvironments can alter the pK_a of lysine side-chains, making them more nucleophilic and reactive (Wu et al., 2017). Isom et al. reported pKa values as low as 5.3 for some internal Lys residues within staphylococcal nuclease (Isom et al., 2011). Perturbed pKa values and hyper-reactivity were similarly observed for lysines in the active sites of other enzymes (Ho et al., 2009; Ward et al., 2017). Although in NeissLock, the lysine residues targeted by the anhydride are predominantly surface-exposed, we expect that the presence of other ionizable groups close to the lysine and the protein's local dielectric constant will

also influence the lysine's pK_a and protonation state. An acidic microenvironment will favour formation of a positively charged lysine side-chain and therefore the apparent pK_a of the ε-amine would be higher than its intrinsic value (Harms et al., 2009). Conversely, a basic microenvironment will favour the neutral state i.e., a lowered pK_a value relative to its intrinsic value (Fitch et al., 2002). For the NeissLock system, we wanted to ensure its compatibility with physiological conditions to allow *in vivo* crosslinking in future applications.

It was observed that cleavage of SPM was favoured at lower pH and cleavage rate decreased as pH increased (Figure 4.5b). Similarly, the rate of formation of the crosslinked product was highest at pH 7.0 and 6.5 (Figure 4.5a). Crosslinking was slowest at pH 8.5. To obtain an estimate of crosslinking efficiency that is not due to a difference in cleavage rate at different pH, I calculated the ratios of the band intensities of crosslinked product (OAZ:ODC) to cleavage product (P-SPM) at the 60 min time point under different pH conditions. I observed little difference in the ratios of crosslinking to cleavage at different pH values (Figure 4.5c).

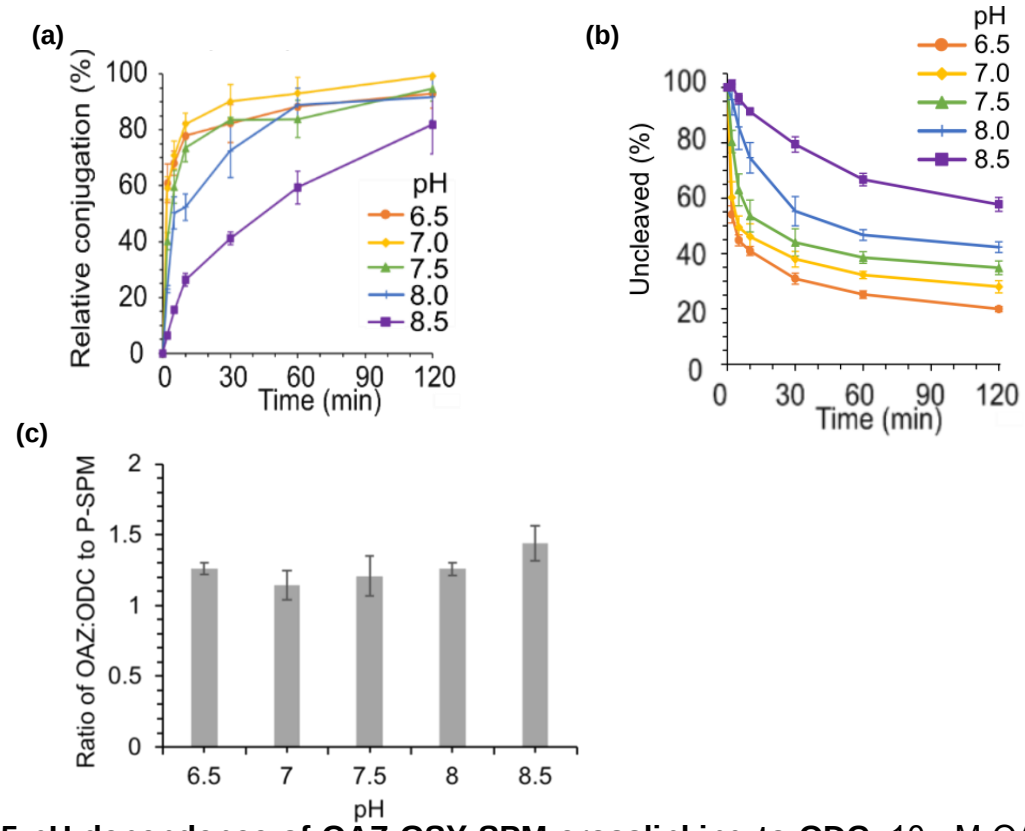


Figure 4.5 pH-dependence of OAZ-GSY-SPM crosslinking to ODC. 10 µM OAZ-GSY-SPM was incubated with 10 µM ODC in reaction buffer with 10 mM $CaCl_2$ and additional 50 mM MES, at 37 °C. Buffers spanned pH 6.5 to 8.5. At the specified time points, the reaction was stopped by boiling in 1×SDS loading buffer with 15 mM EDTA. Samples were run on SDS-PAGE and Coomassie-stained for densitometry analysis giving (a) Time-course for relative conjugation at different pH (mean ± 1 S.D., n=3) (b) Fraction of uncleaved OAZ-GSY-SPM at different pH (mean ± 1 S.D., n=3). (c) Ratio of crosslinked (OAZ:ODC) to cleavage product (P-SPM) after incubation at 37 °C for 60 min (mean ± 1 S.D., n=3).

Results from these experiments show that the NeissLock is compatible with the physiological pH of extracellular fluid (~7.4). Also, surface-exposed lysines can be sufficiently nucleophilic even at pH 7.5 for crosslinking with the anhydride generated from the SPM.

4.2.2. Calcium-dependence of FrpC SPM coupling

From previous literature, the cleavage rate of FrpC is enhanced at high concentrations of free calcium, with a minimum of 0.1 mM Ca^{2+} required to induce cleavage of FrpC (Osicka et al., 2004). With other SPM-fusion proteins, it was shown that 2 mM Ca^{2+} was sufficient to induce rapid cleavage, with a $t_{\frac{1}{2}}$ of approximately 5 min and maximum cleavage yield of 80% in 2 h (Sadilkova et al., 2008).

I investigated the calcium-dependence of NeissLock coupling by measuring the conjugation of OAZ to ODC after incubating the 2 proteins with between 0.5 to 10 mM Ca^{2+} in HEPES-based reaction buffer at pH 7.4. I observed that higher calcium concentrations led to faster conjugation and an overall higher conjugation extent after 2 h but 0.5 mM Ca^{2+} was sufficient for efficient conjugation (Figure 4.6). After 2 h, the conjugation yield with 0.5 mM Ca^{2+} was 78% that of the conjugation yield at 10 mM Ca^{2+} (Figure 4.6).

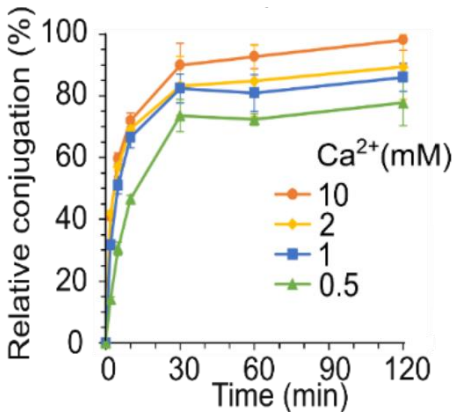


Figure 4.6 Calcium-dependence of OAZ-GSY-SPM crosslinking to ODC. 10 µM OAZ-GSY-SPM was incubated with 10 µM ODC in reaction buffer with 0.5, 1, 2 or 10 mM $CaCl_2$ at pH 7.4, 37 °C. At the specified time points, the reaction was stopped by boiling in 1×SDS loading buffer with 15 mM EDTA. Samples were run on SDS-PAGE and Coomassie-stained for densitometry analysis giving the time-course for the relative formation of crosslinked products at different calcium concentrations (mean ± 1 S.D., n=3).

Intracellular Ca^{2+} concentration in the cytosol under resting conditions is maintained at about 100 nM while extracellular Ca^{2+} concentration is at about 1 mM (Bagur and Hajnóczky, 2017). This suggests that physiological extracellular calcium concentrations would be sufficient to induce NeissLock coupling.

4.2.3. Buffer-dependence of FrpC SPM coupling

Thus far, we have been using HEPES-buffered saline (HBS) as the reaction buffer for Neisslock. The insolubility of calcium phosphate renders phosphate-buffered saline (PBS) unsuitable for NeissLock. I decided to test NeissLock's compatibility with Tris-buffered saline (TBS), another commonly used buffer. Comparing the amount of crosslinked product formed in HBS and TBS at pH 7.4 and 37 °C, I found that the conjugation rates and yields were similar (Figure 4.7). After 2 h, the conjugation yield in TBS was 95% that in HBS (Figure 4.7).

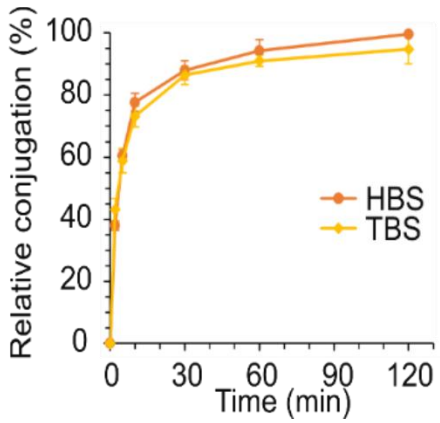


Figure 4.7 Buffer-dependence of OAZ-GSY-SPM crosslinking to ODC. 10 μM OAZ-GSY-SPM was incubated with 10 μM ODC in HEPES or Tris-buffered saline reaction buffer with 10 mM $CaCl_2$ at pH 7.4, 37 °C. At the specified time points, the reaction was stopped by boiling in 1×SDS loading buffer with 15 mM EDTA. Samples were run on SDS-PAGE and Coomassie-stained for densitometry analysis giving the time-course for the relative formation of crosslinked products in the different buffers (mean ± 1 S.D., n=3).

4.2.4. Effect of lyophilisation on FrpC SPM coupling

For more long-term storage of proteins, lyophilisation is used in some cases. I tested if lyophilisation impacts NeissLock by measuring the conjugation to ODC post-lyophilisation and reconstitution of OAZ-GSY-SPM. The conjugation yield of the lyophilised sample after 2 h (51 ± 3% yield, n = 3) was similar to the non-lyophilised sample (48 ± 3% yield, n = 3) (Figure 4.8). This suggests that even with lyophilisation, the structure of SPM was maintained, leading to a similar coupling reaction rate after reconstitution and calcium activation (Figure 4.8).

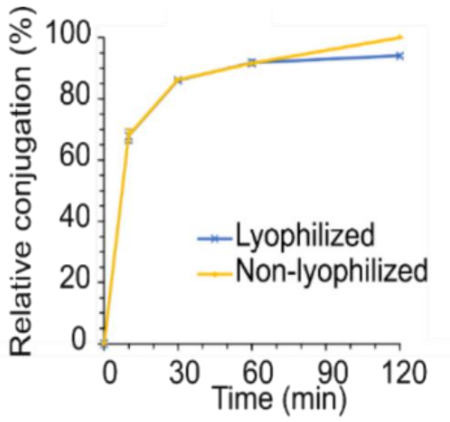


Figure 4.8 Effect of lyophilisation. 10 µM ODC was incubated with 10 µM lyophilized or non-lyophilized OAZ-GSY-SPM in reaction buffer with 10 mM $CaCl_2$ at pH 7.4, 37 °C. At the specified time points, the reaction was stopped by boiling in 1×SDS loading buffer with 15 mM EDTA. Samples were run on SDS-PAGE and Coomassie-stained for densitometry analysis giving the time-course for relative conjugation (mean ± 1 S.D., n=3).

4.3. Effect of C-terminal truncation on FrpC SPM cleavage rate

When the autoproteolytic activity of FrpC from *Neisseria meningitidis* was first reported, the self-processing module was found to be located within positions 414-657

of FrpC (Osicka et al., 2004). This was determined from a series of deletion mutants. A couple of other studies (Matyska Liskova et al., 2016; Sadilkova et al., 2008; Sviridova et al., 2017) following the initial study used a shorter version of the SPM, specifically, positions 414 to 591. This shorter variant is referred as “SPM591” here.

Multiple sequence alignment of FrpC SPM using Jpred showed little sequence conservation and occupancy amongst SPM homologues from position 584 onwards, which may be the rationale for the truncation to SPM591. To verify the domain boundaries of SPM for optimal autoproteolytic activity, I generated 2 other intermediate length variants of SPM based on secondary structure prediction. SPM613 and SPM635 were fused to OAZ with the same -GSY linker. “SPM635” spans positions 414-635 and excludes the α-helix at the C-terminus of SPM at positions 638-653 predicted by Jpred (Drozdetskiy et al., 2015) (Figure 4.9a). “SPM613” spans positions 414-613 of FrpC and excludes 2 further α-helices predicted at positions 616-621 and 624-634 respectively (Figure 4.9a).

After expression of OAZ-GSY-SPM with length variants of SPM, I observed that truncation of SPM led to decreased purity and yields compared to full-length OAZ-GSY-SPM (Figure 4.9b). The same purification protocol was used for all the SPM variants. The reduced purity observed for OAZ-GSY-SPM591, OAZ-GSY-SPM613 and OAZ-GSY-SPM635 may be due to exposure of hydrophobic residues that are normally buried in the core of the protein domain in the full-length version. The yields for OAZ-fusion proteins containing the truncated SPM variants ranged from 1-5 mg per litre of *E. coli* culture. The protein yields for truncated SPM variants were reduced as compared to the 7 mg of purified protein obtained per litre of *E. coli* culture for full-length SPM.

I found that truncation of SPM to the intermediate length variants led to no change in the cleavage rate of OAZ-GSY-SPM while the shortest variant had a slightly reduced cleavage rate (Figure 4.9c). The reduced purity and yield of the truncated variants (Figure 4.9b) as well as slightly slower cleavage rate of the shortest variant (Figure 4.9c) led to the decision to retain the original domain boundaries i.e., 414-657 of FrpC for use in further experiments.

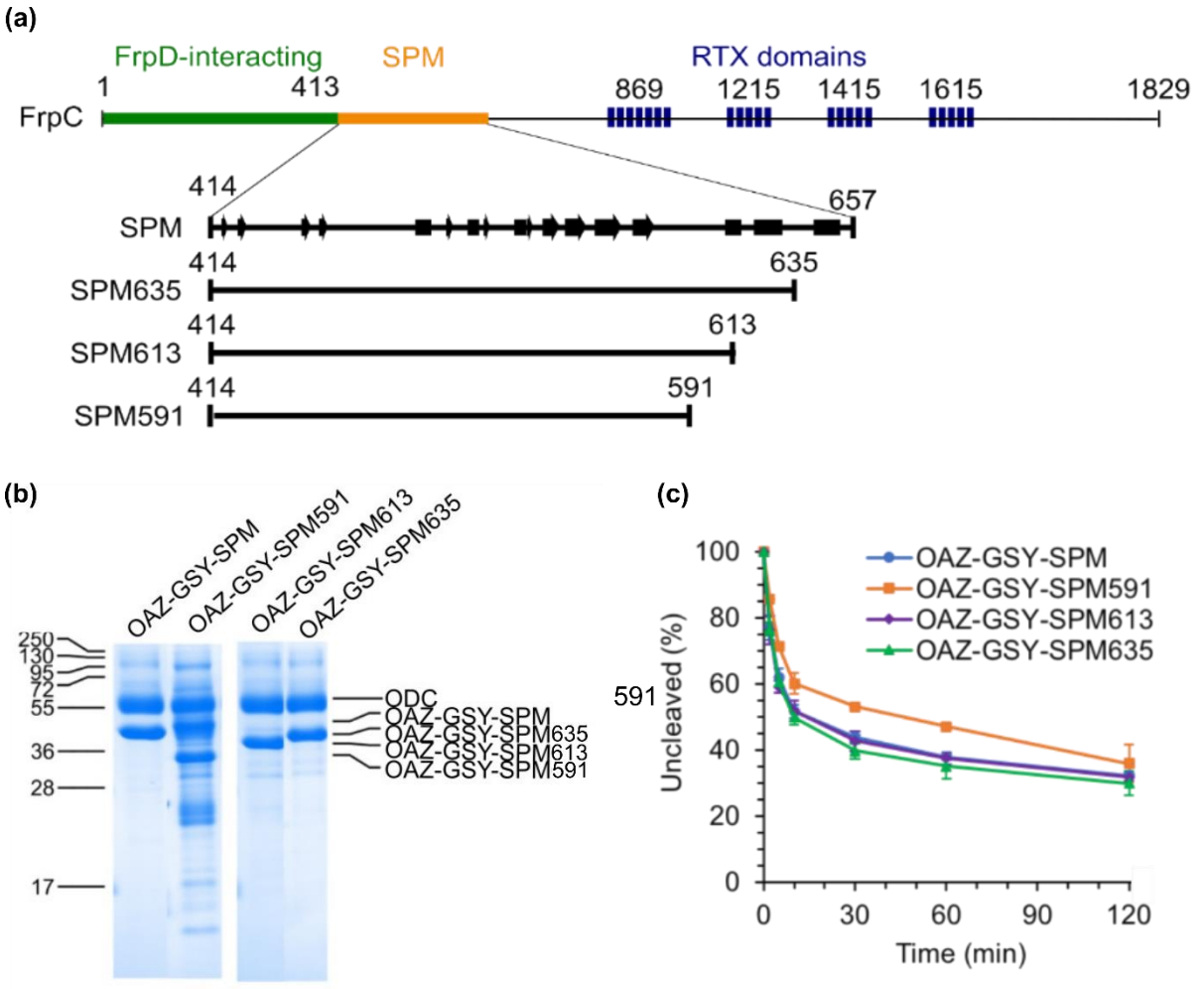


Figure 4.9 Effect of SPM truncation on cleavage activity. (a) Schematic of wild-type and truncated variants of SPM with residue positions in FrpC indicated. α-helices are shown as black rectangles and β-strands as black arrows. Secondary structure is according to PDB 6SJW for residues 1-591 or Jpred prediction for residues 591-657. (b) Coomassie-stained SDS-PAGE gel of 10 μM ODC with 10 μM OAZ-GSY-SPM (wild-type and truncated SPM variants) showing the difference in purity. (c) 10 μM of OAZ-GSY-SPM was incubated with 10 μM ODC at 37 °C in reaction buffer with 10

mM $CaCl_2$, pH 7.4. Reaction was stopped by boiling in 1×SDS loading buffer with 15 mM EDTA. The cleavage of OAZ-GSY-SPM was visualised by Coomassie-stained SDS-PAGE and the band intensity of OAZ-GSY-SPM was quantified (mean ± 1 S.D., n=3).

4.4. Improving OAZ-GSY-SPM stability

During purification of OAZ-SPM, I often observed protein aggregation during dialysis of the protein. Addition of a reducing agent, as recommended in previous published studies (Hsieh et al., 2011), improved OAZ's stability but did not completely inhibit protein precipitation.

We hypothesised that the surface-exposed cysteine at position 175 of OAZ, illustrated in Figure 4.10b, led to the formation of non-specific disulfide bonds, causing protein aggregation. As such, we decided to mutate the cysteine at position 175 to alanine. The mutation to alanine retains the physicochemical properties of cysteine to aid protein folding while preventing disulfide bond formation.

OAZ-GSY-SPM and OAZ(C175A)-GSY-SPM were expressed and purified by Ni-NTA using the same protocol as before. The C175A mutation did not lead to a difference in the yield of OAZ-GSY-SPM. The eluted fractions of the proteins were then dialysed into reaction buffer at pH 7.4, either with or without 2 mM TCEP. The amount of aggregated protein formed post-dialysis was measured by comparing the size of the pellet obtained after centrifuging the post-dialysis protein at 16, 000 g for 30 min. OAZ(C175A)-GSY-SPM that was dialysed into HEPES-buffered saline containing the reducing agent TCEP showed little to no aggregation. Without the C175A mutation or the addition of TCEP, more protein aggregates were observed.

I also compared the cleavage activity of OAZ(C175A)-GSY-SPM to wild-type OAZ-GSY-SPM. Apart from improving the stability of the fusion protein, the C175A

mutation in OAZ led to a slight increase in cleavage rate (Figure 4.10a). The stability of the fusion protein is possibly an important factor influencing SPM cleavage activity. Therefore, when expressing SPM-fusion proteins, it is advantageous to ensure maximal stability of the recombinant protein for optimal cleavage activity.

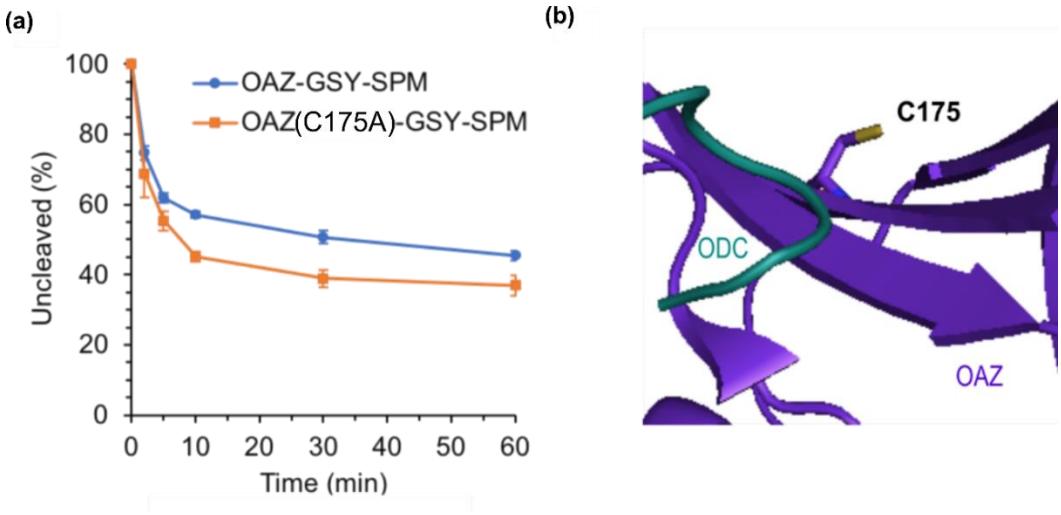


Figure 4.10 C175A mutation to improved OAZ-GSY-SPM stability. (a) OAZ(wild-type)-GSY-SPM or OAZ(C175A)-GSY-SPM at 10 μM each was incubated with 10 mM $CaCl_2$ at 37 °C in reaction buffer, pH 7.4. Reaction was stopped by boiling in 1×SDS loading buffer with 15 mM EDTA. The cleavage of OAZ-GSY-SPM was visualised by Coomassie-stained SDS-PAGE and the band intensity of OAZ-GSY-SPM was quantified (mean ± 1 S.D., n=3). (b) Cartoon of OAZ (purple) and ODC (blue) (PDB 4ZGY). Position of surface-exposed C175 on OAZ indicated as stick residue (yellow).

4.5. Specificity of FrpC SPM coupling

Another crucial question is the specificity of coupling using NeissLock. We needed to ensure that coupling was dependent on the docking of the binder protein to the target e.g., the binding of OAZ to ODC. Previously, the formation of higher molecular weight crosslinked products was observed upon cleavage of SPM in FrpC and in other recombinant proteins (Osicka et al., 2004; Sadilkova et al., 2008). This was attributed to non-specific crosslinking by the anhydride to nucleophilic residues of

other FrpC or SPM-fusion proteins. In the case of using the SPM as a self-excising purification tag, the non-specific crosslinking was inhibited by the addition of thiol compounds such as dithiothreitol (DTT) which react with the anhydride before hydrolysis of the resulting thioester bond (Liu et al., 2014; Sadilkova et al., 2008).

In the case of NeissLock, we hypothesised that the docking of the binding protein to the target protein would direct the anhydride reactivity towards the lysine residue(s) on the target protein in closest proximity to the anhydride. This should reduce the non-specific coupling to irrelevant proteins.

I observed little non-specific crosslinked product formed when OAZ-GSY-SPM was incubated with DogTag-MBP, apart from a faint band that was about 1.4% that of the crosslinked product formed between OAZ and its cognate binding partner ODC (Figure 4.11 lane 6). When OAZ-GSY-SPM was incubated with SpyTag003-sfGFP, there was no detectable crosslinked product formed (Figure 4.11 lane 8). From the examples of these two irrelevant proteins, it suggests there is little to no off-target crosslinking using NeissLock.

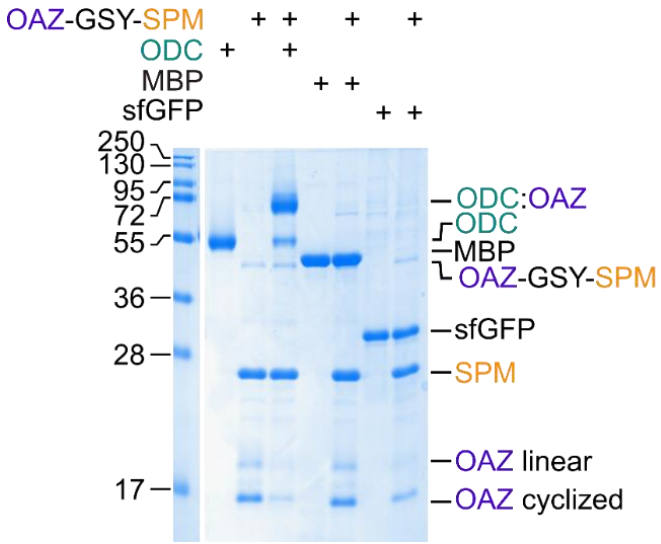


Figure 4.11 Specificity of NeissLock coupling. 10 μM of OAZ-GSY-SPM was incubated in reaction buffer, pH 7.4 with 10 μM ODC or AviTag-DogTag-MBP or SpyTag003-sfGFP. The reaction was incubated at 37 °C for 16 h before stopping by

the addition of 1×SDS loading buffer and 15 mM EDTA followed by boiling. Samples were then run on SDS-PAGE and Coomassie-stained for detection of crosslinked products.

To test the performance of NeissLock in a complex environment with more non-binding proteins, I repeated the test of OAZ-GSY-SPM coupling in mammalian cell lysate, with and without equimolar of its cognate partner ODC doped into the mixture. I performed an anti-His Western blot to identify conjugation partners and found that OAZ conjugated to ODC but did not conjugate to other proteins present in the cell lysate (Figure 4.12a). This suggests that NeissLock coupling retains its specificity even in crowded environments.

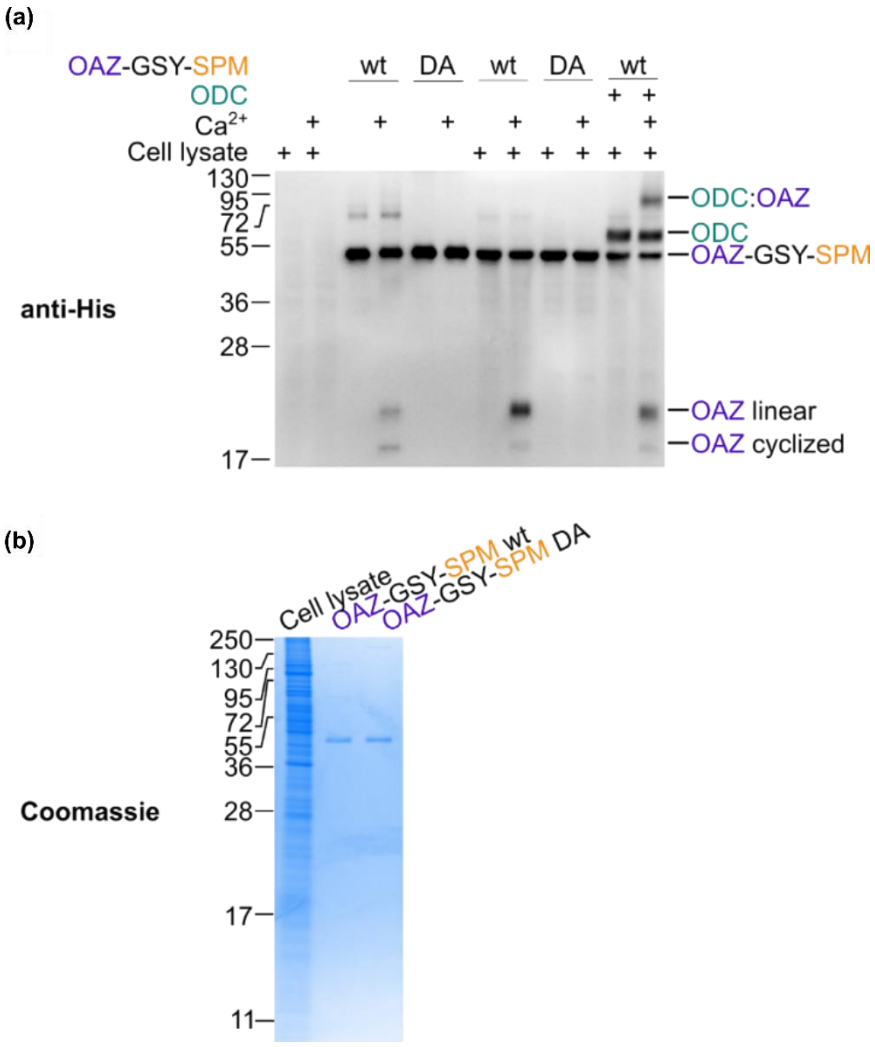


Figure 4.12 Specificity of NeissLock coupling in cell lysate. (a) 1 µM OAZ-GSY-SPM(wild-type or D414A non-reactive mutant) with or without 1 µM ODC was added to A431 cell lysate ± 2 mM Ca^{2+}. The reaction was carried out at 25 °C for 10 min. Samples were then run on SDS-PAGE and blotted against anti-His to detect the N-terminal His_6-tags in OAZ and ODC. (b) Coomassie-stained SDS-PAGE of cell lysate alone, 1 µM OAZ(wild-type)-GSY-SPM or 1 µM OAZ(D414A)-GSY-SPM like in (a).

As a separate negative control, I verified that the crosslinking of OAZ to ODC was dependent on the calcium-induced cleavage of the Asp-Pro bond in SPM by generation of the non-reactive D414A mutant variant of SPM. In FrpC D414A, the Asp-Pro bond was no longer susceptible to cleavage (Osicka et al., 2004).

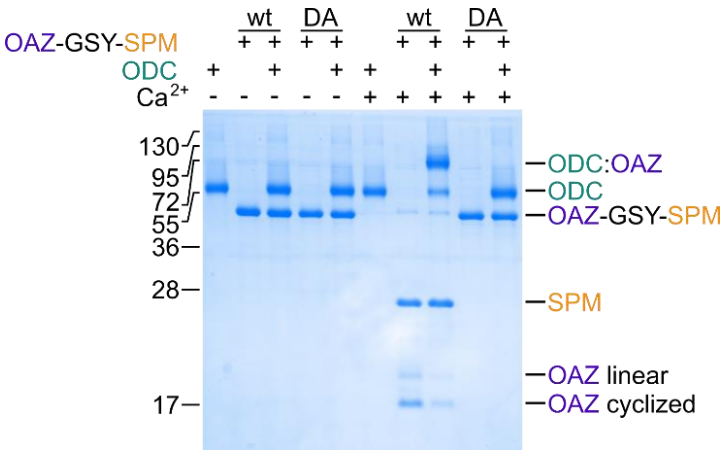


Figure 4.13 Dependence of coupling on cleavage of the D-P bond. OAZ-GSY-SPM(wild-type) or OAZ-GSY-SPM (D414A) was incubated at 10 µM each, with or without 10 µM ODC in reaction buffer ± 10 mM $CaCl_2$, pH 7.4 at 37 °C for 16 h, boiled in 1×SDS loading buffer with 15 mM EDTA, and analysed by Coomassie-stained SDS-PAGE.

As was expected, no cleavage product was observed after 16 h incubation of the D414A variant of SPM in the presence of 10 mM $CaCl_2$ (Figure 4.13). Similarly, when incubated with ODC, cleavage of OAZ-GSY-SPM (D414A) was abolished, and no crosslinked product was detected (Figure 4.13). This confirmed that crosslinking is dependent on Asp-Pro bond cleavage in SPM.

4.6. Effect of binding affinity on FrpC SPM coupling

In the previous section, I have shown that coupling using NeissLock is highly specific. Since coupling is dependent on the initial docking of the binder protein to the target protein i.e., binding of OAZ to ODC, we expect the affinity of the binder protein for the target protein to be a factor that influences the coupling rate and/or extent. We also wanted to better understand the minimum requirements of affinity between the binder protein and target protein for effective coupling.

In order to test the effect of K_d on coupling of OAZ-GSY-SPM to ODC, I generated a series of point mutations in OAZ to disrupt OAZ binding to ODC. These mutation sites on OAZ were identified from a previous study that mapped the binding interface of ODC and OAZ using alanine scanning mutagenesis (Cohavi et al., 2009). OAZ K153A led to a 8.3-fold decrease in binding affinity of OAZ to ODC, the R188A mutation led to a 3.3-fold decrease in binding affinity and the V198A mutation led to a 10-fold decrease in binding affinity (Cohavi et al., 2009).

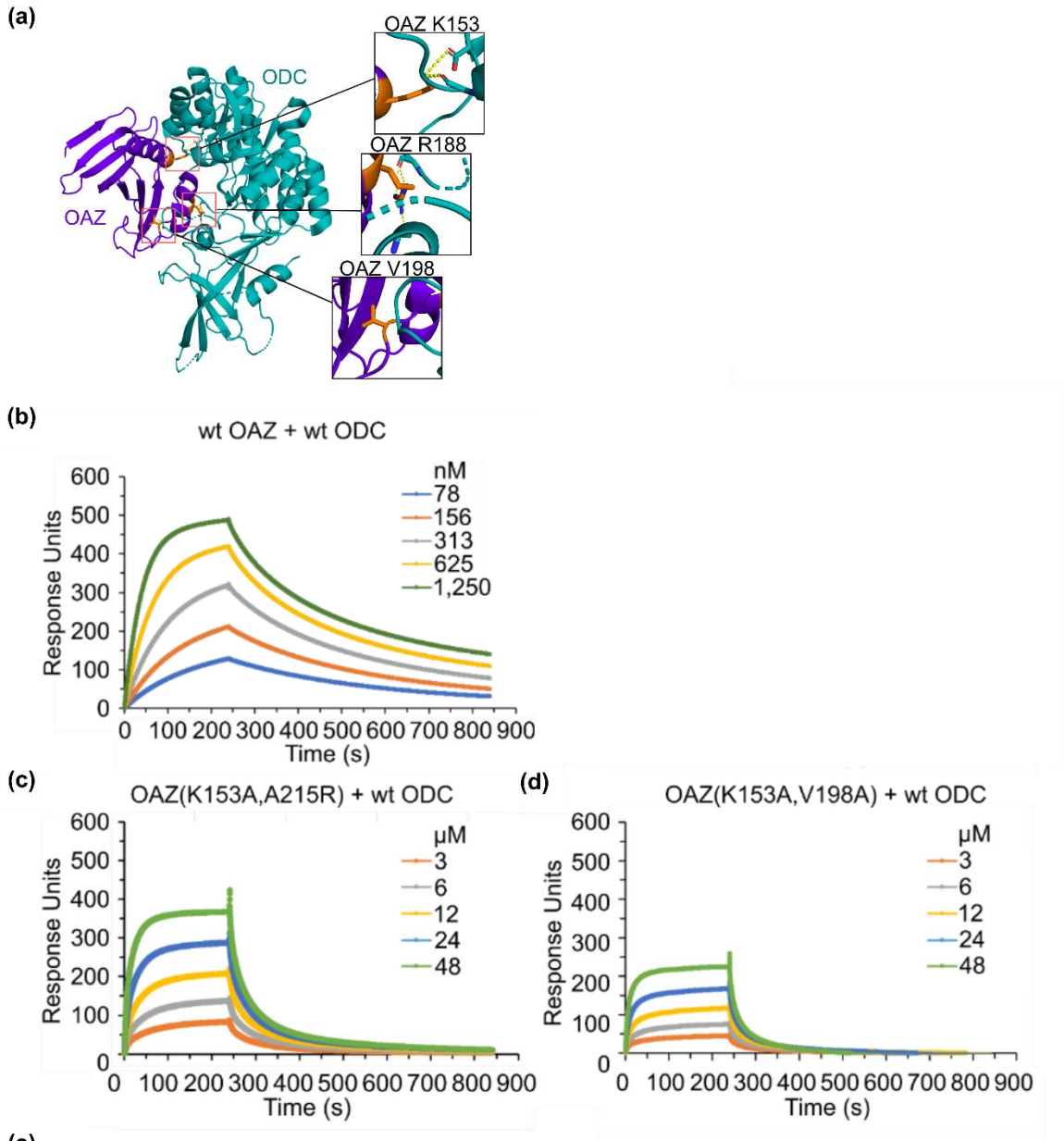
These binding sites are dispersed across the binding interface of OAZ to ODC (Figure 4.14a). To further disrupt the binding interaction, I introduced a reversal of charge at residues 153 and 188 by mutation of both K153 and R188 to glutamate. Based on the crystal structure of the OAZ/ODC complex (PDB 4ZGY), the R188E mutation in OAZ should disrupt the hydrogen bonding to the protein backbone in ODC at positions Y331 and R165 (Figure 4.14a). The K153E mutation in OAZ should disrupt the pre-existing electrostatic interaction with the side-chain of D134 in ODC and the hydrogen bond with the protein backbone at P113 in ODC (Figure 4.14a).

To verify the effect of these rationally-designed mutations on the binding of OAZ to ODC, I used Surface Plasmon Resonance (SPR). I added an AviTag to the N-

terminus of OAZ-GSY-SPM for biotinylation using GST-BirA. This allowed me to immobilise OAZ-GSY-SPM onto the SPR chip using a streptavidin capture reagent. Titrated wild-type ODC analyte was flowed over the SPR chip and measurements were carried out at 25°C.

Using a 1:1 binding model to fit the sensorgram curves, the binding of wild-type OAZ-GSY-SPM to ODC had a K_d of 0.12 μM that was calculated from the ratio of the rate constants. The k_{off} of the interaction of wild-type OAZ-GSY-SPM with ODC was 0.0028 s^{-1} (Figure 4.14e). A faster k_{off} may translate to more potential crosslinking sites on the target protein. The double mutants OAZ(K153A, V198A) and OAZ(K153A, A215R) had a more than 100-fold decrease in binding affinity. OAZ(K153A, V198A) had K_d of 25 μM and OAZ(K153A, A215R) had K_d of 15 μM (Figure 4.14e). Both OAZ double mutants showed approximately a 20-fold decrease in k_{on}. OAZ(K153A, V198A) showed a 12-fold increase in k_{off} while OAZ(K153A, A215R) showed a 6-fold increase in k_{off} (Figure 4.14e). Since V198A mutation disrupted binding to a larger extent than A215R and both mutations are located proximal to each other, I used V198A rather than A215R in the triple mutant. Introducing the charge reversals at positions 153 and 188 in the triple mutant OAZ (K153E, V198A, R188E) further disrupted OAZ/ODC binding, and no interaction was detected by SPR even after titrating higher concentrations of ODC (Figure 4.14e). The individual SPR sensorgrams can be found in Figures 4.14(b-d).

The SPR analysis revealed the tight binding interaction between wild-type OAZ and wild-type ODC even after truncation of OAZ to E219 and genetic fusion of SPM to the C-terminus of OAZ. It also verified that the introduction of mutations to OAZ at the binding interface with ODC disrupted binding and can be used to investigate the effect of K_d change on coupling yield.



(e)

OAZ	ODC	k_{on} ($M^{-1}s^{-1}$)	k_{off} (s^{-1})	K_d (μM)
wt	wt	2.3×10^4	0.0028	0.12
K153A, V198A	wt	1.4×10^3	0.034	25
K153A, A215R	wt	1.1×10^3	0.017	15
K153E, R188E, V198A	wt	undetectable	undetectable	undetectable

Figure 4.14 Generation of OAZ binding mutants. (a) Crystal structure of OAZ (purple) and ODC (blue) with K153, R188 and V198 (orange) and their interacting residues at the binding interface highlighted as stick residues. Polar interactions are illustrated as yellow dashed lines (PDB 4ZGY). (b-d) SPR sensorgrams for wild-type (wt) and mutant variants of OAZ-GSY-SPM binding to ODC. The sensorgrams were

used to fit a 1:1 binding model and K_d in (c) were derived from ratios of the kinetic constants. (e) Table summarising binding affinities of OAZ-GSY-SPM to ODC, including mutant variants of OAZ.

There was reduced crosslinked product formed for the double mutants as compared to wild-type OAZ-GSY-SPM (Figure 4.15). The band intensity of the crosslinked products for the double binding mutants were 39.4 % to 52.7 % that of wild-type OAZ-GSY-SPM after 16 h (Figure 4.15a). As for the triple mutant OAZ-GSY-SPM, crosslinking was 6.89 % that of wild-type and there was barely any observable crosslinked product formed (Figure 4.15a). Lower initial starting concentrations of OAZ-GSY-SPM and ODC were used in the reaction due to the low K_d of wild-type OAZ binding to wild-type ODC i.e., high binding affinity.

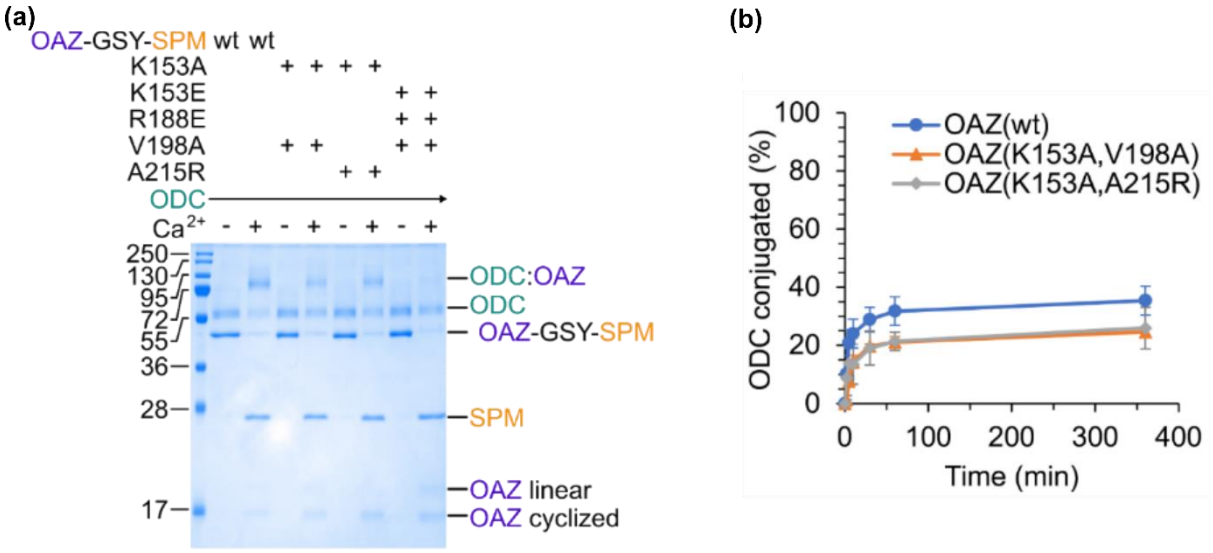


Figure 4.15 Effect of K_d on coupling extent of OAZ to ODC. (a) 0.5 μM OAZ-GSY-SPM (wild-type or mutant variant) was incubated with 0.5 μM ODC in reaction buffer, pH 7.4 at 37 °C for 16 h before boiling in 1× SDS loading buffer with 15 mM EDTA. Samples were analysed by Coomassie-stained SDS-PAGE. (b) Time-course of wild-type or mutant OAZ conjugation to ODC from densitometry analysis of Coomassie-stained SDS-PAGE gels (mean ± 1 S.D., n=3).

The observation of crosslinking occurring even with the double mutants suggests that crosslinking using NeissLock does not require a very tight interaction between the binder protein and target protein. With a K_d of 15 to 25 μM, coupling was still achieved, although the coupling yield was lower than with wild-type OAZ-GSY-SPM. The affinity-dependence was further shown with the triple mutant where if the K_d is sufficiently high i.e., binding interaction is too weak, crosslinking is almost completely abolished.

4.7. Identifying crosslinking sites on ODC

I next sought to identify the nucleophilic sites on ODC that are capable of reaction with the anhydride formed at the C-terminus of OAZ upon SPM cleavage. It is unclear what the distance range of the anhydride reactivity is and the selectivity of lysines to which crosslinking occurs. Although LC-MS/MS data previously generated by Arne Scheu confirmed the primary crosslinking site as K92 on ODC (Scheu et al., 2021), mutation of K92 to arginine did not abolish crosslinking but led to the formation of two distinct crosslinked products (Figure 4.17). One of these two secondary crosslinking sites was identified by LC-MS/MS as K121, which is the next most proximal lysine residue to E219 of OAZ (Figure 4.16). K92 is 3.5 Å away while K121 is 14.2 Å away from the most distal resolved C-terminal residue OAZ (Figure 4.16). The 14.2 Å distance between K121 and the C-terminus of OAZ suggests that the distance requirement for coupling using the anhydride is possibly less stringent. However, it is important to note that an additional three residue flexible -GSY linker was used between SPM and E219, so the anhydride is likely less than 14 Å away from K121.

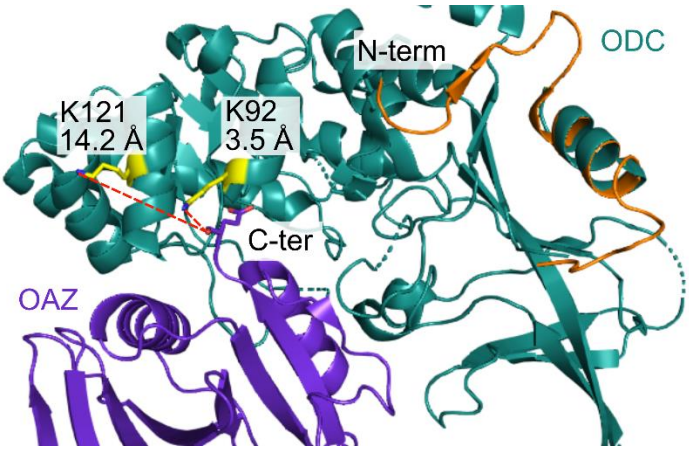


Figure 4.16 Potential crosslinking sites on ODC. K92 and K121 highlighted as stick residues (yellow) are lysines on ODC (blue) in closest proximity to E219 (C-ter) on OAZ (purple). Alignment of N-term of ODC from ODC dimer (PDB 1D7K) (orange) to N-term of ODC in OAZ/ODC complex (PDB 4ZGY) shows an additional β-strand that loops towards OAZ C-terminus.

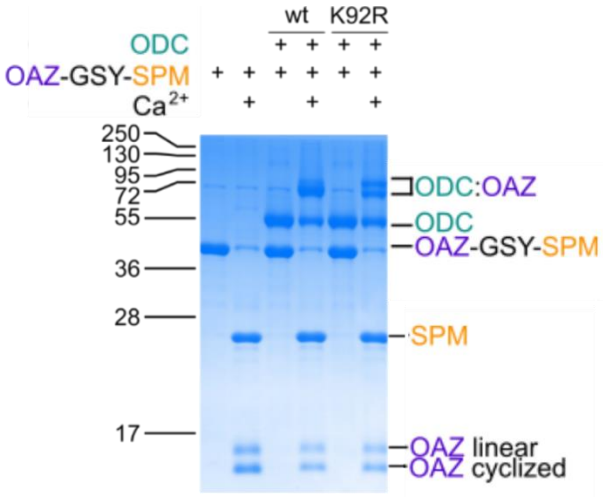


Figure 4.17 Presence of at least two other crosslinking sites apart from K92. 10 μM OAZ-GSY-SPM was incubated with 10 μM ODC (wild-type or K92R) in reaction buffer with 10 mM $CaCl_2$, pH 7.4 at 37 °C for 16 h before boiling in 1×SDS loading buffer with 15 mM EDTA. Samples were analysed by Coomassie-stained SDS-PAGE. Data from Arne Scheu, Howarth laboratory.

Apart from nucleophilic ε-amines, another possible nucleophilic site on ODC is the N-terminal α-amine which typically has a pK_a of 6.8 to 9.1 (Oregioni et al., 2017). The α-amine was initially overlooked as a potential crosslinking site because in the

crystal structure of the OAZ/ODC complex (PD 4ZGY), the most distal resolved N-terminal residue of ODC was 39.4 Å from E219 (Figure 4.16). However, in the crystal structure of the ODC dimer (PDB 1D7K), more residues at the N-terminus of ODC were resolved, revealing an additional β-strand that loops back towards the binding interface between OAZ and ODC (Figure 4.16). This β-strand, the remaining 6 unresolved N-terminal residues and the additional His_6 tag implied that the α-amine of ODC is a lot closer to the anhydride than 39.4 Å and may be susceptible to reaction.

Therefore, I generated a truncated mutant of ODC that had the N-terminal His_6 tag as well as first 9 N-terminal residues, including the six unresolved residues deleted, henceforth referred to as "ΔN". There were no lysines within the truncated 9 N-terminal residues that could react with the anhydride, limiting the only nucleophilic site to the α-amine. I also decided to mutate other proximal ε-amines i.e., K74 and K78. K74 is 21.6 Å from the C-terminus of OAZ while K78 is 28.1 Å from the C-terminus of OAZ. Together with K92R and K121R, these four lysine to arginine mutations are henceforth referred to as "4KR". I then individually re-introduced either the ε-amine at each position or the α-amine of ODC in the "ΔN, 4KR" background ODC and tested for rescue of crosslinking (Figure 4.18a).

Effective coupling of OAZ to wild-type ODC was observed (Figure 4.18a, lane 2). With ΔN and the four lysine to arginine mutations, the amount of crosslinked product formed was reduced to 14% relative to wild-type (Figure 4.18a, lane 4). I verified that the reduced crosslinking was not due to a change in protein conformation or decrease in binding affinity. SPR showed that ΔN 4KR ODC retained tight binding to wild-type OAZ-GSY-SPM (K_d of 0.022 μM for ΔN 4KR ODC compared to 0.12 μM for wild-type ODC).

Reintroduction of the N-terminal residues but retaining the 4 KR mutations led to an observable increase in crosslinked product (Figure 4.18a, lane 6), providing confirmation that the third crosslinking site in ODC is the α-amine. Reintroduction of either K92 or K121 also rescued crosslinking, providing confirmation that the ε-amines of K92 and K121 on ODC are key crosslinking sites (Figure 4.18a). The increased crosslinked product in ΔN,3KR,R92K compared to ΔN,3KR,R121K also suggests a possible proximity-dependence, since K92 is closer to the predicted site of anhydride generation than K121 (Figure 4.16). The observation that crosslinking was not completely abolished for ODC(ΔN, 4KR) suggested that other possible crosslinking sites were present.

Given that the last 40 C-terminal residues of ODC were not resolved in the crystal structure of the OAZ/ODC complex (PD 4ZGY) and there are multiple nucleophilic residues in this region that are potential crosslinking sites, I generated an ODC variant with an additional C-terminal truncation from the last resolved residue i.e., deletion of residues 422-460 of ODC, on top of the "ΔN,4KR" background. This mutant is referred to as "ΔN,ΔC,4KR". The ΔN,ΔC,4KR ODC mutant showed only 7% coupling relative to the wild-type, a decrease from the 15% coupling relative to wild-type observed for the ΔN,4KR mutant (Figure 4.18b, lane 5 and 7). The reduction in coupling suggests that additional crosslinking sites lie in the unresolved C-terminus of ODC. Reintroduction of K92 to ΔN,ΔC,4KR rescued coupling (Figure 4.18b, lane 9), suggesting that the ΔN,ΔC,4KR ODC mutant was still able to bind to OAZ and the reduction in coupling extent was not due to a decrease in binding interaction.

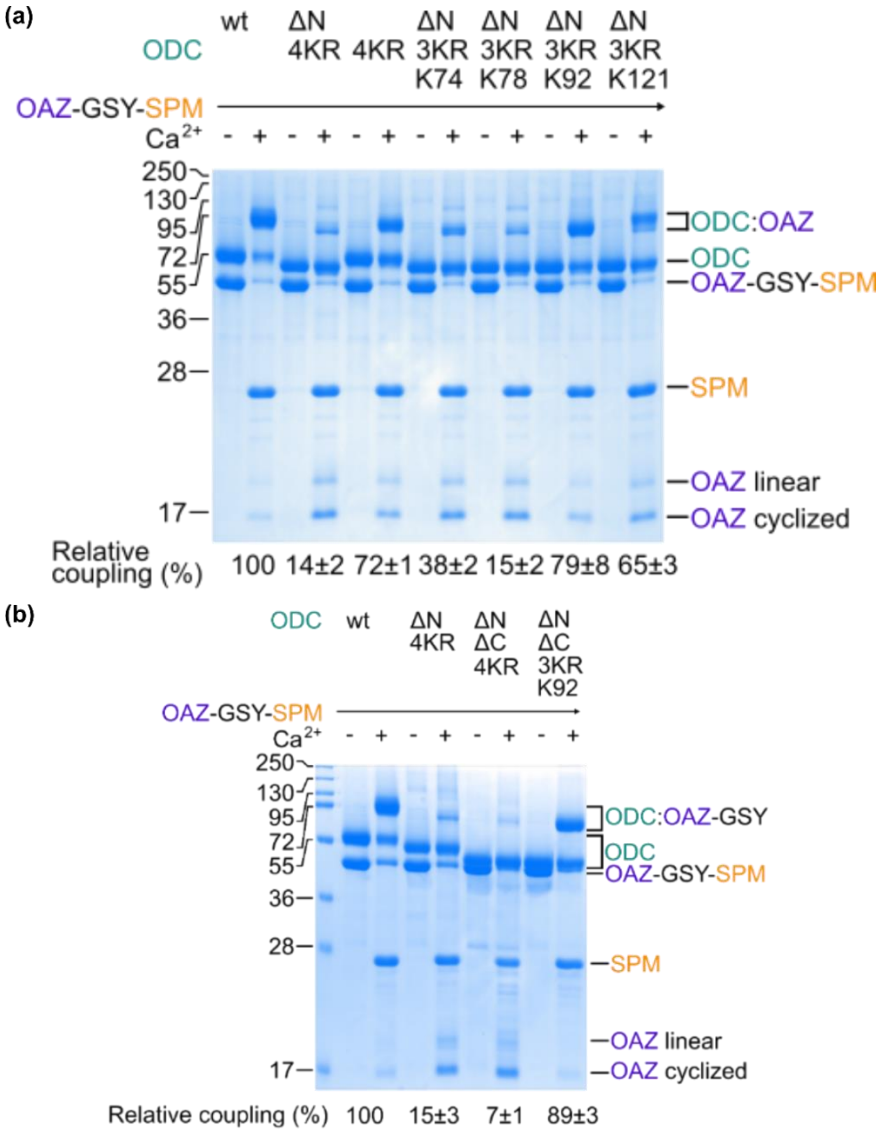


Figure 4.18 Identifying crosslinking sites on ODC. (a) 10 μM OAZ-GSY-SPM was incubated with 10 μM ODC (wild-type or mutant variant as indicated) in reaction buffer with 10 mM $CaCl_2$, pH 7.4 at 37 °C for 16 h before boiling in 1×SDS loading buffer with 15 mM EDTA. Samples were analysed by Coomassie-stained SDS-PAGE (mean ± 1 S.D., n=3). Truncation of the first 9 amino acids and N-terminal His-tag (ΔN), together with K92R, K12R, K74R and K78R (4KR) mutations in ODC reduced conjugation. Re-introduction of the original N-terminus, or K92 or K121 rescued coupling. (b) Same reaction conditions as in (a). Truncation of the C-terminus of ODC on top of ΔN 4 KR mutations (ΔN ΔC 4 KR) further reduced coupling (mean ± 1 S.D., n=3).

4.8. Crosslinking to other nucleophiles

In the OAZ/ODC complex, I observed coupling to α-amine and ε-amines. Multiple nucleophilic residues including cysteine, threonine and serine reside in the C-terminus of ODC where crosslinking sites were shown to lie, suggesting of the potential of the anhydride crosslinking to other residues apart from lysines. To test our hypothesis that NeissLock can be used to couple to other residues apart from lysines, I developed an *in vitro* assay for the anhydride's reactivity with different small molecule nucleophiles that mimic the nucleophilic sides chains of various amino acids (Figure 4.19a). An *in vitro* plate-based assay was chosen over running SDS-PAGE gels to detect for conjugation with the biotinylated nucleophiles because some of the expected bonds to be formed like thioesters and esters may be more labile and susceptible to hydrolysis with heating. In addition, I was uncertain if the small molecular weight of the nucleophiles may render the detection of a gel-shift difficult (Figure 4.19b).

Based on the plate assay, I observed crosslinking between the anhydride and biotin-PEG-SH, biotin-PEG-alkoxyamine, biocytinamide, biotin pentylamine as well as biotin tyramide (in order of decreasing reactivity) (Figure 4.19c). The anhydride's reactivity with biotin-PEG-SH suggests of the possibility of utilising NeissLock for crosslinking to cysteine residues that contain a thiol group as well. The reactivity with biocytinamide and biotin pentylamine concurs with the crosslinking we have observed to the α-amine and ε-amine in ODC. The reduced reactivity with biotin tyramide compared to the other nucleophiles was expected given the lower reactivity of the phenolic hydroxyl group (Szijj et al., 2020). A dose-dependence in reactivities was observed across all the biotinylated nucleophiles tested (Figure 4.19c).

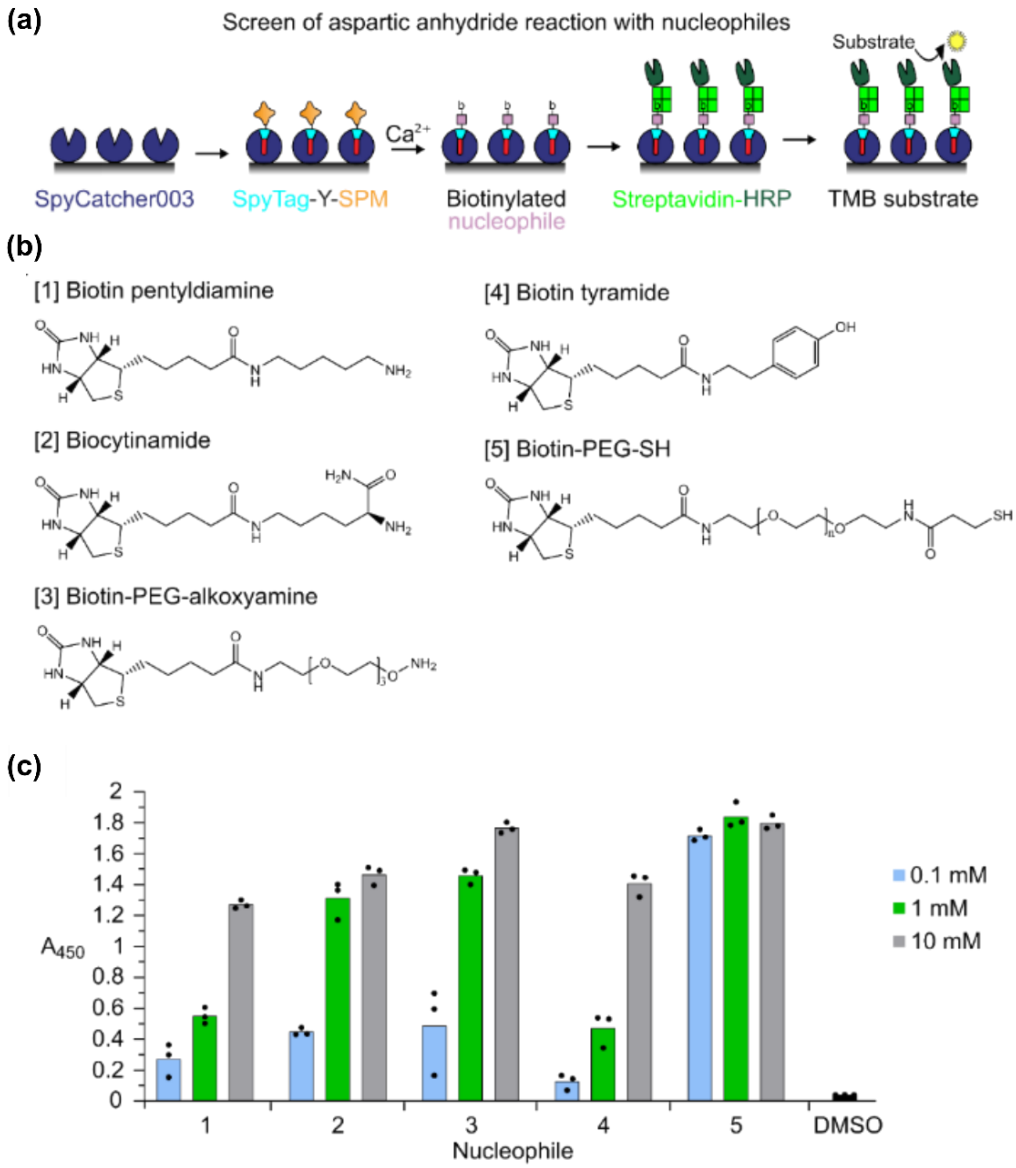


Figure 4.19 Crosslinking to other nucleophiles. (a) Schematic showing plate-based assay for anhydride's reactivity with various nucleophiles. SpyCatcher003 (dark blue) is immobilised on a plate before SpyTag (cyan) covalently reacts with SpyCatcher003, leading to anchorage of SPM (orange). Calcium induces SPM activation and generation of the aspartic anhydride that is available for reaction with the biotinylated nucleophiles (purple). The reacted biotinylated nucleophiles bind streptavidin-HRP (green), enabling detection by addition of a TMB substrate. (b) Structures of biotinylated nucleophiles (c) Aspartic anhydride's reactivity with the biotinylated nucleophiles, each at the indicated concentrations. The biotinylated nucleophiles were incubated in HEPES reaction buffer with 10 mM $CaCl_2$, pH 7.4 at 37 °C for 30 min and conjugation was determined by the absorbance at 450 nm (mean ± 1 S.D., n=3).

4.9. Other complexes

4.9.1. Crosslinking to Antizyme Inhibitor (AZI)

Thus far, beyond the OAZ/ODC complex, the applicability of NeissLock has yet to be demonstrated. I wanted to test if the OAZ-SPM fusion could be used to couple to a different target protein. The second target protein chosen was Antizyme Inhibitor (AZI), another known binding partner of OAZ (Albeck et al., 2008). AZI is an ODC homologue that exhibits a higher binding affinity for OAZ than ODC but is catalytically inactive (Albeck et al., 2008). Like ODC, AZI contains a lysine at position 92 that is close to the C-terminus of OAZ (Figure 4.20a). I hypothesised that K92 in AZI would similarly be a good crosslinking site for the anhydride at the C-terminus of OAZ (Figure 4.20a).

I observed substantial *in vitro* conjugation between OAZ and AZI when the proteins were at 2.5 μM each (Figure 4.20b). After overnight incubation, the overall conjugation yield was 80%, higher than what was observed for ODC (Figure 4.20b). Conjugation was rapid, with more than 40% AZI conjugated in less than 5 min (Figure 4.20c)

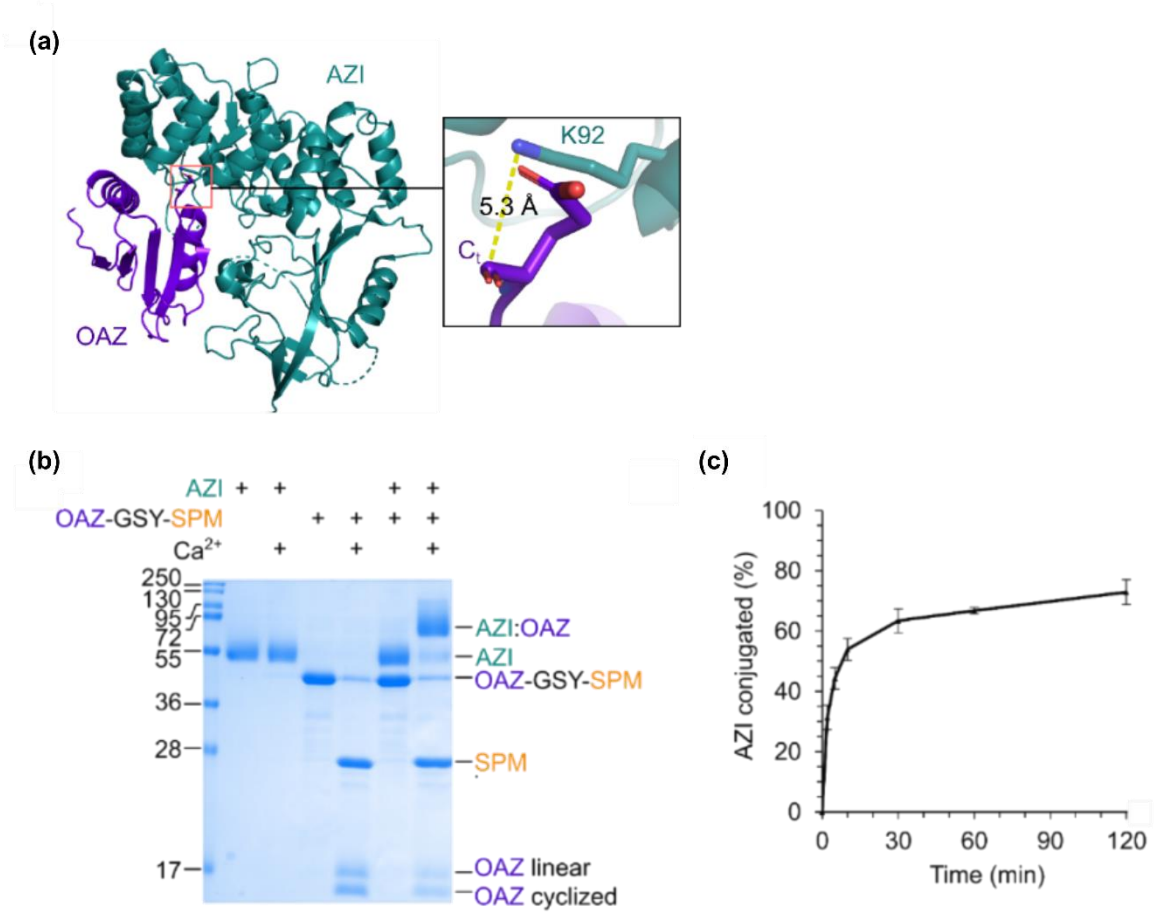


Figure 4.20 Crosslinking to Antizyme Inhibitor (AZI). (a) Crystal structure of OAZ/AZI complex. C-terminus of OAZ (purple) and K92 or AZI (blue) shown as stick residues. (PDB 4ZGZ) (b) 2.5 μM of AZI was incubated with 2.5 μM OAZ-GSY-SPM in HEPES reaction buffer with 10 mM $CaCl_2$, pH 7.4 at 37 °C for 18 h (c) Time-course of AZI conjugation from densitometry analysis of Coomassie-stained SDS-PAGE as in (b) (mean ± 1 S.D., n=3).

4.9.2. Effect of TGFα crosslinking to EGFR on downstream signalling

The third complex we tested NeissLock coupling on was Transforming Growth Factor α (TGFα)/ Epidermal Growth Factor Receptor (EGFR). The C-terminus of TGFα is approximately 5.6 Å from K465 of the ectodomain of EGFR in the TGFα/EGFR complex (PDB 1MOX) and the complex was identified as another promising candidate for NeissLock coupling from the NeissDist database search (Scheu et al., 2021).

SPM was fused to the C-terminus of TGFα with a flexible -GSY linker between the two domains. In order to test NeissLock coupling on live cells, I incubated TGFα-GSY-SPM with A431 cells, a human cell line with upregulated EGFR expression (Kong et al., 2008). Calcium was added to activate SPM and excess unbound TGFα-GSY-SPM was washed off the cells after 10 minutes. After incubation for different time durations, the A431 cells were harvested, and I used Western blotting to detect for conjugation to endogenous EGFR.

10 min post-addition of TGFα-GSY-SPM, excess unreacted TGFα-GSY-SPM was detectable in the A431 cell lysate, in addition to a higher molecular weight band with a M_w consistent with a TGFα:EGFR covalent complex (Figure 4.21a top panel lane 1-3). There were no other bands detected, suggesting that TGFα coupling was specific and there was no coupling to other cell surface proteins. An excess of the strong nucleophile, hydroxylamine, outcompeted EGFR in reacting with the anhydride, leading to no TGFα:EGFR conjugate formed (Figure 4.21a top panel lane 4-6). Similarly, when the D414A mutation was introduced to SPM in TGFα-GSY-SPM, calcium-induced cleavage of SPM was abolished and no coupling to EGFR was observed (Figure 4.21a top panel lane 7-9).

To probe the time-dependence of TGFα coupling to endogenous EGFR, I blotted against TGFα and EGFR at 1 h intervals up to 5 h. The band intensity of the TGFα:EGFR conjugate decreased over time, suggesting of possible degradation of the conjugate in the cell (Figure 4.21b top panel). There was no clear change in the total level of EGFR (Figure 4.21b middle panel). Due to the small molecular weight of TGFα, the TGFα:EGFR conjugate and unreacted EGFR could not be differentiated on the anti-EGFR Western blot, making it difficult to assay the percentage of endogenous EGFR that was coupled to TGFα (Figure 4.21b middle panel).

To study the effect on downstream signalling in A431 cells after covalent coupling of TGFα to EGFR, I decided to assay the changes in levels of downstream signalling effectors. Amongst the different canonical and non-canonical EGFR signalling pathways, the Signal Transducer and Activator of Transcription (STAT) pathway was of interest. It was previously shown that low-affinity ligand interactions with EGFR but not high-affinity ligand interactions activate non-canonical signalling pathways downstream of EGFR like activation of STATs (Krall et al., 2011). From the anti-pSTAT1 Western blot, STAT1 signalling was potentially more sustained when TGFα was coupled to EGFR on A431 cells. Higher levels of phosphorylated STAT1 were observed even at 60 min and 240 min after TGFα conjugation to EGFR, compared to the controls with hydroxylamine addition or the D414A mutant of SPM where there was only non-covalent binding between TGFα and EGFR (Figure 4.21c top panel).

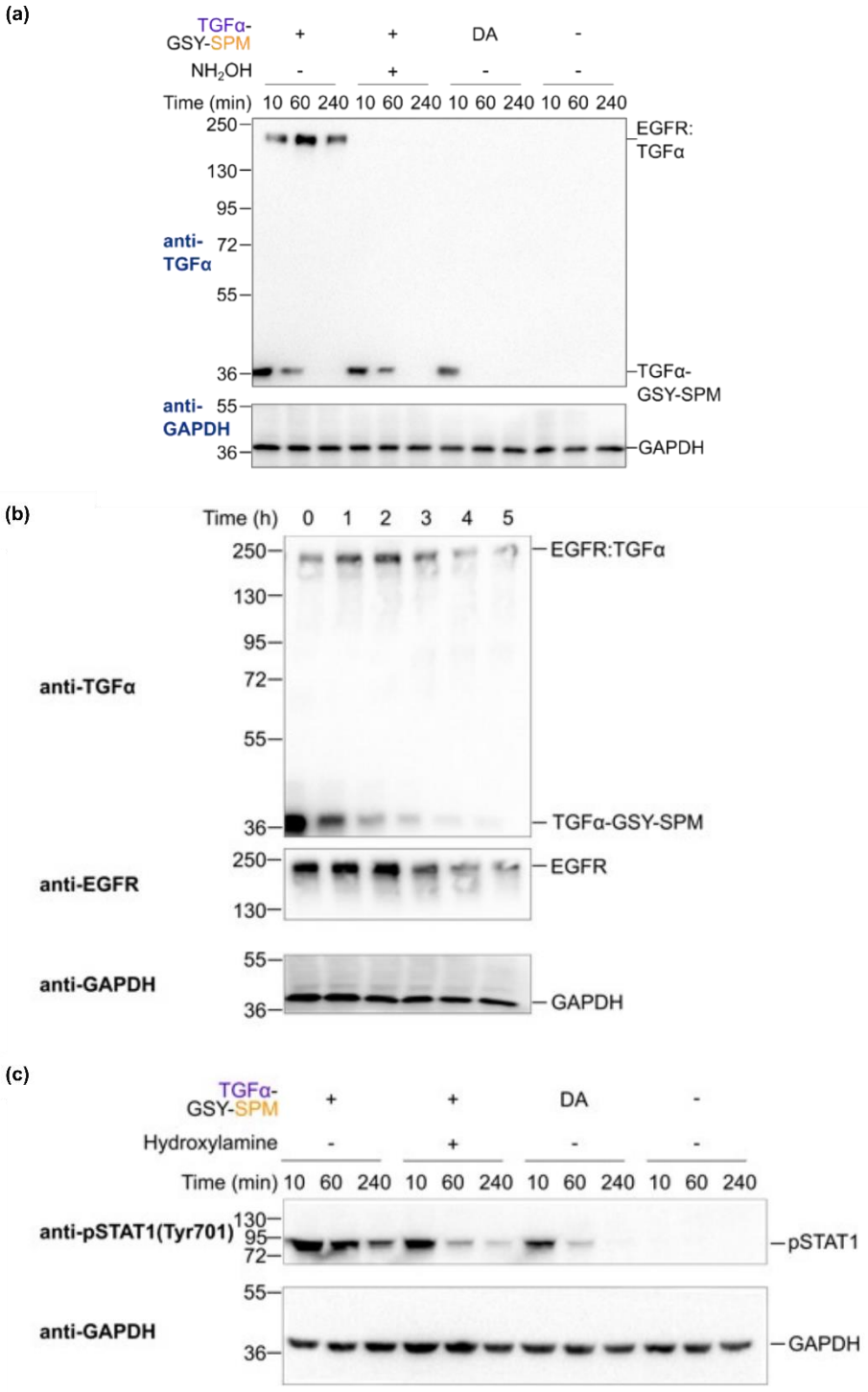


Figure 4.21 TGFα crosslinking to EGFR on A431 cells. (a) 1 μM TGFα-GSY-SPM(wild-type or non-reactive D414A mutant) was incubated with A431 cells at 37°C for 10 min with 2 mM Ca^{2+} ± 5 mM hydroxylamine. After washing off unbound TGFα, cells were incubated for the indicated time at 37°C before cell lysis. Cell lysate was used for anti-TGFα Western blot and anti-GAPDH as sample processing control. (b) Same as in (a) but at 1 h intervals and up to 5 h incubation. Cells were blotted against

EGFR in addition to TGFα. (c) Same as in (a) but cells were blotted against phosphorylated STAT1.

4.10. Discussion

The aim of this chapter was to better understand the anhydride reactivity in NeissLock, using OAZ/ODC as a model system for SPM-enabled coupling. Previously, even though coupling was achieved between OAZ and ODC, very little was known about the conditions required for fast and high-yield coupling, impeding the development of a more general platform using NeissLock.

In the first part of this chapter, by probing the pH-dependence of the crosslinking reaction, I have established the compatibility of NeissLock with pH 7.4. The compatibility of NeissLock with physiological extracellular calcium concentrations and different buffers further supports its potential broader applicability.

The specific pK_a values of individual lysine residues can neither be easily predicted using existing computational tools (Lee et al., 2016) nor determined using experimental approaches like NMR in a high-throughput manner (Andre et al., 2007; Isom et al., 2011). A group previously identified over 300 hyper-reactive lysines out of ~4,000 lysines from a range of proteins using chemical proteomics (Hacker et al., 2017). However, the predicted pK_a values of lysines in ODC where coupling occurred suggest that lysines do not need to be hyper-reactive for reaction with the anhydride (Scheu et al., 2021), suggesting of the possibly wider range of targets for NeissLock coupling.

Another possible factor contributing to the high crosslinking yields despite the higher intrinsic pK_a values of ε-amines is the relatively long half-life of the anhydride before hydrolysis. Bunton et al., measured a $t_{½}$ of 4 min for hydrolysis of carboxylic anhydrides in water at 25 °C (Bunton et al., 1963). The minute-scale half-life of the

anhydride may allow for rapid reaction with deprotonated lysines upon contact, even if the lysines are protonated for most of the time. The observation that there was little change in the preference for crosslinking over hydrolysis with pH change was surprising. Between pH 6.5 and 8.5, majority of the lysines that react with the anhydride would likely still be protonated. It is possible that the strong electrophilic nature of the anhydride means that once the lysines are deprotonated, the anhydride reacts with the lysines rapidly.

I also explored other factors influencing the coupling efficiency. This included improving the cleavage activity of the SPM by introducing the C175A mutation to OAZ to minimise non-specific disulfide bonding between the surface-exposed cysteines. The C175A mutation improved the SPM-fusion protein's stability and led to an increase in cleavage rate compared to wild-type OAZ-GSY-SPM, highlighting the importance of maximal solubility of NeissLock probes for optimal activity. I showed that, although truncated variants of SPM retained cleavage activity, their reduced yields and purity render the longest variant, i.e., residues 414-657 of FrpC, preferable for use in NeissLock probes. Given that SPM is cleaved following calcium activation to leave only a single aspartate residue at the crosslinking site, the additional residues at the C-terminus of SPM are unlikely to negatively impact the function of the crosslinked product.

In this chapter, I have also established that the coupling using NeissLock has little to no non-specific coupling. DogTag-MBP and SpyTag003-sfGFP that are examples of proteins that do not bind to OAZ showed little to no coupling to OAZ. The dependence of coupling on the affinity between the binding and target protein was further illustrated by the generation of OAZ variants with varying binding affinities to ODC. Surprisingly, the double mutants of OAZ still retained crosslinking to ODC

despite K_d values of 15-25 μM. This suggests that a very high affinity between the NeissLock probe and the target protein, as in wild-type OAZ and ODC, is not necessary for coupling to occur. Nevertheless, the reduced coupling in the double mutant of OAZ-GSY-SPM and almost completely abolished coupling in the triple mutant serve as evidence of the affinity-dependence in NeissLock and the specificity the affinity-dependence imparts. Specific coupling between OAZ and ODC was observed even in a crowded environment like cell lysate.

In the last part of this chapter, I probed the selectivity of NeissLock in terms of crosslinking sites. In the OAZ/ODC model system, I identified the key crosslinking sites as the ε-amines of K92, K121, the α-amine of ODC, as well as nucleophilic residue(s) in the C-terminus of ODC. Although not all the crosslinking sites have been identified, the identification of K92 and K121, which are also in closest proximity to the predicted site of anhydride generation, as the key crosslinking sites was in line with our initial predictions. The increased nucleophilicity of the α-amine compared to ε-amine may explain the crosslinking observed at the N-terminus of ODC upon mutation of K92 and K121. It is, however, difficult to the pinpoint the exact relative contributions of each nucleophilic site in wild-type OAZ-GSY-SPM crosslinking to ODC and whether changes to pH would alter the relative contributions. It is likely K92 was still the primary crosslinking site, given identification of this crosslinking site through LC-MS/MS but not K121 and the α-amine. The crosslinking at K121 that was about 15 Å away from the distal resolved residue of OAZ suggested that the distance requirement between the anhydride and the nucleophilic site on the target protein is possibly less stringent. The distance between the anhydride and target nucleophile can be modulated by modification of the flexible linker between SPM and the binder protein whilst finding a balance between reactivity and specificity of the NeissLock probe.

The identification of the α-amine as another crosslinking site helps inform future design of NeissLock probes, since we now know that the α-amine of either the binding protein or the target protein may compete with the designated target lysine for crosslinking. This is especially important if crosslinking at a single nucleophilic site is desired for a homogeneous crosslinked product. In the case of antibody-drug conjugates, their heterogeneity may lead to poor pharmacokinetics and reduced efficacy (Junutula et al., 2008). Alternatively, if the exact crosslinking site is of lesser importance, α-amine crosslinking provides new opportunities. If the α-amine of the target protein is near the binding interface and C-terminus of the binding protein, the α-amine can serve as a viable designed or additional crosslinking site.

Data from the *in vitro* plate-assay showing the anhydride reactivity with other nucleophiles like thiols and phenols hint at opportunities for using NeissLock to crosslink to other nucleophilic residues like cysteines and tyrosines. Although we do not yet have direct evidence of crosslinking to other amino acids apart from lysines in a protein complex, it may be interesting to further test NeissLock's use in complexes where a cysteine residue on the target protein is optimally positioned for reacting with an anhydride at the C-terminus of the binding protein. However, surface-exposed cysteines are less common compared to lysines, making cysteine-targeting more challenging (Hallenbeck et al., 2017). The relative instability of the resulting thioester bonds compared to amide bonds is another potential limitation.

The use of NeissLock for coupling TGFα to endogenous EGFR on live cells suggests future possibilities of using Neisslock for new therapeutic or diagnostic applications. The observed change in pSTAT1 activation with TGFα:EGFR conjugation was surprising, and it will be interesting to further investigate other changes in EGFR downstream signalling events, as well as the basis of these

changes. However, the subsequent degradation of the TGFα:EGFR conjugate in the cells within hours reveals a possible limitation of the NeissLock technology. When using NeissLock to target cell surface receptors, the constant endocytosis and degradation of these receptors (Tomas et al., 2014) may lead to a limited half-life for the conjugate, rendering it necessary for more frequent dosing or selection of an alternative target protein with a longer half-life.

All in all, the insights into anhydride reactivity of NeissLock presented in this chapter enable a more informed design of future NeissLock probes. There is still room for improvement of the system, including increasing the rate of conjugation and minimising self-reaction of the probe. Nevertheless, the potential broader applicability of NeissLock for targeting endogenous proteins has been demonstrated.

Chapter 5: Optimising SPM for an enhanced NeissLock

In this chapter, the key aim is to design a second generation NeissLock with enhanced reaction rate and coupling yield. Although the first generation NeissLock was sufficient for effective coupling, conjugation yield was limited to about 80% and more than 1 h incubation was required for 80% or more cleavage of SPM.

The key indicators for an enhanced second generation NeissLock system are: 1) a more rapid reaction rate (reaction completion within 10 min), 2) high conjugation yield (>80% yield), 3) high specificity in a complex mixture like cell lysate, and 4) mild reaction conditions for *in vivo* labelling (pH 7.4, ≤37 °C). In the first part of this chapter, I use homology searching to identify a naturally-occurring SPM variant with enhanced reactivity for use in NeissLock probes. In the second part of this chapter, I perform initial experiments in developing an intimin-based bacterial surface display system for directed evolution of the SPM domain.

5.1. Cleavage and coupling activity of SPM homologues

The first generation of NeissLock utilized the SPM of FrpC from *Neisseria meningitidis* serotype B strain MC58 (Uniprot: Q9JYV5) for anhydride generation. A question arises as to whether homologues of the SPM exist and if these SPM variants also undergo calcium-induced autoproteolysis. Previously, Osicka *et al.* showed that the SPM from the ApxIVA protein of *Actinobacillus pleuropneumoniae*, an animal pathogen, showed similar calcium-dependent autocatalytic cleavage and coupling as FrpC (Osicka et al., 2004). I performed a BLAST search to identify homologues of FrpC SPM from *Neisseria meningitidis* and other species. SPM homologues from non-pathogenic bacterial species and/or bacterial species whose hosts are non-human were preferentially selected, to reduce the likelihood of pre-existing immunity against

the SPM from prior exposure to similar source organisms. Pre-existing antibodies that recognize the SPM may limit the efficacy of NeissLock's use in therapeutic applications and may lead to adverse events if administered in patients (Gorovits et al., 2016).

I shortlisted four SPM homologues: FrpA SPM based on position 298-543 of FrpA from *Neisseria meningitidis* serotype C (Uniprot: P55126), position 231-476 of haemolysin-type calcium-binding protein related domain-containing protein from *Alysiella filiformis* (GenBank:SOD68839.1), position 190-434 of the bifunctional hemolysin/adenylate cyclase precursor from *Kingella negevensis* (GenBank: SNB83338.1), position 244-491 of the calcium-binding hemolysin protein from *Pseudomonas caricapapayae* (GenBank: RMV74036.1). Apart from FrpA SPM, the other homologues were selected from non-pathogenic bacteria or bacteria with non-human hosts. The homologues were selected to span a range of varying sequence identities compared to the original FrpC SPM sequence.

A. filiformis is a non-pathogenic bacterium that was previously isolated from the mouths of animals such as pigs (Zutic et al., 2013). *K. negevensis* is bacterium isolated from the throat of healthy children (El Houmami et al., 2017). *P. caricapapayae*, on the other hand, is a plant pathogen (Hu et al., 1998). Compared to the original FrpC SPM sequence used in the first generation of NeissLock, the FrpA SPM sequence has 98 % sequence identity, the *A. filiformis* SPM sequence has 71 % sequence identity, the *K. negevensis* SPM sequence has 60 % sequence identity, and the *P. caricapapayae* SPM sequence has 48 % sequence identity (Figure 5.1a).

The SPM homologues were expressed as C-terminal fusions to OAZ whilst retaining the native -1 and -2 residues relative to the cleavable Asp-Pro bond. All the OAZ-SPM fusion constructs were expressed in reasonable yields and purity in *E. coli*,

except for the SPM variant from *P. caricapapayae* that did not express well, and a majority was targeted to insoluble inclusion bodies. OAZ-SPM (FrpA) was purified to a yield of 4 mg/ L, OAZ-SPM (*Alysiella*) was purified to yield of 3 mg/ L and OAZ-SPM (*Kingella*) was purified to yield of 3 mg/ L. Given the poor expression and solubility, I did not proceed with the SPM variant from *P. caricapapayae*.

The SPM homologues fused to OAZ were incubated overnight with 10 µM ODC in HEPES-based reaction buffer containing 10 mM $CaCl_2$ to assay for calcium-induced cleavage and coupling. The reaction was stopped after 16 h incubation at 37 °C. Coomassie-stained SDS-PAGE showed the formation of cleavage products (SPM) and crosslinked products (OAZ:ODC) for all three SPM homologues after but not before calcium addition (Figure 5.1b).

To further investigate the cleavage rates of the SPM homologues, I performed a time-course experiment. Figure 5.1c shows a representative Coomassie-stained SDS-PAGE gel to quantify the cleavage of OAZ-SPM for the FrpA variant at different time-points. In 1 h, approximately 90 % of OAZ-SPM(FrpA) was cleaved, compared to 75% of OAZ-SPM(FrpC), 60% of OAZ-SPM(*Alysiella*) and 15% of OAZ-SPM(*Kingella*) (Figure 5.1d). There was a positive correlation between the level of sequence identity between the SPM homologue to FrpC SPM and the cleavage rate.

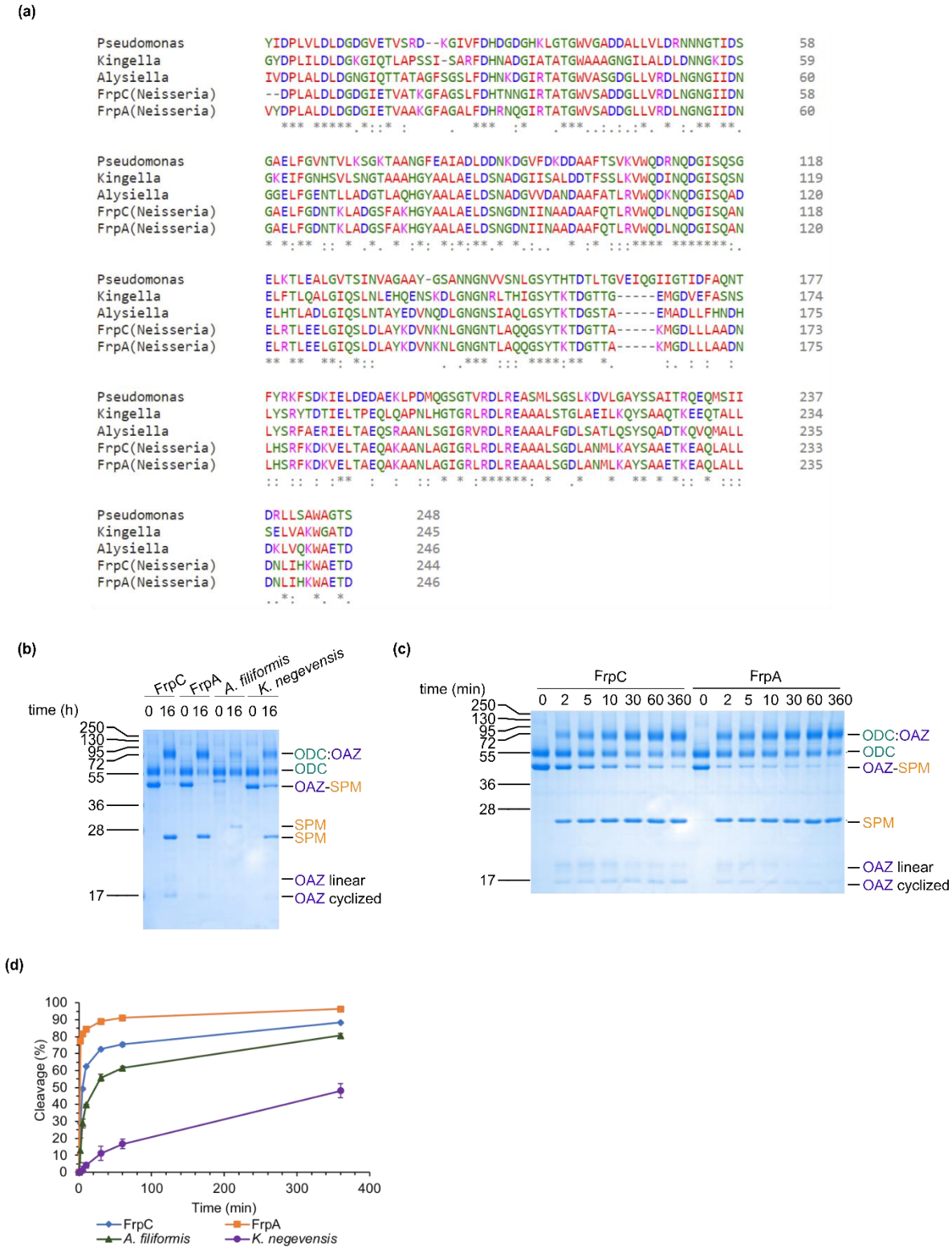


Figure 5.1: Cleavage and coupling activity of SPM homologues. (a) ClustalO sequence alignment of SPM homologues, including the -2 and -1 positions relative to cleavable D-P bond. Red: hydrophobic residues; Blue: acidic residues; Magenta: basic residues; Green: polar residues; * : position with complete residue conservation; : : position with highly similar residues; . : position with weakly similar residues. (b) 5 μM

OAZ-GSY-SPM variants were incubated with 5 μM ODC in HEPES-based reaction buffer with 10 mM $CaCl_2$, pH 7.4 at 37 °C for 16 h. Samples were analysed by Coomassie-stained SDS-PAGE. (c) 5 μM OAZ-GSY-SPM(FrpA) or 5 μM OAZ-GSY-SPM(FrpC) was incubated with 5 μM ODC in HEPES-based reaction buffer with 10 mM $CaCl_2$, pH 7.4 at 37 °C. Samples were analysed by Coomassie-stained SDS-PAGE. (d) Quantification of cleavage rate of OAZ-GSY-SPM(FrpC or homologues) from Coomassie-stained SDS-PAGE as in (b) (mean ± 1 S.D., n=3).

To verify that FrpA SPM is faster reacting than FrpC SPM, I repeated the time-course assay of their relative cleavage and coupling rates but at a lower calcium concentration of 1 mM and a lower temperature of 25 °C. Although the FrpA and FrpC SPM sequences varied only at four residue positions (Figure 5.2a), FrpA SPM showed enhanced cleavage and coupling compared to FrpC SPM even at a lower temperature and calcium concentration (Figure 5.2b). The residues of FrpC SPM vary between the structure in PDB 6SJW and the construct used in this thesis (Figure 5.2a) because sequences from different *Neisseria meningitidis* serogroups were used. Formation of crosslinked product OAZ:ODC was observed within 0.5 minutes of calcium addition for OAZ-SPM(FrpA) but a similar amount of crosslinked product was formed only after 2 to 10 minutes of incubation for OAZ-SPM(FrpC). In 2 min, about 70 % of OAZ-SPM(FrpA) was cleaved, compared to only 20 % of OAZ-SPM(FrpC) (Figure 5.2c).

The time-courses comparing the cleavage reactions of OAZ-GSY-FrpA(SPM) and OAZ-GSY-FrpC(SPM) in Figure 5.1d were fitted to two-phase exponential equations (Figure 5.2d). Based on the residuals, modelling using two-phase exponential equations gave a better fit than single-phase exponential equations (Figure 5.2e and Figure 5.2f). The R-squared value of the two-phase exponential equation model for the cleavage of FrpC SPM was 0.9989 and for FrpA SPM, the R-

squared value was 0.9991. The two-phase exponential model would suggest that the observed cleavage of SPM is a sum of a fast-cleaving first-order reaction and a slow-cleaving first-order reaction. The double exponential kinetics was also previously observed by Elizabeth Tasker, a previous Master's student in the Howarth lab, who studied and modelled the kinetics of FrpA SPM cleavage over 24 h.

Based on the fitted equation model for FrpC SPM, the rate constant of the fast-cleaving reaction was 0.28 min^{-1} and the rate constant of the slow reaction was 0.011 min^{-1}. For FrpA SPM, the rate constant of the fast reaction was 1.5 min^{-1} and the rate constant of the slow reaction was 0.022 min^{-1}. The half-life of the fast cleavage reaction was 2.5 min for FrpC SPM and 0.45 min for FrpA SPM. The derived rate constants and half-lives confirm that the cleavage of SPM is an autocatalytic reaction that occurs more rapidly than uncatalyzed Asp-Pro cleavage. The rate constant of Asp-Pro cleavage in a decapeptide at 40 °C and pH 4.5 was approximately 0.0038 h^{-1}, giving a half-life of 7.6 days (Li et al., 2009). The rate constants of the fast-cleaving reactions of SPM were more than 1000-fold higher than the rate constant of the hydrolysis of the Asp-Pro bond in a decapeptide under acidic conditions.

Apart from a 5.5-fold increase in rate constant of the fast-cleaving reaction for FrpA SPM compared to FrpC SPM, a larger fraction of FrpA SPM (84 %) consisted of the faster-cleaving population compared to FrpC SPM (73 %). Based on the biphasic model, the cleavage of FrpC SPM plateaus at 89 % while the cleavage of FrpA SPM plateaus at 96 %. Since FrpA SPM is faster reacting than FrpC SPM, I decided to proceed with FrpA SPM for further engineering for second generation NeissLock.

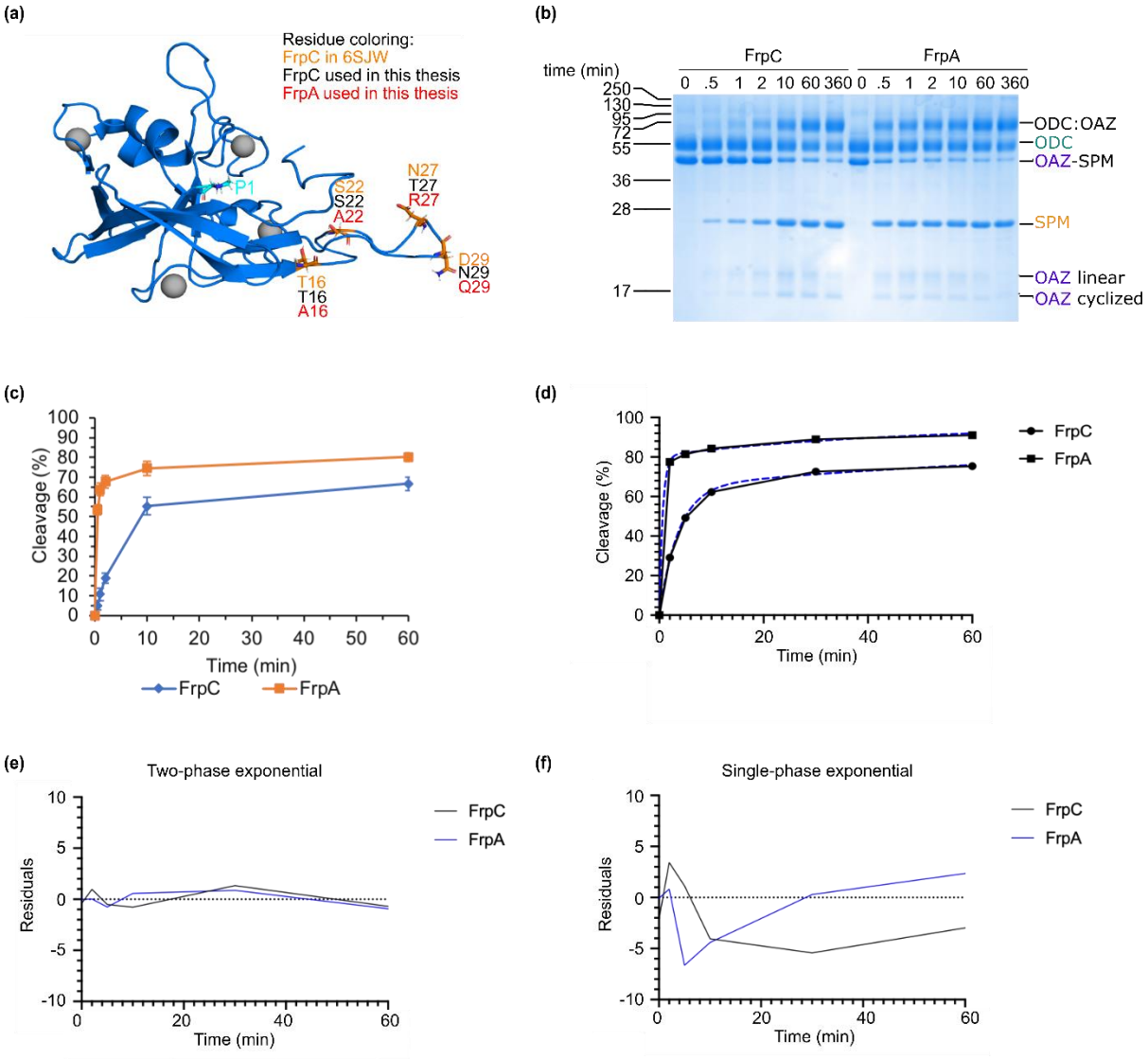


Figure 5.2: FrpA SPM shows enhanced reactivity compared to FrpC SPM. (a) Structure of FrpC SPM post-cleavage with 4 bound calcium ions (grey spheres). The first proline residue is shown as a stick residue (cyan). Residue positions differing between FrpA and FrpC are indicated as stick residues (orange). Residues 16-36 were not resolved but modelled in the NMR structure (PDB 6SJW) (b) 10 µM OAZ-GSY-SPM(FrpA) or 10 µM OAZ-GSY-SPM(FrpC) was incubated with 10 µM ODC in HEPES-based reaction buffer with 1 mM $CaCl_2$, pH 7.4 at 25 °C, before boiling in 1×SDS loading buffer with 15 mM EDTA. Samples were analysed by Coomassie-stained SDS-PAGE. (c) Quantification of cleavage rate of OAZ-GSY-SPM(FrpA) or OAZ-GSY-SPM(FrpC) from Coomassie-stained SDS-PAGE as in (b) (mean ± 1 S.D., n=3). (d) Fitting of time-course of OAZ-GSY-SPM(FrpA) and OAZ-GSY-SPM(FrpC) cleavage over 60 min (10 mM Ca^{2+}, pH 7.4, 37 °C) to a two-phase exponential reaction. The two-phase fittings are shown in blue dashed lines. (e) Residuals from

fitting the cleavage reactions to two-phase exponential equations as in (d). (f) Residuals from fitting the cleavage reactions to single-phase exponential equations.

5.2. Establishing bacterial display system for directed evolution of SPM

To further improve SPM activity beyond what is accessible from natural evolution, I looked towards employing directed evolution. The iterated rounds of gene diversification, followed by screening or selection in the laboratory mimics natural evolution (Packer and Liu, 2015). Directed evolution has been successfully used to identify desirable mutations that enhance or modify the activities of proteins used in medicine, research and industry (Packer and Liu, 2015). One advantage of directed evolution is that pre-existing knowledge of the protein's structure is not necessary. Although a Nuclear Magnetic Resonance (NMR) structure of the calcium-bound FrpC SPM domain was recently elucidated (Kuban *et al.*, 2020), our efforts to use rational design techniques like proline substitution in loops, glycine substitutions and reducing surface hydrophobicity gave limited success in enhancing the reaction yield or thermostability of the domain (Elizabeth Tasker, Howarth Laboratory). In addition, the NMR structure of the SPM domain was incomplete, given an unresolved flexible loop between positions 430 and 450, as well as exclusion of residues 591-657 at the C-terminus (Kuban *et al.*, 2020).

Unlike phage display, cell surface display can be used with high-throughput fluorescence-activated cell sorting (FACS) to screen millions of individual clones (Gai and Wittrup, 2007). I decided on using bacterial surface display over yeast or other eukaryotic cell surface display systems because the high calcium concentrations in the endoplasmic reticulum (ER) (Lam and Galione, 2013) and other compartments of the secretory pathway of eukaryotic cells may lead to premature cleavage of the SPM domain.

Intimin is a membrane protein expressed by pathogenic *E. coli* for adhesion to eukaryotic cell surface receptors (Wentzel et al., 2001). The EaeA intimin of enterohaemorrhagic *E. coli* (EHEC) O157:H7 consists of an N-terminal β-barrel domain that translocates the four C-terminal passenger domains across the bacterial outer membrane (Wentzel et al., 2001) (Figure 5.3).

It was previously shown that deletion of the last 2 carboxyl passenger domains of the EaeA intimin and replacing them with other heterologous recombinant proteins like interleukin 4 (IL4) enabled outer membrane display of the passenger domains on the surface of the *E. coli* K12 strain (Wentzel et al., 2001). The display of the recombinant proteins on the surface of the bacteria provides a genotype-phenotype link required for directed evolution.

As compared to other membrane proteins used as anchoring scaffolds for bacterial surface display like outer membrane A (OmpA) that requires an insertional fusion and peptidoglycan-associated lipoprotein (PAL) that requires an N-terminal fusion, intimin requires C-terminal fusion of the heterologous passenger domains for display (van Bloois et al., 2011). The use of Eae intimin as a scaffold is useful because SPM can only be fused to the C-terminus of the scaffold to retain its cleavage activity. Previously, intimin-based bacterial display was used to identify nanobody binders against human fibrinogen (Salema et al., 2016a) and epidermal growth factor receptor (EGFR) (Salema et al., 2016b).

Felix Metzner, a previous Master's student in the Howarth lab, first made a construct consisting of the Eae intimin fused to the SPM domain from FrpC followed by a C-terminal myc-tag. He found that induced *E. coli* grown in the presence EDTA showed a slight increase in anti-myc fluorescence, confirming the display of intimin-SPM on the cells. With calcium addition, the cells showed a partial reduction in

fluorescence intensity as was expected from SPM cleavage. However, the fluorescence shifts were unclear, possibly due to poor expression and/or staining.

For bacterial display, I decided to start with the FrpA SPM sequence as my wild-type (WT) sequence since FrpA SPM was the fastest-cleaving variant that I have identified so far. I included an intervening HA-tag between the intimin β-barrel, D0 and D00 domains, and the SPM domain to identify bacterial clones that have expressed an SPM variant both pre- and post-cleavage (Figure 5.3). Like the initial intimin-SPM construct, an additional myc-tag was added to the C-terminus of the SPM domain to detect cleavage and loss of the SPM domain. Without a HA-tag, if cells are sorted according to their anti-myc fluorescence levels alone, non-expressing cells cannot be differentiated from cells with SPM domains that are better-cleaving. The anti-HA fluorescence is a function of how well-expressing a variant is, which in turn may suggest the overall stability of the variant (Ren et al., 2019).

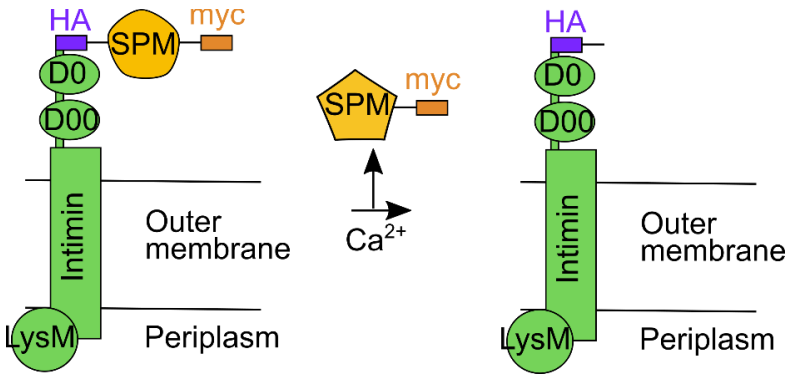


Figure 5.3 Intimin-based bacterial surface display of SPM. Cartoon showing the intimin-fusion used to display SPM on the *E. coli* cell surface. Residues 1-659 of Eae intimin (green) from *E. coli* O157:H7 comprising an N-terminal LysM domain, a transmembrane β-barrel domain, D0 and D00 passenger domains was used as the anchoring motif. FrpA SPM (yellow) was fused to the C-terminus of intimin with additional HA- (purple) and myc- (orange) epitope tags flanking each end of SPM for detection of expression and cleavage.

5.2.1. Validation of display by flow cytometry

After a series of growth trials and experiments to optimise protein expression, I decided to use *E. coli* BL21 (DE3) strain grown in LB media with 2 mM EGTA for expressing the intimin-SPM fusion construct. EGTA, a selective calcium chelator, was added to inhibit premature cleavage of SPM when exposed to calcium present in the media. Protein induction was carried out using 0.42 mM IPTG for maximal induction and at 30 °C to aid folding of the protein.

The cells expressing intimin-HA-SPM-myc were stained with anti-myc and anti-HA antibodies to detect display by flow cytometry. Cells expressing an intimin-SPM construct without the HA- and myc- tags were included as a negative control for specificity of the staining. Cells expressing intimin-HA-SPM-myc showed approximately 11-fold higher anti-HA median fluorescence intensity (MFI) compared to the negative control (Figure 5.4). In the sample where only the secondary antibody was used but no anti-HA primary antibody was added, the anti-HA fluorescence was similar to the negative control, confirming that there was little non-specific binding of the secondary antibody used for anti-HA staining (Figure 5.4). As was expected, for cells expressing intimin-HA-SPM-myc, there was no change in the anti-HA fluorescence with or without incubation with 2 mM Ca^{2+} (Figure 5.4).

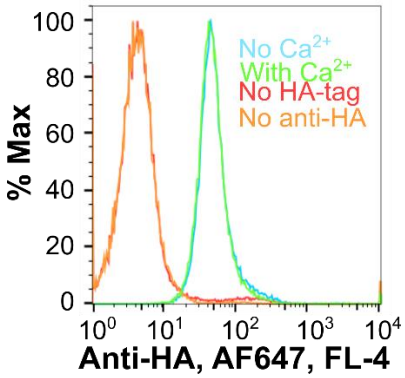


Figure 5.4 Validation of intimin-HA-SPM-myc expression by flow cytometry. Cells expressing intimin-HA-SPM-myc or intimin-SPM-StrepTagII (negative control)

were induced with 0.42 mM IPTG in LB media with 2 mM EGTA at 30 °C for 2.5 h. Cells were stained with anti-myc and anti-HA antibodies, before incubation in HEPES-based reaction buffer with or without 2 mM $CaCl_2$, pH 7.4 at 37 °C for 10 min. The samples were then stained with the corresponding secondary antibodies before flow cytometry.

5.2.2. Calcium-induced cleavage of intimin-HA-FrpA-myc (WT)

To confirm that the SPM domain fused to intimin retains calcium-inducible cleavage activity, I compared the anti-myc fluorescence of cells expressing intimin-HA-SPM-myc, with or without incubation with 2 mM Ca^{2+}. Without the addition of calcium, cells showed a 3.5-fold higher anti-myc MFI as compared to the negative control that expresses an intimin-SPM construct without a myc-tag (Figure 5.5a). When the cells expressing intimin-HA-SPM-myc were incubated with 2 mM Ca^{2+} for 1 min, the anti-myc MFI was 3-fold lower than the sample without Ca^{2+} (Figure 5.5a). The decrease in anti-myc fluorescence suggests that SPM fused to intimin retains its cleavage functionality, leading to the loss of the SPM domain fused to myc-tag upon calcium addition.

The intimin-based bacterial display system and sorting strategy enabled differentiation of faster and slower cleaving SPM variants. Comparing wild-type (WT) FrpA SPM and the L23Y mutant variant of SPM that was found to be slower cleaving (Elizabeth Tasker, Howarth laboratory), the anti-myc MFI of WT SPM was 1.4-fold lower than the L23Y mutant after incubating with calcium for 1 min (Figure 5.5b). The anti-myc MFI of WT and L23Y SPM were similar without calcium addition, suggesting that the difference in MFI after calcium addition was not due to a difference in level of display (Figure 5.5b).

To further verify the anhydride reactivity of SPM displayed on the cell surface, I added biotin-PEG_4-alkoxyamine to the cells expressing intimin-HA-SPM-myc. In the presence of Ca^{2+}, I expected the anhydride generated from SPM activation to react with biotin-PEG_4-alkoxyamine, a strong nucleophile. In the previous chapter, the anhydride reactivity with biotin-PEG_4-alkoxyamine was verified in a plate-based assay (Section 4.8). I removed excess unbound biotin-PEG_4-alkoxyamine from the cells by a series of extensive wash steps. Indeed, I observed a higher MFI with addition of both biotin-PEG_4-alkoxyamine and Ca^{2+}, suggesting that the anhydride reacted with biotin-PEG_4-alkoxyamine (Figure 5.5c). The MFI with biotin-PEG_4-alkoxyamine and Ca^{2+}, was approximately 3-fold higher than without biotin-PEG_4-alkoxyamine (Figure 5.5c). In the sample with the same amount of biotin-PEG-alkoxyamine added but without Ca^{2+}, the MFI was similar to the negative control with no biotin-PEG_4-alkoxyamine added (Figure 5.5c), confirming that there is little residual non-covalent binding of biotin-PEG_4-alkoxyamine to the cells.

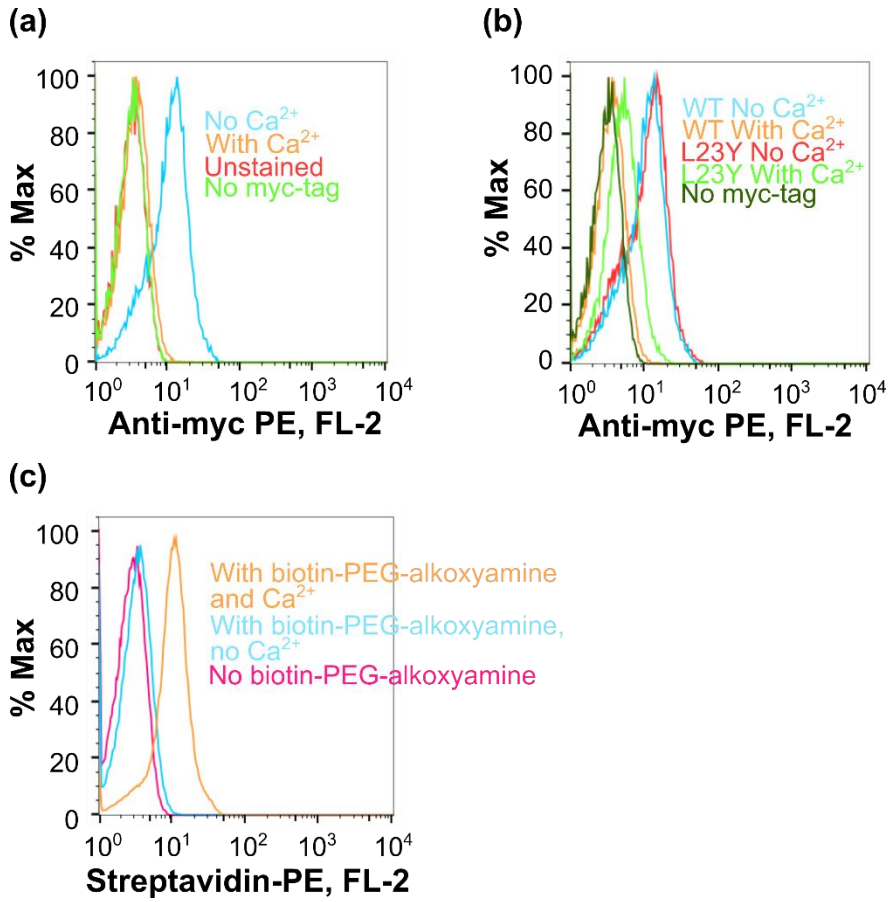


Figure 5.5 Validation of intimin-HA-SPM-myc cleavage activity by flow cytometry. (a) Cells expressing intimin-HA-SPM(WT)-myc or intimin-SPM-StrepTagII (negative control) were induced with 0.42 mM IPTG in LB media with 2 mM EGTA at 30 °C for 2.5 h. Cells were stained with anti-myc PE, before incubation in HEPES-based reaction buffer with or without 2 mM $CaCl_2$, pH 7.4 at 37 °C for 1 min. Flow cytometry was carried out after a series of wash steps. (b) Similar to (a) but with the additional intimin-HA-SPM(L23Y)-myc mutant for comparison. (c) Same induction conditions as in (a). 1 mM biotin-PEG_4-alkoxyamine was added to cells in HEPES reaction buffer with or without 2 mM $CaCl_2$, pH 7.4 at 37 °C for 10 min. Cells were then stained with streptavidin-PE for detection.

5.3. Generation of a library of SPM variants

To generate a non-biased library of SPM variants, I performed error-prone PCR using the wild-type FrpA SPM sequence as the template DNA for random mutagenesis. Forward and reverse primers were designed to anneal to the N- and C-

terminus of the SPM sequence, ensuring that the PCR product retains the Asp-Pro bond at the N-terminus of SPM for cleavage but is amenable to mutation throughout the SPM sequence up till the last base position. Compared to Taq DNA polymerase-dependent error-prone PCR, the Mutazyme II DNA polymerase enables random mutagenesis with more similar mutation rates at purine and pyrimidine bases (Emond et al., 2008). Using an initial template DNA concentration of 50 ng and 30 cycles of amplification during the error prone PCR, I observed a mean mutation frequency of 5 to 6 nucleotide substitutions throughout the SPM sequence when sampling 9 random clones from the naive library, covering approximately 0.01% of the entire library of 10^5 colonies. The mean mutation frequency of the DNA bases translated to 3 to 4 amino acid substitutions throughout the SPM sequence. The mean amino acid substitution frequency is lower than the mean mutation frequency of the DNA sequence due to synonymous mutations, specifically mutations at the third base in codons. Out of the 9 random clones, one of the clones, Clone 7, had a premature stop codon (Figure 5.6).

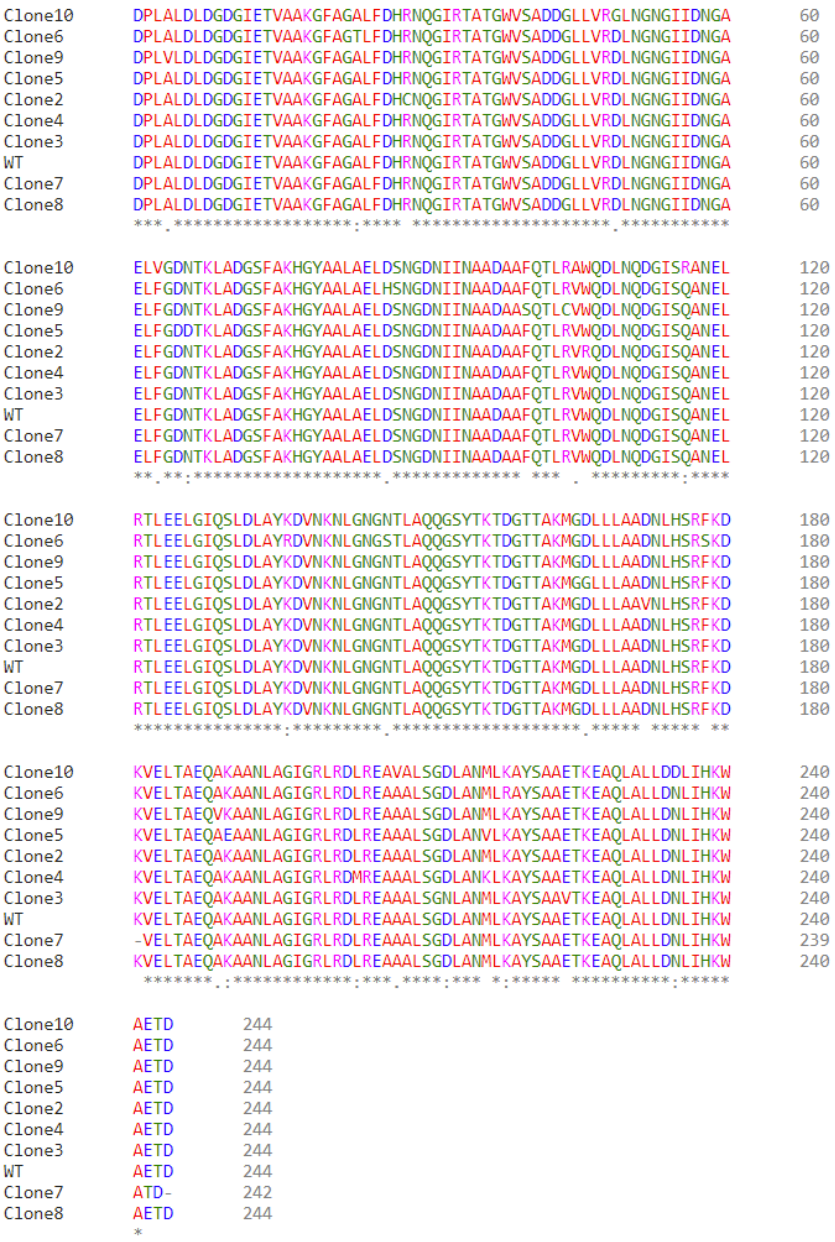


Figure 5.6 ClustalO alignment of amino acid sequences of clones from the naïve library. The plasmids isolated from 9 randomly selected TG1 clones were submitted for Sanger sequencing. The DNA sequences spanning the SPM domain were translated and the translated amino acid sequences were aligned using Clustal Omega.

After error-prone PCR of the SPM sequence, I used Gibson Assembly to insert the SPM sequence into the pET28a vector backbone encoding the flanking intimin as well as HA- and myc- epitope tags. Transformation into electrocompetent *E. coli* TG1

strain gave a library size of approximately 3.7×10^5 cfu. The mixture of clones was harvested from plates and grown overnight before the plasmid DNA was extracted and used for transformation into *E. coli* BL21 (DE3) for expression. A library size of approximately 3.8×10^5 cfu was observed.

5.4. Sorting of the library and evaluation of clones

During the subsequent sorting rounds, I modified the antibody staining procedure to increase the difference in anti-myc fluorescence intensity between cells expressing intimin-HA-SPM(WT)-myc after incubation with calcium, compared to the non-expressing negative control. The change from using a fluorescently labelled anti-myc antibody to using a primary anti-myc antibody followed by a fluorescent secondary antibody increased the fluorescence intensity relative to the negative control, enabling gating for SPM variants that cleave to a larger extent compared to the wild-type (Figure 5.7). Without Ca^{2+}, the anti-myc MFI of cells expressing intimin-HA-SPM(WT)-myc was 347 (Figure 5.7). With Ca^{2+}, the anti-myc MFI decreased approximately 12-fold to 29.1, approximately 7-fold higher than the MFI of non-expressing cells (Figure 5.7).

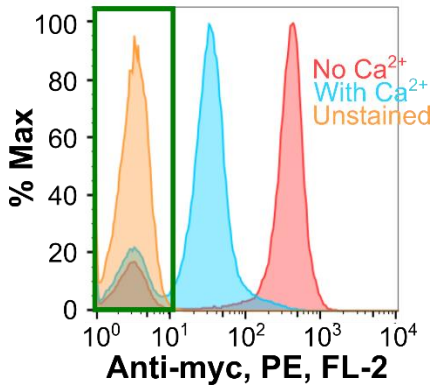


Figure 5.7 Anti-myc fluorescence of cells expressing wild-type FrpA SPM using the optimised staining procedure. Cells were induced with 0.42 mM IPTG in LB media with 2 mM EGTA at 30 °C for 2.5 h. For the sample with Ca^{2+}, cells were incubated in HEPES-based reaction buffer with 2 mM $CaCl_2$, pH 7.4 at 37 °C for 10

min. The gate that can be used to sort variants that cleave to a larger extent than wild-type FrpA SPM is shown in green.

In addition, due to observed cross-reactivity between anti-mouse and anti-rat secondary antibodies with the rat anti-HA and mouse anti-myc primary antibodies, I decided to perform sequential staining i.e., primary antibody A followed by secondary antibody A, before primary antibody B and secondary antibody B. I also changed to the use of a biotinylated anti-HA antibody, so that an AlexaFluor488-labelled streptavidin could be used for detection instead, minimising cross-reactivity with the anti-myc primary antibody (Figure 5.8).

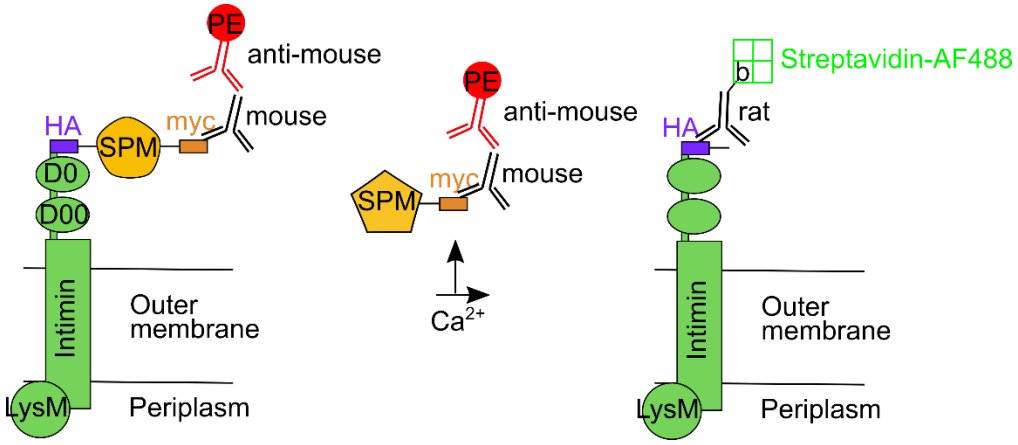


Figure 5.8 Staining procedure for FACS to minimise cross-reactivity. Induced cells are first stained with mouse anti-myc antibody before staining with an anti-mouse PE secondary antibody. The cells are incubated with 2 mM Ca^{2+} at 37 °C for 10 min, leading to cleavage of the C-terminal myc tag. After rounds of washes, the cells are stained with a biotinylated rat anti-HA antibody and detected using AlexaFluor488-labelled streptavidin.

Incorporating the anti-HA stain and using the control that is wild-type with Ca^{2+}, I drew a gate encompassing cells with lower anti-myc fluorescence but whose anti-HA fluorescence was as high as wild-type. This should sort for cells in the naïve library that express as well but cleave to a larger extent after calcium addition (Figure 5.9a

middle panel). Compared to the wild-type control where 0.59% of cells lie in the gate, 2.52% of cells in the library lie within the gate used for sorting in the first round (Figure 5.9a and b). The sorted cells were grown up overnight to amplify the sorted clones and the overnight culture was used to inoculate a new culture for another round of expression and sorting.

However, I observed that in the second round of sorting of the library, all the cells had a high anti-HA but low anti-myc fluorescence even without calcium addition (Figure 5.9c). After sequencing 10 random clones from the library used in the second round of sorting, I found that almost all the clones had premature stop codons in the SPM sequence, leading to deletion of the C-terminal myc-tag. The premature stop codons explain their high anti-HA and low anti-myc fluorescence that was independent of calcium addition.

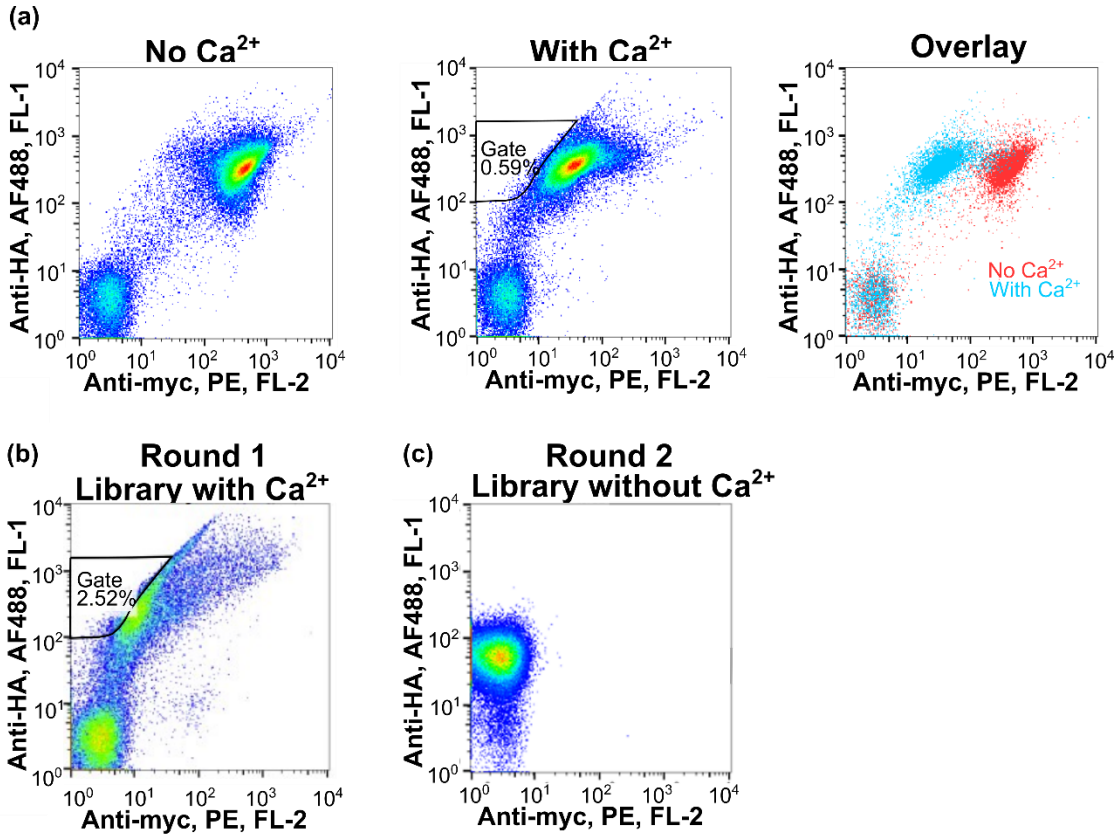


Figure 5.9 FACS analysis of dual-stained cells expressing intimin-HA-SPM-myc. (a) Cells expressing wild-type FrpA SPM as intimin fusion. Cells were induced with

0.42 mM IPTG in LB media with 2 mM EGTA at 30 °C for 4 h. Cells were incubated in HEPES-based reaction buffer with or without 2 mM $CaCl_2$, pH 7.4 at 37 °C for 10 min. (b) Cells expressing the library of SPM variants as intimin fusion after calcium addition. Expression and treatment conditions were same as in (a). Cells that were high in anti-HA and low in anti-myc fluorescence were gated as shown. (c) Cells expressing library of SPM variants in second round of sorting without calcium addition.

Observing that the sorted clones mostly had premature stop codons, I decided to perform sequential sorting instead. Before calcium addition, I first sorted for cells that were high in both anti-HA and anti-myc fluorescence using Gate 1 to ensure that SPM had been fully translated (Figure 5.10a and b). Approximately 2 % of the cells in the library had high anti-HA and anti-myc fluorescence before calcium addition (Figure 5.10b). On the other hand, approximately 74 % of the cells lie within Gate 1 for the wild-type variant (Figure 5.10a), suggesting that a larger proportion of cells for the wild-type variant expressed the full-length construct. After calcium addition to the sorted cells, I performed a second sort using Gate 2 (Figure 5.10a) for cells that have lost the myc-tag. A percentage sort of 0.8 % was observed. After the sorted cells were grown overnight and another round of expression was induced, I found that the anti-myc fluorescence of cells in the Round 2 library was low even without calcium (Figure 5.10c). Unlike in Figure 5.9c, there was a larger proportion of non-expressing cells i.e., low in both anti-HA and anti-myc fluorescence (Figure 5.10c). Again, the sorting did not isolate better cleaving SPM variants, suggesting the need for further optimisation.

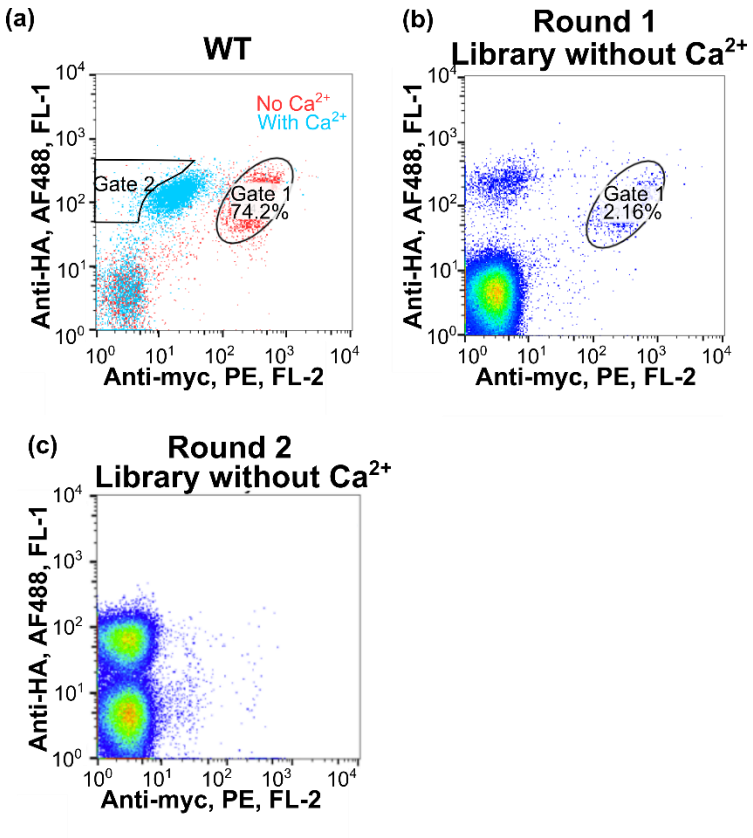


Figure 5.10 FACS analysis of dual-stained cells expressing intimin-HA-SPM-myc during sequential sorting. (a) Cells expressing wild-type FrpA SPM as intimin fusion. Cells were induced with 0.42 mM IPTG in LB media with 2 mM EGTA at 30 °C for 4 h. Cells were incubated in HEPES-based reaction buffer with or without 2 mM $CaCl_2$, pH 7.4 at 37 °C for 10 min. (b) Cells expressing library of SPM variants as intimin fusion under the same conditions as (a). Cells that were high in anti-HA and low in anti-myc fluorescence were sorted using Gate 1 before calcium addition and a second sort using Gate 2. (c) Cells expressing library of SPM variants in the second round of sorting without calcium addition.

5.5. Discussion

The aim of this chapter was to improve on the first generation of NeissLock, specifically, the activity of SPM. In the first generation of NeissLock, the half-life of SPM cleavage was less than 5 min and a maximum cleavage yield of about 80% was observed after 6 h. By screening additional SPM homologues identified through a homology search, I have identified a faster-cleaving SPM variant that is from a closely

related protein, FrpA. Similar to FrpC, FrpA is a protein from *Neisseria meningitidis* containing the RTX domains and both proteins are upregulated under conditions of iron shortage (Thompson and Sparling, 1993). The FrpA SPM sequence that I tested is derived from *Neisseria meningitidis* serogroup C whereas the FrpC SPM sequence used in the first generation of NeissLock was derived from *Neisseria meningitidis* serogroup B. Despite varying in only four amino acid positions compared to FrpC SPM, FrpA SPM exhibited much faster cleavage.

The four amino acids: T16, S22, T27 and N29 (numbering based on PDB 6SJW) in FrpC SPM were mutated to A16, A22, R27 and Q29 in the faster-cleaving FrpA variant (Figure 5.2a). All four residues lie within the unresolved loop between residues 16 - 36 in the NMR structure of calcium-bound FrpC SPM (PDB 6SJW). It was previously suggested by Kuban et al. that residues 16–36 formed an unstructured loop and deletion of the loop did not affect the cleavage of SPM (Kuban *et al.*, 2020). Conversely, the difference in cleavage rates of FrpA and FrpC SPM observed in this thesis suggests that the disordered loop does affect cleavage rate. There are various possible explanations for the role of A16, A22, R27 and Q29. First, the introduction of a positive-charged arginine residue at position 27 may help stabilise the oxyanion tetrahedral transition state formed during cleavage. Second, the mutation of T16 and S22 to aliphatic alanine residues may have abolished self-reaction between the anhydride and the hydroxyl groups. If there was any self-reaction with the hydroxyl groups in FrpC SPM, the observed ‘’uncleaved fraction’’ may have comprised a proportion of cleaved but self-reacted protein from ester bond formation with SPM. However, given the presence of K17 in the vicinity, reaction with the hydroxyl groups is relatively less likely. Third, the four residues may play a role in stabilising the native conformation of the SPM domain, reducing the proportion of misfolded protein.

However. without more structural information about these residues, it is difficult to predict which other residues they interact with. Fourth, the mutation from the bulkier T16 to the smaller alanine may increase the structural flexibility of the loop region that may be crucial for cleavage activity. In other enzymes, loop regions have similarly been found to be important for catalysis (Nestl and Hauer, 2014). Future site-directed mutagenesis experiments would be informative in unveiling the role of these residues in the cleavage mechanism.

The fast-cleaving reaction of FrpA SPM has a half-life of 27 s and an overall cleavage yield of about 96%. This is an improvement from FrpC SPM, whose fast-cleaving reaction half-life is 2.5 min and cleavage yield is 89%.

The biphasic nature of SPM cleavage could be explained by a number of possible reasons, including SPM adopting different conformational states. A difference in chemical structure of the fast and slow-cleaving populations due to post translational modifications (PTMs) is unlikely, since a single peak was observed for OAZ-SPM during RapidFire MS (Scheu et al., 2021).

A population of stable misfolded SPM may lead to the slower cleavage reaction and explain why cleavage never reached 100 %. Another possibility is that cis-trans isomerisation of the Pro in the scissile Asp-Pro bond explains the slow cleaving phase. Although simulations by Kuban *et al.* suggested that the cis isomer of the Asp-Pro amide bond (ω_{414}=0°C) is unlikely to occur (Kuban *et al.*, 2020), the role of cis-trans isomerisation of Pro in SPM cleavage could possibly be tested by co-incubating SPM with calcium and prolyl isomerase to see if there is a change in the cleavage rate and extent, assuming the Pro is accessible to the prolyl isomerase.

There have been various hypotheses for the rate-limiting step in the cleavage of Asp-Pro bonds. Osicka *et al*. proposed that the rate-limiting step in the autocatalytic

cleavage of SPM is the protonation of Pro nitrogen in the scissile bond (Osicka et al., 2004). However, there has been no direct evidence to support this hypothesis; site-directed mutagenesis of Asp residues proximal to the cleavage site purported to promote the protonation ruled out the catalytic role of the Asp residues (Kuban *et al.*, 2020). More work such as NMR studies will be required to investigate the catalytic mechanism and identify the rate-limiting step in the cleavage of SPM.

The enhanced cleavage rate and yield of FrpA SPM should be beneficial when applying NeissLock *in vivo* or *ex vivo* because an enhanced efficacy means that lower concentrations of the binder protein with SPM is required to achieve the same level of conjugation. The rapid activation of SPM also means that a shorter duration of incubation with the target is required. Using lower doses of the binder protein with SPM translates to potentially fewer off-target effects.

The other SPM homologues from *Alysiella filiformis* and *Kingella negevensis* exhibited slower cleavage but retained calcium-inducible cleavage activity, despite overall low sequence homology to FrpC SPM. It is unclear if, like FrpC, these proteins have native functions in bacterium-host adhesion. These alternative SPM homologues serve as potential alternative candidates for use in NeissLock when a short activation time may not be a priority. For instance, these SPM homologues may be more effective in people who are carriers of *Neisseria meningitidis* or who have had past meningococcal infections. The potential pre-existing antibodies against FrpA and FrpC (Osicka et al., 2001) may lead to premature clearance of the SPM fusion protein (Gorovits et al., 2016).

Working towards a cleavage yield of more than 95 % for SPM and to improve the stability of the domain beyond naturally occurring SPM homologues, I set up an intimin-based bacterial surface display system for directed evolution of the SPM

domain. SPM, when fused to the C-terminus of a truncated Eae intimin, was translocated across the outer membrane of *E. coli* and expression of the fusion protein was detected by flow cytometry. The SPM domain retained its cleavage activity based on flow cytometry analysis of epitope tags flanking each end of SPM. Comparing the wild-type and L23Y mutant of SPM, I observed that the surface display system was sensitive to cleavage rate of the SPM domain. The slower cleaving L23Y mutant showed a higher anti-myc fluorescence after incubation with calcium for a short time compared to wild-type SPM incubated with calcium for the same duration.

One limitation I faced was the smaller than expected library size used for sorting. In published studies employing bacterial surface display, the naive library usually consisted of 10^6 - 10^7 clones (Salema et al., 2016a, 2016b). However, my initial library was at least 10-fold smaller. The small initial library size in TG1 cells could be due to toxicity from expression of the intimin-SPM fusions. The addition of glucose to the growth media and agar could reduce leaky expression that is a potential cause of toxicity. Optimising the IPTG concentration for milder induction (Salema and Fernández, 2017) or changing to arabinose induction that is more titratable may help improve cell viability as well (Khlebnikov et al., 2000). Another possible reason for the small library size is the limited transformation efficiency of the *E. coli* strain used during library generation. Using electrocompetent BL21 cells instead of chemically competent BL21 cells may increase the library size.

The sequence diversity may have been further limited when transforming the library isolated from the cloning strain into the expression strain. Alternatively, the mutation frequency of the library could be varied to generate separate low and high mutation frequency libraries that could be separately sorted to increase the chances of identifying beneficial mutations.

The enrichment of non-functional SPM variants rather than better-cleaving variants may have been due to a growth bias of different clones within the library during overnight growth. The growth of clones in the library with lower levels of SPM expression may have been favoured. To circumvent this problem, it may be advantageous to isolate the plasmid from sorted clones and re-transform into a new batch of *E. coli* for the second round of protein induction and cell sorting (Salema et al., 2016b).

One possible explanation for the higher proportion of non-expressing cells in the library compared to cells expressing the wild-type variant observed in Section 5.4 is a lower cell viability. The viability of cells in the library could be optimised by testing different dilutions when setting up the overnight cultures from the pooled library colonies. Additionally, the addition of glucose in the agar plates and liquid media used for the library may reduce leaky expression that is a potential cause of toxicity (Salema et al., 2016b). The viability of clones in the library can be verified by live/dead cell staining using propidium iodide to ensure that the staining observed is not an artefact of dead cells.

Moving forward, the bacterial display system established in this chapter could be further developed for selection of new binder-SPM fusions against arbitrary protein targets. I demonstrated in this chapter that the intimin-SPM construct could conjugate to a nucleophile like biotin-PEG_4-alkoxyamine that serves as a proxy for the target protein in NeissLock coupling. By fusing a binding moiety e.g., peptide in between intimin and SPM domain, randomisation of the binding moiety will generate a library that can be sorted to enrich for covalently reactive binders against the selected target. Figure 5.11 illustrates how bacterial surface display can be used for directed evolution of covalent binders against arbitrary targets.

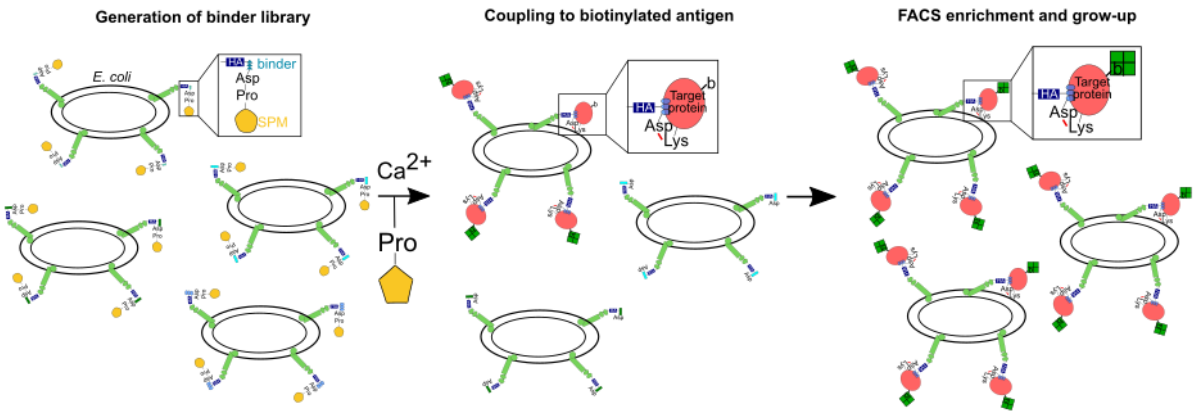


Figure 5.11 Intimin-based bacterial display for selection of covalent binders against arbitrary protein targets. Library of *E. coli* expressing intimin(green)-HA(dark blue)-binder-SPM(yellow) with the binder region randomised is generated. The cells are incubated with the biotinylated bait antigen (red) and calcium is added to induce covalent isopeptide bond formation (red line). Cells are labelled with fluorescently-labelled streptavidin (dark green) that binds the biotinylated target. FACS is used to enrich for cells expressing the binder variants that are covalently reactive.

This sorting strategy capitalises on the use of bacterial display for the sorting of non-covalent binders via high-throughput FACS and the anhydride reactivity of the SPM domain. Extensive wash steps could be used to distinguish between non-covalent binders and covalently reactive binders.

All in all, the intimin-based bacterial display and sorting needs further optimisation to identify beneficial mutations that enhance the stability and cleavage activity of SPM. Nevertheless, establishing the bacterial display system is a step towards developing a second generation NeissLock with enhanced efficacy.

Chapter 6: Extending NeissLock for mammalian binding proteins and red blood cell targets

In this chapter, my goal is to extend the use of NeissLock beyond pre-existing protein complexes from the PDB that were identified through NeissDist. In the first part of this chapter, I develop and characterise the split-SPM system so that binder-SPM fusions can be expressed and purified from mammalian cells. Complex post-translational modifications (PTMs) like disulfide bonds and glycosylation are present on a large number of protein domains and these PTMs are essential for the folding and function of many proteins (Olzscha, 2019). Therefore, establishing a method for eukaryotic expression of binder-SPM fusions with high yields expands the possibilities for future therapeutic and diagnostic applications of NeissLock.

In the second part of this chapter, I make initial steps to develop a modular system for covalent coupling to arbitrary protein targets, by fusing peptides that serve as a binding scaffold to SPM. I fused various human red blood cell-binding peptides to SPM to test for coupling to human red blood cells. Working towards the use of NeissLock for red blood cell decoration, I evaluate various red blood cell surface antigens to identify potential target proteins for coupling.

6.1. Engineering split SPM for mammalian expression

6.1.1. Mammalian expression of SPM

In the first-generation of NeissLock, binder-SPM fusions i.e., OAZ-SPM and TGFα-SPM were expressed and purified solely from bacterial cells. Given that OAZ is an intracellular protein with no complex PTMs like disulfide bonds, OAZ was easily

overexpressed and purified from *E. coli.* In the second protein complex used for NeissLock coupling, TGFα/EGFR, TGFα is a secreted growth factor with three disulfide bonds in its mature soluble form (Singh and Coffey, 2014). Previous studies generally used refolding of inclusion bodies in order to purify recombinant TGFα with high yields and activity from *E. coli* (Boyles et al., 2016; Winkler et al., 1986). However, refolding of inclusion bodies is a time-consuming process that requires careful optimisation to obtain high yields of the biologically-active recombinant protein (Yamaguchi and Miyazaki, 2014). Expressing and purifying TGFα from mammalian cells may be an alternative method for obtaining a larger fraction of the properly folded growth factor.

However, I hypothesised that directly expressing TGFα-SPM in mammalian cells and targeting it for secretion to the extracellular media may lead to premature activation and cleavage of SPM intracellularly and/or in the extracellular media. In a resting state, the cytosolic calcium concentration in eukaryotic cells is kept low at approximately 0.1 μM, while the calcium concentrations in components of the secretory pathway like the Endoplasmic Reticulum (ER) is approximately 0.1-0.8 mM and extracellular calcium concentrations range from 1-2 mM (Bagur and Hajnóczky, 2017; Verkhratsky, 2005).

Therefore, I envisioned the split-SPM system as a modified form of NeissLock that is compatible with mammalian-expression of the binder-SPM fusion. In the split-SPM system, the SPM domain is split into two parts: an N-terminal and a C-terminal fragment. The N-terminal fragment of SPM (SPM_N) is inactive itself unless it is reconstituted with the C-terminal fragment (SPM_C). The binder protein (red) is expressed in mammalian cells as an N-terminal fusion to SPM_N (orange) (Figure 6.1a). When SPM_C (orange) is added, SPM is reconstituted. Similar to full length SPM,

calcium addition generates an aspartic anhydride (yellow star) that is susceptible to hydrolysis or reaction with nucleophiles on the target protein (green) (Figure 6.1a).

To test my initial hypothesis that SPM-fusions could be prematurely activated when expressed in mammalian cells in its full-length form, I transfected Expi293F cells with Superfolder Green Fluorescent Protein (sfGFP) fused to full-length FrpA SPM at the C-terminus and with an additional N-terminal tissue-type plasminogen activator (tPA) signal peptide to target the protein for secretion. sfGFP was used as a simple model protein to ease interpretation of the results. The sfGFP-fusion protein was purified using Ni-NTA from the media 5 days post-transfection. In the left panel of Figure 6.1b, the 30 kDa band observed in the elution fraction likely corresponds to the cleavage product AP-His6-sfGFP while the 56 kDa band likely corresponds to the pre-cleavage form, AP-His6-sfGFP-SPM. It is unclear why other higher molecular weight bands around 70 kDa were observed, since neither N-linked glycosylation consensus sequences nor potential disufide bond formation sites can be found within sfGFP (Aronson et al., 2011; Costantini and Snapp, 2013) or SPM. One possible explanation is that the higher molecular weight band around 70 kDa corresponds to the N-terminal cleavage fragment, AP-His6-sfGFP, covalently coupled to a second protein via the aspartic anhydride. Nevertheless, the observation of a band corresponding to the cleavage fragment verifies that cleavage of the SPM domain may occur when SPM fusion proteins are targeted for secretion from mammalian cells.

For comparison, I fused SPM_N, spanning residues 300 to 318 of FrpA SPM in this case, to the receptor binding domain (RBD) of SARS-CoV-2 spike protein. An additional SpyTag sequence was added at the N-terminus to facilitate purification by Spy&Go (Khairil Anuar et al., 2019). Spy&Go purification of the media from transfected Expi293 cells led to the observation of single major species with molecular weight of

approximately 30 kDa in the elution fraction (Figure 6.1b right). The observed molecular weight for SpyTag-RBD-SPM$_N$ was higher than the expected 27 kDa, likely due to disulfide bridges and glycosylation in the RBD (Rahikainen et al., 2021). A yield of about 677 µg was obtained from a 20 mL Expi293 culture, suggesting that SPM$_N$ fusion proteins can be expressed and purified to high yields from mammalian cells.

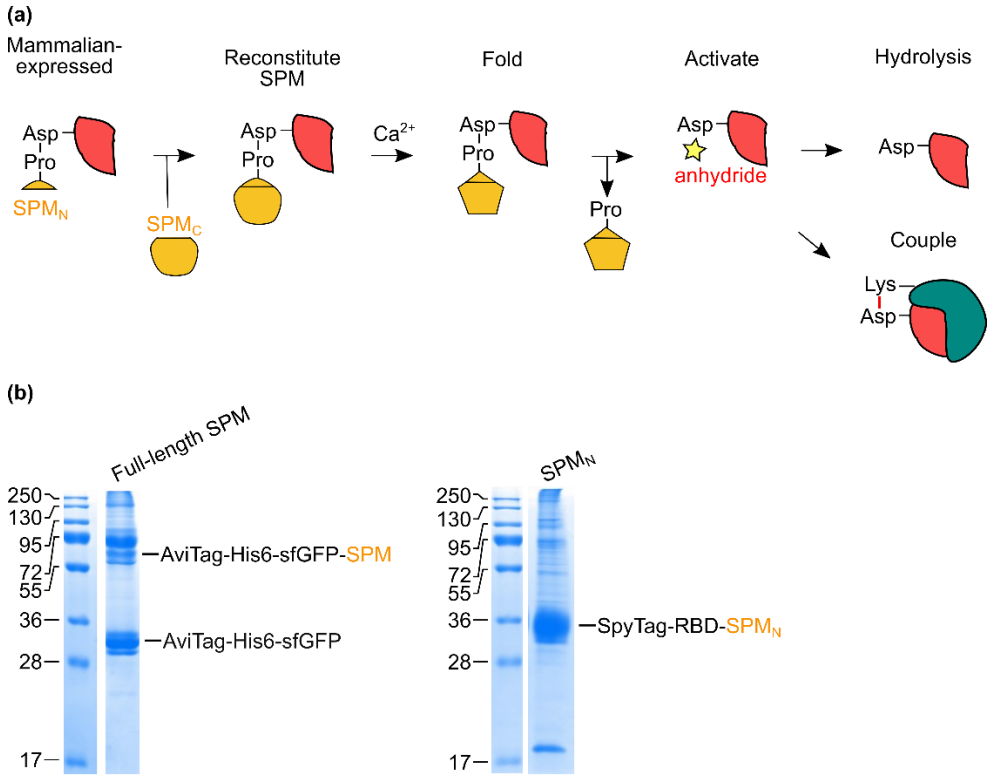


Figure 6.1 Split SPM system for mammalian expression of binder-SPM fusions. (a) Schematic for split SPM system. Binder protein (red)-SPM$_N$ (orange) is expressed in mammalian cells. When reconstituted with bacterial-expressed SPM$_C$ (orange) and in the presence of Ca^{2+}, SPM is activated and autoproteolysis leads to formation of anhydride (yellow star). If the binder protein docks to the target protein (green), coupling can occur. Otherwise, the anhydride can undergo hydrolysis or self-reaction. (b) Coomassie-stained SDS-PAGE showing purified sfGFP-SPM(full-length) and the sfGFP cleavage product from the supernatant of transfected Expi293F cells (left), and purified SpyTag-RBD-SPM$_N$ from the supernatant of transfected Expi293F cells (right).

6.1.2. Optimising the SPM split site

To verify the feasibility of the split-SPM system, I used the simpler prokaryotic expression system to validate SPM reconstitution, as well as calcium-inducible cleavage and conjugation post-reconstitution.

Initially, I decided on the split site in the SPM domain based on the secondary structure prediction by JPred (Drozdetskiy et al., 2015) and decided to have SPM_N span the first two predicted β-strands from start to end i.e., residues 300 to 315. Previous studies have shown that split sites in loop regions are generally better tolerated than in the middle of structured regions (Lim and Wells, 2020a). When the NMR structure of SPM was later elucidated, it was shown that the second β-strand of calcium-bound SPM terminates at position 316 rather than 315 (Figure 6.2b and c). Nevertheless, when SPM_{N315} was fused to the C-terminus of OAZ, I found that the fusion protein expressed with reasonable yield (7 mg/L) and purity. I separately expressed and purified the C-terminal fragment of SPM (SPM_C) using C-tag affinity purification, starting at residue 316 of FrpA SPM.

After overnight incubation with Ca^{2+}, I observed that OAZ-SPM_{N315} alone did not show any sign of cleavage (Figure 6.2a lanes 1 and 2). Conversely, when SPM_{C316}-Ctag was added to OAZ-SPM_{N315}, OAZ-SPM_{N315} was cleaved, leading to the formation of 2 additional bands corresponding to the linear or cyclized OAZ (Figure 6.2a lanes 7 and 8). Surprisingly, when only ODC was added to OAZ-SPM_{N315}, OAZ-SPM_{N315} was cleaved as well (Figure 6.2a lanes 9 and 10), although to a lesser extent than when SPM_{C316}-Ctag was present. When all three proteins, OAZ-SPM_{N315}, SPM_{C316}-Ctag and ODC were present, calcium addition led to the formation of an ODC:OAZ conjugate (Figure 6.2a lanes 11 and 12).

However, conjugation to ODC with the split-SPM system was slow and 50% of the maximum conjugation was observed only after 6 h incubation with the SPM_{N315} and SPM_{C316} pair (Figure 6.2d). To enhance the coupling rate with the split-SPM system, I generated a series of SPM_N and SPM_C variants with different split sites, to identify the split pair that gives the fastest coupling whilst maintaining reasonable stability and expression yields for the individual split fragments (Figure 6.2c). Truncation of SPM_N to SPM_{N312} and SPM_{N309} completely abolished coupling when SPM_{C313} and SPM_{C310} were added and so were excluded from further analysis.

Lower initial starting concentrations of the proteins i.e., 5 μM were used in the assay for coupling rates of different combinations of split pairs, due to the low yields of some of the SPM_C variants. The combination of SPM_{N318} and SPM_{C322} led to an overall highest conjugation yield and an approximately two-fold increase in the conjugation rate compared to the original split pair i.e., SPM_{N315} and SPM_{C316} (Figure 6.2d). It was surprising to observe that exclusion of residues 319-321 in the unresolved loop of SPM led to the fastest conjugation.

The OAZ-SPM_N fusion proteins expressed well, with yields of more than 10 mg /L culture. However, the yields of SPM_C fusion proteins were a lot lower. Initially, I obtained only 800 μg/ L culture of SPM_{C322}-SpyCatcher003 after C-tag purification. To increase the yield, I fused a His_6-SUMO solubility tag to the N-terminus of SPM_{C322}-SpyCatcher003. Following Ni-NTA purification and cleavage by the Ulp1 SUMO protease, the final yield of SPM_{C322}-SpyCatcher003 increased to approximately 4.6 mg/L.

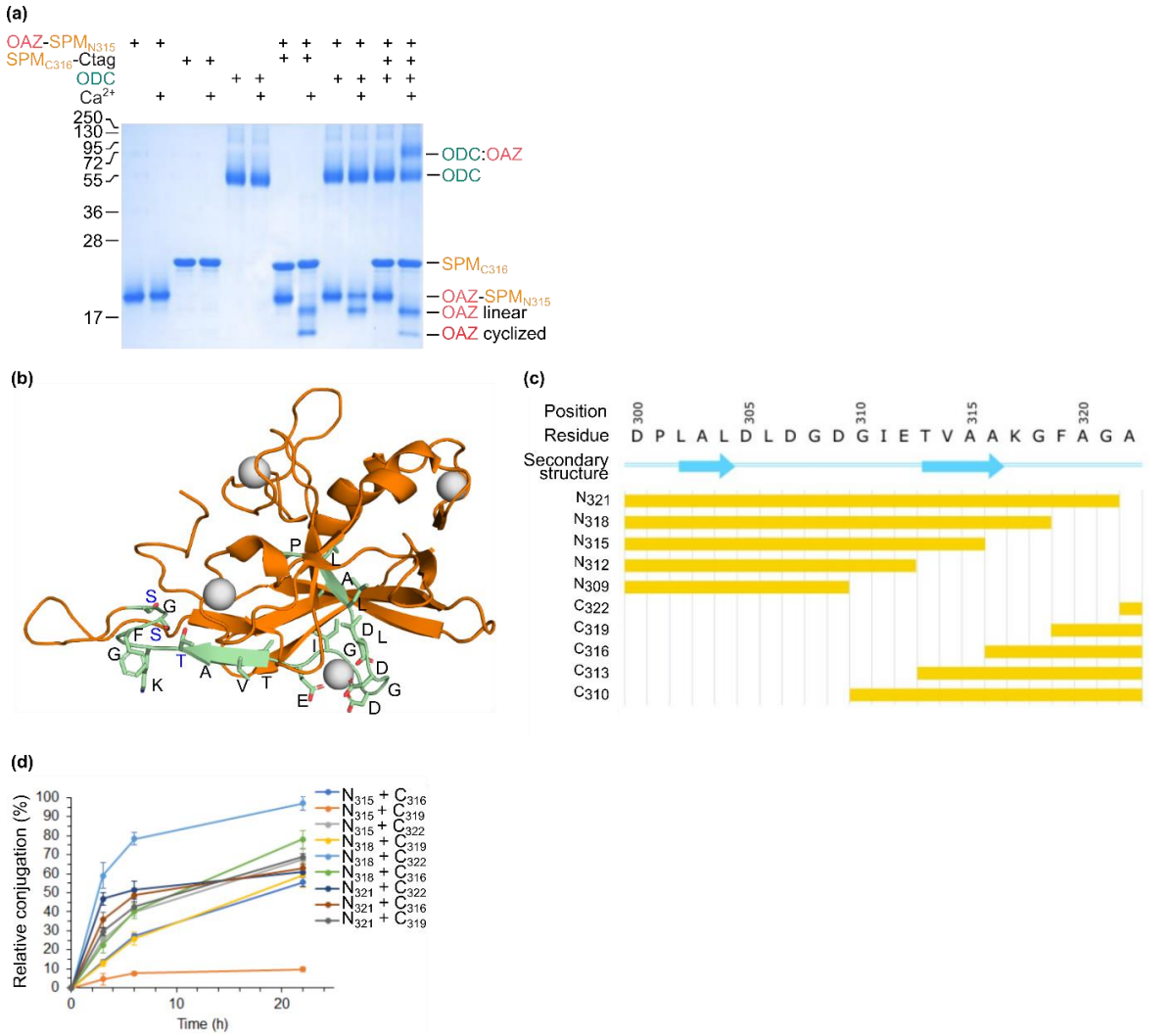


Figure 6.2 Optimising SPM split site for enhanced conjugation rate. (a) 10 μM of OAZ-SPM$_{N315}$ was incubated with 10 μM of SPM$_{C316}$ with or without 10 μM ODC and 10 mM Ca^{2+} in HEPES reaction buffer, pH 7.4, at 37 °C for 16 h. After the reaction was stopped, samples were run on SDS-PAGE and Coomassie-stained. (b) Structure of calcium-bound FrpC SPM (PDB 6SJW) with side-chains of SPM$_N$ (green) highlighted as sticks. Residues in SPM$_N$ are labelled with their one-letter code and residues that differ between FrpC and FrpA are highlighted in blue. Calcium ions are illustrated as grey spheres. (c) Schematic showing alternative SPM$_N$ and SPM$_C$ C- and N-termini respectively. Residues and numbering of positions are based on FrpA SPM. β-strands are illustrated as arrows. (d) Time-course of conjugation to ODC with different combinations of SPM$_N$ and SPM$_C$ shown in (b). 5 μM of OAZ-SPM$_N$ was incubated with 5 μM of SPM$_C$, 5 μM ODC and 10 mM Ca^{2+} in HEPES reaction buffer, pH 7.4, at 37 °C. Conjugation at various time points was based on densitometry analysis of Coomassie-stained SDS-PAGE as in (a) (mean ± 1 S.D., n=3).

6.1.3. Fusion of SpyTag003/SpyCatcher003 to increase rate of activation

A likely reason for the observed slow coupling rate with the split-SPM system is a low binding affinity between the N- and C- terminal fragments of SPM, leading to slow reconstitution. I hypothesised that fusing the individual SPM fragments with a second protein pair with a higher binding affinity for each other may increase the reaction rate. The fusion of split fragments to a second protein pair is analogous to the fusion of dimerization domains to split enzymes so that dimerization of the domains, brings the split enzyme fragments together and activates the enzyme (Lim and Wells, 2020b). I decided to fuse to split-SPM the SpyTag003/SpyCatcher003 peptide-protein pair whose reaction rate is close to the diffusion limit (Keeble et al., 2019). Apart from a faster on-rate, the isopeptide bond formation between SpyTag003 and SpyCatcher003 will bring the split-SPM fragments into close proximity for a longer period of time.

Figure 6.3a shows an illustration of the split-SPM system with fusion of SpyTag003/SpyCatcher003. The binder(red)-SPM_N fusion is expressed with SpyTag003 (cyan) at its C-terminus with an intervening -GSG linker so that the SpyTag003/SpyCatcher003 reaction would not sterically hinder reconstitution of split-SPM. SPM_C (orange), on the other hand, is fused to SpyCatcher003 (dark blue) at the C-terminus, similarly with an intervening -GSG linker between the two domains. When SPM_C-SpyCatcher003 is added to binder-SPM_N-SpyTag003, SpyTag003 rapidly binds to SpyCatcher003 and an isopeptide bond (red) is formed (Figure 6.3a). The SpyTag003/SpyCatcher003 reaction brings the two fragments of SPM into close proximity, enhancing the rate of reconstitution of SPM and downstream SPM activation (Figure 6.3a).

The reaction between SpyTag003 and SpyCatcher003 when fused to the split-SPM fragments was verified by RapidFire intact mass spectrometry. When 10 μM of OAZ-SPM$_{N318}$-SpyTag003 was incubated with an equimolar amount of SPM$_{C322}$-SpyCatcher003 without Ca^{2+}, a single peak corresponding to OAZ-SPM$_{N318}$-SpyTag003:SpyCatcher003-SPM$_{C322}$ was observed, confirming a loss of water from the isopeptide bond formation between Lys in SpyCatcher003 and Asp in SpyTag003 in the conjugate (Expected: 58,853.41; Observed: 58,854.55). When Ca^{2+} was added, the peak corresponding to the OAZ-SPM$_{N318}$-SpyTag003:SpyCatcher003-SPM$_{C322}$ conjugate was no longer present. Instead, the cleavage product from SPM activation, FrpA$_{N318}$-SpyTag003:SpyCatcher003-FrpA$_{C322}$ was observed (Expected: 41,164.77; Observed: 41,164.14). In addition, excess unreacted SPM$_{C322}$-SpyCatcher003 was observed as well (Expected: 37,312.38; Observed: 37,311.43).

The RapidFire MS data verified that fusion to SpyTag003/SpyCatcher003 did not abolish the cleavage activity of the SPM domain after reconstitution. To compare the cleavage rate of SPM with or without fusion to SpyTag003/Catcher003, I performed a time-course assay of cleavage (Figure 6.3c). When OAZ-FrpA$_{N318}$ was incubated with FrpA$_{C322}$, only 54 % of OAZ-FrpA$_{N318}$ was cleaved after 1 h and the cleavage extent increased to 80 % after 16 h (Figure 6.3d). With the fusion of SpyTag003/SpyCatcher003, after 1 h pre-incubation of the proteins without Ca^{2}, almost no remaining OAZ-FrpA$_{N318}$-SpyTag003 was observed while an excess of FrpA$_{C322}$-SpyCatcher003 was observed (Figure 6.3c, with SpyTag/Catcher at 0 min). Based on the band intensity of the OAZ-SPM$_{N318}$-SpyTag003:SpyCatcher003-SPM$_{C322}$ conjugate, 58 % of reconstituted SPM was cleaved within 5 min and up to 94% of reconstituted SPM was cleaved by the end of 16 h (Figure 6.3d).

Having verified that the fusion of SpyTag003/SpyCatcher003 to split-SPM increased the cleavage rate upon reconstitution, I repeated the investigation of the optimal split site for SPM, with the fusion of the SpyTag003/SpyCatcher003 this time. I generated a series of SPM_N and SPM_C variants fused to SpyTag003 and SpyCatcher003 and tested different combinations of the split pairs. Amongst the combinations of split pairs tested, SPM_{N318} incubated with SPM_{C322} led to the fastest cleavage (Figure 6.3e). When 2 μM of OAZ-SPM_{N318} was incubated with 2 μM of SPM_{C322} and 2 mM Ca^{2}, 58 % of SPM cleaved within 5 min (Figure 6.3e). Truncating the last C-terminal residue in SPM_{N318}, i.e., SPM_{N317}, led to little difference in the cleavage rate when combined with SPM_{C322} (Figure 6.3e).

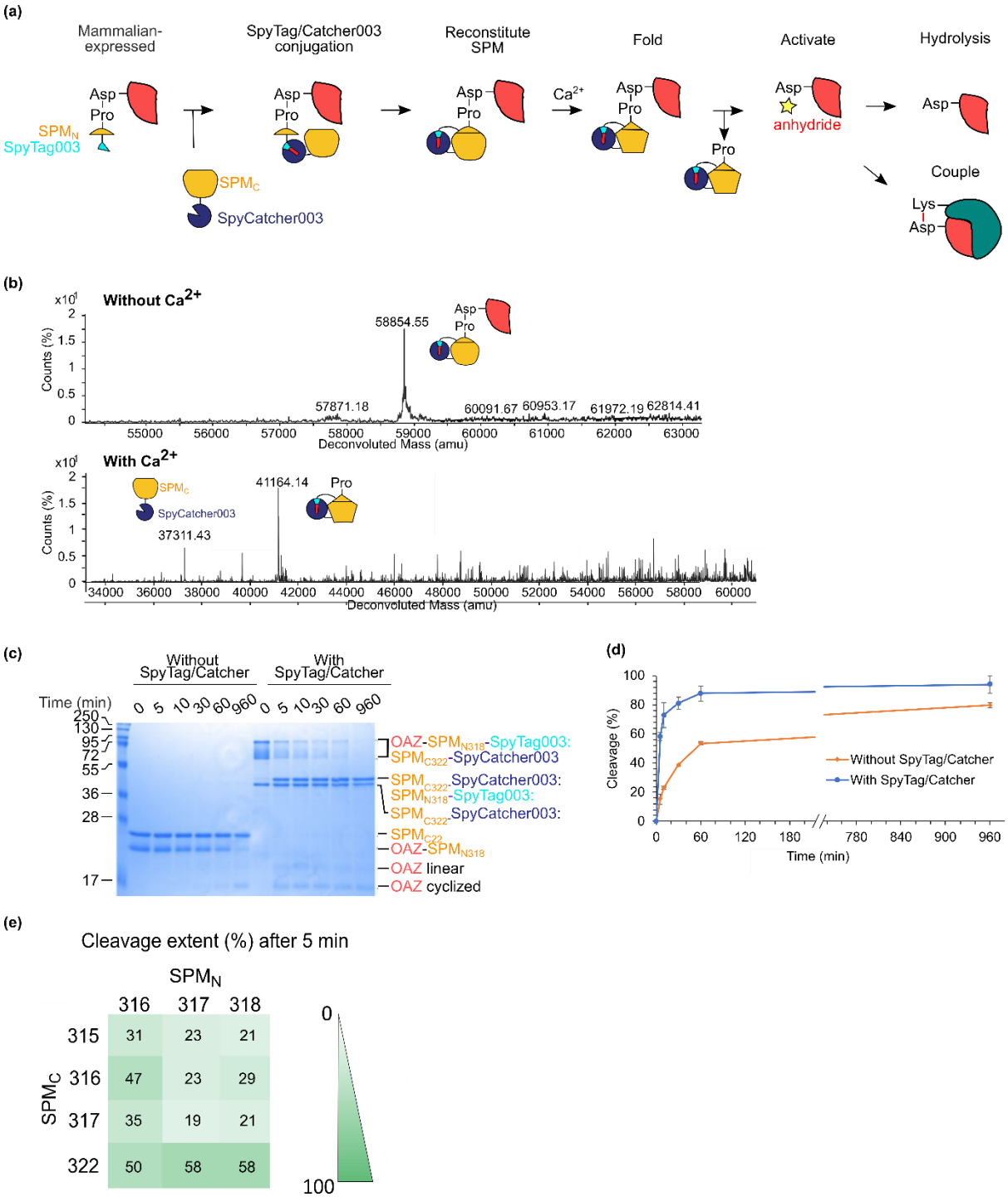


Figure 6.3 Fusion of SpyTag003/SpyCatcher003 to split SPM increase cleavage rate. (a) Schematic for split SPM system with additional SpyTag003/SpyCatcher003 fusion. Binder protein (red)-SPM_N (orange)-SpyTag003 (cyan) can be expressed in mammalian cells. SpyTag003 forms an isopeptide bond (red) with SpyCatcher003 (dark blue) that is fused to SPM_C (orange). After calcium activation of SPM, the anhydride (yellow star) undergoes hydrolysis or couples to a nucleophile on the target protein (green) (b) RapidFire mass spectrometry of 10 μM of OAZ-SPM_{N318}-SpyTag003 incubated with 10 μM of SPM_{C322}-SpyCatcher003 without Ca^{2+} (top panel)

or with 2 mM Ca^{2+} (bottom panel) in HEPES reaction buffer, pH 7.4, at 37 °C. (c) 2 μM of OAZ-SPM_{N318} was incubated with 2 μM of SPM_{C322} with or without SpyTag003/SpyCatcher003 fusion in HEPES reaction buffer, pH 7.4, at 37 °C for 1 h. 10 mM Ca^{2+} was then added and the reaction was stopped at the indicated time points before samples were run on SDS-PAGE and Coomassie-stained. (d) Time-course of SPM cleavage based on densitometry analysis of Coomassie-stained SDS-PAGE gels in (c) (mean ± 1 S.D., n=3). (e) Heat map of cleavage extent of SPM after incubating 2 μM of OAZ-SPM_N-SpyTag003 with 2 μM of SPM_C-SpyCatcher003 and 2 mM Ca^2 in HEPES reaction buffer, pH 7.4, at 37 °C for 5 min following 1 h of pre-incubation without Ca^{2+} (mean ± 1 S.D., n=3).

6.2. Application of split SPM: Conjugation of tPA-TGFα-SPM_N-SpyTag003

In eukaryotic cells, TGFα is synthesised as an inactive propeptide, pre-proTGFα. The N-terminal signal sequence is cleaved during translocation into the ER lumen, producing proTGFα (Wong et al., 1989). proTGFα comprises a N-terminal propeptide, extracellular EGF-like domain, a juxtamembrane domain, a transmembrane domain and a cytoplasmic domain (Singh and Coffey, 2014). Membrane-associated proTGFα undergoes further cleavage by metalloproteases at the N-terminus of the EGF-like domain and the juxtamembrane domain to give the mature, soluble form of TGFα (Singh and Coffey, 2014).

I generated a construct for expressing TGFα-SPM_{N318}-SpyTag003 in mammalian cells by using the coding sequence for mature, soluble TGFα and attaching an N-terminal haemagglutinin (HA) signal sequence to target the protein for secretion. I observed that transfection in ExpiCHO cells gave a low yield of the protein (54 μg/ 25 mL culture) post-Spy&Go purification. However, the yield from transfected ExpiCHO was higher than from Expi293F cells and so I decided to continue with using ExpiCHO cells for expression of TGFα-fusions.

As an example of a therapeutic protein to couple on the surface of cells as cargo for more long-lasting therapeutic delivery, I decided to test the fusion of the Kringle 2 and serine protease domains from human tissue-type plasminogen activator (tPA) to the N-terminus of TGFα-SPM$_{N318}$-SpyTag003 for mammalian expression (Figure 6.5a).

Recombinantly-expressed human tPA, for example Alteplase and Tenecteplase, has been exploited for the treatment of heart attacks and acute ischaemic stroke (Warach et al., 2020). The fibrinolytic activity of tPA stems from its conversion of inactive plasminogen to active plasmin (Warach et al., 2020). Compared to Alteplase, Tenecteplase showed increased resistance to plasminogen activator inhibitor-1 (PAI-1) along with reduced plasma clearance (Paoni et al., 1993). I decided to incorporate the four alanine substitutions in residues 296 to 299 of the protease domain found in Tenecteplase that was shown to increase the half-life of tPA (Keyt et al., 1994).

Recombinant expression of tPA in both prokaryotic (Long et al., 2015; Qiu et al., 1998) and eukaryotic systems such as yeast (Martegani et al., 1992), insect cells (Furlong et al., 1988) and Chinese Hamster Ovary (CHO) cells (Takagi et al., 2000) have been previously reported. The presence of multiple disulfides in tPA, like TGFα, makes *E. coli* expression difficult. Another advantage of expression in CHO cells is that the mammalian glycosylation patterns makes CHO-expressed tPA more similar to endogenous human tPA, reducing the possible immunogenicity of the recombinant protein (Ghaderi et al., 2012).

Transfection of tPA-TGFα-SPM$_{N318}$-SpyTag003 in ExpiCHO, followed by Spy&Go purification of the media post-transfection gave a reasonable yield of about 150 µg / 50 mL culture (Figure 6.4a). During size-exclusion chromatography

purification, apart from some aggregates, the majority of the protein eluted as a single peak, suggesting a homogeneous population (Figure 6.4b).

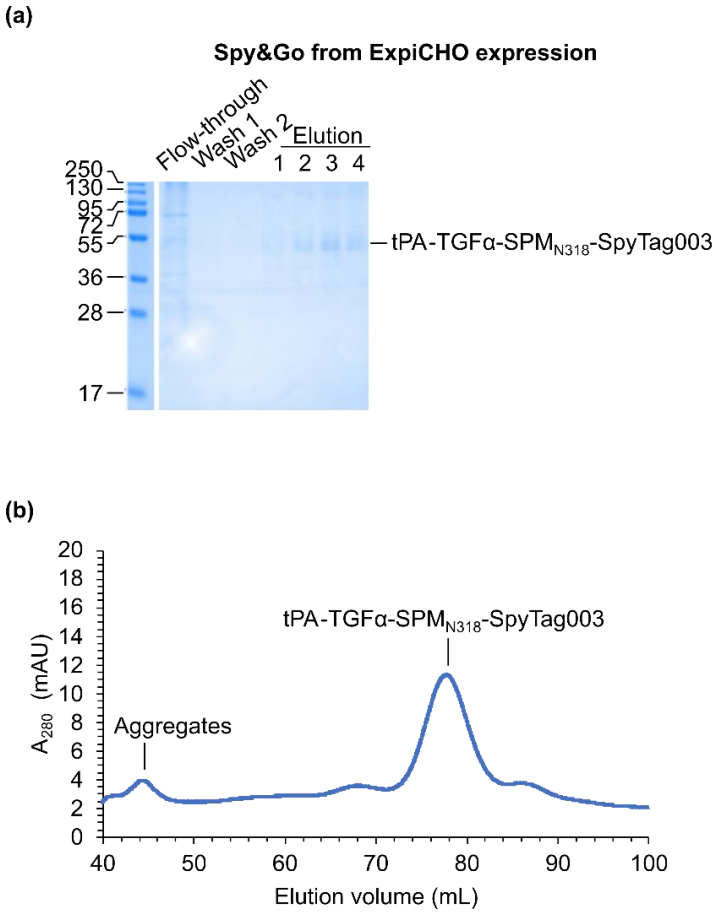


Figure 6.4 Mammalian expression of tPA-TGFα-SPM$_N$-SpyTag003. (a) tPA-TGFα-SPM$_{N318}$-SpyTag003 was purified from the supernatant of transfected ExpiCHO cells using Spy&Go. Fractions were run on SDS-PAGE and then Coomassie-stained. (b) Size-exclusion chromatography profile of tPA-TGFα-SPM$_{N318}$-SpyTag003 after Spy&Go purification. The Superdex 200 16/600 pg column was used.

To test the coupling activity of tPA-TGFα-SPM$_{N318}$-SpyTag003, I incubated tPA-TGFα-SPM$_{N318}$-SpyTag003 with SPM$_N$322-SpyCatcher003 as well as the target protein, the soluble ectodomain of Epidermal Growth Factor Receptor (sEGFR) (Figure 6.5b). A five-fold excess of TGFα to sEGFR was added to increase the proportion of EGFR that is coupled. After the conjugation reaction was stopped, samples were digested with PNGase F that deglycosylates mammalian-expressed

sEGFR as well as tPA-TGFα-SPM$_{N318}$-SpyTag003 to give a sharp band for each protein.

There was no observed change in the proteins, tPA-TGFα-SPM$_{N318}$-SpyTag003, SPM$_N$322-SpyCatcher003 or sEGFR when each protein was incubated alone, with or without Ca^{2+} (Figure 6.5c lanes 2-7). When tPA-TGFα-SPM$_{N318}$-SpyTag003 was incubated with SPM$_N$322-SpyCatcher003, additional higher molecular weight bands were present, likely corresponding to the tPA-TGFα-SPM$_{N318}$-SpyTag003:SpyCatcher003-SPM$_{N322}$ conjugate (Figure 6.5c lane 8). The presence of bands corresponding to residual tPA-TGFα-SPM$_{N318}$-SpyTag003 and SPM$_N$322-SpyCatcher003 suggests that the SpyTag003/SpyCatcher003 reaction did not go to completion (Figure 6.5c lane 8). When calcium was added, the intensity of the bands corresponding to tPA-TGFα-SPM$_{N318}$-SpyTag003:SpyCatcher003-SPM$_{N322}$ complex were reduced and two new bands corresponding to the cleavage products tPA-TGFα and SPM$_{N318}$-SpyTag003:SpyCatcher003-SPM$_{N322}$ were observed, confirming cleavage of SPM. In the last sample with both sEGFR and calcium present, a new band corresponding to the tPA-TGFα:sEGFR covalent complex was formed, accompanied by a decrease in the band intensity of sEGFR (Figure 6.5c lane 11). An anti-EGFR Western blot confirmed the identity of the covalent complex (Figure 6.5d). Based on the decrease in band intensity of EGFR, a conjugation yield of about 59 % was observed.

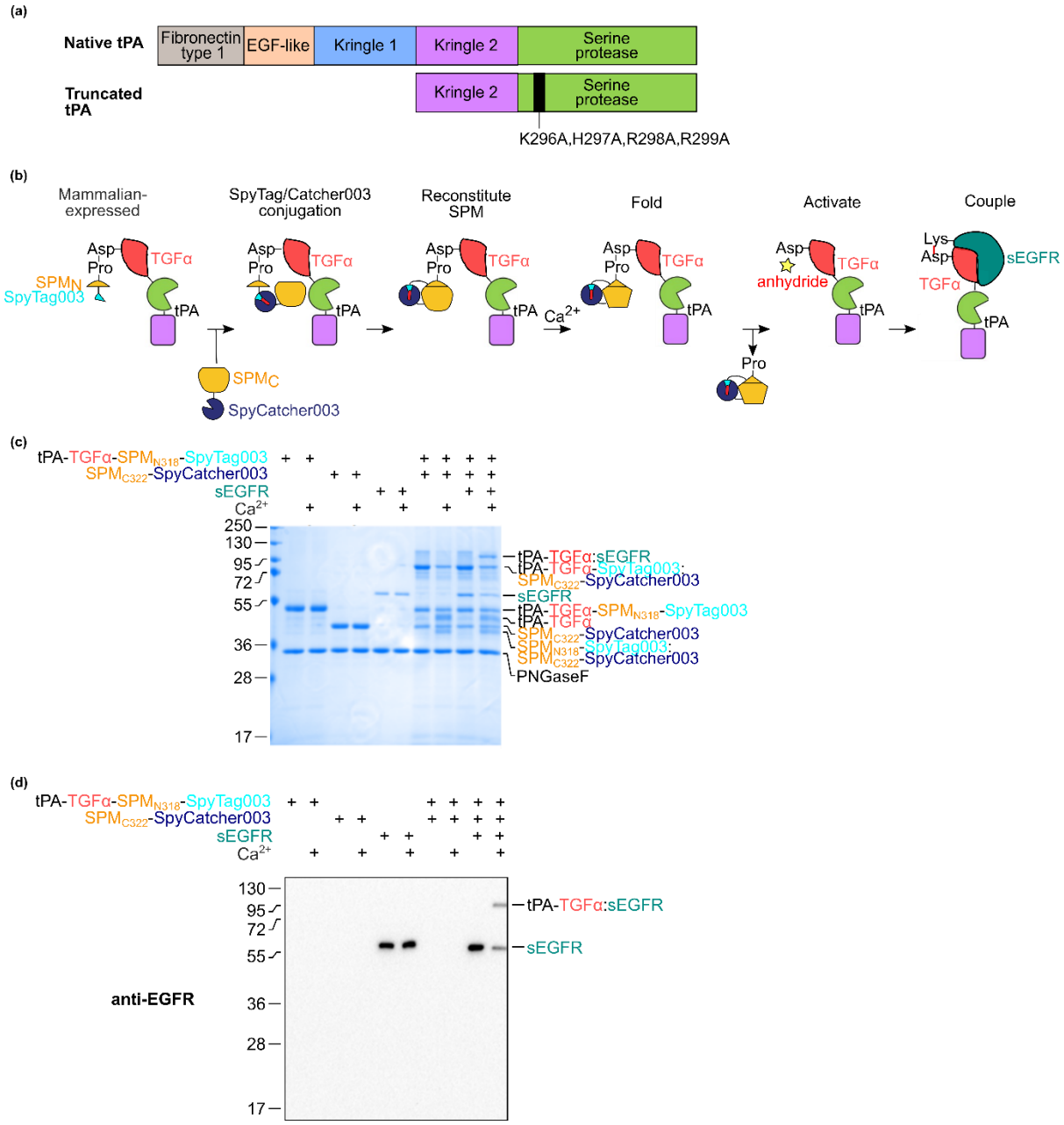


Figure 6.5 Conjugation of mammalian-expressed tPA-TGFα-SPM$_N$-SpyTag003 to sEGFR. (a) Domain structures of native tPA and truncated tPA with tetra-alanine mutations that was used. (b) Cartoon illustrating conjugation of tPA-TGFα-SPM$_{N318}$-SpyTag003 to sEGFR (dark green). Truncated tPA consists of the Kringle 2 domain (pink) and serine protease domain (light green). (c) 7 μM of tPA-TGFα-SPM$_{N318}$-SpyTag003 was incubated with 7 μM of SPM$_{C322}$-SpyCatcher003 in HEPES reaction buffer, pH 7.4, at 37 °C for 1 h. 2 mM Ca^{2+} and 1.4 μM of sEGFR was then added and incubated for another 1 h at 37 °C, before the reaction was stopped by EDTA. Samples were digested with PNGaseF under denaturing conditions before running on SDS-PAGE and Coomassie-staining. (d) Anti-EGFR Western blot of SDS-PAGE gel as in (c).

6.3. Towards a modular system of utilising binding scaffold fused to SPM

In the second part of this chapter, I work towards establishing a modular system for NeissLock. The aim is to enable coupling to any protein, beyond those whose structures are available on the PDB and/or where a binder protein has been identified. Expanding the use of NeissLock beyond complexes shortlisted by NeissDist is an important step towards wider applicability.

A modular system can be achieved by fusing a binding scaffold to the N-terminus of SPM. The binding scaffold would consist of a variable region, analogous to the complementarity-determining (CDR) region of an antibody. The use of an evolvable binding scaffold means that directed evolution and library selections can be used to generate a novel binder for every arbitrary target protein whilst ensuring that the binder simultaneously fulfils the requirements for NeissLock coupling. For example, having a lysine on the target protein near the binding interface whilst avoiding having lysines on the binder protein itself.

However, identifying a suitable binding scaffold was challenging as many of the existing binder scaffolds including affibodies, DARPins, nanobodies and adnectins have C-termini that lie on the opposite face to their binding interfaces with the target proteins (Vazquez-Lombardi et al., 2015). An alternative to using existing antibody-like binding scaffolds is to fuse a short peptide to the N-terminus of SPM. The small size of peptides ensures that the binding interface with the target protein would be in closer proximity to the C-termini of the binding peptides.

When evaluating various potential applications for NeissLock, we identified red blood cell coupling as one relevant and feasible application. Given the extensive studies on the surface antigens of red blood cells (Daniels, 2007; Ravenhill et al.,

2019; Reid and Mohandas, 2004), identifying target proteins for coupling was more straightforward. The longer lifespan of red blood cells (Thiagarajan et al., 2021) meant that the more stable covalent coupling provided by NeissLock could be used to enhance drug delivery whilst using red blood cells as carriers. Coupling to red blood cells can prolong the circulating half-life of the therapeutic protein whilst minimising side-effects (Muzykantov, 2010). For instance, when free tPA was injected into the blood stream, apart from the therapeutic fibrinolysis at the intended site, diffusion of tPA into pre-existing haemostatic clots in the walls of blood vessels may lead to bleeding whilst leakage into other tissues like the central nervous system (CNS) leads to toxicity (Greineder et al., 2013). The adverse side-effects of free tPA are often exacerbated by the high doses used to overcome rapid inactivation by plasminogen activator inhibitor-1 (PAI-1) and clearance by the liver (Greineder et al., 2013). When tPA was coupled to red blood cells, Murciano *et al.* observed less leakage to existing hemostatic clots or other extravascular tissues like the CNS (Murciano et al., 2003). With coupling to red blood cells, the circulation time of tPA was increased from minutes to days (Murciano et al., 2003; Zaitsev et al., 2010).

Other methods like using sortase-mediated transpeptidation (Pishesha et al., 2017), red blood cell-targeting antibodies (Zaitsev et al., 2006), biotin:avidin interactions (Murciano et al., 2003) and nanoparticles (Brenner et al., 2018) have been used for red blood cell surface decoration. There are several potential advantages with using NeissLock compared to other red blood cell decoration methods. Firstly, compared to methods that rely on non-covalent interactions, the covalent conjugation with NeissLock increases the stability of the cell surface decoration. Secondly, with the use of NeissLock to target endogenous proteins, there is no need for modification

of the red blood cell membrane proteins. Thirdly, the minimal scar with NeissLock coupling reduces potential immunogenicity post-coupling.

6.3.1. Validation of covalent coupling of UbqK0-HA-ERY-GSY-SPM

When selecting peptide binders against red blood cells, I first looked at validated peptide binders from the published literature. Several peptides derived from *Plasmodium* proteins were identified to bind to human red blood cells (López et al., 2004; Puentes et al., 2000; Vera Bravo et al., 2000). However, given that the *Plasmodium* species are responsible for malaria, it was unclear if these peptides would be effective binders as foreign sequences. Another criterion that I used to shortlist red blood cell-binders was that the peptide should not contain lysine as the lysine may lead to self-reaction upon SPM activation.

I decided to first test the synthetic human red blood cell-binding peptides identified by the Hubbell lab using phage display (Hubbell et al., 2014). ERY64, ERY123 and ERY141 are peptide ligands found to specifically bind human erythrocytes, but not mouse erythrocytes (Hubbell et al., 2014). All three peptides do not contain lysine. However, the identity of the red blood cell protein(s) the ERY peptides bind to is unknown. As a negative control for red blood cell binding, I included the 1DN2 peptide, which is a 13-residue long synthetic peptide that binds that Fc-region of human IgG1 antibodies (L. et al., 2000). There have been no reports of the 1DN2 peptide binding to human red blood cells.

I fused full-length SPM to the C-terminus of each peptide binder and included an additional modified ubiquitin (UbqK0) at the N-terminus. The additional UbqK0 domain brings the C-terminus of the binder further away from the N-terminal α-amine, since the N-terminus and C-terminus of ubiquitin are on opposite faces of the domain,

reducing the likelihood of the anhydride reacting with the α-amine. In addition, the ubiquitin protein was engineered to be lysine-free (Huang et al., 2014), reducing the likelihood of self-reaction between the anhydride and ε-amines on the binder. An intervening HA-tag that similarly does not contain lysine was added to detect coupling between UbqK0-peptide binder and the target protein.

After calcium addition to UbqK0-HA-ERY-SPM, SPM is activated, and this would allow for coupling between the anhydride generated at the C-terminus of the ERY peptide and a lysine on the target protein (Figure 6.4a).

I tested the cleavage and coupling activity of UbqK0-HA-ERY-SPM by incubating the proteins with calcium, either with or without biotin-PEG_4-alkoxyamine. Since the targets of the ERY peptides are unknown, biotin-PEG_4-alkoxyamine served as a proxy for the nucleophile(s) present on the target protein. Incubation of UbqK0-HA-ERY-SPM with calcium led to the formation of the expected cleavage products, P-SPM as well as UbqK0-HA-ERY-GSY-D (Figure 6.4b), confirming that SPM retained its cleavage activity when fused to a peptide binder. The ERY variants when fused to SPM were found to cleave at different rates. After 10 min of incubation with Ca^{2+}, 19 % of ERY123 was cleaved compared to 42 % cleavage for ERY64 and 49 % cleavage for ERY123. Two separate bands were observed for the proximal cleavage fragment, UbqK0-HA-ERY-GSY-D, suggesting that self-reaction was not completely abolished, despite the use of lysine-free ubiquitin and an HA-tag at the N-terminus that inhibits self-reactivity with ε-amines.

Having verified the cleavage activity of the ERY peptides fused to SPM, the purified proteins were sent to a collaborator, Ash Toye's lab in the University of Bristol to test for covalent conjugation to the Bristol Erythroid Line Adult (BEL-A) cell line, an immortalised human adult erythroid cell line (Trakarnsanga et al., 2017).

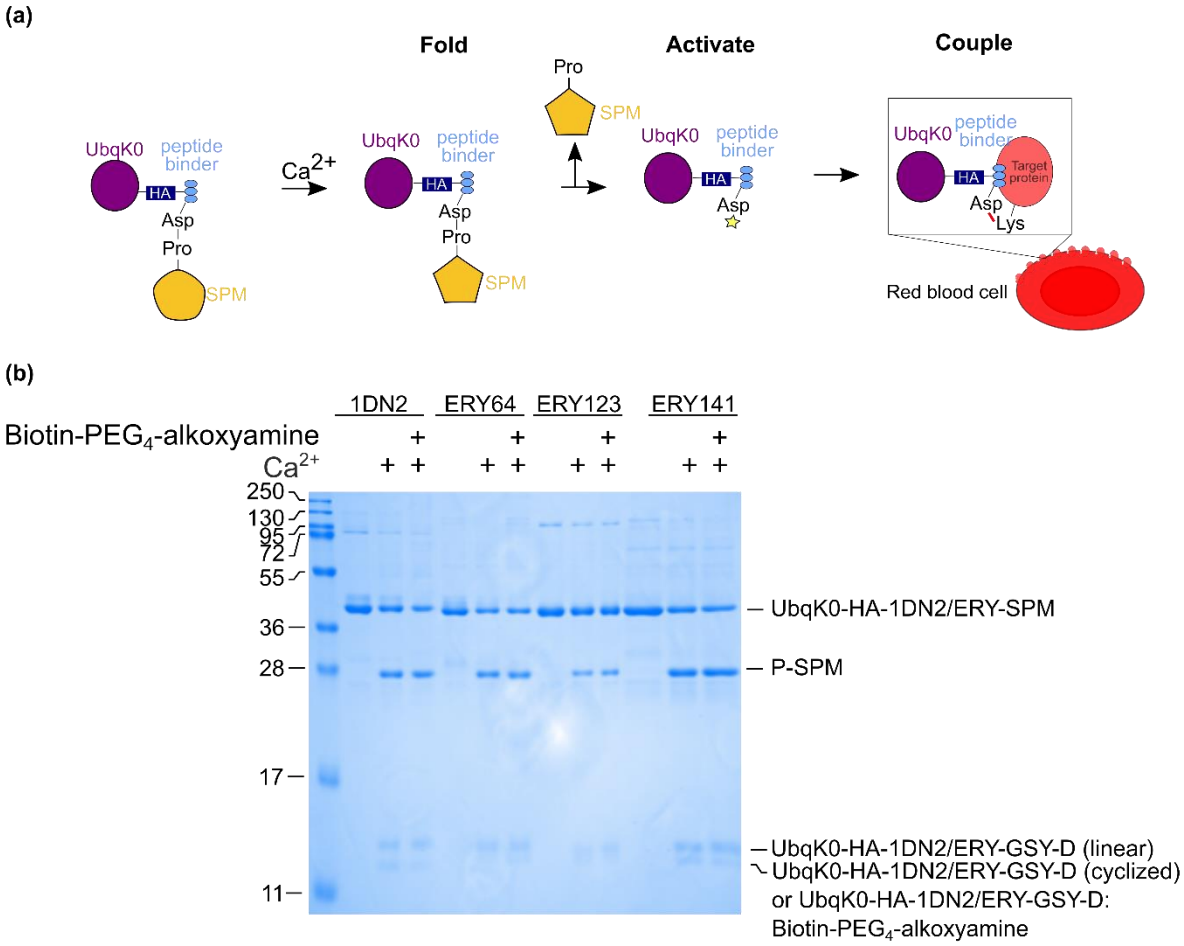


Figure 6.6 Cleavage and conjugation of UbqK0-HA-ERY-GSY-SPM for use in red blood cell coupling. (a) Schematic for use of UbqK0-HA-peptide binder-GSY-SPM for coupling to target proteins on surface on red blood cells. Lysine-free ubiquitin, UbqK0 (purple) is fused to a peptide binder (light blue) and SPM (yellow). Calcium addition leads to SPM activation and generation of an anhydride (yellow star) at the C-terminus of the peptide binder. If the peptide binder is docked onto the target protein (light red) on the red blood cell surface, the anhydride reacts with a proximal lysine on the target, forming an isopeptide bond (red line). (b) 10 µM of each protein, except for UbqK0-HA-ERY141-GSY-SPM that was at 2 µM, was incubated with 2 mM Ca^{2+} with or without 2 mM biotin-PEG_4-alkoxyamine in HEPES-reaction buffer, pH 7.4, 37 °C for 10 min. Additional 2 mM TCEP was included in the buffer for ERY123. After the reaction was stopped, samples were analysed by Coomassie-stained SDS-PAGE.

6.4. Expression of red blood cell bait antigens

To increase the chance of identifying a covalently-reactive peptide binder against human red blood cells, I decided to work towards employing a directed evolution-based strategy, in addition to testing the ERY peptides. In theory, by generating a library of peptide variants using the UbqK0-HA-peptide-SPM scaffold, covalently reactive peptide variants against specific red blood cell bait antigens can be selected for, as illustrated in Figure 6.7. Designating a specific red blood cell antigen for coupling may be favourable since Villa *et al.* previously showed that targeting highly-expressed antigens such as glycophorin A and B and Band3 adversely affected the physiology of red blood cells and led to increased haemolysis (Villa et al., 2018).

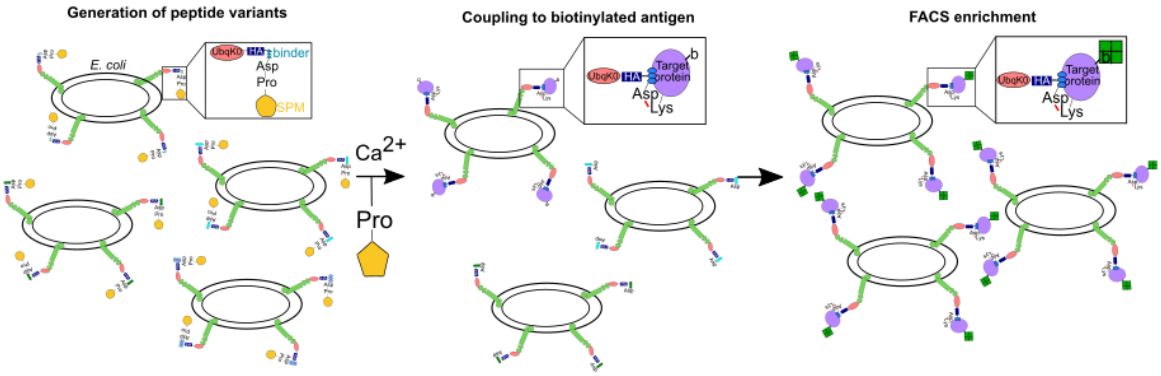


Figure 6.7 Selection of covalent peptide binders against red blood cell bait antigens. Peptide variants fused to lysine-free ubiquitin (light red), HA-tag (dark blue) and C-terminal full-length SPM (yellow) are expressed on the bacterial cell surface. The library of peptide variants is incubated with the biotinylated target bait protein (purple) that is a red blood cell surface protein. Upon calcium activation of SPM, SPM is cleaved from all the variants but only peptides that dock to the target protein and are covalently-reactive form an isopeptide bond (red line) with the target protein. The target protein is detected by fluorescently-labelled streptavidin (dark green) and FACS enables enrichment of bacteria expressing covalently-reactive peptide variants.

The ideal red blood cell surface antigen to target would have:

1) High surface expression on red blood cells to increase the maximum concentration of cargo protein that can be coupled.

2) Low polymorphism so that the same binder can be used even amongst individuals from genetically-distant populations.

3) Minimal function so that conjugation to the target would have minimal impact on red blood cell function.

With these factors in consideration, I identified basigin as a potential target protein. Basigin, also known as CD147, is an Ig-like domain-containing antigen and also a receptor for *Plasmodium falciparum*'s RH5 protein in malaria pathogenesis (Crosnier et al., 2011; Muramatsu, 2016). At least 8 polymorphic forms of basigin have been identified but most occur at low frequencies of less 0.01%. (Crosnier et al., 2011; Latini et al., 2020). The presence of the Ig-like domains in the extracellular region of basigin provides a larger surface area for the binder peptide to dock. Satchwell *et al.* previously showed that the knockout of basigin in BEL-A cells did not abolish reticulocyte viability (Satchwell et al., 2019), suggesting that covalent coupling to basigin is unlikely to abolish red blood cell function.

Given the disulfide bonds and N-glycosylation present in the extracellular domain of basigin, I utilised mammalian Expi293F cells for expression. I generated a basigin construct containing the ectodomain of basigin isoform 1 fused to an AviTag (AP) for biotinylation and a terminal His_6 tag for purification. 5 days post-transfection, I used Ni-NTA to purify basigin from the media and obtained a high yield of the protein (1.74 mg/ 25 mL culture). The high yield confirmed that the ectodomain could be stably expressed and potentially used as a bait antigen in future selections of peptide binders.

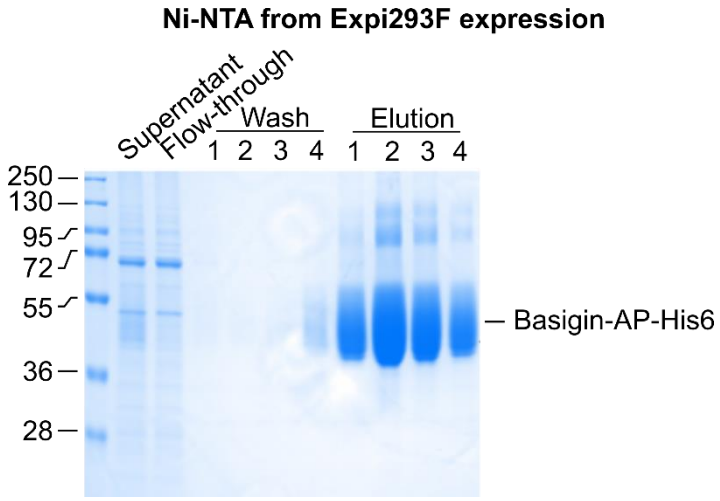


Figure 6.8 Mammalian expression and purification of Basigin. Ni-NTA purification of basigin ectodomain from the media of Expi293F cells 5 days post-transfection. The supernatant, flow-through, wash and elution fractions were analysed by Coomassie-stained SDS-PAGE.

6.5. Discussion

In the first part of this chapter, I established the split-SPM system for mammalian expression of binder-SPM fusions. Splitting SPM into two fragments provided another level of control over SPM activity, in addition to the presence of Ca^{2+}. Expressing the binder protein with the inactive N-terminal fragment of SPM overcomes the problem of premature cleavage of full-length SPM when secreted from mammalian cells.

Apart from using the split-SPM system for NeissLock coupling, the split-SPM system can also be applied to the use of SPM as a purification tag. SPM has been used as a self-cleaving purification tag for bacterial-expressed but not eukaryotic-expressed proteins (Liu et al., 2014; Sadilkova et al., 2008). I have shown that proteins with PTMs, e.g., SARS-CoV-2 RBD, can be expressed as fusions to SPM_N and purified to reasonable yields from the media of transfected mammalian cells. It would be interesting to test the mammalian expression of other recombinant proteins with SPM_N

in tandem with an affinity tag or elastin-like polypeptide (ELP) tag. The cleavage of SPM after affinity purification (Sadilkova et al., 2008) or Inverse Transition Cycling (Liu et al., 2014) enables the isolation of the target protein, as well as cleavage and removal of the purification tags in a single step. There is no longer a need for enzymatic cleavage and a second chromatography step for removal of the purification tags, simplifying the purification process. In the case of an affinity tag, SpyTag003, as shown in this chapter, can be used so that Spy&Go purification can be employed to purify the protein-of-interest (Khairil Anuar et al., 2019). Instead of SpyDock alone, which is an unreactive SpyCatcher variant immobilised on the resin (Khairil Anuar et al., 2019), SPM_C can be immobilised on the resin as a fusion to SpyCatcher003 (Figure 6.9). The downstream calcium activation of cleavage post-binding and addition of DTT during elution to prevent self-reaction (Figure 6.9) will be similar to previous applications that have used SPM as a self-cleaving purification tag (Liu et al., 2014; Sadilkova et al., 2008). Although DTT reacts with the anhydride, the thioester bond was found to be rapidly hydrolysed, leaving an unmodified C-terminal aspartate (Sadilkova et al., 2008). It remains to be tested if the SPM-based purification methods are compatible with disulfide-linked proteins given the use of thiol compounds to inhibit self-reaction.

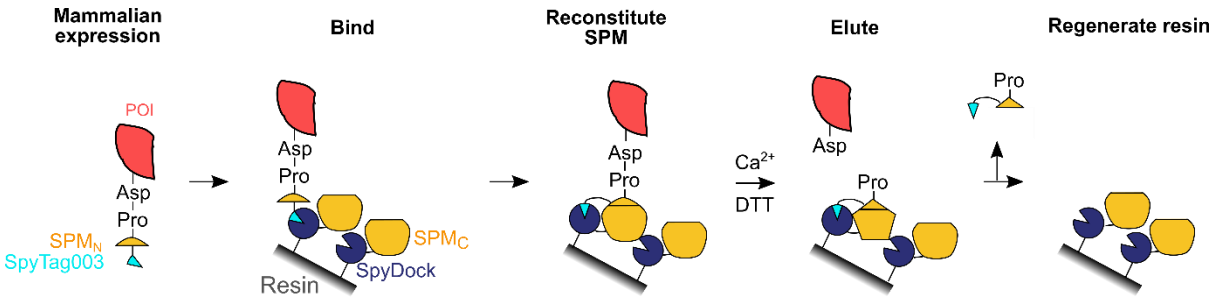


Figure 6.9 Using split-SPM for purification of 'tag-less' mammalian protein. The recombinant protein-of-interest (POI) (red) is expressed and secreted from mammalian cells as a fusion to SPM_N (yellow) and SpyTag003 (cyan). When the clear supernatant is applied to the resin that has immobilised SpyDock (dark blue) fused to

SPM_C (yellow), POI-SPM_N-SpyTag003 binds to the resin due to SpyTag003's affinity for SpyDock/SpyCatcher. The split fragments of SPM are brought into close proximity, enabling reconstitution. Addition of Ca^{2+} activates SPM cleavage, leading to release of the POI. DTT is added to quench the anhydride reactivity and prevent self-reaction, following which hydrolysis of the thioester bond leaves a C-terminal Asp in the POI. The resin is regenerated by removal of non-covalently bound SPM_N-SpyTag003.

Using OAZ as a model system, I optimised the split site of SPM and further improved the rate of split-SPM activation by fusing a second peptide/protein pair, SpyTag003/SpyCatcher003, to the split SPM fragments. Using the combination of SPM_{N318}-SpyTag003 and SPM_{C322}-SpyCatcher003, rapid cleavage was achieved (58 % of reconstituted SPM cleaved within 5 min and a maximum cleavage yield of 94 % was observed). The faster activation of split-SPM is necessary so that NeissLock coupling can take place without long incubation times that may not be practical or feasible for certain applications. For instance, in the case *ex vivo* red blood cell decoration, increasing the length of incubation at 25 °C or 37 °C for coupling to take place may lead to increased physiological changes in the cells. Even when stored in traditional anticoagulant preservative solutions, differences in levels of haemolysis, morphology and metabolic activities of red blood cells have been observed depending on the storage time, potentially impacting the effectiveness of downstream transfusion (García-Roa et al., 2017). The current maximum shelf-life of red blood cells, when refrigerated at 1-6 °C, is approximately 42 days (Hess et al., 2000). However, when stored at 25 °C, the metabolic rate of the red blood cells increases by ten-fold (Ruddell et al., 1998), reducing ATP levels, increasing oxidative stress and reducing overall cell viability (Orlov and Karkouti, 2015). Therefore, it would be preferable for NeissLock coupling to take place rapidly, even with the use of split-SPM.

I applied the split-SPM system to the expression and conjugation of tPA-TGFα-SPM_N-SpyTag003. The observed coupling to sEGFR suggests that mammalian-expressed TGFα retained binding to sEGFR even with an additional N-terminal tPA cargo protein and verified that split-SPM could be used for coupling post-reconstitution. However, the observation that tPA-TGFα-SPM_N-SpyTag003 did not react with SPMc-SpyCatcher003 to completion suggests that it is possible that a fraction of either protein was misfolded or that a fraction of tPA-TGFα-SPM_N-SpyTag003 lost the C-terminal SpyTag003. The harsh elution conditions of 2.5 M imidazole used for Spy&Go purification of tPA-TGFα-SPM_N-SpyTag003 may be one possible reason for the unfolding and/or aggregation of a fraction of tPA-TGFα-SPM_N-SpyTag003. To verify that eukaryotic expression of TGFα-SPM_N leads to a higher proportion of folded TGFα, it would be necessary in the future to compare the specific activities of mammalian-expressed TGFα to its bacterial-expressed counterpart. For instance, a competitive binding assay with EGF for binding to EGFR on live A431 cells could be used to compare the binding affinities of mammalian- and bacterial-expressed TGFα for EGFR (Winkler et al., 1986). Similarly, to verify the folding of recombinant tPA fused to TGFα-SPM_N, an assay of tPA activity should be employed. Nonetheless, the *in vitro* coupling of tPA-TGFα to sEGFR was promising and is a step towards expressing more complex but functional binder-SPM fusions with therapeutic cargo proteins attached. It also showed that attachment of a cargo protein to the N-terminus of the binder protein may not necessarily lead to self-reaction with the cargo protein outcompeting coupling to the target protein.

As mentioned in Chapter 4, the possibly limited lifetime of a TGFα:EGFR conjugate on live cells due to receptor endocytosis (Tomas et al., 2014) curtails the effectiveness of NeissLock coupling for long-term decoration of cells. Requiring

genetic modification of non-EGFR expressing cells to stably express EGFR would negate the benefit of using NeissLock to target endogenous proteins. With these factors in mind, I envisioned a modular system using a binding scaffold fused to SPM so that covalent binders against any arbitrary target protein could be selected for, even for target proteins without known structures.

To develop the modular system, I generated a construct consisting of lysine-free ubiquitin (UbqK0) fused to a HA-tag and binding peptide, followed by SPM at the C-terminus. I verified the cleavage activity of UbqK0-HA-ERY-SPM. The observation of a band corresponding to self-reacted cyclized UbqK0-HA-1DN2/ERY, despite the absence of lysines in the sequence, could be due to the anhydride's reactivity with other nucleophilic residues, like tyrosines. We observed reactivity with tyrosines in the plate-based assay in Chapter 4. UbqK0 contains 1 tyrosine while the HA-tag contains 2 tyrosines. ERY64 and ERY141 each contain 2 tyrosines and ERY123 contains a single cysteine. Site-directed mutagenesis could be employed to identify the coupling sites and reduce self-reactivity.

The decision to test the use of peptides like the human-red blood cell binding ERY peptides resulted from the hypothesis that the C-terminus of a peptide, compared to other antibody-like scaffolds, would be closer to a lysine on the target protein near the binding interface. However, without knowledge of the red blood cell surface antigen(s) the ERY peptides bind to, it may be a long and tedious trial-and-error process before a good peptide binder that fulfils the requirements for NeissLock coupling can be identified.

Therefore, moving forward, the focus would be on employing a directed evolution-based strategy. I identified basigin as another possible promising candidate target protein for red blood cell decoration by virtue of basigin's surface abundance,

low levels of polymorphism, and ease of expression of the ectodomain. The bacterial display of UbqK0-HA-peptide binder-SPM for selection of covalent binders against arbitrary targets may open up many other possibilities in the future.

In the longer term, once covalent peptide binders are identified, it may be necessary to test the immunogenicity of the peptides due to their foreign nature. Activation of an immune response may reduce the efficacy of the coupled biotherapeutic protein and/or raise additional safety concerns (Paul et al., 2013). The immunogenicity of the peptides would likely be difficult to predict without *in vivo* studies, although various tools for *in silico* prediction of T-cell epitopes have been developed (Doneva et al., 2021). In the case of surface decoration of mature red blood cells, the low to no expression of MHC molecules on red blood cells (Landsverk et al., 2009) and previous examples of induced immune tolerance to antigenic peptides conjugated to red blood cells (Pishesha et al., 2017) suggest the possibility of the coupled peptides avoiding immune activation.

All in all, the establishment of the split-SPM system and initial work on utilising a peptide binder fused to SPM in this chapter contribute to the application of NeissLock for conjugating biotherapeutic proteins to red blood cells for enhanced drug delivery.

Chapter 7: Summary and Discussion

The goal of this thesis was to develop a method for covalent conjugation to unmodified endogenous proteins. Proximity-directed ligation methods obviate the need for a recognition motif on the target protein. However, photo-activated proximity-based labelling faces limitations such as phototoxicity (Chen and Topp, 2019) while traceless affinity labelling methods like ligand-directed tosyl chemistry may be difficult to implement for targets without grooves for small-molecule binding (Hayashi and Hamachi, 2012). We decided to utilise protein binders and covalent probes based on the standard genetic code to expand the range of targetable proteins and to simplify the labelling process. The generation of an anhydride during the autocleavage of the Asp-Pro bond in several different domains (Bell et al., 2013; Johansson et al., 2009; Osicka et al., 2004) gave rise to the idea of utilising the anhydride generated as an electrophilic warhead. By attaching the autoproteolytic domain to binder proteins, the binder proteins could covalently couple via the aspartic anhydride to their cognate protein partner upon induction of cleavage.

In Chapter 3, I explored the use of the RGM domains for covalent coupling using the anhydride formed by autoproteolysis. However, the use of additives, alternative cell lines and alternative expression systems did not help isolate the pre-cleavage forms. It was not entirely surprising that the expression of the functional, properly folded RGM domains required the workings of the eukaryotic translational machinery and ER-Golgi secretory pathway. Nonetheless, I observed coupling to biotin-PEG_4-alkoxyamine by the N-terminal cleavage fragment of RGMC, which was promoted under acidic pH conditions. The *in vitro* cleavage confirmed the autoproteolytic nature of the Asp-Pro bond cleavage in the RGM domain. The conjugation to the primary amine underlines the potential of harnessing the

electrophilic anhydride generated from autoproteolysis for conjugation to nucleophilic residues like lysines on target proteins.

Autoproteolysis has been observed in the VWD domains of other proteins including mucins (Johansson et al., 2009), zonadhesin (Bi et al., 2003), IgG Fc-binding protein (FCGBP) (Johansson et al., 2009), crossveinless 2 (CV2) (Serpe et al., 2008) and hemolectin (Goto *et al.*, 2001). Similar to the RGMs, the anhydride generated from cleavage of CV2 and zonadhesin is rapidly hydrolysed (Bi et al., 2003; Serpe et al., 2008). It is possible that, although the cleavage in the various VWD domains occurs at an Asp-Pro bond, the cleavage mechanisms are dependent on the specific microenvironments. Amongst the different VWD domains, there is low sequence homology (Lang et al., 2007, 2016) and varying structural features (Bell et al., 2013; Biterova et al., 2018; Dong et al., 2019). For most of these domains, the Asp-Pro bond is in the middle of the domain. After fusing these autoproteolytic domains to the C-terminus of binder proteins, the fragment N-terminal to the Asp-Pro bond would be retained post-cleavage, leading to a larger scar post-conjugation (Figure 3.1). It is possible that the VWD domains can be truncated to a minimal domain required for autoproteolysis, minimising the fragment N-terminal to the Asp-Pro bond that is not part of the binder protein, so that the post-conjugation scar is minimised.

In Chapter 4, I shifted my focus to developing the NeissLock technology that uses the SPM domain from the FrpC protein of *Neisseria meningitidis*. When investigating the biochemical properties of NeissLock coupling, including its pH-dependence, calcium-dependence and buffer-dependence, I found that coupling was rapid under physiological pH and a coupling yield of up to 80 % was observed. Using the OAZ/ODC model complex, I showed that coupling was not promiscuous, even in a crowded environment like cell lysate. However, sfGFP and MBP that were used as

irrelevant non-binding proteins are not proteins that the binder-SPM fusions would likely encounter. It may be more informative to test off-target coupling whilst titrating different concentrations of serum albumin instead, given that albumin is the most abundant protein in blood (Merlot et al., 2014). Data on the non-specific reactivity with albumin may help inform the choice of buffer components for NeissLock coupling as well as future *in vivo* applications.

The observation that SPM can be activated by physiological extracellular calcium concentrations (1-2 mM) (Osicka et al., 2004; Scheu et al., 2021) suggests that NeissLock can be used *in vivo*. Cleavage of the SPM domain leaves a single amino acid residue scar (Sadilkova et al., 2006). The short sequence of the scar reduces the potential immunogenicity of the conjugate. However, if the entire binder-SPM fusion is to be administered *in vivo*, immune cells would still encounter the SPM domain both pre- and post-cleavage. Therefore, the potential immunogenicity of the SPM domain that is derived from the pathogenic *Neisseria meningitidis* remains a concern. Apart from employing alternative SPM homologues that are potentially less immunogenic, a focus on *ex vivo* rather than *in vivo* applications may be more practical.

In the OAZ/ODC complex, self-reactivity with the binder protein was observed from the formation of the cyclized form of OAZ. Self-reactivity can be minimised by side-directed mutagenesis of lysines proximal to the C-terminus of the binder and/or truncation of the N-terminus if there is self-reactivity with the α-amine. Given that the anhydride in OAZ reacted with K121 in ODC that is approximately 14 Å away from the C-terminus, lysines on the binder at the same distance or less away from the C-terminus may compete for reaction. If the structure of the binder/target complex is available, molecular dynamics simulations may help predict the position of the aspartic

anhydride, especially with the addition of the -GS linker used to increase the rate of coupling. A more accurate positioning of the anhydride would help to identify coupling site(s) on both the binder and target. In *ex vivo* applications, self-reactivity may be less of a concern as excess uncoupled binder proteins can be removed by extensive washes.

The assay of anhydride reactivity in Chapter 4 suggested that, apart from amines, thiols and hydroxyls can react with the aspartic anhydride as well. To test for coupling to alternative nucleophiles, a NeissDist search (Scheu et al., 2021) could be refined to identify candidate protein complexes with a cysteine or tyrosine residue optimally positioned from the C-terminus of a binder. The formation of a thioester or ester bond in the resulting conjugate can be verified by intact mass spectrometry.

With covalent coupling of TGFα to EGFR, I observed changes in the activation of STAT1 compared to the non-covalent control. More studies can be done to investigate the changes to cell signalling with covalent conjugation to cell surface receptors like EGFR. An assay of the activation of other effector proteins like Akt and ERK may reveal insights into how changes to the binding affinity of TGFα to EGFR and/or residency time affects canonical and non-canonical signalling pathways (Krall et al., 2011). Alternatively, other EGFR ligands like epigen, EGF or epiregulin that have different affinities for EGFR could also be fused to SPM to test if their covalent conjugation to EGFR affects downstream signalling (Freed et al., 2017).

In Chapter 5, the discovery of a faster-cleaving variant, the SPM domain from the FrpA protein of *Neisseria meningitidis*, furthers the development of an enhanced NeissLock. However, SPM cleavage does not reach completion even with the FrpA variant. Modelling of the kinetics of OAZ-SPM cleavage suggested that it is a two-phase exponential decay reaction (Figure 5.2d). The biphasic nature of the cleavage

reaction could be due to SPM adopting different stable conformational states, as has been observed in other proteins (Fox et al., 1986). To test the hypothesis that the slow-cleaving population is due to SPM adopting a stable misfolded conformation, the unreacted fraction of OAZ-SPM can be isolated, denatured, and then refolded. If the initially unreacted fraction was indeed a stable misfolded form, we expect that refolding would lead to a new fraction of properly folded SPM that would undergo further cleavage. If a stable misfolded conformation of SPM is identified, negative design can be used to disrupt interactions or strengthen repulsive interactions in the misfolded state (Berezovsky et al., 2007).

In Chapter 5, I worked on optimising the bacterial surface display of SPM for directed evolution of the SPM domain, to develop an enhanced NeissLock with higher coupling yield. I validated the surface display of intimin-HA-SPM-myc and showed that SPM retained its cleavage and coupling activity when fused to intimin. Fast and slow-cleaving variants of SPM could be differentiated by a short incubation with calcium, followed by flow cytometry analysis of the anti-myc signal. The goal was to sort for a SPM variant that cleaves to a larger extent than the FrpA variant and one that is better expressing.

However, after generating an error-prone library of SPM variants and two rounds of sorting, I found that the sorted variants were mostly non-functional with premature stop codons in the SPM sequence. Therefore, I decided to employ sequential sorting where I first sorted for cells with a high anti-myc signal without calcium to minimise the enrichment of variants with nonsense codons. However, I observed that the library had lower surface display levels compared to cells expressing wild-type SPM. The sorting again led to the isolation of non-functional variants,

suggesting that the viability and display levels of cells in the library require further optimisation.

The observed display levels of the intimin-SPM fusion in the library could be enhanced by testing the use of other *E. coli* strains such as the K12 strains with *fim* operon deletions that have been used for the bacterial surface display of nanobodies (Salema and Fernández, 2017). In strains without deletions in the *fim* operon like BL21 (DE3) used in this thesis, the presence of type 1 fimbriae may sterically obstruct the binding of antibodies to the epitope tags in the fusion protein on the outer membrane (Salema and Fernández, 2017). The low display levels of the intimin-SPM fusion could also be due to low viability. If the library shows a low growth and viability post-induction, the concentration of IPTG may have to be adjusted for milder induction (Salema and Fernández, 2017). Alternatively, changing to the more titratable arabinose induction may help improve overall cell viability and expression levels (Khlebnikov et al., 2000). Currently, induction is carried out at 30 °C but in the longer term, to evolve for more thermostable SPM variants with higher folding efficiency, protein expression at higher temperatures like 37 °C or 42 °C could be tested. More stable variants are likely to have higher levels of surface expression, as has been observed in yeast display systems (Lui and Cochran, 2011). If the extended incubation in buffer is a cause for low viability of the cells in the library during sequential sorting, magnetic-activated cell sorting (MACS) could be used as an alternative to flow cytometry for faster sorting (Mattanovich and Borth, 2006) before calcium addition, reducing the time that the cells are incubated in buffer.

To improve on the directed evolution of the SPM domain, another method is to generate a more focused library of variants (Packer and Liu, 2015). Random mutagenesis throughout the entire domain is likely to introduce more destabilising

mutations (Reidhaar-Olson and Sauer, 1988). Starting with the NMR structure of calcium-bound SPM domain, algorithms like Rosetta could be used to predict the differences in Gibbs free energy, ΔΔG, when scanning all possible single amino acid substitution at every position (Kellogg et al., 2011). The modelling could help shortlist regions of the domain that are more likely to accommodate mutations and lead to a better quality library (Packer and Liu, 2015). Alternatively, phylogenetic analysis of SPM homologues could help to identify beneficial mutations through consensus mutations. Consensus mutations have been found to improve the thermostability and catalytic activity of various enzymes (Lehmann et al., 2000). The combination of some aspects of rational design with the directed evolution of the SPM domain may speed up the identification of beneficial mutations.

The development of the split-SPM system in Chapter 6 expanded the range of binder proteins that can be expressed as SPM fusions, including proteins with complex post-translational modifications (PTMs) like tPA that has multiple N-glycans. By optimising the split position and fusing SpyTag003/SpyCatcher003 to the split SPM fragments, I improved the rate of SPM reconstitution by more than 10-fold. In the assays using the split-SPM system, I pre-incubated the fusion proteins containing split SPM fragments fused to SpyTag003/SpyCatcher0003 at 37 °C for 1 h, before addition of the target and calcium. It would be informative to have a more detailed analysis of the binding affinities of the split fragments using SPR and the rate of SPM reconstitution with the fusion of SpyTag003/SpyCatcher003. It would also be worth studying if and how the calcium-dependence of the split-SPM system differs from the use of full-length SPM.

In the latter part of Chapter 6, I demonstrated coupling of tPA-TGFα to sEGFR *in vitro* using the split-SPM system. The next step would be to demonstrate coupling

of tPA-TGFα using the split-SPM system on cells. I only tested the use of a -GSG linker between SPM_N and SpyTag003, as well as between SPM_C and SpyCatcher003. It is possible that increasing the linker length may reduce the steric clash encountered by the split SPM fragments and the SpyTag003/SpyCatcher003 pair during reconstitution. Alternatively, a directed evolution approach can be taken to further enhance the reconstitution rate of the split-SPM fragments.

One current limitation of the split-SPM system is the low yields of some mammalian-expressed fusion proteins like TGFα-SPM_N-SpyTag003. Nevertheless, the routine use of CHO cells in the industrial production of recombinant therapeutic proteins suggests that with optimisation e.g., changing the vector and/or signal peptide, engineering of the CHO cell line etc., the production of the binder-SPM fusions from mammalian cells can be further optimised for increased yields (Tripathi and Shrivastava, 2019).

To develop a covalent binder against human red blood cells, I fused the ERY peptide binders to SPM. Using known peptide binders against human red blood cells was the most direct approach to apply NeissLock for red blood cell decoration. The peptide binders can be randomised to evolve binders against any target protein by directed evolution. The long-term goal was to develop a modular platform for the selection of binders against any target protein, even in cases where the structures are unknown or where there is no known binder.

Instead of directly fusing binder peptides to SPM, the peptides can be grafted onto a larger scaffold like the disulfide-rich knottins to increase the stability of the peptides. Knottins are polypeptides comprising a cystine-knot and β-sheet fold with anti-parallel β-strands (Pallaghy et al., 1994). Antibodies or derivatives of antibodies like Fabs were not suitable as binding scaffolds for NeissLock because the C-terminus

of the binder would be too far from the binding interface (Chiu et al., 2019). Unlike antibodies, the C-terminus of *Ecballium elaterium* trypsin inhibitor II (EETI-II), a knottin, is close to the trypsin-binding region (Figure 7.1). The trypsin-binding region of EETI-II is amenable to amino acid substitutions and insertions, and has been engineered to bind the thrombopoietin receptor (Krause et al., 2007) and integrins (Kimura et al., 2009a; Reiss et al., 2006).

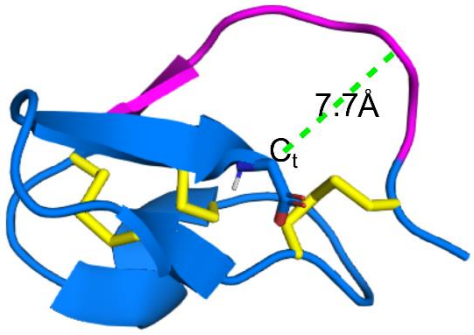


Figure 7.1 Structure of *Ecballium elaterium* trypsin inhibitor II (EETI-II) (PDB 2IT7). The trypsin-binding region that is amenable to substitutions is highlighted in magenta and intramolecular disulfide bonds are shown in yellow. The shortest distance between the C_α of the C-terminus (C_t) and the trypsin binding region is approximately 7.7 Å.

It would be interesting to test covalent coupling to the integrins when the knottin variants with low nanomolar affinity for integrins (Kimura et al., 2009a) are fused to SPM. In fact, given that the OAZ mutant that binds to ODC with a moderate affinity (K_d of 25 μM) still had a coupling yield that was 40 % that of wild-type OAZ, the low nanomolar affinity is not likely a pre-requisite and other variants with weaker binding affinities may still couple to a reasonable extent e.g. 20 % or higher coupling yield. An additional spacer domain, like Ubiquitin K0 mentioned in Chapter 6, may have to be added to the N-terminus of EETI-II to reduce self-reaction between the anhydride and the α-amine. When using NeissLock to couple a therapeutic protein to the N-terminus of the binder, the use of a lysine-free spacer domain in between the therapeutic protein

and binder would reduce the likelihood of self-conjugation to the therapeutic protein, even if the therapeutic protein contains multiple lysines.

As cell-surface receptors with multiple solvent-exposed lysines, integrins are promising targets for NeissLock coupling. The presence of multiple disulfides in knottins confers proteolytic resistance and increased thermal resilience to the binding loop regions (Craik et al., 2001; Kimura et al., 2009a). It was suggested that the knottin scaffold is non-immunogenic as well (Kimura, 2021). These properties further support the use of knottins as an evolvable binding scaffold to fuse to SPM.

Bacterial surface display has been used to display knottins for the selection of peptide binders from combinatorial libraries (Christmann et al., 1999). Alternatives for displaying knottins include the use of yeast display (Kimura et al., 2009b) or mammalian surface display system. Yeast or mammalian display is preferable for complex disulfide-containing proteins (Cherf and Cochran, 2015). However, the surface display of SPM in eukaryotes would require use of the split-SPM system developed in Chapter 6 to prevent premature activation of the anhydride (Figure 7.2).

Figure 7.2 illustrates how the split-SPM system can be used for the eukaryotic display of SPM to evolve covalent knottin binders against any target protein. The library of knottin-SPM variants with randomised trypsin-binding loops can be expressed as a fusion to a yeast membrane protein e.g., Aga1p and Aga2p (Kimura et al., 2009b) or a mammalian membrane protein e.g., Transferrin Receptor (TfR) (Keeble et al., 2019). The sorting strategy would be like that described in Chapters 6 and 7 for evolving covalent binders against arbitrary targets using bacterial surface display. The key difference is in the incubation with SPM$_C$-SpyCatcher003 to reconstitute SPM before addition of the target protein and calcium (Figure 7.2). High-throughput screening using flow cytometry can be employed to detect binding to the

target protein. More work needs to be done to see if the inclusion of harsh wash steps like low pH washes can lead to the differentiation of non-covalent and covalent coupling to the target (Zakeri et al., 2012). Covalent coupling can be further verified by comparisons to a negative control, that is the binder fused to the non-reactive D414A mutant of SPM (Scheu et al., 2021).

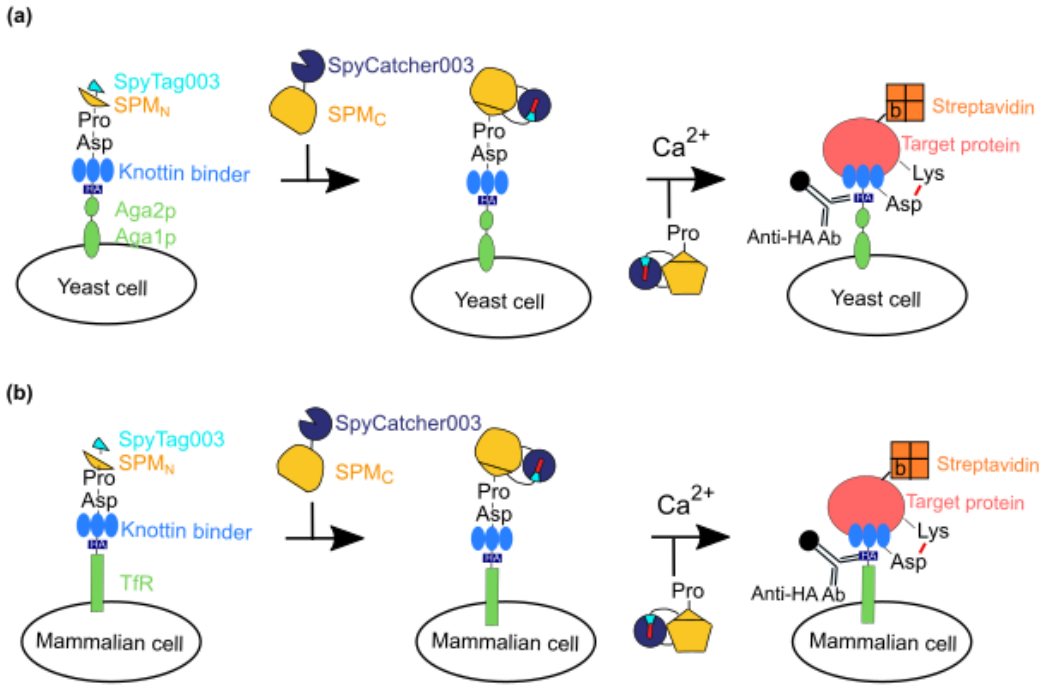


Figure 7.2 Schematic for (a) yeast surface display, or (b) mammalian surface display of knottin-SPM. Knottin(blue)-SPM$_N$(yellow)-SpyTag003(cyan) is expressed as a fusion to Aga1p and Aga2p (green) on the yeast surface. Introduction of SPM$_C$(yellow)-SpyCatcher003(dark blue) leads to formation of an isopeptide bond (red). Upon calcium addition, SPM is cleaved. If the knottin binder was docked to the target protein (red), the anhydride reacts with lysine on the target protein, forming an isopeptide bond (red line). Flow cytometry can be used to detect the biotinylated target protein using a fluorescently-labelled streptavidin. Expression of the knottin binder is detectable using an anti-HA antibody. (b) In mammalian surface display, the knottin-SPM$_N$-SpyTag003 is fused to a membrane protein like the transferrin receptor (TfR) (green rectangle) instead.

The development of a covalently reactive knottin binder presents various opportunities. Given the small size of knottins, they are rapidly removed from circulation by renal clearance (Kintzing and Cochran, 2016). Covalent coupling to the target protein may improve the half-life of knottins. Apart from uses in molecular imaging (Kimura et al., 2009a), knottins have been used as drug conjugates for cancer treatment (Cox et al., 2016; Currier et al., 2016) and immunotherapy (Maaß et al., 2015).

Moving forward, one of the key project directions is in applying NeissLock to *ex vivo* red blood cell surface decoration. In Chapter 6, tPA was used as a model therapeutic enzyme that can be coupled to the cell surface. Other therapeutic enzymes are of interest as well (Bax, 2020). The benefits of red blood cell coupling include prolonged catalysis until clearance of the red blood cells by macrophages (de Back et al., 2014). The average lifetime of red blood cells is approximately 120 days (Shemin and Rittenberg, 1946). Red blood cell coupling reduces the need for repeated intravenous dosing like in traditional enzyme replacement therapy (Bax, 2020).

Coupled enzymes may be more susceptible to cleavage by plasma proteases compared to encapsulated enzymes (Bourgeaux et al., 2016). On the other hand, coupled enzymes do not depend on the diffusion or transport of the substrates into the red blood cells. It was observed that using the biotin/avidin system to couple uricase to the red blood cells led to more efficient metabolism of blood uric acid compared to encapsulation (Magnani et al., 1992). More recently, L-asparaginase was encapsulated in red blood cells to starve tumour cells of circulating asparagine (Domenech et al., 2011). A circulation half-life of about a month was observed for the asparaginase-loaded red blood cells; asparaginase-loaded red blood cells are currently tested in clinical trials against several cancer types (Bachet et al., 2015;

Domenech et al., 2011). It could be interesting to study how NeissLock-coupled asparaginase would compare to asparaginase-loaded red blood cells and biotin/avidin-coupled asparaginase in terms of the safety and efficacy in lowering plasma asparagine concentration and improving overall patient survival.

To test the use of ERY-SPM as covalent probes for red blood cell coupling, our first goal would be to validate their non-covalent and covalent binding to human red blood cells. Once binding and the formation of a new covalent conjugate has been validated, the target protein that the ERY peptide couples to can be identified by Western blot, panning against different red blood cell surface antigens. Alternatively, a tryptic digest of the conjugate followed by liquid chromatography with tandem mass spectrometry (LC MS/MS) can be used to identify the target protein.

One of the key measures of safety in using NeissLock to couple therapeutic proteins to red blood cells is the function of the red blood cells post-coupling. The use of an immortalised human adult erythroid cell line like the BEL-A cell line circumvents the need to collect autologous blood (Trakarnsanga et al., 2017) and may lead to more consistent coupling. For instance, it may allow a standardised level of decoration across different individuals that does not vary according to expression levels of the target protein on the donor cells. To verify that the normal physiology of red blood cells is not impaired, the levels of haemolysis, efficacy of oxygen transport (Di Caprio et al., 2015) as well as mechanical properties such as the deformability of the red blood cells (Trakarnsanga et al., 2017) can be measured post-coupling. In the longer term, further *in vivo* studies can be done to investigate the biocompatibility of the coupled red blood cells post-infusion (Trakarnsanga et al., 2017).

NeissLock is unique in utilising a calcium-inducible anhydride to couple to target endogenous proteins. In this thesis, I have shown that Neisslock coupling is rapid,

specific and requires mild conditions. The identification of FrpA SPM as a faster cleaving variant was an advance from the first-generation of NeissLock. Initial experiments on developing a bacterial display platform for directed evolution of the SPM domain pave the way for even higher-yielding conjugation. Establishing the split-SPM system ensured compatibility with mammalian expression systems, broadening the utility of NeissLock. Having demonstrated NeissLock coupling on live cells, being able to utilise NeissLock to couple therapeutic proteins like tPA to red blood cells in the future is foreseeable. Apart from fusing known binders like the ERY peptides to SPM, developing a platform for directed evolution of binders would be a major step forward. Directed evolution would facilitate the identification of novel covalent binders against arbitrary target proteins and so presents many other exciting opportunities.

Bibliography

Albeck, S., Dym, O., Unger, T., Snapir, Z., Bercovich, Z., and Kahana, C. (2008). Crystallographic and biochemical studies revealing the structural basis for antizyme inhibitor function. Protein Sci. *17*, 793–802.

Almrud, J.J., Oliveira, M.A., Kern, A.D., Grishin, N. V, Phillips, M.A., and Hackert, M.L. (2000). Crystal structure of human ornithine decarboxylase at 2.1 A resolution: structural insights to antizyme binding. J. Mol. Biol. *295*, 7–16.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.

Andre, I., Linse, S., and Mulder, F.A.A. (2007). Residue-specific pKa determination of lysine and arginine side chains by indirect 15N and 13C NMR spectroscopy: application to apo calmodulin. J. Am. Chem. Soc. *129*, 15805–15813.

Aronson, D.E., Costantini, L.M., and Snapp, E.L. (2011). Superfolder GFP Is Fluorescent in Oxidizing Environments When Targeted via the Sec Translocon. Traffic *12*, 543–548.

Axelsson, M.A.B., Karlsson, N.G., Steel, D.M., Ouwendijk, J., Nilsson, T., and Hansson, G.C. (2001). Neutralization of pH in the Golgi apparatus causes redistribution of glycosyltransferases and changes in the O-glycosylation of mucins. Glycobiology *11*, 633–644.

Bachet, J.-B., Gay, F., Maréchal, R., Galais, M.-P., Adenis, A., MsC, D.S., Cros, J., Demetter, P., Svrcek, M., Bardier-Dupas, A., et al. (2015). Asparagine Synthetase Expression and Phase I Study With L-Asparaginase Encapsulated in Red Blood Cells in Patients With Pancreatic Adenocarcinoma. Pancreas *44*, 1141–1147.

Back, N., and Soinila, S. (1996). Dose-dependent effects of chloroquine on secretory granule formation in the melanotroph. Acta Anat. (Basel). *156*, 307–314.

de Back, D.Z., Kostova, E.B., van Kraaij, M., van den Berg, T.K., and van Bruggen, R. (2014). Of macrophages and red blood cells; a complex love story. Front. Physiol. *5*, 9.

Bagur, R., and Hajnóczky, G. (2017). Intracellular Ca(2+) Sensing: Its Role in Calcium Homeostasis and Signaling. Mol. Cell *66*, 780–788.

Banerjee, A., and Howarth, M. (2018). Nanoteamwork: covalent protein assembly beyond duets towards protein ensembles and orchestras. Curr. Opin. Biotechnol. *51*, 16–23.

Bax, B.E. (2020). Erythrocytes as Carriers of Therapeutic Enzymes. Pharmaceutics *12*, 435.

Bell, C.H., Healey, E., van Erp, S., Bishop, B., Tang, C., Gilbert, R.J.C., Aricescu, A.R., Pasterkamp, R.J., and Siebold, C. (2013). Structure of the Repulsive Guidance

Molecule (RGM)—Neogenin Signaling Hub. Science *341*, 77–80.

Benink, H.A., and Urh, M. (2015). HaloTag technology for specific and covalent labeling of fusion proteins. Methods Mol. Biol. *1266*, 119–128.

Berezovsky, I.N., Zeldovich, K.B., and Shakhnovich, E.I. (2007). Positive and Negative Design in Stability and Thermal Adaptation of Natural Proteins. PLOS Comput. Biol. *3*, e52.

Bi, M., Hickox, J.R., Winfrey, V.P., Olson, G.E., and Hardy, D.M. (2003). Processing, localization and binding activity of zonadhesin suggest a function in sperm adhesion to the zona pellucida during exocytosis of the acrosome. Biochem. J. *375*, 477–488.

Biterova, E., Esmaeeli, M., Alanen, H.I., Saaranen, M., and Ruddock, L.W. (2018). Structures of Angptl3 and Angptl4, modulators of triglyceride levels and coronary artery disease. Sci. Rep. *8*, 6752.

Björnfot, A.-C., Lavander, M., Forsberg, Å., and Wolf-Watz, H. (2009). Autoproteolysis of YscU of Yersinia pseudotuberculosis Is Important for Regulation of Expression and Secretion of Yop Proteins. J. Bacteriol. *191*, 4259 LP – 4267.

van Bloois, E., Winter, R.T., Kolmar, H., and Fraaije, M.W. (2011). Decorating microbes: surface display of proteins on Escherichia coli. Trends Biotechnol. *29*, 79–86.

Bosques, C.J., Collins, B.E., Meador, J.W. 3rd, Sarvaiya, H., Murphy, J.L., Dellorusso, G., Bulik, D.A., Hsu, I.-H., Washburn, N., Sipsey, S.F., et al. (2010). Chinese hamster ovary cells can produce galactose-alpha-1,3-galactose antigens on proteins. Nat. Biotechnol. *28*, 1153–1156.

Bourgeaux, V., Lanao, J.M., Bax, B.E., and Godfrin, Y. (2016). Drug-loaded erythrocytes: on the road toward marketing approval. Drug Des. Devel. Ther. *10*, 665–676.

Boyles, J.S., Atwell, S., Druzina, Z., Heuer, J.G., and Witcher, D.R. (2016). Structural basis of selectivity and neutralizing activity of a TGFα/epiregulin specific antibody. Protein Sci. *25*, 2028–2036.

Branon, T.C., Bosch, J.A., Sanchez, A.D., Udeshi, N.D., Svinkina, T., Carr, S.A., Feldman, J.L., Perrimon, N., and Ting, A.Y. (2018). Efficient proximity labeling in living cells and organisms with TurboID. Nat. Biotechnol. *36*, 880.

Brenner, J.S., Pan, D.C., Myerson, J.W., Marcos-Contreras, O.A., Villa, C.H., Patel, P., Hekierski, H., Chatterjee, S., Tao, J.-Q., Parhiz, H., et al. (2018). Red blood cell-hitchhiking boosts delivery of nanocarriers to chosen organs by orders of magnitude. Nat. Commun. *9*, 2684.

Buck, E., and Wells, J.A. (2005). Disulfide trapping to localize small-molecule agonists and antagonists for a G protein-coupled receptor. Proc. Natl. Acad. Sci. U. S. A. *102*, 2719 LP – 2724.

Buck, E., Bourne, H., and Wells, J.A. (2005). Site-specific disulfide capture of agonist and antagonist peptides on the C5a receptor. J. Biol. Chem. *280*, 4009–4012.

Buldun, C.M., Jean, J.X., Bedford, M.R., and Howarth, M. (2018). SnoopLigase Catalyzes Peptide-Peptide Locking and Enables Solid-Phase Conjugate Isolation. J. Am. Chem. Soc. *140*, 3008–3018.

Bunton, C.A., Fuller, N.A., Perry, S.G., and Shiner, V.J. (1963). 542. The hydrolysis of carboxylic anhydrides. Part III. Reactions in initially neutral solution. J. Chem. Soc. 2918–2926.

C. McIlvaine, T. (1921). A Buffer Solution for Colorimetric Comparison.

Di Caprio, G., Stokes, C., Higgins, J.M., and Schonbrun, E. (2015). Single-cell measurement of red blood cell oxygen affinity. Proc. Natl. Acad. Sci. *112*, 9984 LP – 9989.

Chen, C.-L., and Perrimon, N. (2017). Proximity-dependent labeling methods for proteomic profiling in living cells. WIREs Dev. Biol. *6*, e272.

Chen, Y., and Topp, E.M. (2019). Photolytic Labeling and Its Applications in Protein Drug Discovery and Development. J. Pharm. Sci. *108*, 791–797.

Chen, X.-H., Xiang, Z., Hu, Y.S., Lacey, V.K., Cang, H., and Wang, L. (2014). Genetically encoding an electrophilic amino acid for protein stapling and covalent binding to native receptors. ACS Chem. Biol. *9*, 1956–1961.

Cherf, G.M., and Cochran, J.R. (2015). Applications of Yeast Surface Display for Protein Engineering. Methods Mol. Biol. *1319*, 155–175.

Chiu, M.L., Goulet, D.R., Teplyakov, A., and Gilliland, G.L. (2019). Antibody Structure and Function: The Basis for Engineering Therapeutics. Antibodies (Basel, Switzerland) *8*, 55.

Chmura, A.J., Orton, M.S., and Meares, C.F. (2001). Antibodies with infinite affinity. Proc. Natl. Acad. Sci. U. S. A. *98*, 8480–8484.

Christmann, A., Walter, K., Wentzel, A., Krätzner, R., and Kolmar, H. (1999). The cystine knot of a squash-type protease inhibitor as a structural scaffold for Escherichia coli cell surface display of conformationally constrained peptides. Protein Eng. *12*, 797–806.

Cohavi, O., Tobi, D., and Schreiber, G. (2009). Docking of antizyme to ornithine decarboxylase and antizyme inhibitor using experimental mutant and double-mutant cycle data. J. Mol. Biol. *390*, 503–515.

Cole, N.B. (2013). Site-specific protein labeling with SNAP-tags. Curr. Protoc. Protein Sci. *73*, 30.1.1-30.1.16.

Costantini, L.M., and Snapp, E.L. (2013). Fluorescent proteins in cellular organelles: serious pitfalls and some solutions. DNA Cell Biol. *32*, 622–627.

Cox, N., Kintzing, J.R., Smith, M., Grant, G.A., and Cochran, J.R. (2016). Integrin-Targeting Knottin Peptide–Drug Conjugates Are Potent Inhibitors of Tumor Cell Proliferation. Angew. Chemie Int. Ed. *55*, 9894–9897.

Craik, D.J., Daly, N.L., and Waine, C. (2001). The cystine knot motif in toxins and implications for drug design. Toxicon *39*, 43–60.

Crosnier, C., Bustamante, L.Y., Bartholdson, S.J., Bei, A.K., Theron, M., Uchikawa, M., Mboup, S., Ndir, O., Kwiatkowski, D.P., Duraisingh, M.T., et al. (2011). Basigin is a receptor essential for erythrocyte invasion by Plasmodium falciparum. Nature *480*, 534–537.

Currier, N. V, Ackerman, S.E., Kintzing, J.R., Chen, R., Filsinger Interrante, M., Steiner, A., Sato, A.K., and Cochran, J.R. (2016). Targeted Drug Delivery with an Integrin-Binding Knottin{\textendash}Fc{\textendash}MMAF Conjugate Produced by Cell-Free Protein Synthesis. Mol. Cancer Ther. *15*, 1291–1300.

Daniels, G. (2007). Functions of red cell surface proteins. Vox Sang. *93*, 331–340.

DeLano WL, Ultsch MH, de Vos AM, Wells JA. (2000). Convergent Solutions to Binding at a Protein-Protein Interface. Science (80-.). 287, 1279–1283.

Demaurex, N., Furuya, W., D'Souza, S., Bonifacino, J.S., and Grinstein, S. (1998). Mechanism of Acidification of the trans-Golgi Network (TGN): IN SITU MEASUREMENTS OF pH USING RETRIEVAL OF TGN38 AND FURIN FROM THE CELL SURFACE . J. Biol. Chem. *273*, 2044–2051.

Domenech, C., Thomas, X., Chabaud, S., Baruchel, A., Gueyffier, F., Mazingue, F., Auvrignon, A., Corm, S., Dombret, H., Chevallier, P., et al. (2011). l-asparaginase loaded red blood cells in refractory or relapsing acute lymphoblastic leukaemia in children and adults: results of the GRASPALL 2005-01 randomized trial. Br. J. Haematol. *153*, 58–65.

Doneva, N., Doytchinova, I., and Dimitrov, I. (2021). Predicting Immunogenicity Risk in Biopharmaceuticals. Symmetry (Basel). *13*, 388.

Dong, X., Leksa, N.C., Chhabra, E.S., Arndt, J.W., Lu, Q., Knockenhauer, K.E., Peters, R.T., and Springer, T.A. (2019). The von Willebrand factor D'D3 assembly and structural principles for factor VIII binding and concatemer biogenesis. Blood *133*, 1523–1533.

Drozdetskiy, A., Cole, C., Procter, J., and Barton, G.J. (2015). JPred4: a protein secondary structure prediction server. Nucleic Acids Res. *43*, W389–W394.

Emond, S., Mondon, P., Pizzut-Serin, S., Douchy, L., Crozet, F., Bouayadi, K., Kharrat, H., Potocki-Véronèse, G., Monsan, P., and Remaud-Simeon, M. (2008). A novel random mutagenesis approach using human mutagenic DNA polymerases to generate enzyme variant libraries. Protein Eng. Des. Sel. *21*, 267–274.

Erlanson, D.A., Braisted, A.C., Raphael, D.R., Randal, M., Stroud, R.M., Gordon, E.M., and Wells, J.A. (2000). Site-directed ligand discovery. Proc. Natl. Acad. Sci. *97*, 9367 LP – 9372.

Evans, M.J., and Cravatt, B.F. (2006). Mechanism-Based Profiling of Enzyme Families. Chem. Rev. *106*, 3279–3301.

Fairhead, M., and Howarth, M. (2015). Site-specific biotinylation of purified proteins using BirA. Methods Mol. Biol. *1266*, 171–184.

Ferris, H.U., Furukawa, Y., Minamino, T., Kroetz, M.B., Kihara, M., Namba, K., and Macnab, R.M. (2005). FlhB regulates ordered export of flagellar components via autocleavage mechanism. J. Biol. Chem. *280*, 41236–41242.

Fitch, C.A., Karp, D.A., Lee, K.K., Stites, W.E., Lattman, E.E., and Garcia-Moreno E, B. (2002). Experimental pK(a) values of buried residues: analysis with continuum methods and role of water penetration. Biophys. J. *82*, 3289–3304.

Fox, R.O., Evans, P.A., and Dobson, C.M. (1986). Multiple conformations of a protein demonstrated by magnetization transfer NMR spectroscopy. Nature *320*, 192–194.

Freed, D.M., Bessman, N.J., Kiyatkin, A., Salazar-Cavazos, E., Byrne, P.O., Moore, J.O., Valley, C.C., Ferguson, K.M., Leahy, D.J., Lidke, D.S., et al. (2017). EGFR Ligands Differentially Stabilize Receptor Dimers to Specify Signaling Kinetics. Cell *171*, 683-695.e18.

Fujimoto, K., Sasago, S., Mihara, J., and Nakamura, S. (2018). DNA Photo-cross-linking Using Pyranocarbazole and Visible Light. Org. Lett. *20*, 2802–2805.

Fujimoto, K., Yang-Chun, H., and Nakamura, S. (2019). Strong Inhibitory Effects of Antisense Probes on Gene Expression through Ultrafast RNA Photocrosslinking. Chem. Asian J. *14*, 1912–1916.

Furlong, A.M., Thomsen, D.R., Marotti, K.R., Post, L.E., and Sharma, S.K. (1988). Active Human Tissue Plasminogen Activator Secreted from Insect Cells Using a Baculovirus Vector. Biotechnol. Appl. Biochem. *10*, 454–464.

Gai, S.A., and Wittrup, K.D. (2007). Yeast surface display for protein engineering and characterization. Curr. Opin. Struct. Biol. *17*, 467–473.

García-Roa, M., Del Carmen Vicente-Ayuso, M., Bobes, A.M., Pedraza, A.C., González-Fernández, A., Martín, M.P., Sáez, I., Seghatchian, J., and Gutiérrez, L. (2017). Red blood cell storage time and transfusion: current practice, concerns and future perspectives. Blood Transfus. *15*, 222–231.

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R.D., and Bairoch, A. (2003). ExPASy: The proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res. *31*, 3784–3788.

Gentile, D.R., Rathinaswamy, M.K., Jenkins, M.L., Moss, S.M., Siempelkamp, B.D., Renslo, A.R., Burke, J.E., and Shokat, K.M. (2017). Ras Binder Induces a Modified Switch-II Pocket in GTP and GDP States. Cell Chem. Biol. *24*, 1455-1466.e14.

Ghaderi, D., Zhang, M., Hurtado-Ziola, N., and Varki, A. (2012). Production platforms for biotherapeutic glycoproteins. Occurrence, impact, and challenges of non-human

sialylation. Biotechnol. Genet. Eng. Rev. *28*, 147–175.

Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison, C.A., and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat. Methods *6*, 343–345.

Gorovits, B., Clements-Egan, A., Birchler, M., Liang, M., Myler, H., Peng, K., Purushothama, S., Rajadhyaksha, M., Salazar-Fontana, L., Sung, C., et al. (2016). Pre-existing Antibody: Biotherapeutic Modality-Based Review. AAPS J. *18*, 311–320.

Goto,A., Kumagai, T., Kumaigai, C., Hirose, J., Narita, H., Mori, H., Kadowaki, T., Beck, K., and Kitagawa, Y. (2001). A Drosophila haemocyte-specific protein, hemolectin, similar to human von Willebrand factor. Biochem. J. *359*, 99 LP – 108.

Greineder, C.F., Howard, M.D., Carnemolla, R., Cines, D.B., and Muzykantov, V.R. (2013). Advanced drug delivery systems for antithrombotic agents. Blood *122*, 1565–1575.

Hacker, S.M., Backus, K.M., Lazear, M.R., Forli, S., Correia, B.E., and Cravatt, B.F. (2017). Global profiling of lysine reactivity and ligandability in the human proteome. Nat. Chem. *9*, 1181–1190.

Hallenbeck, K.K., Turner, D.M., Renslo, A.R., and Arkin, M.R. (2017). Targeting Non-Catalytic Cysteine Residues Through Structure-Guided Drug Discovery. Curr. Top. Med. Chem. *17*, 4–15.

Han, X., Wang, C., and Liu, Z. (2018). Red Blood Cells as Smart Delivery Systems. Bioconjug. Chem. *29*, 852–860.

Harms, M.J., Castaneda, C.A., Schlessman, J.L., Sue, G.R., Isom, D.G., Cannon, B.R., and Garcia-Moreno E, B. (2009). The pK(a) values of acidic and basic residues buried at the same internal location in a protein are governed by different factors. J. Mol. Biol. *389*, 34–47.

Hayashi, T., and Hamachi, I. (2012). Traceless affinity labeling of endogenous proteins for functional analysis in living cells. Acc. Chem. Res. *45*, 1460–1469.

Hayashi, T., Sun, Y., Tamura, T., Kuwata, K., Song, Z., Takaoka, Y., and Hamachi, I. (2013). Semisynthetic Lectin–4-Dimethylaminopyridine Conjugates for Labeling and Profiling Glycoproteins on Live Cell Surfaces. J. Am. Chem. Soc. *135*, 12252–12258.

Hayashi, T., Yasueda, Y., Tamura, T., Takaoka, Y., and Hamachi, I. (2015). Analysis of Cell-Surface Receptor Dynamics through Covalent Labeling by Catalyst-Tethered Antibody. J. Am. Chem. Soc. *137*, 5372–5380.

He, D., Xie, X., Yang, F., Zhang, H., Su, H., Ge, Y., Song, H., and Chen, P.R. (2017). Quantitative and Comparative Profiling of Protease Substrates through a Genetically Encoded Multifunctional Photocrosslinker. Angew. Chem. Int. Ed. Engl. *56*, 14521–14525.

Hess, J.R., Rugg, N., Knapp, A.D., Gormas, J.F., Silberstein, E.B., and Greenwalt, T.J. (2000). Successful storage of RBCs for 9 weeks in a new additive solution.

Transfusion *40*, 1007–1011.

Hino, N., Okazaki, Y., Kobayashi, T., Hayashi, A., Sakamoto, K., and Yokoyama, S. (2005). Protein photo-cross-linking in mammalian cells by site-specific incorporation of a photoreactive amino acid. Nat. Methods *2*, 201–206.

Ho, M.-C., Menetret, J.-F., Tsuruta, H., and Allen, K.N. (2009). The origin of the electrostatic perturbation in acetoacetate decarboxylase. Nature *459*, 393–397.

Holm, L., Moody, P., and Howarth, M. (2009). Electrophilic affibodies forming covalent bonds to protein targets. J. Biol. Chem. *284*, 32906–32913.

El Houmami, N., Bakour, S., Bzdrenga, J., Rathored, J., Seligmann, H., Robert, C., Armstrong, N., Schrenzel, J., Raoult, D., Yagupsky, P., et al. (2017). Isolation and characterization of Kingella negevensis sp. nov., a novel Kingella species detected in a healthy paediatric population. Int. J. Syst. Evol. Microbiol. *67*, 2370–2376.

Hsieh, J.-Y., Yang, J.-Y., Lin, C.-L., Liu, G.-Y., and Hung, H.-C. (2011). Minimal Antizyme Peptide Fully Functioning in the Binding and Inhibition of Ornithine Decarboxylase and Antizyme Inhibitor. PLoS One *6*, e24366.

Hu, F.P., Young, J.M., and Fletcher, M.J. (1998). Preliminary description of biocidal (syringomycin) activity in fluorescent plant pathogenic Pseudomonas species. J. Appl. Microbiol. *85*, 365–371.

Huang, T., Li, J., and Byrd, R.A. (2014). Solution structure of lysine-free (K0) ubiquitin. Protein Sci. *23*, 662–667.

Hubbell, J., Kontos, S., and Dane, K. (2014). Erythrocyte-binding therapeutics.

Isom, D.G., Castañeda, C.A., Cannon, B.R., and García-Moreno E., B. (2011). Large shifts in pK_a values of lysine residues buried inside a protein. Proc. Natl. Acad. Sci. *108*, 5260 LP – 5265.

Jain, J., Veggiani, G., and Howarth, M. (2013). Cholesterol loading and ultrastable protein interactions determine the level of tumor marker required for optimal isolation of cancer cells. Cancer Res. *73*, 2310–2321.

Johansson, D.G.A., Macao, B., Sandberg, A., and Hard, T. (2008). SEA domain autoproteolysis accelerated by conformational strain: mechanistic aspects. J. Mol. Biol. *377*, 1130–1143.

Johansson, M.E. V, Thomsson, K.A., and Hansson, G.C. (2009). Proteomic Analyses of the Two Mucus Layers of the Colon Barrier Reveal That Their Main Component, the Muc2 Mucin, Is Strongly Bound to the Fcgbp Protein. J. Proteome Res. *8*, 3549–3557.

Junutula, J.R., Raab, H., Clark, S., Bhakta, S., Leipold, D.D., Weir, S., Chen, Y., Simpson, M., Tsai, S.P., Dennis, M.S., et al. (2008). Site-specific conjugation of a cytotoxic drug to an antibody improves the therapeutic index. Nat. Biotechnol. *26*, 925–932.

Kaczmarczyk, A., Thuveson, M., and Fries, E. (2002). Intracellular coupling of the heavy chain of pre-alpha-inhibitor to chondroitin sulfate. J. Biol. Chem. *277*, 13578–13582.

Keeble, A.H., Turkki, P., Stokes, S., Khairil Anuar, I.N.A., Rahikainen, R., Hytönen, V.P., and Howarth, M. (2019). Approaching infinite affinity through engineering of peptide–protein interaction. Proc. Natl. Acad. Sci. *116*, 26523 LP – 26533.

Keeble, A.H., Yadav, V.K., Ferla, M.P., Bauer, C.C., Chuntharpursat-Bon, E., Huang, J., Bon, R.S., and Howarth, M. (2021). DogCatcher allows loop-friendly protein-protein ligation. Cell Chem. Biol.

Kellogg, E.H., Leaver-Fay, A., and Baker, D. (2011). Role of conformational sampling in computing mutation-induced changes in protein structure and stability. Proteins *79*, 830–838.

Keyt, B.A., Paoni, N.F., Refino, C.J., Berleau, L., Nguyen, H., Chow, A., Lai, J., Peña, L., Pater, C., and Ogez, J. (1994). A faster-acting and more potent form of tissue plasminogen activator. Proc. Natl. Acad. Sci. U. S. A. *91*, 3670–3674.

Khairil Anuar, I.N.A., Banerjee, A., Keeble, A.H., Carella, A., Nikov, G.I., and Howarth, M. (2019). Spy&Go purification of SpyTag-proteins using pseudo-SpyCatcher to access an oligomerization toolbox. Nat. Commun. *10*, 1734.

Khlebnikov, A., Risa, O., Skaug, T., Carrier, T.A., and Keasling, J.D. (2000). Regulatable arabinose-inducible gene expression system with consistent control in all cells of a culture. J. Bacteriol. *182*, 7029–7034.

Kim, D.I., and Roux, K.J. (2016). Filling the Void: Proximity-Based Labeling of Proteins in Living Cells. Trends Cell Biol. *26*, 804–817.

Kim, J.K., Yang, I.S., Shin, H.J., Cho, K.J., Ryu, E.K., Kim, S.H., Park, S.S., and Kim, K.H. (2006). Insight into autoproteolytic activation from the structure of cephalosporin acylase: a protein with two proteolytic chemistries. Proc. Natl. Acad. Sci. U. S. A. *103*, 1732–1737.

Kimura, T. (2021). Stability and Safety of Inhibitor Cystine Knot Peptide, GTx1-15, from the Tarantula Spider Grammostola rosea. Toxins *13*.

Kimura, R.H., Cheng, Z., Gambhir, S.S., and Cochran, J.R. (2009a). Engineered knottin peptides: a new class of agents for imaging integrin expression in living subjects. Cancer Res. *69*, 2435–2442.

Kimura, R.H., Levin, A.M., Cochran, F. V, and Cochran, J.R. (2009b). Engineered cystine knot peptides that bind alphavbeta3, alphavbeta5, and alpha5beta1 integrins with low-nanomolar affinity. Proteins *77*, 359–369.

Kintzing, J.R., and Cochran, J.R. (2016). Engineered knottin peptides as diagnostics, therapeutics, and drug delivery vehicles. Curr. Opin. Chem. Biol. *34*, 143–150.

Kong, A., Calleja, V., Leboucher, P., Harris, A., Parker, P.J., and Larijani, B. (2008). HER2 oncogenic function escapes EGFR tyrosine kinase inhibitors via activation of

alternative HER receptors in breast cancer cells. PLoS One *3*.

Koshi, Y., Nakata, E., Miyagawa, M., Tsukiji, S., Ogawa, T., and Hamachi, I. (2008). Target-Specific Chemical Acylation of Lectins by Ligand-Tethered DMAP Catalysts. J. Am. Chem. Soc. *130*, 245–251.

Krall, J.A., Beyer, E.M., and MacBeath, G. (2011). High- and Low-Affinity Epidermal Growth Factor Receptor-Ligand Interactions Activate Distinct Signaling Pathways. PLoS One *6*, e15945.

Krall, N., da Cruz, F.P., Boutureira, O., and Bernardes, G.J.L. (2016). Site-selective protein-modification chemistry for basic biology and drug development. Nat. Chem. *8*, 103–113.

Krause, S., Schmoldt, H.-U., Wentzel, A., Ballmaier, M., Friedrich, K., and Kolmar, H. (2007). Grafting of thrombopoietin-mimetic peptides into cystine knot miniproteins yields high-affinity thrombopoietin antagonists and agonists. FEBS J. *274*, 86–95.

Kuban, V., Macek, P., Hritz, J., Nechvatalova, K., Nedbalcova, K., Faldyna, M., Sebo, P., Zidek, L., and Bumba, L. (2020a). Structural Basis of Ca(2+)-Dependent Self-Processing Activity of Repeat-in-Toxin Proteins. MBio *11*.

Kuninger, D., Kuns-Hashimoto, R., Nili, M., and Rotwein, P. (2008). Pro-protein convertases control the maturation and processing of the iron-regulatory protein, RGMc/hemojuvelin. BMC Biochem. *9*, 9.

Lam, A.K.M., and Galione, A. (2013). The endoplasmic reticulum and junctional membrane communication during calcium signaling. Biochim. Biophys. Acta - Mol. Cell Res. *1833*, 2542–2559.

Landsverk, O.J.B., Bakke, O., and Gregers, T.F. (2009). MHC II and the Endocytic Pathway: Regulation by Invariant Chain. Scand. J. Immunol. *70*, 184–193.

Lang, K., and Chin, J.W. (2014). Cellular Incorporation of Unnatural Amino Acids and Bioorthogonal Labeling of Proteins. Chem. Rev. *114*, 4764–4806.

Lang, T., Hansson, G.C., and Samuelsson, T. (2007). Gel-forming mucins appeared early in metazoan evolution. Proc. Natl. Acad. Sci. U. S. A. *104*, 16209–16214.

Lang, T., Klasson, S., Larsson, E., Johansson, M.E. V, Hansson, G.C., and Samuelsson, T. (2016). Searching the Evolutionary Origin of Epithelial Mucus Protein Components—Mucins and FCGBP. Mol. Biol. Evol. *33*, 1921–1936.

Latini, A., Agolini, E., Novelli, A., Borgiani, P., Giannini, R., Gravina, P., Smarrazzo, A., Dauri, M., Andreoni, M., Rogliani, P., et al. (2020). COVID-19 and Genetic Variants of Protein Involved in the SARS-CoV-2 Entry into the Host Cells. Genes (Basel). *11*.

Lee, J., Miller, B.T., and Brooks, B.R. (2016). Computational scheme for pH-dependent binding free energy calculation with explicit solvent. Protein Sci. *25*, 231–243.

Lehmann, M., Pasamontes, L., Lassen, S.F., and Wyss, M. (2000). The consensus concept for thermostability engineering of proteins. Biochim. Biophys. Acta *1543*, 408–415.

Li, N., Fort, F., Kessler, K., and Wang, W. (2009). Factors affecting cleavage at aspartic residues in model decapeptides. J. Pharm. Biomed. Anal. *50*, 73–78.

Lidell, M.E., and Hansson, G.C. (2006). Cleavage in the GDPH sequence of the C-terminal cysteine-rich part of the human MUC5AC mucin. Biochem. J. *399*, 121 LP – 129.

Lidell, M.E., Johansson, M.E. V, and Hansson, G.C. (2003). An Autocatalytic Cleavage in the C Terminus of the Human MUC2 Mucin Occurs at the Low pH of the Late Secretory Pathway. J. Biol. Chem. *278*, 13944–13951.

Lim, S.A., and Wells, J.A. (2020a). Chapter Twelve - Split enzymes: Design principles and strategy. In Enzyme Engineering and Evolution: Specific Enzyme Applications, D.S.B.T.-M. in E. Tawfik, ed. (Academic Press), pp. 275–296.

Lim, S.A., and Wells, J.A. (2020b). Split enzymes: Design principles and strategy. Methods Enzymol. *644*, 275–296.

Lins, L., Thomas, A., and Brasseur, R. (2003). Analysis of accessible surface of residues in proteins. Protein Sci. *12*, 1406–1417.

Liu, W.-J., Wu, Q., Xu, B., Zhang, X.-Y., Xia, X.-L., and Sun, H.-C. (2014). Single-step purification of recombinant proteins using elastin-like peptide-mediated inverse transition cycling and self-processing module from Neisseria meningitides FrpC. Protein Expr. Purif. *98*, 18–24.

Lobingier, B.T., Hüttenhain, R., Eichel, K., Miller, K.B., Ting, A.Y., von Zastrow, M., and Krogan, N.J. (2017). An Approach to Spatiotemporally Resolve Protein Interaction Networks in Living Cells. Cell *169*, 350-360.e12.

Long, X., Gou, Y., Luo, M., Zhang, S., Zhang, H., Bai, L., Wu, S., He, Q., Chen, K., Huang, A., et al. (2015). Soluble expression, purification, and characterization of active recombinant human tissue plasminogen activator by auto-induction in E. coli. BMC Biotechnol. *15*, 13.

Lopez, J.E., Haynes, S.E., Majmudar, J.D., Martin, B.R., and Fierke, C.A. (2017). HDAC8 Substrates Identified by Genetically Encoded Active Site Photocrosslinking. J. Am. Chem. Soc. *139*, 16222–16227.

López, R., Valbuena, J., Curtidor, H., Puentes, A., Rodríguez, L.E., García, J., Suárez, J., Vera, R., Ocampo, M., Trujillo, M., et al. (2004). Plasmodium falciparum: red blood cell binding studies using peptides derived from rhoptry-associated protein 2 (RAP2). Biochimie *86*, 1–6.

Lui, B.H., and Cochran, J.R. (2011). 4.412 - Rational and Combinatorial Methods to Create Designer Protein Interfaces. P.B.T.-C.B. Ducheyne, ed. (Oxford: Elsevier), pp. 161–179.

Maaß, F., Wüstehube-Lausch, J., Dickgießer, S., Valldorf, B., Reinwarth, M., Schmoldt, H.-U., Daneschdar, M., Avrutina, O., Sahin, U., and Kolmar, H. (2015). Cystine-knot peptides targeting cancer-relevant human cytotoxic T lymphocyte-associated antigen 4 (CTLA-4). J. Pept. Sci. *21*, 651–660.

Madeira, F., Park, Y.M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A.R.N., Potter, S.C., Finn, R.D., et al. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Res. *47*, W636—W641.

Magnani, M., Mancini, U., Bianchi, M., and Fazi, A. (1992). Comparison of uricase-bound and uricase-loaded erythrocytes as bioreactors for uric acid degradation. Adv. Exp. Med. Biol. *326*, 189–194.

Mains, R.E., and May, V. (1988). The role of a low pH intracellular compartment in the processing, storage, and secretion of ACTH and endorphin. J. Biol. Chem. *263*, 7887–7894.

de Marco, A. (2009). Strategies for successful recombinant expression of disulfide bond-dependent proteins in Escherichia coli. Microb. Cell Fact. *8*, 26.

Martegani, E., Forlani, N., Mauri, I., Porro, D., Schleuning, W.D., and Alberghina, L. (1992). Expression of high levels of human tissue plasminogen activator in yeast under the control of an inducible GAL promoter. Appl. Microbiol. Biotechnol. *37*, 604–608.

Martell, J.D., Deerinck, T.J., Sancak, Y., Poulos, T.L., Mootha, V.K., Sosinsky, G.E., Ellisman, M.H., and Ting, A.Y. (2012). Engineered ascorbate peroxidase as a genetically encoded reporter for electron microscopy. Nat. Biotechnol. *30*, 1143–1148.

Matsuo, K., and Hamachi, I. (2017). Ligand-Directed Tosyl and Acyl Imidazole Chemistry. Chemoselective Bioorthogonal Ligation React. 147–163.

Mattanovich, D., and Borth, N. (2006). Applications of cell sorting in biotechnology. Microb. Cell Fact. *5*, 12.

Matyska Liskova, P., Fiser, R., Macek, P., Chmelik, J., Sykora, J., Bednarova, L., Konopasek, I., and Bumba, L. (2016). Probing the Ca(2+)-assisted pi-pi interaction during Ca(2+)-dependent protein folding. Soft Matter *12*, 531–541.

Mayo, C.B., and Cole, J.L. 1038_nmeth752. ri. (2017). Interaction of PKR with single-stranded RNA. Sci. Rep. *7*, 3335.

Merlot, A.M., Kalinowski, D.S., and Richardson, D.R. (2014). Unraveling the mysteries of serum albumin-more than just a serum protein. Front. Physiol. *5*, 299.

Miroux, B., and Walker, J.E. (1996). Over-production of proteins in Escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. J. Mol. Biol. *260*, 289–298.

Miseta, A., and Csutora, P. (2000). Relationship Between the Occurrence of Cysteine in Proteins and the Complexity of Organisms. Mol. Biol. Evol. *17*, 1232–

1239.

Monnier, P.P., Sierra, A., Macchi, P., Deitinghoff, L., Andersen, J.S., Mann, M., Flad, M., Hornberger, M.R., Stahl, B., Bonhoeffer, F., et al. (2002). RGM is a repulsive guidance molecule for retinal axons. Nature *419*, 392–395.

Muramatsu, T. (2016). Basigin (CD147), a multifunctional transmembrane glycoprotein with various binding partners. J. Biochem. *159*, 481–490.

Murciano, J.-C., Medinilla, S., Eslin, D., Atochina, E., Cines, D.B., and Muzykantov, V.R. (2003). Prophylactic fibrinolysis through selective dissolution of nascent clots by tPA-carrying erythrocytes. Nat. Biotechnol. *21*, 891–896.

Muzykantov, V.R. (2010). Drug delivery by red blood cells: vascular carriers designed by mother nature. Expert Opin. Drug Deliv. *7*, 403–427.

Nestl, B.M., and Hauer, B. (2014). Engineering of Flexible Loops in Enzymes. ACS Catal. *4*, 3201–3211.

Nguyen, V.D., Hatahet, F., Salo, K.E.H., Enlund, E., Zhang, C., and Ruddock, L.W. (2011). Pre-expression of a sulfhydryl oxidase significantly increases the yields of eukaryotic disulfide bond containing proteins expressed in the cytoplasm of E.coli. Microb. Cell Fact. *10*, 1.

Oh, B.-K., Nam, J.-M., Lee, S.W., and Mirkin, C.A. (2006). A fluorophore-based bio-barcode amplification assay for proteins. Small *2*, 103–108.

Olzscha, H. (2019). Posttranslational modifications and proteinopathies: how guardians of the proteome are defeated. Biol. Chem. *400*, 895–915.

Oregioni, A., Stieglitz, B., Kelly, G., Rittinger, K., and Frenkiel, T. (2017). Determination of the pKa of the N-terminal amino group of ubiquitin by NMR. Sci. Rep. *7*, 43748.

Orlov, D., and Karkouti, K. (2015). The pathophysiology and consequences of red blood cell storage. Anaesthesia *70*, 29-e12.

Osicka, R., Kalmusová, J., Krízová, P., and Sebo, P. (2001). Neisseria meningitidis RTX protein FrpC induces high levels of serum antibodies during invasive disease: polymorphism of frpC alleles and purification of recombinant FrpC. Infect. Immun. *69*, 5509–5519.

Osicka, R., Prochazkova, K., Sulc, M., Linhartova, I., Havlicek, V., and Sebo, P. (2004). A novel “clip-and-link” activity of repeat in toxin (RTX) proteins from gram-negative pathogens. Covalent protein cross-linking by an Asp-Lys isopeptide bond upon calcium-dependent processing at an Asp-Pro bond. J. Biol. Chem. *279*, 24944–24956.

Ostrem, J.M., Peters, U., Sos, M.L., Wells, J.A., and Shokat, K.M. (2013). K-Ras(G12C) inhibitors allosterically control GTP affinity and effector interactions. Nature *503*, 548–551.

Packer, M.S., and Liu, D.R. (2015). Methods for the directed evolution of proteins. Nat. Rev. Genet. *16*, 379–394.

Pallaghy, P.K., Nielsen, K.J., Craik, D.J., and Norton, R.S. (1994). A common structural motif incorporating a cystine knot and a triple-stranded beta-sheet in toxic and inhibitory polypeptides. Protein Sci. *3*, 1833–1839.

Paoni, N.F., Keyt, B.A., Refino, C.J., Chow, A.M., Nguyen, H. V, Berleau, L.T., Badillo, J., Peña, L.C., Brady, K., Wurm, F.M., et al. (1993). A Slow Clearing, Fibrin-Specific, PAI-1 Resistant Variant of t-PA (T103N, KHRR 296-299 AAAA). Thromb Haemost *70*, 307–312.

Papanikolaou, G., Samuels, M.E., Ludwig, E.H., MacDonald, M.L.E., Franchini, P.L., Dube, M.-P., Andres, L., MacFarlane, J., Sakellaropoulos, N., Politou, M., et al. (2004). Mutations in HFE2 cause iron overload in chromosome 1q-linked juvenile hemochromatosis. Nat. Genet. *36*, 77–82.

Park, J., Koh, M., Koo, J.Y., Lee, S., and Park, S.B. (2016). Investigation of Specific Binding Proteins to Photoaffinity Linkers for Efficient Deconvolution of Target Protein. ACS Chem. Biol. *11*, 44–52.

Paroutis, P., Touret, N., and Grinstein, S. (2004). The pH of the Secretory Pathway: Measurement, Determinants, and Regulation. Physiology *19*, 207–215.

Paul, S., Kolla, R. V, Sidney, J., Weiskopf, D., Fleri, W., Kim, Y., Peters, B., and Sette, A. (2013). Evaluating the Immunogenicity of Protein Drugs by Applying *In Vitro* MHC Binding Data and the Immune Epitope Database and Analysis Resource. Clin. Dev. Immunol. *2013*, 467852.

Petruccelli, S., Otegui, M.S., Lareu, F., Tran Dinh, O., Fitchette, A.-C., Circosta, A., Rumbo, M., Bardor, M., Carcamo, R., Gomord, V., et al. (2006). A KDEL-tagged monoclonal antibody is efficiently retained in the endoplasmic reticulum in leaves, but is both partially secreted and sorted to protein storage vacuoles in seeds. Plant Biotechnol. J. *4*, 511–527.

Pham, N.D., Parker, R.B., and Kohler, J.J. (2013). Photocrosslinking approaches to interactome mapping. Curr. Opin. Chem. Biol. *17*, 90–101.

Pishesha, N., Bilate, A.M., Wibowo, M.C., Huang, N.-J., Li, Z., Deshycka, R., Bousbaine, D., Li, H., Patterson, H.C., Dougan, S.K., et al. (2017). Engineered erythrocytes covalently linked to antigenic peptides can protect against autoimmune disease. Proc. Natl. Acad. Sci. U. S. A. *114*, 3157–3162.

Piszkiewicz, D., Landon, M., and Smith, E.L. (1970). Anomalous cleavage of aspartyl-proline peptide bonds during amino acid sequence determinations. Biochem. Biophys. Res. Commun. *40*, 1173–1178.

Popp, M.W., Antos, J.M., Grotenbreg, G.M., Spooner, E., and Ploegh, H.L. (2007). Sortagging: a versatile method for protein labeling. Nat. Chem. Biol. *3*, 707–708.

Prochazkova, K., Osicka, R., Linhartova, I., Halada, P., Sulc, M., and Sebo, P. (2005). The Neisseria meningitidis outer membrane lipoprotein FrpD binds the RTX

protein FrpC. J. Biol. Chem. *280*, 3251–3258.

Puentes, A., Garcia, J., Vera, R., Lopez, Q.R., Urquiza, M., Vanegas, M., Salazar, L.M., and Patarroyo, M.E. (2000). Serine repeat antigen peptides which bind specifically to red blood cells. Parasitol. Int. *49*, 105–117.

Qiu, J., Swartz, J.R., and Georgiou, G. (1998). Expression of active human tissue-type plasminogen activator in Escherichia coli. Appl. Environ. Microbiol. *64*, 4891–4896.

Rahikainen, R., Rijal, P., Tan, T.K., Wu, H.-J., Andersson, A.-M.C., Barrett, J.R., Bowden, T.A., Draper, S.J., Townsend, A.R., and Howarth, M. (2021). Overcoming Symmetry Mismatch in Vaccine Nanoassembly through Spontaneous Amidation. Angew. Chemie Int. Ed. *60*, 321–330.

Ramanathan, M., Majzoub, K., Rao, D.S., Neela, P.H., Zarnegar, B.J., Mondal, S., Roth, J.G., Gai, H., Kovalski, J.R., Siprashvili, Z., et al. (2018). RNA-protein interaction detection in living cells. Nat. Methods *15*, 207–212.

Ravenhill, B.J., Kanjee, U., Ahouidi, A., Nobre, L., Williamson, J., Goldberg, J.M., Antrobus, R., Dieye, T., Duraisingh, M.T., and Weekes, M.P. (2019). Quantitative comparative analysis of human erythrocyte surface proteins between individuals from two genetically distinct populations. Commun. Biol. *2*, 350.

Reid, M.E., and Mohandas, N. (2004). Red blood cell blood group antigens: structure and function. Semin. Hematol. *41*, 93–117.

Reidhaar-Olson, J.F., and Sauer, R.T. (1988). Combinatorial cassette mutagenesis as a probe of the informational content of protein sequences. Science (80-.). *241*, 53–57.

Reiss, S., Sieber, M., Reiss, S., Sieber, M., Oberle, V., Wentzel, A., Spangenberg, P., Claus, R., Kolmar, H., and Lösche, W. (2006). Inhibition of platelet aggregation by grafting RGD and KGD sequences on the structural scaffold of small disulfide-rich proteins. Platelets *17*, 153–157.

Ren, C., Wen, X., Mencius, J., and Quan, S. (2019). Selection and screening strategies in directed evolution to improve protein stability. Bioresour. Bioprocess. *6*, 53.

Rhee, H.-W., Zou, P., Udeshi, N.D., Martell, J.D., Mootha, V.K., Carr, S.A., and Ting, A.Y. (2013). Proteomic mapping of mitochondria in living cells via spatially restricted enzymatic tagging. Science *339*, 1328–1331.

Roux, K.J., Kim, D.I., Raida, M., and Burke, B. (2012). A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. J. Cell Biol. *196*, 801–810.

Roux, K.J., Kim, D.I., Burke, B., and May, D.G. (2018). BioID: A Screen for Protein-Protein Interactions. Curr. Protoc. Protein Sci. *91*, 19.23.1-19.23.15.

Ruan, Q., Zhao, C., Ramsay, C.S., and Tetin, S.Y. (2018). Characterization of

Fluorescently Labeled Protein with Electrospray Ionization-MS and Fluorescence Spectroscopy: How Random is Random Labeling? Anal. Chem. *90*, 9695–9699.

Ruddell, J.P., Lippert, L.E., Babcock, J.G., and Hess, J.R. (1998). Effect of 24-hour storage at 25 degrees C on the in vitro storage characteristics of CPDA-1 packed red cells. Transfusion *38*, 424–428.

Sadilkova, L., Osicka, R., Linhartova, I., Prochazkova, K., and Sebo, P. (2006). A Novel Self-Processing Module Derived from the FrpC Protein of Neisseria meningitidis for a Single-Step Purification of Free Recombinant Proteins. FASEB J. *20*, A960–A961.

Sadilkova, L., Osicka, R., Sulc, M., Linhartova, I., Novak, P., and Sebo, P. (2008). Single-step affinity purification of recombinant proteins using a self-excising module from Neisseria meningitidis FrpC. Protein Sci. *17*, 1834–1843.

Salema, V., and Fernández, L.Á. (2017). Escherichia coli surface display for the selection of nanobodies. Microb. Biotechnol. *10*, 1468–1484.

Salema, V., López-Guajardo, A., Gutierrez, C., Mencía, M., and Fernández, L.Á. (2016a). Characterization of nanobodies binding human fibrinogen selected by E. coli display. J. Biotechnol. *234*, 58–65.

Salema, V., Mañas, C., Cerdán, L., Piñero-Lambea, C., Marín, E., Roovers, R.C., Van Bergen En Henegouwen, P.M.P., and Fernández, L.Á. (2016b). High affinity nanobodies against human epidermal growth factor receptor selected on cells by E. coli display. MAbs *8*, 1286–1301.

Satchwell, T.J., Wright, K.E., Haydn-Smith, K.L., Sánchez-Román Terán, F., Moura, P.L., Hawksworth, J., Frayne, J., Toye, A.M., and Baum, J. (2019). Genetic manipulation of cell line derived reticulocytes enables dissection of host malaria invasion requirements. Nat. Commun. *10*, 3806.

Scheu, A.H.A., Lim, S.Y.T., Metzner, F.J., Mohammed, S., and Howarth, M. (2021). NeissLock provides an inducible protein anhydride for covalent targeting of endogenous proteins. Nat. Commun. *12*, 717.

Schultz, J. (1967). No. In Methods in Enzymology, p. 255.

Serpe, M., Umulis, D., Ralston, A., Chen, J., Olson, D.J., Avanesov, A., Othmer, H., O'Connor, M.B., and Blair, S.S. (2008). The BMP binding protein Crossveinless 2 is a short-range, concentration-dependent, biphasic modulator of BMP signaling in Drosophila. Dev. Cell *14*, 940–953.

Shah, N.H., and Muir, T.W. (2014). Inteins: Nature's Gift to Protein Chemists. Chem. Sci. *5*, 446–461.

Shemin, D., and Rittenberg, D. (1946). The life span of the human red blood cell. J. Biol. Chem. *166*, 627–636.

Sheng, Z., Wu, K., Carraway, K.L., and Fregien, N. (1992). Molecular cloning of the transmembrane component of the 13762 mammary adenocarcinoma sialomucin

complex. A new member of the epidermal growth factor superfamily. J. Biol. Chem. *267*, 16341–16346.

Shi, J., Kundrat, L., Pishesha, N., Bilate, A., Theile, C., Maruyama, T., Dougan, S.K., Ploegh, H.L., and Lodish, H.F. (2014). Engineered red blood cells as carriers for systemic delivery of a wide array of functional probes. Proc. Natl. Acad. Sci. *111*, 10131 LP – 10136.

Shieh, P., and Bertozzi, C.R. (2014). Design strategies for bioorthogonal smart probes. Org. Biomol. Chem. *12*, 9307–9320.

Siebold, C., Yamashita, T., Monnier, P.P., Mueller, B.K., and Pasterkamp, R.J. (2017). RGMs: Structural Insights, Molecular Regulation, and Downstream Signaling. Trends Cell Biol. *27*, 365–378.

Silvestri, L., Pagani, A., and Camaschella, C. (2008). Furin-mediated release of soluble hemojuvelin: a new link between hypoxia and iron homeostasis. Blood *111*, 924–931.

Singh, B., and Coffey, R.J. (2014). From wavy hair to naked proteins: the role of transforming growth factor alpha in health and disease. Semin. Cell Dev. Biol. *28*, 12–21.

Smith, E., and Collins, I. (2015). Photoaffinity labeling in target- and binding-site identification. Future Med. Chem. *7*, 159–183.

Snapp, E. (2005). Design and use of fluorescent fusion proteins in cell biology. Curr. Protoc. Cell Biol. *Chapter 21*, 21.4.1-21.4.13.

Spence, E.F., Dube, S., Uezu, A., Locke, M., Soderblom, E.J., and Soderling, S.H. (2019). In vivo proximity proteomics of nascent synapses reveals a novel regulator of cytoskeleton-mediated synaptic maturation. Nat. Commun. *10*, 386.

Steffen, W., Ko, F.C., Patel, J., Lyamichev, V., Albert, T.J., Benz, J., Rudolph, M.G., Bergmann, F., Streidl, T., Kratzsch, P., et al. (2017). Discovery of a microbial transglutaminase enabling highly site-specific labeling of proteins. J. Biol. Chem. *292*, 15622–15635.

Stevens, A.J., Sekar, G., Shah, N.H., Mostafavi, A.Z., Cowburn, D., and Muir, T.W. (2017). A promiscuous split intein with expanded protein engineering applications. Proc. Natl. Acad. Sci. 201701083.

Stornaiuolo, M., Lotti, L. V, Borgese, N., Torrisi, M.-R., Mottola, G., Martire, G., and Bonatti, S. (2003). KDEL and KKXX retrieval signals appended to the same reporter protein determine different trafficking between endoplasmic reticulum, intermediate compartment, and Golgi complex. Mol. Biol. Cell *14*, 889–902.

Suchanek, M., Radzikowska, A., and Thiele, C. (2005). Photo-leucine and photo-methionine allow identification of protein-protein interactions in living cells. Nat. Methods *2*, 261–268.

Sviridova, E., Rezacova, P., Bondar, A., Veverka, V., Novak, P., Schenk, G.,

Svergun, D.I., Kuta Smatanova, I., and Bumba, L. (2017). Structural basis of the interaction between the putative adhesion-involved and iron-regulated FrpD and FrpC proteins of Neisseria meningitidis. Sci. Rep. *7*, 40408.

Szijj, P.A., Kostadinova, K.A., Spears, R.J., and Chudasama, V. (2020). Tyrosine bioconjugation – an emergent alternative. Org. Biomol. Chem. *18*, 9018–9028.

Takagi, M., Hayashi, H., and Yoshida, T. (2000). The effect of osmolarity on metabolism and morphology in adhesion and suspension chinese hamster ovary cells producing tissue plasminogen activator. Cytotechnology *32*, 171–179.

Tamura, T., and Hamachi, I. (2019). Chemistry for Covalent Modification of Endogenous/Native Proteins: From Test Tubes to Complex Biological Systems. J. Am. Chem. Soc. *141*, 2782–2799.

Tamura, T., Tsukiji, S., and Hamachi, I. (2012). Native FKBP12 Engineering by Ligand-Directed Tosyl Chemistry: Labeling Properties and Application to Photo-Cross-Linking of Protein Complexes in Vitro and in Living Cells. J. Am. Chem. Soc. *134*, 2216–2226.

Tamura, T., Song, Z., Amaike, K., Lee, S., Yin, S., Kiyonaka, S., and Hamachi, I. (2017). Affinity-Guided Oxime Chemistry for Selective Protein Acylation in Live Tissue Systems. J. Am. Chem. Soc. *139*, 14181–14191.

Tamura, T., Ueda, T., Goto, T., Tsukidate, T., Shapira, Y., Nishikawa, Y., Fujisawa, A., and Hamachi, I. (2018). Rapid labelling and covalent inhibition of intracellular native proteins using ligand-directed N-acyl-N-alkyl sulfonamide. Nat. Commun. *9*, 1870.

Thiagarajan, P., Parker, C.J., and Prchal, J.T. (2021). How Do Red Blood Cells Die? Front. Physiol. *12*, 318.

Thompson, S.A., and Sparling, P.F. (1993). The RTX cytotoxin-related FrpA protein of Neisseria meningitidis is secreted extracellularly by meningococci and by HlyBD+ Escherichia coli. Infect. Immun. *61*, 2906–2911.

Thuveson, M., and Fries, E. (2000). The Low pH in Trans-Golgi Triggers Autocatalytic Cleavage of Pre-α-inhibitor Heavy Chain Precursor. J. Biol. Chem. *275*, 30996–31000.

Tomas, A., Futter, C.E., and Eden, E.R. (2014). EGF receptor trafficking: Consequences for signaling and cancer. Trends Cell Biol.

Torrent, M., Llompart, B., Lasserre-Ramassamy, S., Llop-Tous, I., Bastida, M., Marzabal, P., Westerholm-Parvinen, A., Saloheimo, M., Heifetz, P.B., and Ludevid, M.D. (2009). Eukaryotic protein production in designed storage organelles. BMC Biol. *7*, 5.

Trakarnsanga, K., Griffiths, R.E., Wilson, M.C., Blair, A., Satchwell, T.J., Meinders, M., Cogan, N., Kupzig, S., Kurita, R., Nakamura, Y., et al. (2017). An immortalized adult human erythroid line facilitates sustainable and scalable generation of functional red cells. Nat. Commun. *8*, 14750.

Tripathi, N.K., and Shrivastava, A. (2019). Recent Developments in Bioprocessing of Recombinant Proteins: Expression Hosts and Process Development. Front. Bioeng. Biotechnol. *7*, 420.

Tsukiji, S., Miyagawa, M., Takaoka, Y., Tamura, T., and Hamachi, I. (2009). Ligand-directed tosyl chemistry for protein labeling in vivo. Nat. Chem. Biol. *5*, 341–343.

Tzakoniati, F., Xu, H., Li, T., Garcia, N., Kugel, C., Payandeh, J., Koth, C.M., and Tate, E.W. (2020). Development of Photocrosslinking Probes Based on Huwentoxin-IV to Map the Site of Interaction on Nav1.7. Cell Chem. Biol. *27*, 306-313.e4.

Varnaitė, R., and MacNeill, S.A. (2016). Meet the neighbors: Mapping local protein interactomes by proximity-dependent labeling with BioID. Proteomics *16*, 2503–2518.

Vazquez-Lombardi, R., Phan, T.G., Zimmermann, C., Lowe, D., Jermutus, L., and Christ, D. (2015). Challenges and opportunities for non-antibody scaffold drugs. Drug Discov. Today *20*, 1271–1283.

Vera Bravo, R., Marín, V., García, J., Urquiza, M., Torres, E., Trujillo, M., Rosas, J., and Patarroyo, M.E. (2000). Amino terminal peptides of the ring infected erythrocyte surface antigen of Plasmodium falciparum bind specifically to erythrocytes. Vaccine *18*, 1289–1293.

Verkhratsky, A. (2005). Physiology and Pathophysiology of the Calcium Store in the Endoplasmic Reticulum of Neurons. Physiol. Rev. *85*, 201–279.

Villa, C.H., Pan, D.C., Johnston, I.H., Greineder, C.F., Walsh, L.R., Hood, E.D., Cines, D.B., Poncz, M., Siegel, D.L., and Muzykantov, V.R. (2018). Biocompatible coupling of therapeutic fusion proteins to human erythrocytes. Blood Adv. *2*, 165–176.

Wang, Y., Prosen, D.E., Mei, L., Sullivan, J.C., Finney, M., and Vander Horn, P.B. (2004). A novel strategy to engineer DNA polymerases for enhanced processivity and improved performance in vitro. Nucleic Acids Res. *32*, 1197–1207.

Warach, S.J., Dula, A.N., and Milling, T.J. (2020). Tenecteplase Thrombolysis for Acute Ischemic Stroke. Stroke *51*, 3440–3451.

Ward, C.C., Kleinman, J.I., and Nomura, D.K. (2017). NHS-Esters As Versatile Reactivity-Based Probes for Mapping Proteome-Wide Ligandable Hotspots. ACS Chem. Biol. *12*, 1478–1483.

Wentzel, A., Christmann, A., Adams, T., and Kolmar, H. (2001). Display of passenger proteins on the surface of Escherichia coli K-12 by the enterohemorrhagic E. coli intimin EaeA. J. Bacteriol. *183*, 7273–7284.

Winkler, M.E., Bringman, T., and Marks, B.J. (1986). The purification of fully active recombinant transforming growth factor alpha produced in Escherichia coli. J. Biol. Chem. *261*, 13838–13843.

Wofsy, L., Metzger, H., and Singer, S.J. (1962). Affinity Labeling—a General Method

for Labeling the Active Sites of Antibody and Enzyme Molecules*. Biochemistry *1*, 1031–1039.

Wong, S.T., Winchell, L.F., McCune, B.K., Earp, H.S., Teixidó, J., Massagué, J., Herman, B., and Lee, D.C. (1989). The TGF-alpha precursor expressed on the cell surface binds to the EGF receptor on adjacent cells, leading to signal transduction. Cell *56*, 495–506.

Wood, D.W., and Camarero, J.A. (2014). Intein applications: from protein purification and labeling to metabolic control methods. J. Biol. Chem. *289*, 14512–14519.

Wu, H., and Kohler, J. (2019). Photocrosslinking probes for capture of carbohydrate interactions. Curr. Opin. Chem. Biol. *53*, 173–182.

Wu, H.-Y., Chen, S.-F., Hsieh, J.-Y., Chou, F., Wang, Y.-H., Lin, W.-T., Lee, P.-Y., Yu, Y.-J., Lin, L.-Y., Lin, T.-S., et al. (2015). Structural basis of antizyme-mediated regulation of polyamine homeostasis. Proc. Natl. Acad. Sci. U. S. A. *112*, 11229–11234.

Wu, X., Lee, J., and Brooks, B.R. (2017). Origin of pKa Shifts of Internal Lysine Residues in SNase Studied Via Equal-Molar VMMS Simulations in Explicit Water. J. Phys. Chem. B *121*, 3318–3330.

Xiang, Z., Ren, H., Hu, Y.S., Coin, I., Wei, J., Cang, H., and Wang, L. (2013). Adding an unnatural covalent bond to proteins through proximity-enhanced bioreactivity. Nat. Methods *10*, 885–888.

Xue, L., Karpenko, I.A., Hiblot, J., and Johnsson, K. (2015). Imaging and manipulating proteins in live cells through covalent labeling. Nat. Chem. Biol. *11*, 917–923.

Yamaguchi, H., and Miyazaki, M. (2014). Refolding techniques for recovering biologically active recombinant proteins from inclusion bodies. Biomolecules *4*, 235–251.

Yan, Y., Jiang, K., Liu, P., Zhang, X., Dong, X., Gao, J., Liu, Q., Barr, M.P., Zhang, Q., Hou, X., et al. (2016). Bafilomycin A1 induces caspase-independent cell death in hepatocellular carcinoma cells via targeting of autophagy and MAPK pathways. Sci. Rep. *6*, 37052.

Zaitsev, S., Danielyan, K., Murciano, J.-C., Ganguly, K., Krasik, T., Taylor, R.P., Pincus, S., Jones, S., Cines, D.B., and Muzykantov, V.R. (2006). Human complement receptor type 1-directed loading of tissue plasminogen activator on circulating erythrocytes for prophylactic fibrinolysis. Blood *108*, 1895–1902.

Zaitsev, S., Spitzer, D., Murciano, J.-C., Ding, B.-S., Tliba, S., Kowalska, M.A., Bdeir, K., Kuo, A., Stepanova, V., Atkinson, J.P., et al. (2010). Targeting of a mutant plasminogen activator to circulating red blood cells for prophylactic fibrinolysis. J. Pharmacol. Exp. Ther. *332*, 1022–1031.

Zakeri, B., Fierer, J.O., Celik, E., Chittock, E.C., Schwarz-Linek, U., Moy, V.T., and Howarth, M. (2012). Peptide tag forming a rapid covalent bond to a protein, through

engineering a bacterial adhesin. Proc. Natl. Acad. Sci. U. S. A. *109*, E690-7.

Zhang, M., and Vogel, H.J. (1993). Determination of the side chain pKa values of the lysine residues in calmodulin. J. Biol. Chem. *268*, 22420–22428.

Zhang, A.-S., West, A.P.J., Wyman, A.E., Bjorkman, P.J., and Enns, C.A. (2005). Interaction of hemojuvelin with neogenin results in iron accumulation in human embryonic kidney 293 cells. J. Biol. Chem. *280*, 33885–33894.

Zhang, M., Lin, S., Song, X., Liu, J., Fu, Y., Ge, X., Fu, X., Chang, Z., and Chen, P.R. (2011). A genetically incorporated crosslinker reveals chaperone cooperation in acid resistance. Nat. Chem. Biol. *7*, 671–677.

Zhang, T., Hatcher, J.M., Teng, M., Gray, N.S., and Kostic, M. (2019). Recent Advances in Selective and Irreversible Covalent Ligand Development and Validation. Cell Chem. Biol. *26*, 1486–1500.

Zilberstein, D., Agmon, V., Schuldiner, S., and Padan, E. (1984). Escherichia coli intracellular pH, membrane potential, and cell growth. J. Bacteriol. *158*, 246–252.

Zutic, J., Radosavljevic, V., Radanovic, □, Ivetic, V., Pavlovic, I., and Zutic, M. (2013). A Case Report: Isolation of Alysiella filiformis from Pig's Lungs [1]. Kafkas Univ. Vet. Fak. Derg. *19*, 691–693.