Harnessing Bioreactivity for Protein Covalent Probes against Endogenous Targets

Arne Hagen August Scheu

Department of Biochemistry
Lincoln College
University of Oxford

Trinity 2021

This thesis is submitted to the Department of Biochemistry at the University of Oxford in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Abstract

Various small-molecule drugs that covalently modify their targets have been discovered and designed to date. Covalent reaction has also been identified in the interaction between the cells of the human body and microorganisms, both from the immune system to pathogens and from pathogens to host cells. Harnessing bioreactivity found in covalent pathogen adhesion, in this work I have demonstrated the use of a self-activating protein domain from *Neisseria meningitidis* for covalent reaction against endogenous targets.

N. meningitidis iron-regulated protein C (FrpC) contains a 'self-processing module' (SPM); subject to calcium-activation, SPM cleaves the protein backbone at an Asp-Pro bond, forming a reactive anhydride from the Asp-Pro cleavage site. This protein anhydride can facilitate protein-protein crosslinking via reaction with lysine residues. SPM can also be introduced recombinantly to introduce a reactive anhydride to non-FrpC proteins. In this thesis I established 'NeissLock', a method for targeted protein-protein conjugation after SPM-mediated biochemical activation. First, I designed and implemented 'NeissDist', a tool to identify protein complexes in which a binder protein activated by SPM is suitably positioned for covalent attachment to a target protein of interest. I identified the complexes of Ornithine Decarboxylase Antizyme (OAZ) binding to Ornithine Decarboxylase (ODC) and Transforming Growth Factor alpha (TGFα) binding to Epidermal Growth Factor Receptor (EGFR) as model systems to study NeissLock conjugation.

NeissLock-mediated conjugation of OAZ to ODC shows effective (>50% yield) and specific coupling to the target protein, over a broad pH range. Using the OAZ/ODC system, I characterised conjugation to multiple target residues, showing flexibility in the site of reaction, and gained insight into linker design for efficient anhydride formation. Having shown specific conjugation of TGFα to EGFR on cells, NeissLock technology may allow broad compatibility with further protein-protein complexes as well as a range of therapeutic and diagnostic applications.

Declaration of Authorship

I declare that this thesis constitutes my own work, other than for instances described herein. This thesis has not previously been submitted for any other degree. Parts of this thesis, including figures with or without modification, were published in (Scheu *et al.*, 2021).

Extended experiments on pH tolerance and dependence of OAZ/ODC NeissLock conjugation on complex affinity were conducted by Sheryl Lim as indicated, with protein samples purified by Sheryl Lim.

Mass spectrometry of intact proteins was performed by Anthony Tumber.

In-gel tryptic digestion for mass spectrometry was performed by or with Sabrina Liberatori and Melissa Webby. Mass spectrometry of tryptic fragments as well as data analysis of these experiments were performed by the research group of Shabaz Mohammed or by Shabaz Mohammed.

SpyCatcher003:DyLight protein conjugate was provided by Samuel Stokes.

Figure 3.2 incorporates graphics retrieved from www.flaticon.com, see (Scheu *et al.*, 2021) for license information.

Acknowledgments

I am grateful to Professor Mark Howarth for his vision and guidance with NeissLock, for entrusting me to set a foundation in this project, encouraging me take to big leaps, helping me to assemble a complete story, and for fostering my curiosity in entrepreneurship. I am also grateful for Professor Shabaz Mohammed for his fruitful collaboration, including various helpful discussions.

I want to thank the members of the Howarth Group for providing a nurturing and supportive environment. I am grateful for advice of Robert Wieduwild, James Ross, and Matteo Ferla. I am thankful to Sheryl Lim for her collaboration towards the publication of the NeissLock technology. Both Irsyad Khairil and Sheryl Lim were a fantastic source of support, within and outside the research.

I am most grateful to The Rhodes Trust and Rhodes House, for funding my studies and for support far beyond that. I want to thank the community of Rhodes Scholars, especially Jory Fleming, James & Casey Pavur, Matthew Rogers, Kristiana Yao, Jung-Kian Un, Simone Delzin, Beatrice Yeung, Philipp Kerth, and Daisy Fleming, for making my life at Oxford unique. I also want to thank my peers at Lincoln College for their friendship, especially Sofie Behluli and James Freeman. I want to thank Lincoln College as well as the members of the College for their support, especially Ioannis Vakonakis. I also want to thank Jonas Koeppel, Marc Etri, Simon Pelczer and Martin Millan Jochum for lifelong Freundschaft, and I want to encourage Jonas Koeppel to keep striving for the highest achievement.

Finally, I want to give my most heartfelt thanks to my family: to Sven, Fenja, Neele, for the closest bond; to my parents, for unwavering trust and support; to my grandparents, who I always hoped to make proud. I dedicate my work to Rolf Würtele, whose memory provides me with strength and inspiration.

Abbreviations

A_{280}	Absorbance at 280 nm wavelength
AzI	Antizyme Inhibitor
BLA	β-lactamase
CAPP	Cleavage after protein photo-cross-linking
disCrawl	Distance Crawler
DMEM	Dulbecco's modified eagle medium
DOPA	3,4-dihydroxy-L-phenylalanine
DTT	Dithiothreitol
EBA	N,N'-ethylene bisacrylamide
EDTA	Ethylenediaminetetraacetic acid
EGFR	Epidermal Growth Factor Receptor
FrpC	Iron-regulated protein C
FrpD	Iron-regulated protein D
GAS	Group A *Streptococcus pyogenes*
HBS	HEPES-buffered saline
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid)
HRP	Horseradish peroxidase
IMAPP	(MS)-label transfer after protein photo-cross-linking
K_A	Association constant
K_D	Dissociation constant
k_{on}	Rate of association
k_{off}	Rate of dissociation
k_{irr}	Rate of irreversible conjugation
LC-MS	Liquid chromatography with mass spectrometry
LC-MS/MS	Liquid chromatography with tandem mass spectrometry
LDT	ligand-directed tosyl chemistry
LplA	Lipoic acid ligase
MALDI-TOF	Matrix-assisted laser desorption/ionization
MB	Megabyte
MBP	Maltose-binding protein
2-ME	2-Mercaptoethanol
MIAA	'Metal ion-inducible autocleavage'
NeissDist	Neisslock Distance Database
Nu	Nucleophile
Nε	Lysine Atom NZ
OAZ	Ornithine Decarboxylase Antizyme, also $OAZ^{95\text{-}219}$
OD_{600}	Optical density at 600 nm wavelength
ODC	Ornithine Decarboxylase
OH	Tyrosine Atom OH
P-(P)ALM	Post-(photo)affinity labeling modification
PAGE	Poly-acrylamide gel electrophoresis
PAL	Photo-affinity labelling
PBS	Phosphate-buffered saline
PDB	Protein Data Bank
PIC	Photoinduced covalent crosslinking
PMSF	Phenylmethylsulfonyl fluoride
PNGase F	Peptide:N-glycosidase F
PVDF	Polyvinylidene fluoride
RF-MS	Rapidfire mass spectrometry
RGMB	Repulsive Guidance Molecule B

RTX	Repeat-in toxin
SC	SpyCatcher
SDS	Sodium dodecylsulfate
SDS-LB	SDS loading buffer
sEGFR501	Soluble EGFR fragment 1-501
SPM	Self-processing module
ST	SpyTag
ST3	SpyTag003
TGFα	Transforming Growth Factor alpha
UAA	Unnatural amino acid
Wt	wildtype
X (atom)	Specified target atom of interest, e.g. Nε
X (amino acid)	Any amino acid, single letter code
Xaa	Any amino acid, three letter code

Contents

1. Introduction 11
1.1. (Photo)chemical activation in protein crosslinking 14
1.2. Proximity-enabled protein ligation 17
1.2.1. Chemical fine-tuning: proximity-enhanced bioreactivity in the Z_{SPA} model 21
1.2.2. Applications of proximity-enabled crosslinking 23
1.3. Foreign protein conjugation and covalent adhesion 24
1.4. The Neisseria meningitidis self-processing module 29
1.4.1. FrpC function in Neisseria meningitidis biology and pathogenicity 29
1.4.2. FrpC SPM structure and activity 31
1.4.3. Biotechnological applications of FrpC SPM 33
1.5. Thesis aims 35
2. Materials and Methods 37
2.1. The NeissDist Database 37
2.1.1. Data retrieval and biological assembly 37
2.1.2. NeissDist database assembly 37
2.1.3. NeissDist analysis 38
2.1.4. Candidate selection 39
2.1.5. Rational design of candidate protein point mutations 39
2.2. Plasmid design and cloning 40
2.2.1. Gene and oligonucleotide synthesis 41
2.2.2. DNA amplification and purification 41
2.2.3. Gibson assembly 41
2.2.4. Transformation of *Escherichia Coli* 42
2.2.5. Plasmid preparation 43
2.3. Bacterial protein expression and purification 43
2.3.1. Standard bacterial protein expression 43
2.3.2. Initial purification of ODC, OAZ-SPM and TGFα-SPM 44
2.3.3. Final purification conditions 44
2.3.4. Purification of Affibody-SPM 46
2.3.5. High-throughput preparation of SpyTag-X-SPM variants 46
2.4. Preparation of sEGFR501 47
2.5. General protein analysis 48
2.5.1. Estimation of protein concentration 48
2.5.2. SDS-PAGE 48
2.5.3. Western Blot 49

2.6. SPM activity and in-vitro conjugation assays 49
2.6.1. Affibody-SPM anhydride reactivity tests 49
2.6.2. SpyTag-X-SPM activity screen 50
2.6.3. OAZ-SPM:ODC and variant conjugation 50
2.6.4. TGFα-SPM:sEGFR and variant conjugation 52
2.7. Mass spectrometry 53
2.7.1. Intact mass spectrometry 53
2.7.2. Enzymatic digest analysis 54
2.8. Cellular assays 55
2.8.1. Cell culture 55
2.8.2. Immunostaining 55
2.8.3. TGFα-GSY-SPM cell conjugation 56
3. NeissDist database: Distance screen of the Protein Data Bank 58
3.1. NeissLock principle 58
3.2. Considerations for NeissLock candidates 60
3.3. disCrawl and the NeissDist database 61
3.3.1. Overview over the NeissDist pipeline 61
3.3.2. Implementation of disCrawl processing 62
3.3.3. Overview of NeissDist 68
3.3.4. Selection of NeissLock model complexes 71
3.3.5. Exploration of additional NeissDist complexes 74
3.4. Alternative disCrawl implementations and uses of NeissDist 77
3.4.1. Search for SPM alternatives 77
3.4.2. PDB dipeptide analysis 78
3.5. Discussion 79
4. Establishing the NeissLock principle 82
4.1. The Ornithine Decarboxylase / Antizyme in vitro model complex 82
4.1.1. Introduction to Ornithine Decarboxylase / Antizyme biology 83
4.2. Initial evidence of OAZ/ODC Neisslock-conjugation 85
4.3. Optimization of conjugation rate and yield 87
4.3.1. Kinetics of anhydride activation 87
4.3.2. Saturation mutagenesis of cleavage site with SpyTag-X-SPM 90
4.3.3. Purification conditions 92
4.3.4. Spacer insertion and steric constraints 95
4.3.5. pH-dependence 97
4.4. Intact mass spectrometry of OAZ:ODC conjugation 99

4.5. Mapping of OAZ:ODC crosslinking sites 103
4.5.1. Mutation of the target site 103
4.5.2. Introduction of alternative sites 103
4.5.3. Enzymatic digest and LC-MS/MS 105
4.6. Assessment of specificity 107
4.6.1. Design of binding mutants 107
4.6.2. Nonspecific protein conjugation 109
4.7. Discussion 110
5. Cellular NeissLock conjugation 117
5.1. Targeting the Epidermal Growth Factor Receptor with NeissLock 117
5.2. *In vitro* conjugation of EGFR/TGFα 119
5.2.1. Initial validation of TGFα-SPM cleavage activity 119
5.2.2. Preparation of soluble EGFR for *in vitro* study 120
5.2.3. Initial evidence of EGFR:TGFα conjugation 121
5.2.4. sEGFR501 deglycosylation and *in vitro* conjugation yield 123
5.2.5. Size exclusion chromatography and TGFα-GSY-SPM folding 125
5.3. Characterisation of sEGFR501:TGFα-GSY conjugate 129
5.3.1. Mapping of crosslinking sites 129
5.3.2. *In vitro* Western blot 130
5.4. Cellular NeissLock-conjugation 132
5.4.1. TGFα-GSY-SPM cell binding 132
5.4.2. Screening of cell conjugation conditions 133
5.4.3. Sensitivity of cellular conjugation to TGFα-GSY-SPM mutant design 134
5.4.4. Initial evidence of TGFα/EGFR conjugate fate 135
5.5. Discussion 137
6. Summary and Future Work 145
7. References 151
8. Appendix 173

1. Introduction

Specific covalent modification of endogenous proteins is challenging, but of high interest. Although non-covalent interactions can produce remarkably low dissociation rates (e.g. for biotin/streptavidin (Chivers *et al.*, 2011)), the transience of non-covalent interactions places limits on applications. By covalent conjugation to its target, the pharmacodynamics and pharmacokinetics of a drug can be improved and side effects can be reduced, e.g. by ensuring stable localisation of a drug and thereby reducing the need for prolonged heightened systemic concentrations (Singh *et al.*, 2011). At an early time in drug development, despite a lack of mechanistic understanding, covalent drugs were utilized: acetylsalicylic acid covalently modifies cyclooxygenase by acetylation, leading to the inactivation of the enzyme (Vane and Botting, 2003).

Covalent modification of endogenous proteins presents a unique challenge. Recombinantly produced proteins can be genetically modified to incorporate specific chemistry or recognition tags (e.g. transglutaminase (Lin and Ting, 2006), sortase (Proft, 2010), SpyCatcher (Zakeri *et al.*, 2012)), or small molecules such as FlAsH-tag (Griffin *et al.*, 1998)), or can be manipulated in a controlled manner as purified protein. A variety of useful reactions are available to introduce new functionalities after translation of a recombinant protein (Hoyt *et al.*, 2019). Both cysteine and lysine are commonly labelled, with cysteine offering high reactivity with good specificity due to its low abundance (Tjong *et al.*, 2007; Takaoka *et al.*, 2013). Disulfide-bonding still provides a convenient method for reversible modification of cysteines, wherein the bond can be broken by reduction (Saito *et al.*, 2003). The covalent modification of cysteine groups with maleimide derivatives is a popular irreversible modification (Takaoka *et al.*, 2013), e.g. to introduce a fluorescent label into a purified protein before imaging experiments. Modern approaches enable the specific modification of a wide array of amino acids (Hoyt *et al.*, 2019).

To modify only a specific type of endogenous protein within a mixed environment, i.e. without prior purification, it is necessary to introduce selectivity (Tamura and Hamachi, 2019). A target protein can be modified using chemical properties highly specific to the target protein, for

instance in the aforementioned modification of cyclooxygenase by acetylsalicylic acid (Vane and Botting, 2003) or in the inhibition of a protease active site with phenylmethylsulfonyl fluoride (PMSF) (Gold and Fahrney, 1964). Specific motifs in a protein can be recognised for covalent modification by an enzyme or domain with catalytic activity, for instance by transglutaminases (Lin and Ting, 2006), and aforementioned recombinant fusion tags have been derived from such motifs. However, most endogenous proteins do not offer conveniently exploitable activity or unique chemistry. To derive a generally applicable approach for modification of endogenous protein therefore presents a unique problem: such an approach needs to be compatible with common chemistry, i.e. reactions with the side-chains of the 20 proteogenic amino acids, the N-terminal amine, or the protein backbone, while being specific only to the protein of interest (Tamura and Hamachi, 2019). To introduce selectivity into endogenous protein conjugation in mixed samples, a protein-binding ligand can be fused to a reactive probe to preferentially react with the targeted protein (Tamura and Hamachi, 2019). Two prominent approaches are *in situ* activation of a (photo)chemical probe and proximity-enabled labelling techniques.

In (photo)chemical labelling, an endogenous target protein is bound by a molecule incorporating an inactive probe; then, a highly reactive probe is generated *in-situ* by (photo)chemical activation (Preston and Wilson, 2013; Murale *et al.*, 2017). In this strategy, after activation of the probe (e.g. by UV radiation) it crosslinks nearby targets (Preston and Wilson, 2013; Murale *et al.*, 2017) (Figure 1.1a). Similarly, a small molecule catalyst (Wang *et al.*, 2011) or enzyme (e.g. in BioID (Roux *et al.*, 2012), APEX (Martell *et al.*, 2012; Rhee *et al.*, 2013; Lam *et al.*, 2014)) can facilitate localised reactivity.

In proximity based labelling, a weak chemical probe is used which reacts with endogenous protein nucleophiles, such as cysteines or lysines, when brought in proximity in a bound complex, driven by the increase in local concentration (Wang, 2017; Hoyt *et al.*, 2019; Tamura and Hamachi, 2019) (Figure 1.1b).

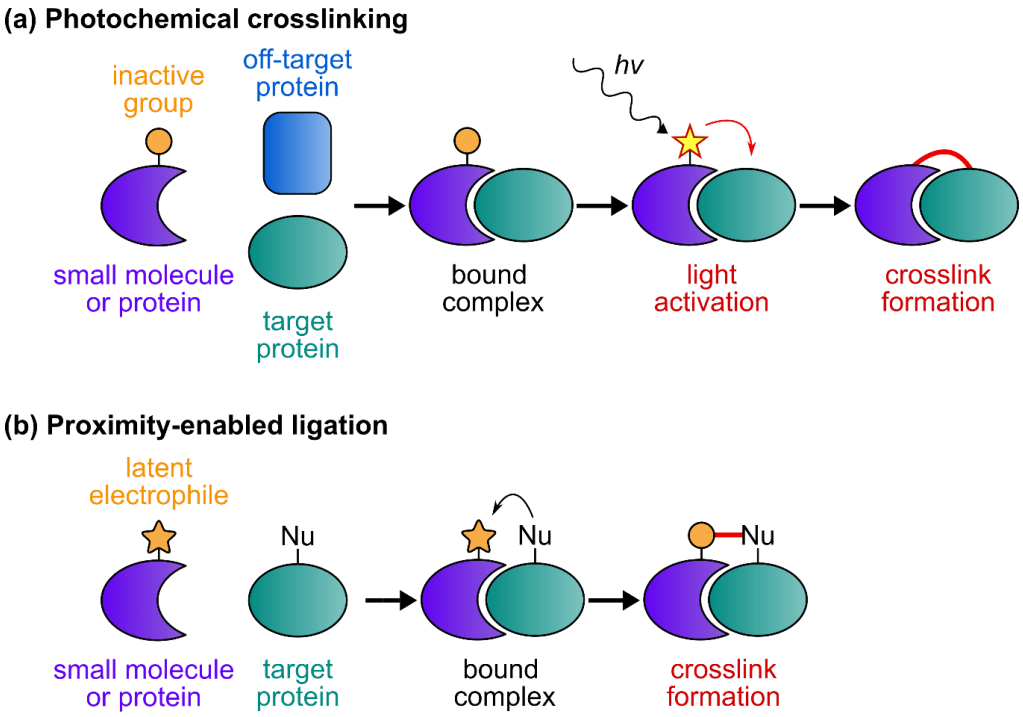


Figure 1.1. Strategies for specific modification of endogenous proteins in a mixed environment. (a) (Photo)chemical crosslinking. A small molecule or binding protein is equipped with a (photo)activatable reactive group. Specific binding of this probe with the target protein differentiates it in a mixed protein environment. Upon activation of the inactive group (e.g. by UV illumination), the bound target protein is preferentially modified. (b) Proximity-enabled ligation. A weak electrophile is introduced to a small molecule or binding protein. Upon target protein binding, nucleophiles (Nu) on the target protein are brought into proximity, enabling nucleophilic attack and covalent conjugation.

For both photochemical labelling and proximity-enabled labelling, a chemical probe can be incorporated into a recombinant protein to enable crosslinking of recombinant proteins to endogenous proteins (Hoyt *et al.*, 2019). Reactive groups can be introduced to recombinant proteins by various methods in addition to classical chemical modification, such as catalytic activation (Isenegger and Davis, 2020) or incorporation of unnatural amino acids (Hoyt *et al.*, 2019). In recent years, unnatural amino acids (UAAs) for coupling of recombinant peptides and proteins to endogenous targets have gained traction (Wang, 2017; Hoyt *et al.*, 2019). A wide array of UAAs with distinct chemistries are available for incorporation into proteins (Liu and Schultz, 2010; Nödling *et al.*, 2019). In the most common approach, UAAs are genetically encoded with a tRNA recognizing a stop codon or quadruplet codon; this tRNA is in turn loaded with the UAA by a corresponding aminoacyl-tRNA synthetase (Liu and Schultz, 2010). Incorporation of an unnatural amino acid into a protein of interest enables a variety of

photoreactive residues and bioorthogonal chemistries for protein conjugation such as “click” chemistry (e.g. Azide-Alkyne chemistry) (Liu and Schultz, 2010). Selected UAAs used in conjugation to endogenous proteins are described in Chapter 1.1 (photochemical UAAs) and Chapter 1.2.1 / Figure 1.5 (UAAs for proximity-enabled ligation).

1.1. (Photo)chemical activation in protein crosslinking

“Photoinduced covalent crosslinking” (PIC) (Preston and Wilson, 2013) or “photo-affinity labelling” (PAL) (Murale *et al.*, 2017) have been used extensively to study biomolecular interactions (Preston and Wilson, 2013; Murale *et al.*, 2017). In photoinduced protein-protein conjugation, reactivity is not driven by proximity *per se*; rather, upon light excitation of a photochemical probe a reactive species is formed which can crosslink to nearby proteins and thereby capture spatiotemporal dynamics about the probe environment (Preston and Wilson, 2013) (Figure 1.1a). The most popular choices for photoinduced crosslinking are aryl azides, diazirines, and benzophenones (Figure 1.2) (Preston and Wilson, 2013). These can have high reactivity, capable of conjugation to backbone residues (Preston and Wilson, 2013) and are therefore not limited by the availability of surface nucleophilic residues in a target protein.

Photochemical probes can be introduced into recombinant proteins using unnatural amino acids: The diazirine probes photo-leucine or photo-methionine can be directly incorporated by the translational machinery (Suchanek *et al.*, 2005). Alternatively, codon suppression can be used to genetically encode various photoreactive unnatural amino acids, such as azido derivatives (*p*-azido-L-phenylalanine, Figure 1.2a) (Chin, Santoro, *et al.*, 2002), diazirine derivatives (Figure 1.2b) (Tippmann *et al.*, 2007) or benzoyl derivatives (*p*-benzoyl-L-phenylalanine, Figure 1.2c) (Chin, Martin, *et al.*, 2002) of phenylalanine. For instance, *p*-benzoyl-L-phenylalanine has been used to capture a specific protein-DNA interaction by incorporation of *p*-benzoyl-L-phenylalanine into a DNA-binding protein (Lee *et al.*, 2009). Photoreactive probes can even be introduced by posttranslational enzymatic modification (Baruah *et al.*, 2008): A reengineered lipoic acid ligase (LplA) from *Escherichia coli* can

introduce aryl azides at a recombinant peptide tag (LAP, 'LplA acceptor peptide') (Baruah *et al.*, 2008).

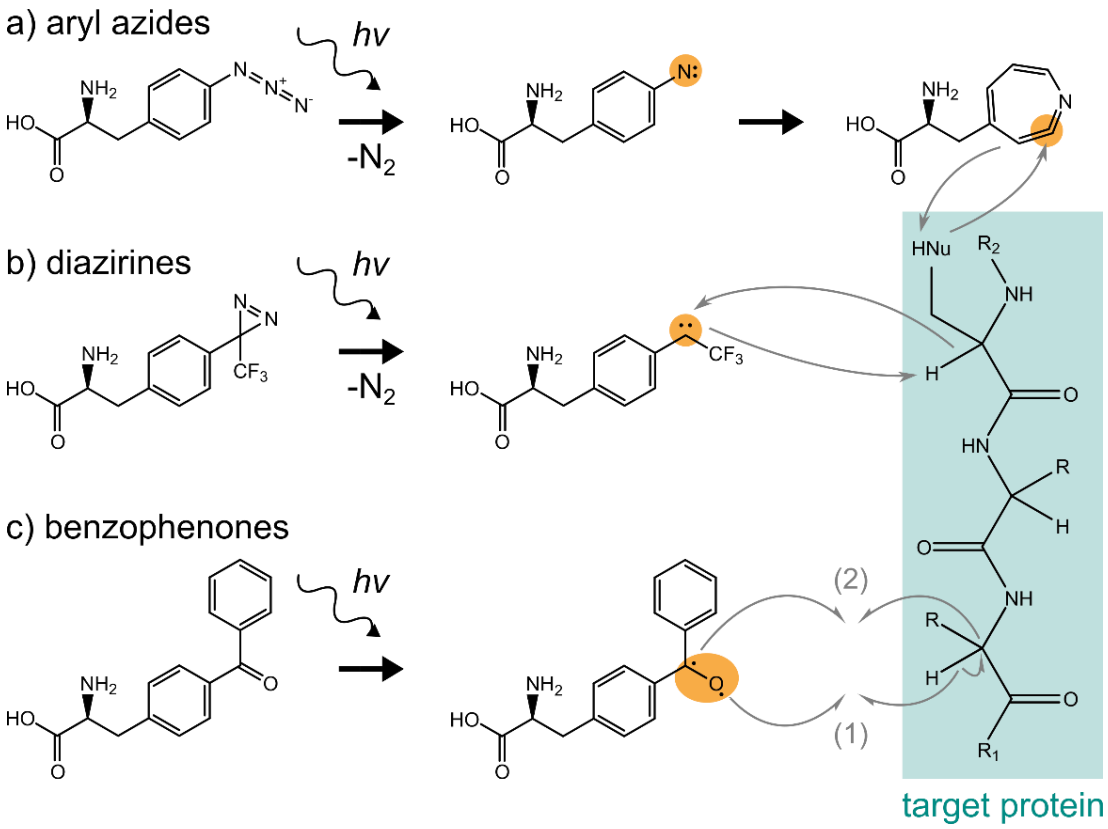


Figure 1.2. Common photochemistries for protein-protein crosslinking have been derived as UAAs. Example of aryl azide (Chin, Santoro, *et al.*, 2002), diazirine (Tippmann *et al.*, 2007) and benzophenone (Chin, Martin, *et al.*, 2002) phenylalanine-derived UAAs, representing commonly used photocrosslinkers (Preston and Wilson, 2013). Mechanism of photoactivation and target protein conjugation based on (Preston and Wilson, 2013) (*hv*: light, Nu: nucleophile).

Photoreactive probes and UAAs can also be modified to include additional functionality (Figure 1.3a), for instance in photoreactive UAAs containing a diazirine fused to a cleavable linker (Lin *et al.*, 2014; Yang *et al.*, 2016; Y. Yang *et al.*, 2017). An affinity-tagged protein incorporating these amino acids can bind to an endogenous target protein and crosslink, allowing for co-purification of the target protein before cleavage, optionally transferring a tag onto the target protein for mass spectrometry (Y. Yang *et al.*, 2017) ('Cleavage after protein photo-cross-linking' (CAPP) (Lin *et al.*, 2014) or '(MS)-label transfer after protein photo-cross-linking' (IMAPP) (Yang *et al.*, 2016)). These UAAs are conceptually related to the small-molecule approach of 'Post-(photo)affinity labeling modification' (P-(P)ALM), wherein a photo-crosslinking label (first generation) or an electrophilic proximity label (epoxide, second

generation) is used (Hayashi and Hamachi, 2012). After crosslinking, the ligand could be substituted for a secondary chemical probe, either after ligand release (in P-PALM, after disulfide reduction), or releasing the ligand directly (in P-ALM, via nucleophilic substitution) (Hayashi and Hamachi, 2012) (Figure 1.3a).

Providing an intermediate in reactivity between traditional photoreactive probes and more stable proximity-enabled probes (see Chapter 1.2.), photocaged quinone methide has been implemented as a precursor to Michael-acceptors for conjugation to a broad range of protein nucleophiles (including Gln, Arg and Asn) (J. Liu, Cai, *et al.*, 2019; J. Liu, Li, *et al.*, 2019). Instead of *in situ* photoactivation, *in situ* chemical activation can also be utilized. For example, by oxidation to the quinone using $NaIO_4$ and subsequent nucleophilic attack, 3,4-dihydroxy-L-phenylalanine (DOPA) was utilized to crosslink synthetic peptides incorporating DOPA to proteins (Burdine *et al.*, 2004). DOPA oxidation has been shown to be compatible for conjugation to the α-amine as well as Lys, His and Cys residues (Liu *et al.*, 2006) and has been implemented in profiling protein-protein interactions (Liu *et al.*, 2007). DOPA can be introduced chemically or as an unnatural amino acid (Alfonta *et al.*, 2003), and genetically encoded DOPA has been used to crosslink Sortase A dimers (Umeda *et al.*, 2009). Another notable modification lies in dehydroalanine. Dehydroalanine can be chemically derived from cysteine, or derived with high specificity by using rare or synthetically incorporated precursors (Dadová *et al.*, 2018). For instance, the UAA phenylselenocysteine can be directly incorporated into proteins, providing a precursor for dehydroalanine upon oxidation (Wang *et al.*, 2007). In itself, dehydroalanine can serve as a Michael acceptor for active site cysteine (Dadová *et al.*, 2018) and has been implicated in protein-protein crosslinking of aging proteins (Z. Wang *et al.*, 2014). More importantly, however, dehydroalanine serves as a platform for the versatile incorporation of a wide range of modifications, e.g. via reaction with thiols, amines, or radicals (Dadová *et al.*, 2018). Dehydroalanine therefore provides a post-translational system for protein derivatization orthogonal to the translational incorporation of UAAs, e.g. to probe the effect of post-translational protein modifications or to explore the effect

of (potentially complex) protein modification by “posttranslational mutagenesis” (Wright *et al.*, 2016; Dadová *et al.*, 2018). Recently, the use of proximity-enabled reactivity in the generation of dehydroalanine has also been described (using the Tyr-derived fluorosulfonate UAA ‘FSY’, also see Chapter 1.2 and Figure 1.5g), with potential implications to *in situ* protein crosslinking (Yang *et al.*, 2019).

1.2. Proximity-enabled protein ligation

The concept of proximity-enabled conjugation or affinity-based labelling relies on the increase in local concentration during formation of a bound complex, such as in protein-protein binding or binding of a small molecule by a protein of interest (Wang, 2017). Local concentrations of reactive group increase as complex formation brings the proteins into proximity to each other – a common theme in nature (Oehler and Müller-Hill, 2010). DNA-templated synthesis is an example application of this strategy for organic synthesis in a complex environment (Li and Liu, 2004). For protein conjugation, this enables the use of a chemical probe that is normally poorly reactive towards proteins in solution, but reacts effectively with a target protein upon binding (Wang, 2017). As amino acid sidechains are generally either inert or nucleophilic, most of these bio-reactive probes are electrophiles.

Affinity-based labelling has a long history for conjugation of small molecules to proteins (Wofsy *et al.*, 1962). However, irreversible anchoring at relevant binding sites, e.g. the active site of an enzyme, usually inactivates the target protein (Hayashi and Hamachi, 2012). In traceless affinity labelling, this approach was modified to release the ligand and free up the binding site (Hayashi and Hamachi, 2012). The initial versions of ‘Post-(photo)affinity labeling modification’ (P-(P)ALM) involve a two-step or three-step process to ligand release and substitution (Hayashi and Hamachi, 2012) (Chapter 1.1, Figure 1.3a). In an improved version using ‘ligand-directed tosyl’ (LDT) chemistry, ligand, electrophile and probe were part of a single reagent, and a protein nucleophile could substitute the ligand from the probe in a single step (Tsukiji *et al.*, 2009; Hayashi and Hamachi, 2012) (Figure 1.3b). LDT has been utilized to label proteins at cysteine, histidine, tyrosine and even glutamate and aspartate (Tsukiji and Hamachi, 2014).

Ultimately limited by the low reactivity of the tosyl probe, newer conjugation chemistries have since been described (Tamura *et al.*, 2018).

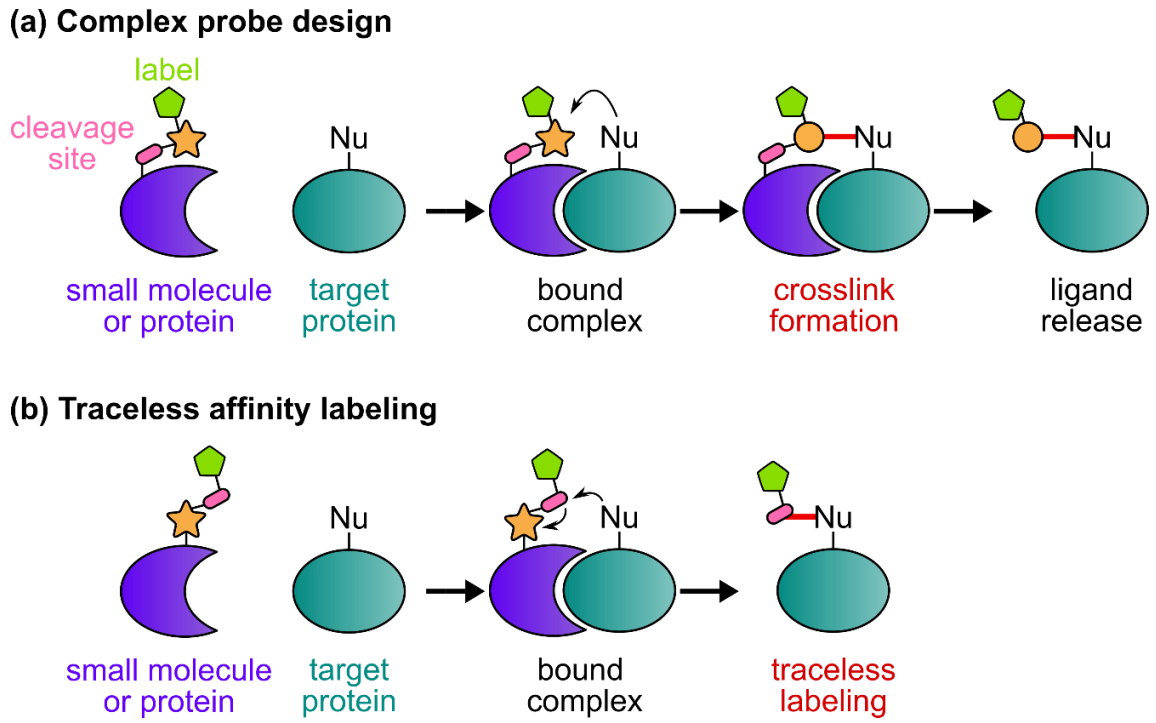


Figure 1.3. Covalent probe designs can incorporate additional features. (a) A cleavage site allows release after conjugation of a small-molecule ligand in Post-(Photo)affinity Labeling Modification [P-(P)ALM] (Hayashi and Hamachi, 2012) or of a protein ligand in 'Cleavage after protein photo-cross-linking' (CAPP) (Lin *et al.*, 2014) and '(MS)-label transfer after protein photo-cross-linking' (IMAPP) (Y. Yang *et al.*, 2017). Inversely, a 'click'-handle has been used to introduce additional features after conjugation (B. Yang *et al.*, 2017). (b) In 3rd generation traceless-affinity labelling (i.e. "ligand-directed tosyl" LDT chemistry, developed from P-ALM), a suitable leaving group enables ligand release upon covalent conjugation (Hayashi and Hamachi, 2012).

While the release of chemical probes from protein actives sites is powerful, an alternative approach can avoid conjugation at the active site altogether. Instead of relying on specific binding sites in a target protein, as in the binding of small molecules, covalent probes have been derived from larger biomolecules engineered to bind arbitrary targets: In a defining paper by Chmura et al., the use of proximity-based reactivity was pioneered with the anti-chelate antibody CHA255, wherein they elegantly demonstrate the impact of unimolecular reactions on local concentration (Chmura *et al.*, 2001; Butlin and Meares, 2006). In this study, electrophilic functional groups were introduced to a benzyl-EDTA(In) chelate recognized by the CHA255 antibody. A cysteine was introduced to the antibody in a favourable position for conjugation to the chelate electrophile (as judged by the crystal structure considering surface

exposure, ligand interface, and distance) (Chmura *et al.*, 2001). They observed efficient conjugation for both acrylamidobenzyl-EDTA(In) and chloroacetamidobenzyl-EDTA(In) to CHA255 with cysteine at position 95 (S95C) (Chmura *et al.*, 2001). Interestingly, they did not observe conjugation for cysteine at position 96 (CHA255 S96C) (Chmura *et al.*, 2001). Since then, various biomolecules have been functionalised with electrophilic probes to crosslink to endogenous proteins.

A first generalised approach for proximity-enabled crosslinking of protein probes to a target protein was described by Holm, Moody and Howarth (Holm *et al.*, 2009). Here, they introduced the Michael acceptor acrylamide as a general electrophile into an affibody binding protein with the aim to crosslink to nucleophilic residues in the cognate Z_{SPA} protein. They choose D36 on the affibody for mutation to cysteine and subsequent chemical conjugation to N,N'-ethylene bisacrylamide (EBA) as the electrophile and N6 on Z_{SPA} for mutation to a nucleophilic residue (Figure 1.4) (Holm *et al.*, 2009).

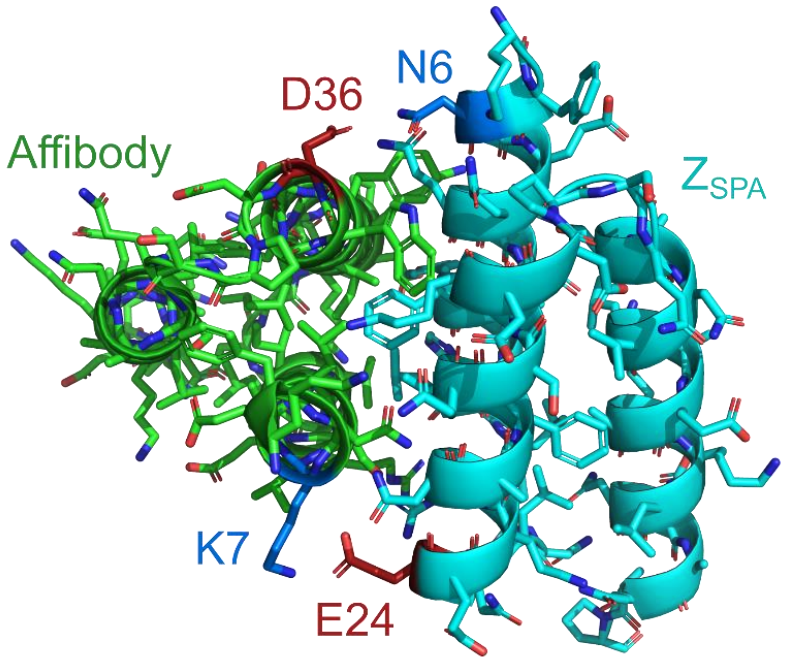


Figure 1.4. The affibody/Z_{SPA} complex used for investigation of electrophilic protein probes for protein-protein conjugation. Red: Residues at which electrophilic reactive groups were introduced. Blue: Nucleophilic residue (K7) or residue at which a nucleophilic residue was introduced (N6C). PDB ID 1lp1, visualized in PyMOL (Högbom *et al.*, 2003).

Using this approach, they reported a yield of around 90% towards cysteine at a 4:1 ratio of affibody-EBA to Z_{SPA} (fused to maltose-binding protein, MBP-Z_{SPA}), as well as crosslinking to

lysine and histidine, although at starkly reduced yield. Furthermore, they observed residual crosslink formation to wild type Z_{SPA}, possibly by linking to more distant residues (Holm *et al.*, 2009). Regarding the usability of this method, it should be mentioned that lysine-derived acrylamide functionality can now be introduced as an unnatural amino acid (Lee *et al.*, 2013).

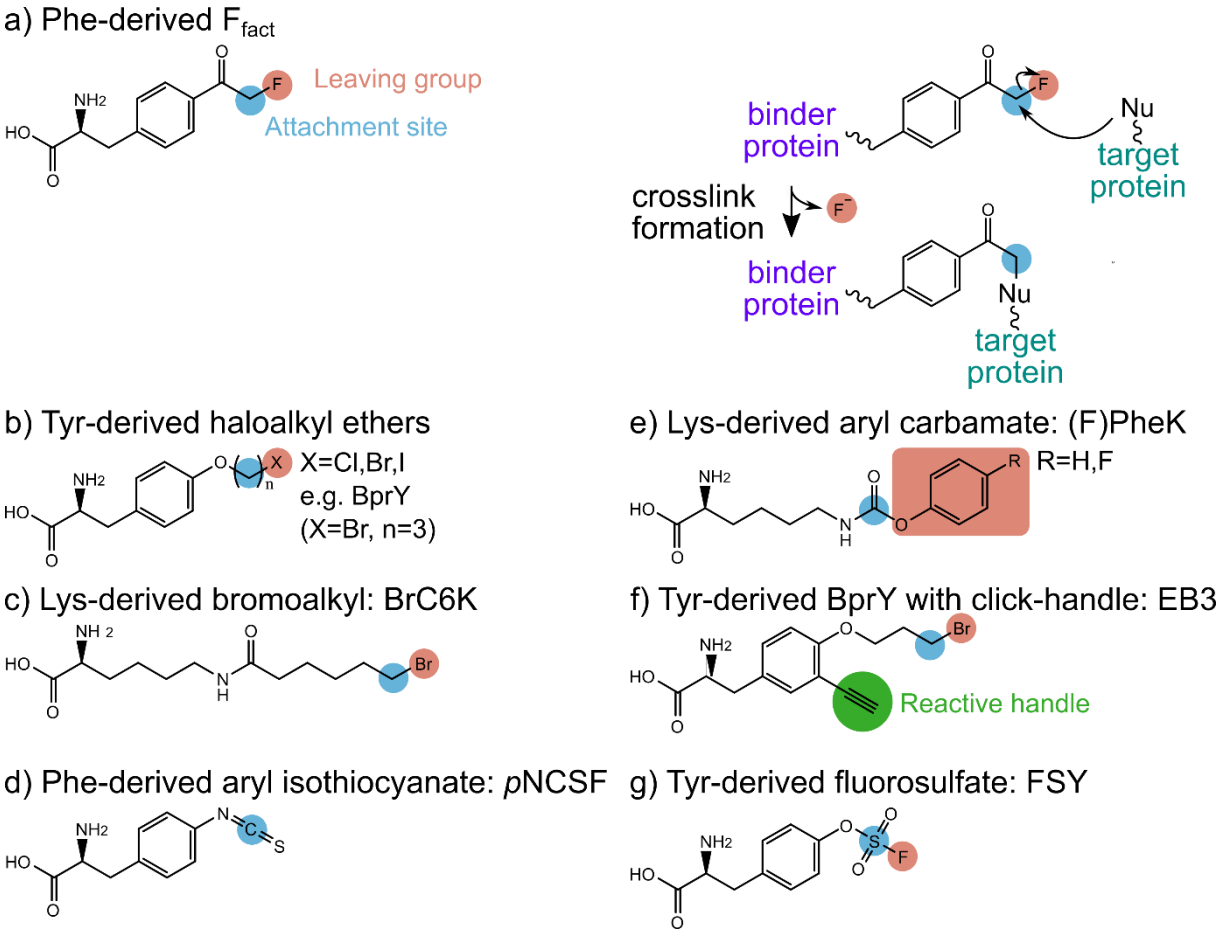


Figure 1.5. Overview of proximity-enabled UAAs studied in the Z_{SPA}/Affibody model system. a) Phe-derived F_{fact} (Xiang et al., 2013) and example conjugation mechanism (Nu: nucleophile), b) Tyr-derived haloalkyl ethers including O-bromopropyl-L-tyrosine (BprY) (Xiang *et al.*, 2014), c) Lys-derived bromoalkyl BrC6K (Chen *et al.*, 2014), d) Phe-derived aryl isothiocyanate *p*NCSF (Xuan *et al.*, 2016), e) Lys-derived aryl carbamate (with R=H PheK, R=F FPheK) (Xuan *et al.*, 2017), f) BprY with additional click-handle EB3 (B. Yang *et al.*, 2017), g) Tyr-derived fluorosulfate FSY (Wang *et al.*, 2018). *Also compare to* L. Wang (Wang, 2017).

The anti-Z_{SPA} affibody / Z_{SPA} model system has been the subject of extensive study of proximity-enabled UAAs by the research groups of Lei Wang and Peter Schultz, at two different target sites (Figure 1.4). An overview of these UAAs is provided in Figure 1.5. An extensive comparison of the underlying work is provided in Chapter 1.2.1, highlighting the

impact of electrophile chemistry and probe design on crosslinking efficiency in proximity-enabled ligation.

1.2.1. Chemical fine-tuning: proximity-enhanced bioreactivity in the Z_{SPA} model

In 2013, the research group of Lei Wang described the use of an electrophilic *p*-2′-fluoroacetylphenylalanine UAA (F_{fact}, Figure 1.5a) in intra- and intermolecular protein crosslinking with endogenous cysteine residues; herein, they also coined the term 'proximity-enhanced bioreactivity', translating affinity-based conjugation to protein-protein crosslinking (Xiang *et al.*, 2013). The electrophilic halogen functionality of F_{fact} is susceptible to nucleophilic substitution by nearby cysteines, generating a crosslink (Xiang *et al.*, 2013). The research group of Lei Wang initially modified the same residue on the affibody for electrophile incorporation as Holm et al., that is D36 (Holm *et al.*, 2009; Xiang *et al.*, 2013) (Figure 1.4). Xiang et al. showed that incorporation of the F_{fact} UAA into the affibody (D36UAA) can crosslink to Z_{SPA}-MBP N6C with about 63% yield (ZSPA-MBP N6C: ZSPA fused to maltose binding protein, with nucleophile at N6, see Figure 1.4; reaction at 4:1 ratio, concentrations as for Holm et al.).

They then derived O-chloro-, -bromo-, and -iodoalkyl ethers from tyrosine and incorporated these into the affibody as above (Figure 1.5b) (Xiang *et al.*, 2014). At a ratio of 2:1 affibody D36UUA to Z_{SPA} N6C, Xiang *et al.* report a crosslinking efficiency of 24%, 42% and 46% for affibodies incorporating O-chloro-, -bromo- and -iodopropyl-L-tyrosine respectively (Figure 1.5b, X=Cl/Br/I, n=3), reflecting halide nucleofugicity ('a result consistent with the order of halide leaving ability in S_N2 reactions') (Xiang *et al.*, 2014). In this work, they also varied alkyl-linker lengths during incorporation. Incorporation efficiency of bromoalkyl UAAs appeared to decrease with increasing linker length from O-bromoethyl-L-tyrosine (Figure 1.5b, X=Br, n=2) to O-bromopentyl-L-tyrosine (n=5), wherein O-bromoethyl-L-tyrosine (n=2) and O-bromopropyl-L-tyrosine (n=3) had similar crosslinking efficiency (42%, 43% respectively) (Xiang *et al.*, 2014).

In another paper, the research group of Lei Wang described a lysine-derived bromoalkyl UAA (BrC6K, Figure 1.5c) using the same affibody system as above (Chen *et al.*, 2014). They observed an increase in crosslinking efficiency with the lysine-derived bromoalkyl BrC6K (73%; Figure 1.5c) compared to O-bromoethyl-L-tyrosine (57%; Figure 1.5b) or O-iodoethyl-L-tyrosine (66%; Figure 1.5b) used previously (Xiang *et al.*, 2014), which they attribute to the higher length and flexibility as well as lower hydrophobicity of BrC6K (Chen *et al.*, 2014). These experiments were conducted at a 4:1 ratio of electrophilic affibody to Z_{SPA} (Chen *et al.*, 2014), more closely matching conjugation conditions used in other publications (Holm *et al.*, 2009; Xiang *et al.*, 2013) and giving increased conjugation yield compared to their initial report of these UAAs at a 2:1 ratio (Xiang *et al.*, 2014). Finally, they show that BrC6K can crosslink to affibodies incorporating lysine and histidine instead of cysteine (Chen *et al.*, 2014).

The research group of Peter Schultz has also utilized the affibody/Z_{SPA} interaction for proof of concept, although inverting the probe/target pair by introducing an electrophile at Z_{SPA} E24 to conjugate to the native K7 residue in the anti-Z_{SPA} affibody (Figure 1.4). *(Schultz* et al. *use a different numbering system than is used in the Z_{SPA} / affibody crystal structure PDB ID 1lp1* (Högbom *et al.*, 2003) *and by Holm* et al. and *Wang* et al. *For comparison, their residue numbering is adjusted herein from Z_{SPA} E25 to E24 and anti-Z_{SPA} affibody K8 to K7.)* Herein, they developed genetically encoded aryl isothiocyanate (as phenylalanine-derived UAA, pNCSF, Figure 1.5d) (Xuan *et al.*, 2016) and aryl carbamate (as lysine-derived UAA, (F)PheK, Figure 1.5e) (Xuan *et al.*, 2017) functionalities as Z_{SPA} E24UAA. For Z_{SPA} incorporating the aryl isothiocyanate UAA, they achieved about 90% crosslinking to lysine (after 3h at a 3:1 ratio of electrophilic Z_{SPA} to affibody). Not only does this efficiency match the one observed by Holm *et al.* (Holm *et al.*, 2009), but, due to the instability of other adducts, isothiocyanates are selective towards amines (Hermanson, 2013; Xuan *et al.*, 2016). The aryl carbamate FPheK (Figure 1.5e) provided increased reactivity and could react with cysteine, lysine and tyrosine with high efficiency (of which crosslinking at lysine was shown with affibody/Z_{SPA}) (Xuan *et al.*, 2017).

This ‘inverted’ affibody/Z_{SPA} system (Figure 1.4) was then also used by Lei Wang *et al.* to demonstrate additional unnatural chemistry. They have incorporated a click-handle into their tyrosine bromoalkyl ether UAA (Figure 1.5b) to make EB3 (Figure 1.5f) (B. Yang *et al.*, 2017). After *in vivo* crosslink and protease digest of Z_{SPA} E24UAA conjugated to affibody K7, the click handle allowed the introduction of biotin thereby the enrichment of crosslinked fragments for Mass Spectrometry (B. Yang *et al.*, 2017). The authors propose this method to aid the identification of protein-protein interactions in cells (compare to Figure 1.3a) (B. Yang *et al.*, 2017). The research group of Lei Wang also demonstrated a sulfur-fluoride exchange probe as a bioreactive UAA (FSY, Figure 1.5g) (Wang *et al.*, 2018). Upon mutagenesis of the target residue K7, they confirmed reaction with affibody K7H and K7Y in addition to wt (K7), but not for other nucleophilic residues (K7A, K7C, K7S, K7T) (Wang *et al.*, 2018). From this probe, they later derived a covalent inhibitor of cell-cell signalling (Li *et al.*, 2020)

1.2.2. Applications of proximity-enabled crosslinking

UAAs for proximity-enabled protein-protein crosslinking have now been applied in diverse contexts, e.g. to stabilize a protein-protein complex for crystallography (Cigler *et al.*, 2017), to stabilize a dimeric enzyme (Li *et al.*, 2018), to discover or further understand protein-protein interactions (B. Yang *et al.*, 2017; Böttke *et al.*, 2020), to conjugate an affibody to a membrane receptor on mammalian cells (Chen *et al.*, 2014; Furman *et al.*, 2014), and for covalent inhibition of cell-cell signalling (Li *et al.*, 2020). In an interesting variation, Hoppmann *et al.* described the photoswitchable bioreactive unnatural amino acids Cl-PSCaa and F-PSCaa which change cis/trans-conformation upon irradiation and can thereby even alter protein structure (Hoppmann *et al.*, 2014, 2015).

Small peptide scaffolds have also been used for proximity-enabled conjugation. Xia *et al.* described various interesting proximity-based crosslinking strategies, which utilize concepts such as coiled-coil peptide-peptide interaction (via a peptide tag) (J. Wang *et al.*, 2014), binding based folding (Liu *et al.*, 2017), and protein-peptide binding (Lu *et al.*, 2014; Yu *et al.*, 2017). The peptide-protein binding probes provide a practical example wherein bioreactive

chemistry is introduced during peptide synthesis (i.e. PDZ-SH3 (Lu *et al.*, 2014) or Grb2-Sos1 (Yu *et al.*, 2017)). Wang et al. also described the use of reactive DNA aptamers to crosslink to proteins of interest, which notably offer the possibility to select binders from large aptamer libraries (Wang *et al.*, 2016). Here, they incorporate modified nucleotides by chemical synthesis, targeting endogenous lysines (Wang *et al.*, 2016).

In summary, efficient chemical probes for affinity-based endogenous protein conjugation are available to selectively crosslink to cysteine, lysine, and even a wider spectrum of nucleophilic residues. Here, drastic differences in reported crosslinking efficiencies highlight the importance of fine-tuned chemistry. These probes can be implemented to enable various biotechnological applications.

1.3. Foreign protein conjugation and covalent adhesion

It has become apparent in recent years that covalent crosslinking to endogenous proteins is a strategy of both the innate immune system and in adhesion of pathogens to host cells, utilizing specialized self-activating domains to provide a reactive handle. Complement proteins covalently mark pathogens for recognition by other components of the innate immune system (Law and Dodds, 1997). For bacteria, adhesion to host cells is often a key requirement for successful colonisation (Stones and Krachler, 2016).

An internal, reactive Cys-Gln thioester implicated in covalent binding has first been reported in complement proteins C3 and C4 (Law and Dodds, 1997). The Cys-Gln thioester is part of a four-residue ring which is occluded in the inactive complement proteins (Law and Dodds, 1997) (Figure 1.6 (Janssen *et al.*, 2005)). Upon cleavage-activation, C3b and C4b undergo a conformational change and the labile Cys-Gln thioester bond becomes sensitive to nucleophilic attack (Law and Dodds, 1997). Through reaction with nucleophiles on a cellular surface, C3b and C4b are involved as part of the innate immune system in marking bacterial cells (Law and Dodds, 1997). C4 has two isotypes, C4A and C4B, wherein the thioester ring in isotype A C4b is directly subject to nucleophilic attack (Dodds *et al.*, 1996; Law and Dodds, 1997). After proteolytic activation, C3b and isotype B C4b feature a histidine nearby the Cys-

Gln thioester which has been proposed to substitute the thiol from the thioester bond to form a reactive acyl-imidazole intermediate (Dodds *et al.*, 1996; Law and Dodds, 1997; Gadjeva *et al.*, 1998). C3, C4a, C4b thereby derive differential reactivities from a Cys-Gln thioester, wherein isotype A C4b preferentially reacts with amines, while C3b and isotype B C4b efficiently react with hydroxyls (Law and Dodds, 1997). In 2005 and 2006, C3 and C3b crystal structures supplemented early experimental studies to increase the understanding of thioester-protection in uncleaved (pre-activated) C3, showing how the thioester is protected from hydrolysis by occlusion of water from the reactive site and prevention of acyl-imidazole formation before C3 cleavage (Figure 1.6) (Janssen *et al.*, 2005, 2006; Wiesmann *et al.*, 2006). After cleavage to C3b, the thioester then becomes accessible to His 1104 and highly solvent-exposed, primed for covalent conjugation (Janssen *et al.*, 2005, 2006; Wiesmann *et al.*, 2006).

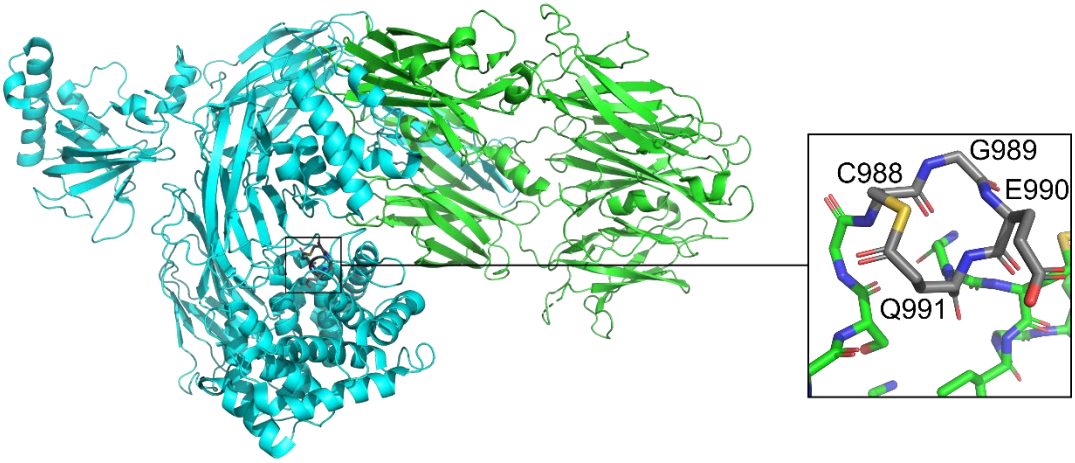


Figure 1.6. The crystal structure of uncleaved complement protein C3. C3 α-chain in teal, C3 β-chain in green, crystal structure from PDB ID 2a73 (Janssen *et al.*, 2005), visualized in PyMOL. Inset: The Cys-Gln thioester in C3 is part of a four-residue ring.

The C3 and C4 complement proteins therefore provide an interesting example of covalent adhesion to a foreign cell surface via a Cys-Gln bond, as well as mechanisms for the regulation of that activity. In the complement system, the host proteins covalently modify pathogen cells (Law and Dodds, 1997). More recently, the inverse has also been found.

Multiple species of pathogenic *Streptococcus* bacteria feature pili involved in host cell adhesion (Telford *et al.*, 2006), including Group A *Streptococcus pyogenes* ('GAS') (Mora *et al.*, 2005; Abbot *et al.*, 2007). After a Cys-Gln thioester bond was found in the C-terminal region of the 'GAS' pilus adhesin protein Spy0125 (at Cys426-Gln575), a role for reactive thioesters in pathogen adhesion was first proposed (Pointon *et al.*, 2010). Walden *et al.* reported that the Cys426-Gln575 thioester was not involved in stabilization of the pilus protein (Walden *et al.*, 2014), and showed that streptococcal SfbI can covalently cross-link to fibrinogen via lysine *in vitro*, blood plasma, and at the cell surface (Walden *et al.*, 2015). Interestingly, Spy0125 serotype homologue Cpa was shown to contain an additional thioester in the N-terminal domain, homologous to the C-terminal region, making the protein bivalent (Linke-Winnebeck *et al.*, 2014). Upon BLAST searches of the N-terminal or C-terminal thioester containing domains of Cpa, both Linke-Winnebeck *et al.* and Walden *et al.* suspected that such 'thioester-containing domains' were a commonly occurring feature (Linke-Winnebeck *et al.*, 2014; Walden *et al.*, 2015).

Another reactive group relevant to pathogen adhesion is the formation of aspartic anhydrides. Asp-Pro is a particularly unstable peptide bond, a fact that has become apparent due to its fragmentation in peptide analysis; in particular, unlike other bonds formed with aspartate, it is liable to cleavage under acidic conditions (Piszkiewicz *et al.*, 1970). Asp-Pro bonds are also less stable than other peptide bonds under ionizing conditions, including Asn-Pro (Mák *et al.*, 1998). Notable self-cleaving activity at an Asp-Pro bond has been reported in a diverse set of proteins, including the human mucins MUC2 (Lidell *et al.*, 2003) and MUC5AC (Lidell and Hansson, 2006), muscarinic toxin MTX2 (Ségalas *et al.*, 1995), the 'metal ion-inducible autocleavage' (MIIA) domain (in NopE1 (Schirrmeister *et al.*, 2011), VIC_001052 (Schirrmeister *et al.*, 2013) and MdCE (Durán *et al.*, 2018)), in SO1698 (DUF1888) (Osipiuk *et al.*, 2012), and within a 'self-processing module' (SPM) in repeat-in-toxin (RTX) proteins FrpC and ApxIVA (Osička *et al.*, 2004).

Asp-Pro cleavage activity is commonly pH-dependent. Mucins MUC2 and MUC5AC have a self-cleavage activity that is accelerated by or dependent on acidic pH (~6) in the secretory pathway; for both molecules, Lidell *et al.* suspected anhydride formation and showed reactivity towards primary amines (Lidell *et al.*, 2003; Lidell and Hansson, 2006). Interestingly, MUC2 and MUC5AC form polymeric gels, therefore Lidell *et al.* suggested crosslinking of mucins via lysine. However, Recktenwald *et al.* later showed that MUC2 is instead polymerized by transglutaminase-catalysed isopeptide bonds (Recktenwald and Hansson, 2016). pH-dependent self-cleavage activity at an Asp-Pro bond was also observed in pre-α-inhibitor heavy chain 3 precursor (Thuveson and Fries, 1999, 2000), 'Repulsive Guidance Molecule B' RGMB (Bell *et al.*, 2013), and SO1698 (Osipiuk *et al.*, 2012). The crystal structure of SO1698 is particularly notable, as it features formation of an intramolecular Lys-Asp bond with Asp-Pro cleavage (PDB ID 3n55, also see Chapter 3.4, Figure 3.8a) (Osipiuk *et al.*, 2012).

Aside from pH-activation, another mechanism for control of autoproteolytic Asp-Pro processing is calcium induction: Although unrelated in sequence, cleavage is calcium-induced in both the 'metal ion-inducible autocleavage' (MIAA) domain (e.g. in NopE1, VIC_001052, MdCE) and the 'self-processing module' (SPM) found in various RTX proteins (e.g. FrpC, ApxIVA), with differing tolerances for activation by non-calcium ions (Osička *et al.*, 2004; Schirrmeister *et al.*, 2011; Matyska Liskova *et al.*, 2016; Durán *et al.*, 2018). Asp to Glu mutations (DP to EP) of NopE1 (Schirrmeister *et al.*, 2011) and FrpC (Osička *et al.*, 2004) were shown to retain cleavage activity, whereas substitution of Asp to Glu abolished cleavage activity in the pH-dependent self-processing of H3 (Thuveson and Fries, 2000), hinting at mechanistic difference. There is now strong evidence for a role of anhydride formation in pathogen adhesion: With the 'self-processing module' found in RTX proteins, mutation of SPM in ApxIVA has been shown to affect infectivity of *Actinobacillus pleuropneumoniae* (Kuban *et al.*, 2020), and FrpC has been shown to facilitate anhydride-mediated crosslinking to host cells in cell culture (Sviridova *et al.*, 2017). The RTX 'self-processing module' is discussed in detail in chapter 1.4.

Related to proteins with autocatalytic cleavage activity at Asp-Pro, some proteins also show self-processing activity at Asn-Pro. An overview of proteins self-processing at Asn is provided by Rawlings *et al.* (Rawlings *et al.*, 2011). In particular, the *E. coli* Type 3 Secretion System protein 'switch' EscU shows an Asn-Pro cleavage mechanism via a succinimide intermediate (Zarivach *et al.*, 2008), as do various of its homologues (e.g. SpaS (Zarivach *et al.*, 2008); also see Chapter 3.4, Figure 3.8b). Interestingly, following spontaneous succinimide formation of Asn/Asp side chains with the protein backbone, both Asn and Asp have now also been implicated in cross-link formation in aging proteins (Friedrich *et al.*, 2018).

Inteins are a particularly prominent example of proteins with self-processing activity at Asn, again usually proceeding through succinimides (Shah and Muir, 2014). However, formation of an anhydride intermediate has been implicated in processing of Asn to Asp variants at both the N-extein (Amitai *et al.*, 2004) and C-extein (Amitai *et al.*, 2004). To understand processing of an Asn to Asp N-extein mutant found in an earlier study (Mills *et al.*, 2006) in more detail, Minteer *et al.* mutated *Pyrococcus abyssi* PolII Intein to abrogate processing activity at both the N-terminal N-extein bond (Cys to Ala) and the C-terminal C-extein bond (Gln to Ala in intein, Cys to Ala in C-extein fragment), preventing thioester formation or thioester transfer as well as bond cleavage (Minteer *et al.*, 2017). Upon mutation of the C-terminal asparagine to aspartate in the N-extein fragment, the intein gained an atypical folding-dependent cleavage activity at the N-extein/intein site at low pH, for which strong evidence of intermediate anhydride formation was provided (Minteer *et al.*, 2017).

In summary, both reactive thioester bonds and protein anhydrides have been identified as reactive sites for protein-protein crosslinking in covalent pathogen adhesion as well as in other contexts. In this work, the use of such activity of the *Neisseria meningitidis* FrpC protein was explored for targeted protein-protein conjugation.

1.4. The *Neisseria meningitidis* self-processing module

1.4.1. FrpC function in *Neisseria meningitidis* biology and pathogenicity

Neisseria meningitidis is a Gram-negative human pathogen that is also known as 'meningococcus' (Virji, 2009). It is commonly ("about 10% of [United Kingdom population]" (Virji, 2009)) "found in the nasopharynx of healthy individuals, but can be the cause of life-threatening meningitis (Virji, 2009). Various adhesion mechanisms are involved in interaction of *N. meningitidis* with host cells, including pili structures (Virji, 2009). In addition to other virulence factors, adhesion mechanisms can differ substantially between *N. meningitidis* populations (Virji, 2009).

'Iron-regulated protein C' (FrpC) is a secretory RTX protein from *N. meningitidis*, which is produced in iron-depleted conditions (Thompson and Sparling, 1993; Thompson *et al.*, 1993) and has been shown to be immunogenic during *N. meningitidis* infection (Osička *et al.*, 2001). Three functional regions of full-length FrpC have been identified: an N-terminal region (Frpc1-414), a 'self-processing module' (SPM, delineated by Frp414/415-657), as well as a C-terminal 'repeat in toxin' (RTX) domain (FrpC863-1829) (Figure 1.7) (Osička *et al.*, 2004; Sadilkova *et al.*, 2008; Sviridova *et al.*, 2017) .

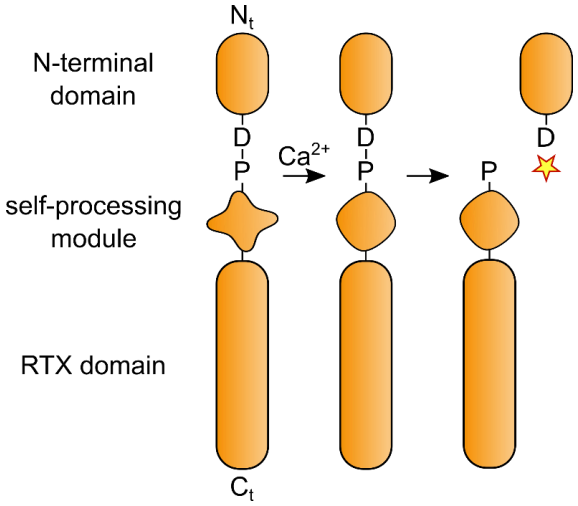


Figure 1.7. Simplified FrpC domain overview. FrpC contains of an N-terminal domain, a calcium-activated self-processing module with self-cleaving activity, and a C-terminal RTX domain involved in protein secretion.

The function of these FrpC regions is best understood in inverse order: First, the C-terminal RTX domain is involved in secretion of FrpC (Linhartová et al., 2010). RTX domains contain repetitive calcium-binding motifs for secretion via the type I secretion pathway (Linhartová *et al.*, 2010). RTX repeats are unfolded in the cytoplasm and fold upon exposure to calcium in the extracellular medium, driving secretion (Bumba *et al.*, 2016).

Second, Osicka *et al.* identified that FrpC has a calcium-dependent self-cleavage activity at Asp414-Pro415, and they observed similar activity for the FrpC-like protein FrpA and another homologous protein, ApxIVA (from *Actinobacillus pleuropneumoniae*) (Osička *et al.*, 2004). After cleavage, Osicka *et al.* further reported *in vitro* formation of stable multimers for FrpC, FrpA and ApxIVA (Osička *et al.*, 2004). For FrpC, they showed that protein-protein crosslinking occurs via Asp-Lys isopeptide formation with the N-terminal FrpC fragment (Osička *et al.*, 2004). This cleavage activity has been implicated in FrpC and ApxIVA host cell interaction (Sviridova *et al.*, 2017; Kuban *et al.*, 2020). A proposed mechanism (Osička *et al.*, 2004; Sadilkova *et al.*, 2008) is shown in Figure 1.8.

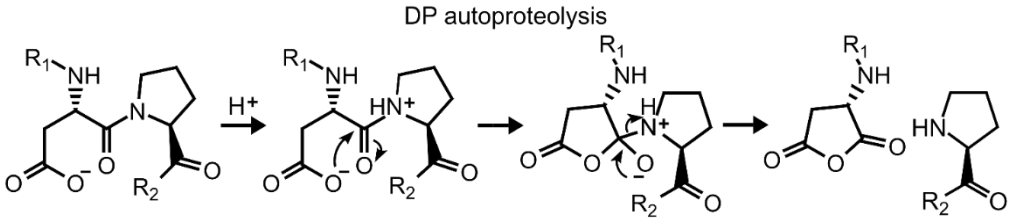


Figure 1.8. Proposed mechanism (Osička et al., 2004; Sadilkova et al., 2008) for FrpC autoproteolysis. Protonated proline acts as the leaving group to facilitate circularisation of aspartate to an aspartyl anhydride, resulting in backbone cleavage.

Finally the third region, FrpC1-414, appears to be the region involved in mediating the interaction of *N. meningitidis* to the host cell. Herein, the membrane protein FrpD binds the N-terminal FrpC fragment (Prochazkova *et al.*, 2005). An FrpC1-414 fragment covalently linked to the host cell surface could thereby provide an anchor for the *N. meningitidis* membrane protein FrpD (Sviridova *et al.*, 2017). FrpC appears to be involved in interaction of *N. meningitidis* with endothelial surfaces (Sviridova *et al.*, 2017). For ApxIVA, pathogenicity could be tested directly and the ApxIVA SPM domain was shown to be crucial to (pig) host

invasion (Kuban *et al.*, 2020). In the following, the biochemical characteristics of SPM known to date are described in detail.

1.4.2. FrpC SPM structure and activity

FrpC self-processing activity can be localized to a short region called the 'self-processing module' (SPM). Osicka *et al.* showed that FrpC retains self-cleaving activity after deletion of the C-terminal RTX region (FrpC∆RTX) and further delineated SPM to FrpC400-657 by deletion mutagenesis (Osička *et al.*, 2004). In 2008, the same research group showed that FrpC414-657 was sufficient for self-processing in recombinant fusion proteins (Sadilkova *et al.*, 2008). The minimal reported active region spans residues 414-591 (Kubáň *et al.*, 2015; Matyska Liskova *et al.*, 2016; Kuban *et al.*, 2020). However, unpublished work by Felix Metzner in the Howarth laboratory indicated reduced cleavage rate of this construct (Metzner, 2017), which was confirmed by Sheryl Lim in the context of this work (Scheu *et al.*, 2021).

The self-cleavage activity of SPM is resistant to protease inhibitors, but efficiently inhibited by EDTA (Osička *et al.*, 2004). Zn^{2+}, Co^{2+}, Mn^{2+}, Ni^{2+}, Mg^{2+} and Cu^{2+} cannot significantly activate FrpC at 2 mM (Osička *et al.*, 2004), but significant cleavage has been reported for Cd^{2+} (Osička *et al.*, 2004) and Tb^{3+} (Matyska Liskova *et al.*, 2016) in addition to Ca^{2+}. FrpC self-processes well under 'human body' conditions: For the FrpC∆RTX construct, a half time of about 5 minutes was determined at 2 mM Ca^{2+}, with a maximum cleavage yield of ~85% after 30 min, an optimal temperature of 35-45 °C and optimal pH between 5.5 and 8.5 (Osička *et al.*, 2004). Similar yields have been reported for other fusion proteins with FrpC414-657 (Sadilkova *et al.*, 2008; Liu *et al.*, 2014).

After cleavage, crosslinking of FrpC∆RTX via lysine is effectively inhibited by cysteine, dithiothreitol (DTT), and 2-mercaptoethanol at 10 mM (Sadilkova *et al.*, 2008). Addition of thiol nucleophiles increased release of fusion proteins upon cleavage from affinity beads, potentially by preventing protein crosslinking (Sadilkova *et al.*, 2008). For DTT, the authors did not detect an adduct during MALDI-TOF mass spectrometry, suggesting efficient hydrolysis

(Sadilkova *et al.*, 2008). However, if the thiol is cysteine a stable adduct is formed, presumably by S,N-acyl shift (Sadilkova *et al.*, 2008).

Bumba *et al.* studied the structural transition of the FrpC415-591 fragment (Kubáň *et al.*, 2015; Matyska Liskova *et al.*, 2016). The calcium-free protein is mostly disordered, containing two α-helical and two β-structural elements (Kubáň *et al.*, 2015). Based on far-UV circular dichroism data, they report two phases of structural transitions upon calcium binding, with initial changes at ~5 μM and a sharp structural transition at ~150 μM Ca^{2+} (Matyska Liskova *et al.*, 2016).

A series of point mutations detrimental to SPM activity have also been determined. Regarding mutations at the DP cleavage site, D414A or P415A result in catalytically dead SPM, whereas D414E is active at a reduced rate (Osička *et al.*, 2004). Oscika *et al.* also tested if mutations at putative calcium binding sites negatively affected cleavage rate. Herein, SPM cleavage activity was reduced or abolished by various mutations (D499K, D510A, ΔD510, D521K, E532A, and D462K ΔA511) (Osička *et al.*, 2004). Bumba *et al.* further provided evidence for a calcium-binding sites at W451 and W519, including the effects of W451F and W519F in NMR analysis (Matyska Liskova *et al.*, 2016).

The NMR structure of SPM (as FrpC414-591) was recently reported (Kuban *et al.*, 2020). As expected (Kubáň *et al.*, 2015; Matyska Liskova *et al.*, 2016), SPM folding is highly calcium-dependent, with four calcium ions in the post-processed state (Figure 1.9) (Kuban *et al.*, 2020). In the post-processed state, Pro415 residue is partially enveloped by the folded SPM (Kuban *et al.*, 2020). Judging from the NMR structure, it appears to confirm the role of most experimentally predicted residues in calcium binding (PDB ID 6sjw (Kuban *et al.*, 2020)). All aforementioned residues appear to be either directly in contact with calcium (D462, D499, D510, W519, D521, E532), interacting with calcium-site residues (W451 interacting with W519) or to be directly adjacent to the binding site (A511) (Figure 1.9). W451 is also at an appropriate distance to form a cation-π interaction with calcium, although it appears more obstructed by D521 than predicted by Liskova *et al.* (Matyska Liskova *et al.*, 2016).

Nevertheless, as predicted by Liskova *et al.* (Matyska Liskova *et al.*, 2016), W451 and W519 are involved in the same binding site as D521, E532, wherein W451 appears to stabilize W519 (PDB ID 6sjw (Kuban *et al.*, 2020), Figure 1). Kuban *et al.* also propose a slightly different mechanism for SPM processing than Osička *et al.* (Figure 1.8) (Osička *et al.*, 2004), in which calcium stabilizes a hydroxide ion (Kuban *et al.*, 2020).

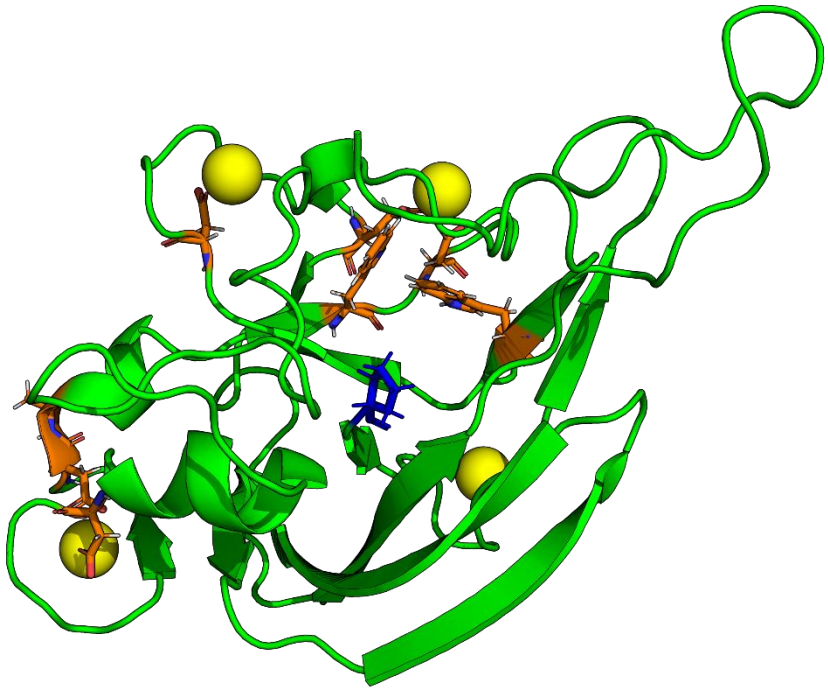


Figure 1.9. NMR structure of FrpC 'self-processing module' (SPM). FrpC415-593 structure determined by Kuban *et al.*, PDB ID 6sjw, visualized in PyMOL (Kuban *et al.*, 2020). Pro415 in blue, in stick format. Yellow: Calcium ions. Orange: Residues studied by mutagenesis in prior studies, in stick format.

1.4.3. Biotechnological applications of FrpC SPM

FrpC SPM has been utilized for biotechnological applications. In principle, two types of activity can be of interest: FrpC SPM as a self-cleaving module, and FrpC SPM as a means of chemical protein activation.

The self-cleaving activity of SPM has been used for protein release after column purification (Sadilkova *et al.*, 2008). For protein purification, anhydride formation is an unwanted side effect which can cause crosslinking and inhibit elution of protein from a column (Sadilkova *et al.*, 2008). The anhydride can be quenched effectively by thiol addition, suppressing protein-protein crosslinking (Sadilkova *et al.*, 2008). By further incorporating elastin-like polypeptides,

an intriguing system for column-free, protease-free protein production and purification has been proposed (Liu *et al.*, 2014). FrpC SPM itself is heat-resistant, which was used for purification of FrpC415-591 (70°C, 15 min) (Kubáň *et al.*, 2015) and could provide a means of purification for heat-stable SPM-conjugated proteins.

In a second set of applications, SPM cleavage is utilized to activate a protein for chemical conjugation. The activated C-terminal species is reactive towards thiols, which can be used to out-compete bioreactive crosslinking at sufficient concentration (e.g. 10 mM DTT or cysteine) (Sadilkova *et al.*, 2008). Based on the presumed formation of an aspartic anhydride the chemical properties of succinic anhydride or, perhaps more appropriately, N-protected aspartic anhydride give an indication for possible reactions. Due to their use in peptide synthesis (Isidro-Llobet *et al.*, 2009), various modifications of N-protected aspartic and glutamic anhydrides have been described, although these are usually not compatible with an aqueous solution. In water at pH 7, such anhydrides usually have a half-life time of a ~4-5 minutes (Bunton *et al.*, 1963). However, the SPM aspartic anhydride is highly reactive towards cysteine, with apparently complete reaction at few mM concentration (Sadilkova *et al.*, 2008).

Finally, SPM is involved in the attachment of pathogens to host cells (Sviridova *et al.*, 2017; Kuban *et al.*, 2020), presumably via covalent attachment of the N-terminal cleavage fragment to host cell membrane protein (Sviridova *et al.*, 2017). It has previously been shown that adequately placed electrophiles can be used for the rationally designed conjugation of protein-protein complexes (Holm *et al.*, 2009) (also see chapter 1.2.). Furthermore, SPM remains active in various recombinant fusion proteins (Sadilkova *et al.*, 2008; Liu *et al.*, 2014). Therefore, it is likely that SPM could be used to facilitate rationally designed protein-protein crosslinking.

1.5. Thesis aims

Various methods for conjugation of recombinant to endogenous proteins have been described, with a particular interest in proximity-enabled protein ligation (Chapter 1.1, Chapter 1.2). However, traditional means of introducing reactivity either involve post-translational modification, or incorporation of reactivity via UAAs. The aim of this thesis is to derive a method for conjugation of recombinant to endogenous proteins without the need for complex modifications during or after protein production, solely relying on canonical amino acids. Conjugation to endogenous proteins has been observed in natural proteins, particularly in FrpC SPM (Chapter 1.3, Chapter 1.4). Akin to its natural function, I aimed to investigate the potential of FrpC SPM-mediated protein activation for protein-protein conjugation.

Herein, SPM would be genetically fused to a protein binding another protein target. If residues appropriate for reaction are available on the target protein, the binding proteins could crosslink upon calcium-induced cleavage (Figure 1.10). This concept was called 'NeissLock' (from an SPM fusion design in preliminary work in the same research group (Metzner, 2017)).

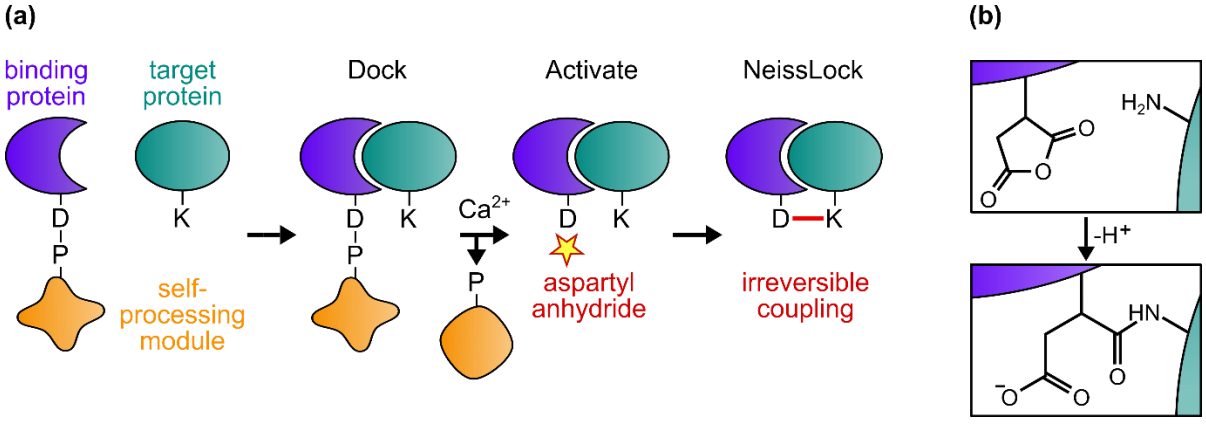


Figure 1.10. Overview of NeissLock conjugation. (a) SPM can be introduced C-terminally of a binding protein to enable calcium-induced protein activation. The binding protein itself specifically binds to a target protein. After SPM-processing, the binding protein features a C-terminal aspartic anhydride which can react with nucleophiles on the target protein to form a stable covalent bond. (b) Illustration of covalent bond formation between target protein (cyan) and activated binding protein (purple).

SPM has already been successfully applied in biotechnology (Sadilkova *et al.*, 2008; Liu *et al.*, 2014). However, the concept of targeted protein conjugation using SPM had not been validated. The first goal was to prove that this is possible in principle. To achieve this, I decided to first identify 'ideal' model complexes based on hypothesised parameters for protein-protein

conjugation (Chapter 3, compare to Figure 1.4 and Chapter 1.2.1) and then test for protein-protein conjugation after SPM-mediated activation (Chapter 4, Chapter 5). Given successful conjugation, the second goal was to identify parameters which could influence conjugation efficiency, as well as map where crosslinking would occur on the target protein. It is known that crosslinking via lysine is possible (Osička *et al.*, 2004); unpublished results by Felix Metzner on head-to-tail crosslinking in BLA further suggest that the primary amine on the N-terminus can be a suitable target for anhydride reaction (Metzner, 2017). I utilized site-directed mutagenesis, mass spectrometry after enzymatic digest, and incorporation of linker residues to identify target residues and parameters relevant to cleavage rate and protein-protein conjugation (Chapter 4, Chapter 5). Finally, I aimed to identify a model complex for NeissLock-mediated conjugation to mammalian cells, to test broader compatibility and provide an initial indication of therapeutic translatability (Chapter 3, Chapter 5). The main findings of this works are published in (Scheu *et al.*, 2021).

2. Materials and Methods

2.1. The NeissDist Database

2.1.1. Data retrieval and biological assembly

Files containing structural information were retrieved from the Protein Data Bank (PDB), either from the PDBe (ftp.ebi.ac.uk/pub/databases/pdb/) or wwPDB (ftp.wwpdb.org) repositories. In initial iterations, the pdb file format (containing the asymmetric unit) or pdb1 file format (containing the first biological assembly) were used. As the pdb format is obsolete (Adams *et al.*, 2019), the final workflow involved retrieval of the mmCIF file format. Herein, I deployed a script provided by wwPDB (https://mmcif.wwpdb.org/docs/sw-examples/python/html/assemblies.html) to generate biological assemblies from mmCIF asymmetric units. Biological assembly files were primarily processed, if no biological assemblies were successfully generated from a given asymmetric unit file, the asymmetric unit file corresponding to that structure was processed instead.

2.1.2. NeissDist database assembly

The NeissDist database was assembled using a script called 'disCrawl' (for 'distance crawler') written in the Python programming language (Python Software Foundation, https://www.python.org/). disCrawl was deployed in two main versions, wherein the second version represented a near complete rewrite. The final version is provided in a public repository (https://github.com/arnescheu/disCrawl) and briefly described hereafter. First, biological assembly files or asymmetric unit files were incorporated into a task list for analysis. This task list was then split and passed to separate processes to enable parallel processing.

Structural information was parsed using the BioPython (Hamelryck and Manderick, 2003; Cock *et al.*, 2009) module. For each chain in a given structure, residues were iterated to identify the first ('N-terminal') and last ('C-terminal') resolved canonical amino acid residues, as well as pre-defined 'target' residues (e.g. Lys and Tyr). If available, distances were calculated between multiple atoms from all 'C-terminal' to all 'target' residues (i.e. atoms X/CA/N to atoms C/CA/N, wherein X is a side-chain nucleophile). For side-chain target

residues (which were not also N-terminal residues), C-terminal atom C to target atom X was the ‘primary’ distance, if available. For N-terminal residues (which were not also side-chain target residues), C-terminal atom C to target atom N was the ‘primary’ distance, if available. For residues which were both N-terminal and a side-chain target residue, the shorter available distance of C-terminal atom C to target atom X and C-terminal atom C to target atom N was chosen. If none of these atom pairs were available, primary distances were assigned in priority order as shown in Table 3.1. This process was repeated for each model in a given structure, to a limit of 10 models. Distances and corresponding metadata were deposited in a ‘distance’ table. To simplify filtering, an additional ‘summary’ table included the highest ranked intermolecular distance per structure. Additional description is provided in Chapter 3.

2.1.3. NeissDist analysis

To provide an overview of distance distributions and estimate the amount of eligible structures, the latest iteration of the NeissDist database was interrogated further. For each structure, the shortest intramolecular distance, intermolecular distance between homomers and intermolecular distance between heteromers was selected from corresponding entries in the ‘distance’ table (as defined from associated metadata). To simplify analysis, only ‘distance’ table entries containing specific distances were considered per analysis (i.e. C-terminal C to Nε [Lys] for Figure 3.4a-b, C-terminal C to OH [Tyr] for Figure 3.4c-d, C-terminal C to N [N-terminal] for Figure 3.4e-f). Other distances (e.g. N to CA) were not considered. Once these ‘shortest’ distances have been identified, they were compared between biological assemblies corresponding to the same structure to identify a single distance per type of intramolecular distance, intermolecular distance between homomers and intermolecular distance between heteromers. The dataset of these distances was then analysed in Python using the SQLalchemy module to determine the number of structures with a shortest distance per category between a given threshold, in 0.1 Å intervals. A histogram was created from this data, as shown in Figure 3.4a,c,e. For each structure, the distances per category were further compared with each other to identify an overall shortest distance. This data was interrogated

in a similar manner, and a histogram was created as shown in Figure 3.4b,d,f. The corresponding SQL statements and python code are deposited in https://github.com/arnescheu/disCrawl. Code and results can also be compared to the similar analysis published in the publication corresponding to this work (Scheu *et al.*, 2021) (https://github.com/arnescheu/NeissDist (Scheu, 2020)). Therein, only primary distances were considered instead of all distances of a specific type (i.e. representing an alternative analysis of N-terminal target residues, as only Lysine residues were considered (Scheu *et al.*, 2021)).

2.1.4. Candidate selection

The NeissDist database allows for the sorting of structures by intermolecular distance between a C-terminal residue of a given chain to a target residue. The final iteration of NeissDist furthermore annotates multiple features of a given distance and structure to allow for simple filtering (e.g. by oligomeric state). An excerpt of NeissDist is provided with the supplementary information of the paper associated with this work (Scheu *et al.*, 2021). Ultimately, NeissDist is a tool to aid the selection and evaluation of potential candidates for protein-protein conjugation via NeissLock. In this work, the Ornithine Decarboxylase (ODC) / Antizyme (OAZ) complex and the Epidermal Growth Factor Receptor (EGFR) / Transforming Growth Factor α (TGFα) complex were chosen as the primary model systems. These model systems were selected from an early version of the NeissDist database based on a combination of structural criteria (e.g. shortest distance between C-termini and target residue, size of protein components), experimental interest (e.g. ease of expression and handling) and/or biological interest (e.g. receptor targeting). During this selection, protein structures were visualized in PyMOL (Schrödinger). A visualization script for PyMOL was deployed to aid candidate selection, the final version is deposited in https://github.com/arnescheu/disCrawl. Further details of candidate selection are provided in Chapters 3 to 5.

2.1.5. Rational design of candidate protein point mutations

In the course of this work, specific point mutants were introduced into ODC/OAZ and EGFR/TGFα model proteins in an attempt to reduce complex stability or otherwise impact

NeissLock-mediated protein-protein conjugation. To select suitable point mutants, protein structures were visualized in PyMOL (Schrödinger). OAZ mutations K153A and V198 were intended to reduce binding affinity to ODC, based on previously published data (Cohavi *et al.*, 2009). Furthermore, OAZ mutation A215R and R188E were identified after visual inspection of the protein structure (PDB 4zgy (Wu *et al.*, 2015)) as potentially disruptive to ODC-binding. Herein, OAZ A215R was identified during inspection of the structure with James Ross, OAZ R188E together with Sheryl Lim. To investigate the impact of removing the primary covalent crosslinking target site on NeissLock-mediated conjugation, ODC K92R was chosen under the presumption that ODC K92 is the primary target residue. TGFα mutation R42A was selected to reduce binding affinity to EGFR, as described previously (Lazar *et al.*, 1989). Finally, FrpC[D414A] mutants of self-processing module (SPM) ('[DA]SPM') were intended to abolish self-processing activity, as was previously described (Osička *et al.*, 2004).

2.2. Plasmid design and cloning

The following constructs were used in the pET28 backbone for expression in *E. coli*: Anti-HER2 Affibody fused to *N. meningitis* self-processing module (SPM) (H_6-Affibody-SPM); SpyTag fused to SPM (H_6-SpyTag-X-SPM, Xaa: 1 of 20 canonical amino acids); truncated human Ornithine Decarboxylase Antizyme residues 95-219 (OAZ); OAZ fused to SPM (H_6-OAZ-SPM, H_6-OAZ-Y-SPM, H_6-OAZ-GSY-SPM [deposited as Addgene plasmid 163613], H_6-OAZ[ΔE219]-SPM, H_6-OAZ[K153A]- SPM, H_6-OAZ[K153,A215R]- SPM, H_6-OAZ[K153,V198A,ΔE219]- SPM); human Ornithine Decarboxylase (ODC) fused to SPM (H_6-ODC-SPM [deposited as Addgene plasmid 163614], H_6-ODC[K92R]-SPM); Transforming Growth Factor alpha (TGFα) fused to SPM (H_6-TGFα-SPM, H_6-ST3-TGFα-SPM, TGFα-GSY-SPM-H_6 [deposited as Addgene plasmid 163615], TGFα[R42A]-GSY-SPM-H_6, TGFα-GSY-[DA]SPM-H_6). Furthermore, for secretory expression from mammalian cells, the soluble ectodomain of Epidermal Growth Factor Receptor (sEGFR501 (Elleman *et al.*, 2001)) was cloned with a tissue plasminogen activator (tPA) secretion sequence (as tPA-sEGFR501-H_6)

in the pENTR4 backbone. Detailed sequence information for all constructs is provided in the appendix (Chapter 8, Supplementary Table 8.1). Construct preparation is described below.

2.2.1. Gene and oligonucleotide synthesis

DNA Oligonucleotides and gene fragments were designed using Benchling (https://www.benchling.com/). H_6-Affibody-SPM in pET28a was cloned by Felix Metzner, in the same research group (Metzner, 2017), and provided the template for SPM amplification. Genes encoding ODC, OAZ and TGFα proteins were synthesized by Integrated DNA Technologies (IDT). DNA encoding EGFR was sourced from 'pcDNA6A-EGFR ECD' by Dr. Mien-Chie Hung (Addgene plasmid 42666 (Hsu and Hung, 2007)). SpyTag-X-SPM and ST3-TGFα-SPM were further derived from SpyTag-MBP in pET28a (Addgene plasmid 35050) and SpyTag003-MBP in pET28a (Addgene plasmid 133450) from our research group. DNA Oligonucleotides as primers for DNA amplification, including for site-directed mutagenesis, were ordered from IDT.

2.2.2. DNA amplification and purification

Prior to plasmid assembly, plasmid backbone fragment DNA and/or gene insert DNA were amplified by polymerase chain reaction using Q5® High-Fidelity DNA Polymerase (NEB) or KOD Hot Start DNA Polymerase (Merck Millipore). DNA was resolved via agarose gel electrophoresis (0.7%-1% agarose in TBE buffer with SYBR Safe DNA Gel Stain [Invitrogen] at 1:10,000 concentration). Samples were cut from the agarose gel and purified using the Wizard® SV Gel and PCR Clean-Up System (Promega). Commonly, PCR fragments intended for ligation by Gibson assembly were combined (as gel slices or after gel solubilization) before loading them on a single DNA purification column. As necessary, eluted DNA fragments were combined with other DNA fragments (e.g. via PCR and gel filtration or purchased) prior to Gibson assembly.

2.2.3. Gibson assembly

ODC, OAZ, TGFα or EGFR were subcloned into suitable expression vectors using Gibson Assembly (Gibson *et al.*, 2009). For Gibson assembly with the plasmid backbone, IDT-

synthesized DNA was either used directly, or DNA was amplified by polymerase chain reaction and purified from agarose gel electrophoresis. DNA fragments were combined during or after purification and combined with an equal volume of 2x Gibson Assembly Master Mix (prepared in our research group, stored at -20°C). Samples were mixed at 50°C or quickly placed onto 50°C and then incubated for at least 0.5 h (commonly for 5 h). Then, 8 µL Gibson assembly reaction products were transformed to chemically competent NEB Turbo *E. Coli*.

2.2.4. Transformation of *Escherichia Coli*

Gibson assembly reaction products or purified plasmid were transformed to chemically competent *E. Coli* (prepared in our research group, stored at -80°C). In general, cell aliquots were thawed on ice, then DNA was added directly to the cell samples. For Gibson assembly reaction products, 8 µL of mixture were transformed to the NEB Turbo *E. Coli* strain; for purified plasmids, 0.5-1 µL plasmid preparations were transformed to NEB Turbo (for plasmid amplification), BL21-(DE3)-RIPL (for protein induction) or Rosetta-gami 2(DE3) (for protein induction) *E. Coli* strains. DNA was mixed into the cells by ejecting the DNA in a continuous motion and/or gentle flicking and cells were incubated on ice for another 10-30 min. Cells were placed to a 42°C water bath and incubated for 45 s, before quickly placing them back on ice. After 2 min incubation on ice, 300 µL of SOC medium (2% [m/v] Tryptone, 0.5% [m/v] yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM Glucose) was added to cells. The cell suspension was transferred to 15 mL falcon tubes, and cells were allowed to recover for ~1 h at 37°C with 200 rpm shaking. Finally, cells were plated onto prewarmed lysogeny broth (LB) plates supplemented with 50 µg/mL kanamycin (34 µg/mL chloramphenicol may be added for BL21-(DE3)-RIPL or Rosetta-gami 2(DE3) strains). Plates were incubated at 37°C at least until individual colonies could be picked (usually overnight). Individual colonies were picked into LB supplemented with 50 µg/mL kanamycin; for BL21-(DE3)-RIPL or Rosetta-gami 2(DE3) strains, the medium was further supplemented with 34 µg/mL chloramphenicol. Liquid cultures were incubated overnight at 37°C with 200 rpm

shaking. Overnight cultures were then used either for plasmid purification or protein amplification.

2.2.5. Plasmid preparation

For small-scale plasmid preparation, plasmids were prepared from 5-10 mL overnight cultures using QIAprep Spin Miniprep Kit (Qiagen) or GeneJET Plasmid Miniprep Kit (Thermo Scientific); for preparation from 10 mL, twice the recommended buffer volumes were used prior to column loading. Gene inserts from newly assembled plasmids were verified by Sanger sequencing (via Source Bioscience or Eurofins). For sEGFR501 in pENTR4 DNA intended for mammalian cell transfection, plasmid was prepared from 200 mL overnight culture volume (in LB supplemented with Kanamycin) using Zymo Research ZymoPURE™ II Plasmid Maxiprep Kit and sterile filtered at 0.22 µm (Corning™ Costar™ Spin-X™ Centrifuge Tube Filters).

2.3. Bacterial protein expression and purification

2.3.1. Standard bacterial protein expression

Protein encoding plasmids were transformed to BL21-(DE3)-RIPL or Rosetta-Gami 2(DE3) strains as described (Chapter 2.2.4). Individual colonies were picked to 11 mL LB supplemented with 34 µg/mL chloramphenicol and 50 µg/mL kanamycin and incubated as overnight pre-culture at 37°C, 200 rpm in 50 mL falcon tubes. 10 mL of pre-culture were added to 1 L LB supplemented with 34 µg/mL chloramphenicol and 50 µg/mL kanamycin in baffled flasks. Cells were grown at 37°C, 200 rpm shaking to OD_{600} of 0.5-0.7 after which protein production was induced by addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to 0.42 mM. In initial production of ODC in BL21-(DE3)-RIPL, OAZ-SPM in BL21-(DE3)-RIPL and TGFα-SPM in BL21-(DE3)-RIPL or Rosetta-Gami 2(DE3) (see Figure 4.3, Figure 5.2), induction temperature was 18°C; in optimized production of ODC variants and OAZ-SPM variants as well as induction of Affibody-SPM, induction temperature was 25°C. Samples were incubated for 16-19 h at 200 rpm, after which cells were separated from the medium by centrifugation at 4°C. Cells were resuspended in Ni-Nta binding buffer (50 mM TRIS, 300 mM

NaCl) and transferred to 50 mL falcon tubes. Cells were harvested by centrifugation and optionally stored at -80°C before further processing.

2.3.2. Initial purification of ODC, OAZ-SPM and TGFα-SPM

For initial purification of ODC, OAZ-SPM or TGFα-SPM, cell pellets were resuspended in Ni-Nta binding buffer (50 mM TRIS-HCl, 300 mM NaCl, pH 7.8) supplemented with 1 mM phenylmethylsulfonyl fluoride (PMSF) and cOmplete mini EDTA-free protease inhibitor cocktail (Roche; diluted at 3 tablets per 4.5 mL as 100x). Cells were disrupted by sonication (on ice, 3-5 times at output 4-5, 50% duty cycle for 60 s, with 1 min break between cycles). The total lysate was cleared from cell debris by centrifugation with a JA25-50 rotor at 4°C, 17,000 rpm for 30 min. Ni-Nta agarose bead slurry was added to the cleared lysate and samples were incubated at 4°C under rolling (Stuart SRT6, 33 rpm). Ni-Nta beads were separated from lysate by centrifugation at 2000 g for 5 min and washed with 15 mL Ni-Nta binding buffer. The beads were separated again, resuspended with 10 mL Ni-Nta wash buffer 1 (as Ni-Nta binding buffer, with 10 mM Imidazole), and applied to a gravity flow column. Beads were washed an additional time with Ni-Nta wash buffer 2 (as Ni-Nta binding buffer, with 30 mM Imidazole). Finally, protein was eluted 5 times with 1 mL Ni-Nta elution buffer (as Ni-Nta binding buffer, with 200 mM Imidazole). Protein aggregates were removed by spinning in microcentrifuge tubes for 30 min at 16,900 rcf. Protein concentrations were estimated by Nanodrop measurement. Protein was dialysed thrice against 50 mM HEPES, 140 mM NaCl, pH 7.50. Aggregates were again removed by centrifugation before protein was concentrated in Vivaspin spin concentrators (30 kDa cutoff for OAZ-SPM or ODC, 5 kDa cutoff for TGFα-SPM). Samples were flash frozen with dry ice / ethanol bath.

2.3.3. Final purification conditions

For final purification conditions of ODC variants (Figure 4.8), OAZ-(GS)Y-SPM (Figure 4.8) and TGFα-GSY-SPM (Figure 5.6) variants, cell pellets were resuspended in 30 mM TRIS-HCl, 200 mM NaCl, 15 mM Imidazole, 5% (v/v) Glycerol, pH 7.5 supplemented with 1 mM PMSF, cOmplete mini EDTA-free protease inhibitor cocktail, 2 U/mL benzonase (Merck) and 1 mg/mL

lysozyme (Merck). For ODC variants, this buffer was also supplemented with 0.02 mM pyridoxal phosphate. For ODC variants and OAZ-SPM variants (such as OAZ-GSY-SPM), 2-Mercaptoethanol (2-ME) was added before or directly after sonication to 5 mM lysis buffer volume.

Cells were disrupted by sonication (on ice, 3 times at output 4-5, 50% duty cycle for 60 s, with 1 min break between cycles). Samples were applied to pre-equilibrated Ni-Nta agarose beads. For TGFα-GSY-SPM and variants, samples were incubated at 4°C with rolling to facilitate higher resin binding. Beads were applied to a gravity flow column and flowthrough collected. Samples were washed with Ni-Nta wash buffer 1 (supplemented with 5 mM 2-Mercaptoethanol for ODC variants and OAZ-(GS)Y-SPM), Ni-Nta wash buffer 2 (supplemented with 5 mM 2-Mercaptoethanol for ODC variants and OAZ-(GS)Y-SPM), and finally eluted with Ni-Nta elution buffer (supplemented with 5 mM 2-Mercaptoethanol for ODC variants and OAZ-(GS)Y-SPM). Eluted samples were concentrated using Vivaspin spin concentrators.

Particles or aggregates were removed by centrifugation at 16,900 rcf before injection to a pre-equilibrated HiLoad 16/600 Superdex 200 pg column (GE Healthcare) via an ÄKTA pure protein purification system (GE Healthcare). For TGFα-GSY-SPM and variants, the gel filtration buffer was 50 mM HEPES, 150 mM NaCl, pH 7.40. For OAZ-(GS)Y-SPM and variants, the gel filtration buffer was 50 mM HEPES, 150 mM NaCl, 2 mM TCEP, pH 7.40. For ODC, the gel filtration buffer 50 mM HEPES, 150 mM NaCl, 2 mM TCEP and 0.02 mM pyridoxal phosphate, pH 7.40. Gel chromatography fractions were assessed on SDS-PAGE. Appropriate peak fractions were combined and concentrated using Vivaspin spin concentrators.

Figure 4.7. and Figure 4.10. represent an intermediary protein purification process, wherein proteins had initially been thiol-protected by addition of 2 mM 2-ME into modified lysis buffer, then dialyzed into 50 mM HEPES, 150 mM NaCl, pH 7.40 or (for ODC) directly into the buffers indicated in Figure 4.7. OAZ samples were dialyzed an additional time after removing aggregates to assess continuous aggregation in (non-)reducing buffers (Figure 4.7). Protein

from the same purification was used in Figure 4.10. Herein, OAZ-Y-SPM (from 50 mM HEPES, 150 mM NaCl, pH 7.40) or ODC (from 30 mM TRIS, 250 mM NaCl, pH 7.40, 2 mM 2-ME, ~0.02 mM PLP) was resolved by ÄKTA pure protein purification (as in 2.3.3.) with 50 mM HEPES, 150 mM NaCl, pH 7.40 before dialysis to 50 mM HEPES, 150 mM NaCl, 1 mM TCEP, pH 7.40 in the final buffer formulation.

2.3.4. Purification of Affibody-SPM

For purification of Affibody-SPM, cell pellet was resuspended and disrupted as described in Chapter 2.3.2. The total lysate was split to 1.5 mL microcentrifuge tubes and cleared by 20 min centrifugation at 16,900 rcf, 4°C. Ni-Nta agarose beads pre-equilibrated in Ni-Nta binding buffer were added to the combined cleared lysate to a final resin volume of ~1 mL. Beads were incubated with cleared lysate for 1 h under rolling at 33 rpm, 4°C. The sample was applied to a gravity-flow column and washed twice with 10 mL Ni-Nta wash buffer 1 before washing once with 5 mL Ni-Nta wash buffer 2. Protein was eluted with Ni-Nta elution buffer. Protein content of eluates was measured by spectrophotometry, suitable protein eluates were combined and dialyzed thrice against 20 mM Tris-Cl, 150 mM NaCl, pH 7.40 to prepare for purification with CaptureSelect C-tagXL Affinity Matrix (Thermo Fisher Scientific). 2 mL pre-equilibrated C-tagXL resin was applied to dialyzed protein and incubated for 1 h at 4°C with rolling. Beads were washed 5 times with 10 mL 20 mM Tris-HCl, 150 mM NaCl, pH 7.40 before elution with 20 mM Tris-HCl, 2 M $MgCl_2$, pH 7.4 in 2 mL fractions. Eluates were assessed by spectrophotometry. Combined fractions were dialyzed thrice against 50 mM HEPES, 150 mM NaCl, pH 7.40. Finally, purified protein was concentrated with a 30 kDa cutoff Vivaspin spin concentrator.

2.3.5. High-throughput preparation of SpyTag-X-SPM variants

Individual colonies of BL21-(DE3)-RIPL cells transformed with plasmids encoding SpyTag-X-SPM protein variants were picked to 60 mL auto-induction medium (AIM) supplemented with 50 µg/mL Kanamycin and 34 µg/mL Chloramphenicol in 200 mL round-bottom flasks. Samples were incubated for 24 h at 30°C, 200 rpm shaking. Then, samples were split to 50 mL falcon

tubes and centrifuged at 4000 rcf for 10 min at 4°C to pellet cells. Cells were resuspended in Ni-Nta binding buffer supplemented with 1 mM PMSF, cOmplete mini EDTA-free protease inhibitor cocktail, 2 U/mL benzonase (Merck) and 1 mg/mL lysozyme (Merck) and the cell suspension was transferred to a microcentrifuge tube. After 30 min head-over-head rotation at ambient temperature, samples were placed to -80°C to initiate freeze-thawing. Samples were subjected to a total of 6 freeze-thaw cycles (from -80°C to 23°C thermocycler block); herein, the process could be paused by storing the samples at -80°C. Samples were spun at 16,900 rcf to separate cell debris from cleared lysate. Cleared lysate was separated to a new microcentrifuge tube and 100 µL pre-equilibrated Ni-Nta agarose bead slurry (~1:1) was added to each sample. Samples were applied to filtration columns in 96-well plate format (AcroPrep). Flowthrough was collected by centrifugation at 300 rcf for 30 s at 4°C. Samples were washed twice with 300 µL Ni-Nta wash buffer 1 and centrifugation at 300 g for 10 s at 4°C, and then twice more at the same conditions with Ni-Nta wash buffer 2. Finally, protein was eluted in 300 µL Ni-Nta elution buffer. Samples were applied to Vivaspin 500 spin concentrators (5 kDa cutoff) and buffer exchanged into 50 mM HEPES, 150 mM NaCl, pH 7.40 by repeated spin concentration and subsequent dilution (at 4°C). Aggregates were removed by centrifugation and samples were flash frozen in an ethanol / dry ice bath.

2.4. Preparation of sEGFR501

The soluble ectodomain of EGFR was previously described as EGFR1-501 (sEGFR501) (Elleman *et al.*, 2001). sEGFR501 was solubly expressed from pENTR4-tPA-sEGFR501-H_6 in Expi293 cells. First, Expi293 cells were seeded at 2.5-3 x 10^6 cells/mL density in 25 mL Gibco Expi293 Expression Medium (ThermoFisher). Cells were transfected with 25 µg sterile filtered pENTR4-tPA-sEGFR501-H_6 plasmid (prepared by maxiprep) using the ExpiFectamine 293 Transfection Kit (ThermoFisher). Herein, 80 µL ExpiFectamine was pre-incubated with 1.4 mL expression medium for 5 minutes. 1.5 mL expression medium containing plasmid DNA was combined with the ExpiFectamine dilution, and the mixture was incubated for another 20 minutes before adding to Expi293 cells. Optionally, the mannosidase-inhibitor Kifunensine

(Sigma-Aldrich) was added to approximately 5 µM final concentration during transfection (30 µL of 5 mM solution).

Expi293 cells were cultured for 20 h at 37°C, 8% CO_2, 125 rpm before adding 150 µL enhancer 1, 150 µL enhancer 2 and Penicillin/Streptomycin (Gibco) to 50 U/mL. Cells were cultured at 37°C, 8% CO_2, 125 rpm and the cell supernatant was collected after 4 days. 5 mL Ni-Nta binding buffer, cOmplete mini EDTA-free protease inhibitor cocktail (1x, as above) and 1 mM PMSF were added to the supernatant. ~0.5 mL pre-equilibrated Ni-Nta beads in 5 mL Ni-Nta binding buffer were added to the supernatant, and samples were incubated at 4°C with rolling. Beads were applied to a gravity flow column and flowthrough collected. Samples were washed two times with 10 mL Ni-Nta wash buffer 1, and then with 2 mL Ni-Nta wash buffer 2. Protein was eluted with Ni-Nta elution buffer. Finally, protein was buffer exchanged and concentrated into 50 mM HEPES, 150 mM NaCl, pH 7.40 using 30 kDa Vivaspin spin concentrators.

2.5. General protein analysis

2.5.1. Estimation of protein concentration

Protein concentrations were estimated from sample absorption measured at A_{280} using a NanoDrop 1000 (Thermo Scientific) or Nanodrop One (Thermo Scientific) spectrophotometer. Extinction coefficients were predicted using the ExPASY Protparam tool ((Gasteiger *et al.*, 2005), https://web.expasy.org/protparam/). The native reduction state (oxidized for TGFα-derived proteins, sEGFR501; reduced for other proteins) was used for concentration calculations.

2.5.2. SDS-PAGE

Samples were prepared for SDS-PAGE by addition of SDS-loading buffer to 1x final concentration (usually as 6x master mix [34% (v/v) glycerol in 0.33 M Tris-HCl pH 6.8, then supplemented with ~170 µM bromophenol blue and 9.4% (w/v) SDS], or as 5x master mix supplemented with EDTA), with or without reduction (typically with 20 mM dithiothreitol). Generally, samples were that already contained other reducing agents were not reduced (e.g.

2 mM TCEP). Samples were denatured at 95°C for 6 min prior to gel loading. Protein samples were resolved on SDS-PAGE, using 10%, 12%, 16% or 18% acrylamide content. After resolution, gels were stained with InstantBlue (Expedeon). For Figure 5.7, homemade Coomassie and EZBlue (Sigma-Adrich) were used instead. After staining, samples were destained with water. Finally, samples were imaged with a ChemiDox XRS imager (Bio-Rad). For quantification, images were analysed in Image Lab (6.0.1., Bio-Rad).

2.5.3. Western Blot

In general, proteins were transferred to methanol pre-activated PVSD membrane in transfer buffer (96 mM glycine, 119 mM Tris, 10% Methanol). For high-molecular weight cell lysates, transfer conditions were prolonged (e.g. 16 h at 4°C). Secondary antibody was conjugated with horseradish-peroxidase and resolved by incubation with SuperSignal™ West Pico PLUS Chemiluminescent Substrate and Chemiluminescence accumulation in a ChemiDoc XRS imager with ImageLab Software. Further details are described in 2.6.3.1 and in 2.8.3.

2.6. SPM activity and in-vitro conjugation assays

2.6.1. Affibody-SPM anhydride reactivity tests

Anti-HER2 Affibody-SPM was used to estimate anhydride reactivity. To test reactivity with different protein nucleophiles, a selection of nucleophiles was used to mimic natural amino acid reactivities, i.e. Glycine dipeptide (as N-terminal amine analogue; Sigma Aldrich), Nα-Acetyl-L-Lysine (as lysine ε-amine analogue; Sigma Aldrich), dithiothreitol (DTT, as cysteine side chain analogue; Sigma Aldrich), L-cysteine (as cysteine side chain analogue; MP biochemicals), (Dimethylaminomethyl)phenol (as tyrosine side chain analogue; Sigma Aldrich [Mixture of components at different substitution level, lot number 041H04021, certificate of analysis COA: 59.7% monosubstituted, 27.3% phenol]). Nucleophiles were first diluted in 50 mM HEPES, 150 mM NaCl, pH 7.40 (HBS) to 60 mM / 6 mM concentration. 20 µM Affibody-SPM in HBS ± 10 mM CaCl2 were incubated at 1 h 37°C, optionally with 10 mM / 1 mM nucleophile as indicated. The reaction was stopped by addition of 5x SDS-loading buffer (SDS-LB) supplemented with 75 mM EDTA to 1x concentration. Samples were boiled and

resolved on SDS-PAGE. To test anhydride lifetime, 7.5 µM Affibody-SPM were incubated for the indicated time at 37°C before inhibiting self-processing and quenching the reaction with addition of 5 µL 100 mM cysteine in 100 mM EDTA (to a final concentration of 20 mM each). After quenching all reactions, 6x SDS-LB was added to a concentration of 1x, samples were boiled and resolved on 18% SDS-PAGE.

2.6.2. SpyTag-X-SPM activity screen

Per protein sample, SpyTag-X-SPM (X: 1 of 20 canonical amino acids) was diluted to ~10 µM in 50 mM HEPES, 150 mM NaCl, pH 7.40 (HBS) and then supplemented with 0.5 volumes of 40 mM Cysteine in HBS. Samples were split to 15 µL in a single well per timepoint (0 min / 5 min / 15 min / 60 min) in PCR-strips in triplicate. 20 SpyTag-X-SPM proteins were processed across 3 days (up to 8 each), with all triplicates processed together. SpyTag-A-SPM was included as a control protein across all days (wherein the first triplicate was designated for analysis prior to experimentation). Samples were preheated to 37°C before adding 5 µL 40 mM $CaCl_2$ in HBS in timed intervals (pipetting up and down multiple times to mix; to a final concentration of ~5 µM per protein in 50 mM HEPES, 150 mM NaCl, pH 7.40, 10 mM Cysteine, 10 mM $CaCl_2$). After the designated amount of time, 5 µL of 5x SDS-loading buffer supplemented with 75 mM EDTA was added to quickly stop the reaction. For the 0 min timepoint, this stop solution was added prior to addition of $CaCl_2$. Samples were boiled at 95°C for 6 min and stored at -20°C. All replicates of all samples were resolved on SDS-PAGE that were processed in immediate succession (with triplicates split across gels to reduce overlap). SDS-PAGE were stained for ~16h with InstantBlue and de-stained with water before imaging. Images were analysed in Image Lab (6.0.1., Bio-Rad), with 2.0 mm disk size baseline subtraction. Per replicate, the amount of unprocessed SpyTag-X-SPM protein at a given timepoint was calculated relative to the unprocessed protein at 0 min.

2.6.3. OAZ-SPM:ODC and variant conjugation

ODC / OAZ-SPM and variants thereof were used as the main model system to study NeissLock protein-protein conjugation *in vitro*. In general, the protein components were diluted

in a HEPES-based buffer (e.g. 50 mM HEPES, 150 mM NaCl, 2 mM TCEP, pH 7.40) and incubated at 37°C, to which calcium diluted in the same buffer was added to 10 mM calcium as indicated, or buffer without calcium was added. Timecourse reactions were stopped by the addition of 75 mM EDTA in 5x SDS-LB. Samples were boiled before resolving on SDS-PAGE.

For Figures 4.3 and 4.19, ODC and OAZ-SPM were not yet purified under optimized conditions (most prevalently protective reducing agent and size exclusion chromatography), and 6x SDS-LB was supplemented with reducing agent. For Figure 4.10, protein samples were retrieved from the initial optimization of purification conditions (see Chapter 2.3.3.), with 50 mM HEPES, 150 mM NaCl, 1 mM TCEP as the final dialysis buffer. For Figure 4.12, 50 mM HEPES, 150 mM NaCl, 2 mM TCEP at differing pH was prepared in small volumes; calcium was diluted in the buffer corresponding to tested pH. Two samples were excluded from OAZ-GSY-SPM due to an experimental error which was documented before resolving on SDS-PAGE (through difference in sample volume). For Figure 4.16, ODC and OAZ-(GS)Y-SPM had been refrozen once, ODC consisted of a slightly different pooled fraction that other ODC proteins (corresponding to the same peak). In Figure 4.16 and 4.17, double banding was resolved on 18 % SDS-PAGE. For Figure 4.21, Bovine Serum Albumin (Sigma-Aldrich) was dissolved in 50 mM HEPES, 150 mM NaCl, 2 mM TCEP and diluted as indicated, wherein the concentration was estimated based on a molecular weight of 66.4 kDa.

2.6.3.1. Western blot of in vitro protein TGFα-GSY-SPM/sEGFR501 conjugation

For western blot of recombinant protein conjugation in Figure 5.9., protein conjugation samples from Figure 5.5c were used. Samples were reduced by addition of 2-mercaptoethanol before resolving on 18% SDS-PAGE. Proteins were transferred to methanol-activated polyvinylidene fluoride (PVDF) membrane in transfer buffer for 3 h at 35 V. The membrane was blocked overnight at 4°C with 5% skim milk in PBS supplemented with 0.05% Tween 20 (PBS-T). Then, the membrane was incubated with a 1:1000 dilution of mouse Anti-EGFR antibody (LA22, 1 mg/mL Merck) in 5% skim milk PBS-T. After 3h incubation at ambient temperature, the membrane was washed multiple times with PBS-T. Supernatant was

replaced for 1:5000 anti-mouse antibody:horseradish peroxidase conjugate (Sigma-Aldrich A4416) in 5% skim milk PBS-T. The membrane was incubated overnight at 4°C. After washing with PBS-T, secondary antibody was resolved by incubation with SuperSignal™ West Pico PLUS Chemiluminescent Substrate and Chemiluminescence accumulation in a ChemiDoc XRS imager with ImageLab Software.

2.6.4. TGFα-SPM:sEGFR and variant conjugation

In general, TGFα-SPM variants were incubated in non-reducing buffer, i.e. 50 mM HEPES, 150 mM NaCl, pH 7.40 at 37°C with sEGFR501 at the indicated protein and calcium concentrations. Samples were only deglycosylated after proceeding with the reaction, i.e. remaining glycosylation of sEGFR501 was still present during conjugation.

After conjugation of TGFα-GSY-SPM with sEGFR501, samples were optionally deglycosylated by treatment with PNGase F (NEB). Herein, 0.1x initial sample volume of 10x Glycoprotein Denaturing Buffer (to ~1x, NEB) was added. Samples were heated to 100°C for 10 minutes to reduce and denature protein samples. Then, 0.2x initial sample volume of 10x GlycoBuffer 2 (NEB, to 1x), 0.2x initial sample volume of 10% NP-40 (NEB, to 1%), 0.4x initial sample volume Milli-Q H_2O were added, as well as 0.1x initial sample volume PNGase F (NEB, to 25 U/µL) or 0.1x Milli-Q H_2O as indicated. Samples were digested at 37°C for at least 1 h. Finally, SDS-loading buffer was added, samples were boiled for 95°C, 6 min and resolved on SDS-PAGE.

In early experiments with sEGFR501 which had not been expressed under Kifunensine treatment, samples that had not been denatured prior to PNGase F treatment showed incomplete digestion (multiple bands on SDS-PAGE, data not shown). Therefore, I recommend sEGFR501 samples to be denatured prior to PNGase F treatment.

2.6.4.1. SpyCatcher-DyLight detection

For fluorescent detection of NeissLock-mediated TGFα conjugation with sEGFR501, I introduced SpyTag003 to TGFα-SPM, i.e. I prepared ST3-TGFα-SPM (as described in 2.3.2, followed by C-tag purification as in 2.3.4), wherein SpyTag003 can be covalently conjugated

with SpyCatcher protein (Keeble *et al.*, 2019). SpyCatcher003 S49C protein labelled with DyLight 680 ("SpyCatcher003:DyLight") was provided by Samuel Stokes, from the same research group. 20 µM ST3-TGFα-SPM was combined with 5 µM sEGFR501 prepared from cells without Kifunensine treatment, or 7 µM sEGFR501 prepared from cells with Kifunensine treatment, or neither, as well as 1 mM $CaCl_2$ as indicated. After overnight incubation at 37°C, samples were labelled with SpyCatcher-DyLight at 8 nM concentration for 1h at 23°C in the dark. Samples were resolved on SDS-PAGE and imaged via an Odyssey Fc Imaging System (LI-COR Biosciences). Then, SDS-PAGE was stained with InstantBlue and imaged as described.

2.7. Mass spectrometry

2.7.1. Intact mass spectrometry

For intact mass spectrometry, samples were processed with an Agilent Q-TOF 6550 system, which featured a liquid chromatography inlet in parallel to an Agilent RapidFire autosampler. Mass spectrometry was carried out by Anthony Tumber at the Department of Chemistry, University of Oxford.

Prior to submission, ODC/OAZ NeissLock conjugation samples were prepared as follows: ODC and OAZ-(GS)Y-SPM conjugation samples were co-incubated at 37°C with 10 mM $CaCl_2$ in 50 mM HEPES, 150 mM NaCl, pH 7.40 as indicated (Chapter 4). After the designated time frame, the reaction was stopped by addition of EDTA to 15 mM.

2.7.1.1. Mass spectrometry following automated solid phase extraction

Samples designated for mass spectrometry by solid phase extraction were diluted with milli-Q water acidified by addition of 0.1 volumes 10% formic acid. Samples were processed from a 384-well plate using an Agilent RapidFire automated system. Samples were applied to a C4 cartridge and washed with 0.1% formic acid before elution onto the mass spectrometry system. Chromatogram data was processed in MassHunter Qualitative Analysis B.07.00 (Agilent). Chromatograms were deconvoluted from 10,000 – 80,000 Da in a 600 – 5,000 m/z range.

2.7.1.2. Liquid chromatography - mass spectrometry (LC-MS)

Samples designated for mass spectrometry with prior liquid chromatography were diluted at a 1:1 ratio with mass spectrometry-grade water. For liquid chromatography coupled with mass spectrometry, samples were injected to an Agilent 1290 Infinity II LCMS system for reverse-phase C18 ultra high performance liquid chromatography and injected into Agilent Q-TOF 6550 (same system as used for automated mass spectrometry in 2.7.1.1). Chromatogram data was processed in MassHunter Qualitative Analysis B.07.00 (Agilent). For Figure 4.15, mass spectrometry data corresponding to peak positions on the liquid chromatogram was deconvoluted from 10,000 – 80,000 Da in a 600 – 5,000 m/z range.

2.7.2. Enzymatic digest analysis

2.7.2.1. Protein conjugation

To identify the site of attachment, NeissLock-conjugation samples were subjected to enzymatic digest with trypsin followed by identification of peptide species by coupled mass spectrometry / mass spectrometry following liquid chromatography (LC-MS/MS).

For OAZ/ODC conjugation (Figure 4.17), 7.5 µM OAZ-Y-SPM was incubated with 7.5 µM ODC or 7.5 µM ODC K92R and 10 mM $CaCl_2$ in 50 mM HEPES, 150 mM NaCl, 2 mM TCEP, pH 7.4 as indicated. Samples were boiled before reaction products were resolved on 18% SDS-PAGE, leaving a lane between samples to facilitate cutting of gel bands. The gel was briefly stained before imaging and separating the reaction products; for conjugation to ODC K92R, two resolved product bands were processed separately.

For TGFα-GSY-SPM conjugation to sEGFR501, 12.5 µM TGFα-GSY-SPM was incubated with 2.5 µL sEGFR501 and 2 mM $CaCl_2$ in 50 mM HEPES, 150 mM NaCl, pH 7.4. Samples were denatured and deglycosylated with PNGase F (see 2.6.4.) before resolving on SDS-PAGE.

2.7.2.2. Sample preparation and enzymatic digest

ODC/OAZ sample preparation was conducted by Sabrina Liberatori from the research group of Shabaz Mohammed at the Department of Biochemistry, University of Oxford. TGFα/EGFR

sample preparation was conducted using the same protocol, assisted by Melissa Webby. First, cut gel bands were destained at 37°C with 50% (v/v) acetonitrile in 50 mM ammonium bicarbonate. The supernatant was replaced with 10 mM TCEP in 100 mM ammonium bicarbonate. After 30 min at 25°C, the supernatant was removed, and gel slices were dehydrated by incubation with 100% acetonitrile. Cysteines were carbamidomethylated by incubation with 50 mM α-chloroacetamide dissolved in 100 mM ammonium bicarbonate, protected from light at 25°C for 30 min. 100 ng trypsin (Promega, lyophilized reconstituted to 200 ng/µL with 50 mM acetic acid) was added per sample after two washes with 100% acetonitrile. Samples were digested overnight. Supernatant was collected and gel slices were rinsed with 10% (v/v) formic acid (stopping enzymatic digest). Gel slices were further rinsed with acetonitrile, of which the supernatant was vacuum evaporated after transfer to a new tube. Peptides were redissolved in 5% formic acid, 5% DMSO and combined with the previous extraction samples.

2.7.2.3 Liquid chromatography - tandem mass spectrometry (LC-MS/MS)

Samples were processed and analysed by Shabaz Mohammed or the research group of Shabaz Mohammed according to the protocol described in (Scheu *et al.*, 2021). Proteomics data were deposited in the Proteomics Identification Database (PRIDE) as project PXD023073 by Shabaz Mohammed (Perez-Riverol *et al.*, 2019; Scheu *et al.*, 2021).

2.8. Cellular assays

2.8.1. Cell culture

A431 cells were cultured at 37°C, 5% CO_2 in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% fetal bovine serum, 100 U/mL penicillin/streptomycin (Gibco) and 1x GlutaMAX (Gibco).

2.8.2. Immunostaining

For immunostaining in Figure 5.10, cells were seeded to glass-bottom petri dishes (MatTek). Seeding density of A431 cells was $2*10^4$ cells/cm^2. After incubation for 24 h, the cell medium was exchanged for Dulbecco's Modified Eagle Medium (DMEM) supplemented with 100 U/mL

penicillin/streptomycin (Gibco) and 1x GlutaMAX (Gibco) and cells were serum starved for an additional 16-18 h. Samples were washed twice with 50 mM HEPES, 150 mM NaCl, 5 mM $MgCl_2$, pH 7.4 (HBS-$MgCl_2$) at 4°C. Supernatant was substituted to protein dilutions (1% BSA in HBS-$MgCl_2$; TGFα-GSY-SPM and 1% BSA in HBS-$MgCl_2$; TGFα-GSY-[DA]SPM and 1% BSA in HBS-$MgCl_2$; or TGFα[R42A]-GSY-SPM and 1% BSA in HBS-$MgCl_2$) and samples were incubated for 1 h. Cells were washed twice with HBS-$MgCl_2$ before incubation with 1 mL 1:100 Anti-His:Phycoerythrin antibody conjugate (BioLegend 362603) in HBS-$MgCl_2$. Samples were incubated for 1 h at 4°C. After three washes, cells were left in HBS-$MgCl_2$ and taken for imaging. A DV core inverted microscope (Micron) was used for brightfield and fluorescent imaging (filter set: 575/25 nm excitation, 625/45 nm emission).

2.8.3. TGFα-GSY-SPM cell conjugation

2.8.3.1. Screen of conjugation conditions

For Figure 5.11, A431 cells were grown in separate 25 cm^2 flasks, with one flask per condition, and serum staved in DMEM before the experiment. The initial seeding density was $1.5*10^4$ cells/cm^2. After 24 h incubation, cells were washed once and the medium was replaced with serum-free DMEM. Cells were starved for 13 h before replacing the medium with HBS-M. Cells were placed to 37°C or 4°C as indicated. For two dishes, starvation medium was instead replaced for HBS-M supplemented with 80 µM Dynasore, for 30 min incubation at 37°C. Then, 1 mL of 1 µM TGFα-GSY-SPM diluted in HBS-M or HBS-M was added to samples as indicated, and samples were incubated for another 5 min at 37°C or 35 min at 4°C. Samples were washed once with HBS-M as indicated before addition of 1 mL 2 mM $CaCl_2$ in HBS-M or HBS-M as indicated and continued incubation for 15 min at 37°C or 40 min at 4°C. Optionally, samples were not washed before adding 4 mM $CaCl_2$ in HBS-M directly to the protein dilution ('direct'); optionally $CaCl_2$ in HBS-M was added immediately after adding the protein dilution ('co-incubation'). For coincubation at 37°C, incubation time was 15 min. For co-incubation at 4°C, incubation time was 35 min. For each sample, the plates were washed with ice cold HBS-M at the end of the treatment before placing them at -80°C.

Samples were treated with 750 µL hot SDS lysis buffer (1% SDS dissolved in 10 mM TRIS-HCl, 1 mM EDTA pH 8.0) per plate. Lysates were aspirated and sonicated briefly (40 s, 50-60% duty cycle, output 4, 1/8" tip). Samples were boiled at 95°C for 10 min and transferred to microcentrifuge tubes. After centrifugation at 16,900 rcf for 10 min, supernatant was boiled with reducing SDS-loading buffer, resolved on SDS-PAGE and transferred for 16 h at 30 V at 4 °C to activated PVDF membranes. Membranes were blocked with 5% skim milk PBS-T at room temperature, before incubation with 1:1000 dilutions of mouse anti-EGFR antibody (LA22, 1 mg/mL Merck) or mouse anti-TGFα (MF9, Novus Biologicals, reconstituted at 0.2 mg/mL) in 5% skim milk PBS-T. Membranes were washed 3-4 times with PBS-T and incubated for 1-2 h with a 1:5000 dilution of secondary antibody (anti-mouse antibody:horseradish peroxidase conjugate (Sigma-Aldrich A4416)) in 5% skim milk PBS-T. After additional washes, chemiluminescence was measured as previously.

2.8.3.2. Conjugation mutant test

For Figure 5.12, A431 cells were seeded at $2*10^4$ cells/cm^2 density in 25 cm^2 flasks. After 24 h incubation at 37°C, 5% CO2, cells were washed twice and covered with serum-free DMEM supplemented with 100 U/mL penicillin/streptomycin (Gibco) and 1x GlutaMAX (Gibco). Cells were starved for 17-18h before conjugation. Samples were washed before adding 1 mL HBS-M or 1 mL protein dilutions at 1 µM in HBS-M (either TGFα-GSY-SPM, TGFα-GSY-[DA]SPM or TGFα[R42A]-GSY-SPM), immediately followed by addition of 1 mL HBS-M or 4 mM CaCl2 in HBS-M as indicated. Cells were incubated for 15 min at 37°C. The supernatant was collected to 40 µL 0.5 M EDTA (stopping the reaction, for Figure 13c). Samples were rinsed twice with 5 mL HBS-M, once with 1 mL HBS-M and finally cells were lysed by addition of 750 µL 95°C hot SDS lysis buffer. Samples were lysed as in 2.8.3.1., except that they were placed on ice after sonication and after centrifugation (which may have negatively impacted membrane recovery). After sample resolution on reducing SDS-PAGE and transfer to PVDF membrane with transfer buffer for 16 h at 30 V, 4°C, membranes were blocked for 1 h with 5% skim milk PBS-T, incubated for 16 h at 4°C with 1:1000 dilutions of

mouse anti-EGFR antibody (LA22, 1 mg/mL Merck), washed thrice with PBS-T, incubated 1 h at room temperature with a 1:5000 dilution of secondary antibody (anti-mouse antibody:horseradish peroxidase conjugate (Sigma-Aldrich A4416)) in 5% skim milk PBS-T, washed thrice with PBS-T and finally imaged as previously.

2.8.3.3. Preliminary conjugation timecourse

For Figure 5.13, A431 grown in a 24-well plate were starved for 44 h in DMEM medium. From up to 5 h prior to cell lysis in 1 h intervals, cells were conjugated for 15 min at 37°C with 1 µM TGFα-GSY-SPM and 2 mM $CaCl_2$ in HBS-M, after which cells were placed back in DMEM medium. Optionally, the protein was added to DMEM instead of calcium. After conjugation of the final samples, all cells were washed and stored at -80°C. Cells were lysed with hot SDS lysis-buffer; due to the small volumes, samples could not be subjected to sonication (increasing viscosity and thereby decreasing accuracy). Nevertheless, samples were boiled with SDS-loading buffer and resolved on SDS-PAGE before transfer to activated PMSF. After blocking, membranes were sampled with mouse anti-TGFα or mouse anti-EGFR followed by mouse anti-HRP. Luminescence was measured upon incubation with HRP substrate and ladder was imaged with trans-UV imaging.

3. NeissDist database: Distance screen of the Protein Data Bank

3.1. NeissLock principle

The aim of this research project was to establish the feasibility of SPM-mediated anhydride generation for general conjugation of recombinant proteins to endogenous proteins, a concept we have termed ‘NeissLock’. To enable specific protein-protein conjugation at low concentrations, we aimed to utilize the increase of local concentration in a protein-protein complex (Chmura *et al.*, 2001; Oehler and Müller-Hill, 2010) for reaction with a functionalised binding protein, a concept which has also been termed proximity-enabled protein crosslinking (Xiang *et al.*, 2014). In the NeissLock design, a binding protein is fused to SPM, wherein SPM enables the generation of an aspartic anhydride through calcium-induced self-processing (Osička *et al.*, 2004). By design, the binding protein interacts with a specific target protein.

This binding brings reactive nucleophiles on the target protein in proximity to the newly generated aspartic anhydride, facilitating their reaction (Figure 1.10).

An anhydride electrophile could potentially react with various nucleophiles, although producing products with varying stability: Anhydrides have long been used for amine conjugation (Lundblad, 2004), e.g. to test the effects of charge inversion at lysine residues on the behaviour of various proteins (Shiao *et al.*, 1972; Ball and Winn, 1982; Batra *et al.*, 1990). Similarly, the reactivity of SPM-activated proteins has been demonstrated for amines and thiols *in vitro* (Osička *et al.*, 2004; Sadilkova *et al.*, 2008). For protein-protein crosslinking after SPM-processing, only amides have been identified as stable adducts; Sadilkova *et al.* have demonstrated that labile thioesters formed after conjugation with DTT hydrolysed before mass spectrometry experiments (Sadilkova *et al.*, 2008). We therefore focussed our design on amine conjugation. In a complex of anhydride-activated binding protein and target protein, nucleophiles on the target protein could then react with the anhydride, driving covalent bond formation, i.e. conjugation of the C-terminal anhydride of the binding protein to lysine amines or N-terminal amines to form an (iso)peptide bond between binding protein and target protein (Figure 1.10).

In general, amines (such as lysine ε-amines, N-terminal α-amines) and thiols (such as in cysteine) are more nucleophilic than hydroxyls (such as serine, threonine, tyrosine) (Bischoff and Schlüter, 2012) or water. However, the molarity of pure water (M[H_2O] in water) is 55.5 M, whereas protein content in any *in vivo* or *in vitro* application are unlikely to exceed µM range in water. Therefore, specific NeissLock conjugation would rely on the change in local concentration brought about by the complex formation of binding protein to target protein (Figure 1.8, Figure 1.10).

In affinity-based labelling, protein binding also imparts specificity for target residues (Tamura and Hamachi, 2019). Likewise, to enable NeissLock-conjugation, I hypothesized that the target protein would need to feature suitable nucleophiles which are in proximity to the C-terminal anhydride generated on the binding protein. I predicted that molecular distances in

the crystal structure of a protein-protein complex would be a promising indicator for NeissLock conjugation, constituting the first criterion for the selection of NeissLock model complexes.

3.2. Considerations for NeissLock candidates

To readily investigate if and under which conditions NeissLock conjugation was possible, I decided to select a model complex in which I would expect optimal conditions for conjugation and *in vitro* study. I decided to use the Protein Data Bank (PDB) (Berman *et al.*, 2003) as a resource by screening the PDB for such a candidate system in a semi-automated fashion. Before I could perform this search, we decided to outline criteria for selection of a hypothesized "ideal" candidate complex.

Prior to establishing selection criteria for NeissLock conjugation, there was little knowledge on the requirements of SPM-mediated protein-protein crosslinking. FrpC has been shown to undergo self-crosslinking in vitro, as has ApxIVA, a protein containing a domain homologous to SPM (Osička *et al.*, 2004). Furthermore, crosslinking of the N-terminal domain of FrpC to the cell surface has been demonstrated (Sviridova *et al.*, 2017; Kuban *et al.*, 2020) and these studies have shown crosslinking at "physiological" pH (pH 7.40). I was particularly concerned about the distance of the C-terminal residue to amines on the target protein, which I considered a minimal requirement for effective conjugation. In addition, prior experiment in our research group have shown that SPM can cause self-reaction of activated proteins (e.g. in β-lactamase, presumably from the C-terminal anhydride to the N-terminal amine or other nucleophiles on the protein surface (Metzner, 2017); also see Chapter 4, Chapter 5). Therefore, I further aimed to find a model binding protein which does not have amine residues nearby its C-terminus, since this could promote self-reaction after SPM-activation, inactivating the binding protein for conjugation. Furthermore, in addition to simple distance, the C-terminal residue of the binder protein would likely need to be accessible for the target nucleophile (e.g. not blocked by bulky residues) and accessible for SPM processing (e.g. to allow protein-protein interaction and SPM processing). These 'structural' features are illustrated in Figure 3.1.

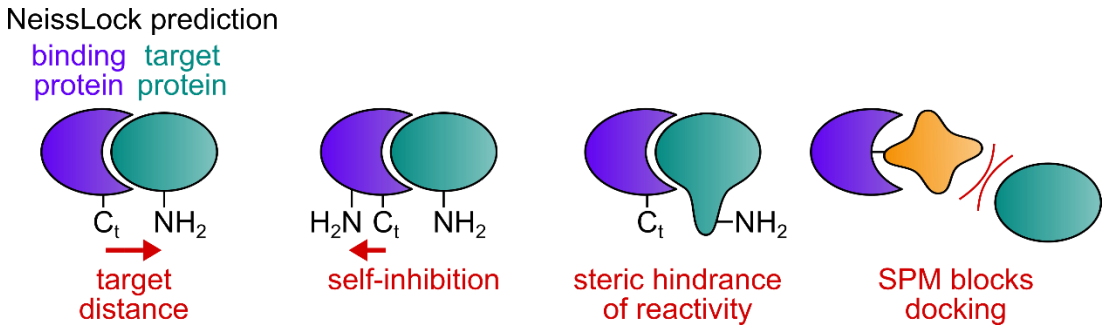


Figure 3.1. Structural considerations for selection of a NeissLock candidate. Various parameters were considered as especially important for selection of a NeissLock candidate. The distance between the C-terminal residue of the binding protein to amines on the target protein was considered the primary requirement for conjugation. Conversely, amines of the binding protein which are close to the C-terminus on the binding protein could be cause of inhibiting self-reaction. Conjugation could also be inhibited by steric factors, e.g. blocking the interaction of the binding protein C-terminal residue with nucleophiles on the target protein.

3.3. disCrawl and the NeissDist database

3.3.1. Overview over the NeissDist pipeline

I wrote an automated tool to calculate the distance from the C-terminal residue of a chain in a given protein structure to protein amines in the same structure, thereby calculating inter-molecular distances (between protein chains) and intra-molecular distances (within the same protein chain). I then iterated this “distance Crawler” (disCrawl) program over a local copy of the Protein Data Bank to make the NeissLock distance database, “NeissDist”. This database could be sorted for structures featuring short distances from the C-terminus to target amines. I also collated certain metadata in NeissDist (e.g. name, multimerization, chain length), which facilitating filtering. After initial selection in NeissDist, I inspected complexes in PyMOL and finally cross-referenced literature information to collate an assessment of candidate structures. The process of candidate selection using NeissDist is illustrated in Figure 3.2.

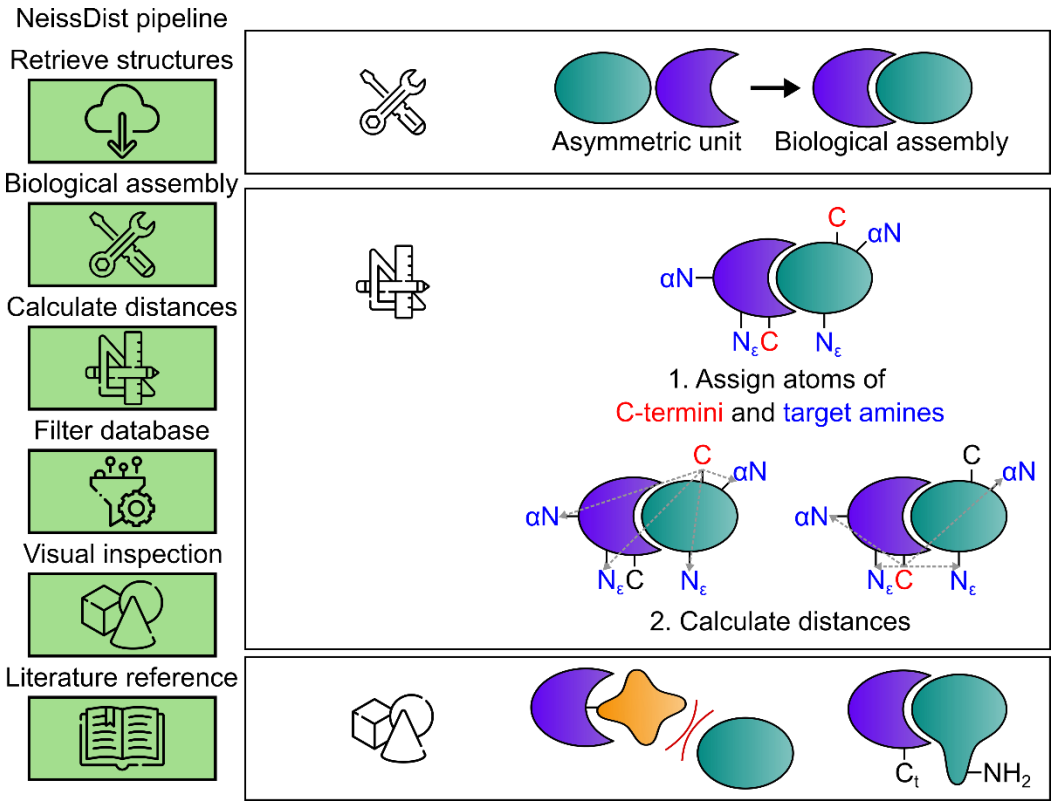


Figure 3.2. The NeissDist pipeline. Overview of NeissLock candidate selection with semi-automated screening of the PDB. Structural data were retrieved from the PDB. Biological assemblies were generated prior to distance assessment. For each protein chain, the distances from the C-terminal residue to target residues in the structure (e.g. N-terminus, lysines) were calculated and deposited with additional metadata in the NeissDist database. NeissDist can be filtered for assessment of protein complexes with specified parameters. Final complex selection is supplemented with rapid manual inspection of the protein structures and literature reference.

3.3.2. Implementation of disCrawl processing

disCrawl was developed in the Python programming language. The tool was written in two main iterations, which are provided online (www.github.com/arnescheu/disCrawl). The first iteration served for the selection of the main complexes used in this study (also see Chapter 4 and Chapter 5). The second iteration of disCrawl was a near-complete rewrite, which allowed for the incorporation of more comprehensive metadata (e.g. annotation of chain identity) and better database structure. Unless otherwise mentioned, the second iteration is described in detail.

Prior to disCrawl analysis, I prepared a local copy of the PDB. The structural information in a protein structure is commonly presented in the asymmetric unit. Whereas the asymmetric unit represents the repeating element in a crystal structure, a biological assembly represents a

tertiary or quaternary protein assembly as it is predicted to occur in a biological context (i.e. “biologically relevant” protein assemblies (Krissinel and Henrick, 2007)). For instance, asymmetric units can contain protein-protein interfaces which are a product of crystallisation, whereas biologically relevant assembly can constitute an assembly of multiple asymmetric units that is not found in the crystal structure (e.g. a virus assembly). To avoid analysis of crystal interfaces which would not occur in solution, biological assemblies were preferentially analysed if available for a given protein structure.

Protein structures were retrieved from the PDB in PDB file format, either as asymmetric units or biological assemblies in PDB format (early versions) or in mmCIF format. Biological assemblies in PDB file format (“.pdb1”, “.pdb2”, …) were initially handled directly. However, the PDB file format is obsolete (Adams *et al.*, 2019) and the deposition of biological assemblies is incomplete (ftp.wwpdb.org/pub/pdb/data/biounit/). The mmCIF format contains the relevant information for the creation of biological assemblies (i.e. orientation and multiplicity of the asymmetric unit elements). For the second iteration, I therefore generated biological assemblies locally (implementing code by wwpdb.org). As this process is resource-intensive, I iterated this script over the local copy of the PDB prior to distance calculation, generating a local database of biological assemblies. These files were then analysed to calculate relevant distances and assemble the NeissDist database.

For a given protein structure, the disCrawl process was designed as follows: the Biopython module was used to interpret structural data. The residues of a given chain were iterated to identify the last resolved standard amino acid residue as the “C-terminal” residue of that chain. Furthermore, the first resolved standard amino acid residue (“N-terminal”) of a chain and specified target residues (e.g. lysines) were identified as “target” residues. Herein, residues were constrained to the 20 standard amino acids, disregarding unnatural amino acids or non-amino acid residues, e.g. nucleotides. This was repeated for all chains in the protein structure (Figure 3.3a,b).

Each combination of C-terminal residues and target residues gives a residue pair for which a "distance" object was created. Both the C-terminal and target residue were assigned up to three atoms each (Figure 3.3c). Multiple distances were calculated for each residue pair to conserve information about their relative orientation (by triangulation) and to provide redundancy (e.g. if atom Nε of a lysine residue was not resolved). For each "distance" object (e.g. representing Chain B C-terminal Glu 219 to Chain A Lys 92), up to nine distances were calculated from and to various atoms in the residues (e.g. Glu 219 atom C to Lys 92 atom Nε, Glu 219 atom C to Lys 92 atom C, …), depending on the resolved atoms and target residue (Figure 3.3e). 'Default' distances were C-terminal atom C to N-terminal atom N or C-terminal atom C to atom Nε on lysine (Figure 3.3d). In addition, any other residue or target atom combination could be specified, e.g. tyrosine residues at hydroxyl oxygen atom.

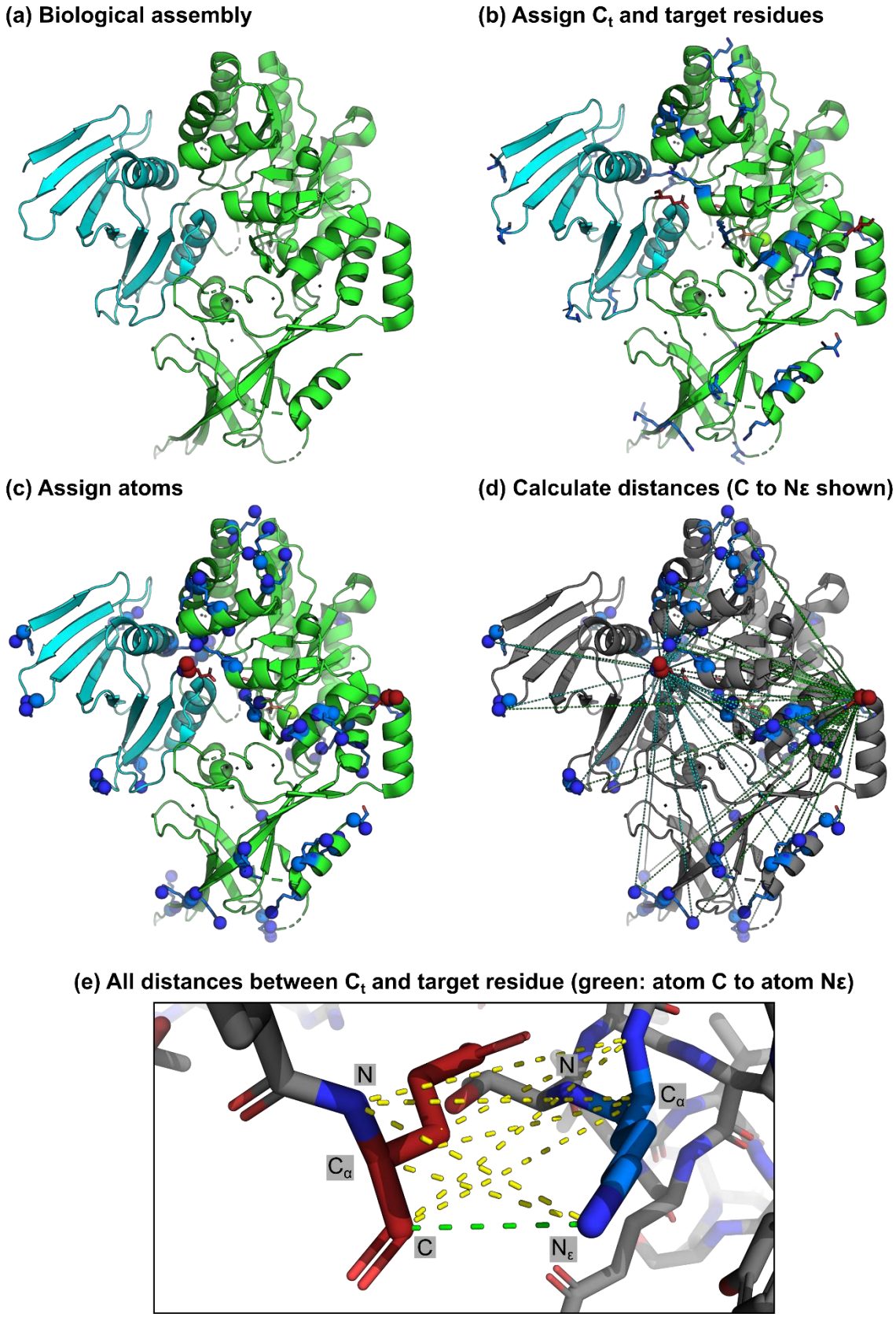


Figure 3.3. Visualization of disCrawl distance calculation. The disCrawl distance calculation algorithm is illustrated with the Ornithine Decarboxylase (ODC) / Antizyme (OAZ) complex (PDB ID 4zgy) visualized in PyMOL (a) The ODC/OAZ crystal structure contains two protein chains, with OAZ in cyan and ODC in green; water molecules as dots. (b) For each chain, the C-terminal residues are

assigned (red). Furthermore, target residues (here: N-termini and lysines) are assigned (blue). (c) For each pair of C-terminal residue and target residue, the distance between multiple atoms is calculated. These atoms are illustrated with spheres. (d) Overview of pairwise distance calculations (cyan dotted line: distances from OAZ C-terminus; green dotted line: distances from ODC C-terminus). For simplicity, only distances from atom C to atom Nε are shown. Illustration of up to nine distances which are calculated between a residue pair, from OAZ C-terminal E219 to ODC K92: C-terminal atoms C/Cα/N to target residue side chain atom "X" (assignable, here Nε for Lys) as well as backbone atoms Cα, N.

Combined with other relevant information (e.g. structure, atoms, chain, residues), these distances form a single distance entry in the NeissDist 'distance' table. An illustration is provided in Table 3.1, summarizing the most important parameters.

Table 3.1. Illustration of NeistDist 'distance' table. For each structure, multiple distances could be identified, which were deposited in full. Here, the top ten entries for PDB ID 4zgy (biological assembly 1) are summarized, showing the C-terminal and target residues as well as up to nine distances associated with each residue pair (distances in Å). Atom X: specified 'target atom', here Nε for lysine residues. In bold are the 'primary' distances for these entries (redundancy for unresolved atoms; order of priority as shown from left to right, unless targeting N-terminus).

PDB	C-terminus	Target	X-C	X-CA	X-N	CA-C	CA-CA	CA-N	N-C	N-CA	N-N
4zgy	Chain A LYS 92 atom NZ	Chain LYS 92 NZ atom B	**3.537**	4.343	4.819	7.564	7.146	7.879	7.698	7.063	7.555
4zgy	Chain A LYS 78 atom NZ	Chain LYS 78 NZ atom A	**11.185**	10.323	10.640	11.995	10.696	10.927	13.335	11.997	12.140
4zgy	Chain A LYS 121 atom NZ	Chain LYS 121 NZ atom B	**14.174**	15.179	15.297	11.019	11.713	11.930	10.771	11.305	11.395
4zgy	Chain A LYS 141 atom NZ	Chain LYS 141 NZ atom B	**14.328**	14.126	13.129	16.550	16.809	16.185	17.892	18.103	17.432
4zgy	Chain A LYS 115 atom NZ	Chain LYS 115 NZ atom B	**16.278**	15.742	15.231	11.941	11.463	10.682	13.094	12.526	11.664
4zgy	Chain A LYS 74 atom NZ	Chain LYS 74 NZ atom A	**17.253**	16.464	17.241	17.615	16.427	16.708	18.784	17.569	17.772
4zgy	Chain B LYS 153 atom NZ	Chain LYS 153 NZ atom B	**19.031**	18.302	17.263	21.210	20.428	19.172	22.416	21.636	20.351
4zgy	Chain A LYS 69 atom NZ	Chain LYS 69 NZ atom B	**19.571**	18.283	17.862	16.721	15.419	15.442	17.067	15.839	15.947
4zgy	Chain A LYS 57 atom NZ	Chain LYS 57 NZ atom A	**21.313**	20.633	19.182	20.837	19.891	18.540	19.677	18.680	17.349
4zgy	Chain A LYS 74 atom NZ	Chain LYS 74 NZ atom B	**21.552**	21.076	22.237	20.596	19.742	20.598	20.374	19.436	20.223

In the second iteration of disCrawl, a 'top' distance was selected for each structure. To preferentially identify "binding protein" / "target protein" complexes, this 'top' distance referred to the shortest intermolecular distance between heteromers if such a distance was available, otherwise the 'top' distance referred to the shortest distance in the structure. Combined with information about the structure (e.g. multimerization state), this 'top' distance formed one entry per biological assembly in the NeissDist database 'summary' table (Table 3.2).

Table 3.2. Illustration of NeissDist database 'summary' table. In addition to all distances deposited in the NeissDist 'distance' table, the 'summary' table provides an overview of relevant information for each analysed protein structure. Herein, a single distance is selected to represent each protein structure. A selection of interesting protein complexes is provided (further illustration of 1mox and 4zgy in Figure 3.6 and 5yqz and 1g0y in Figure 3.7). Chain names and amino acid position numbers as annotated in structure, with resolved residue count as identified by disCrawl.

PDB ID	Complex	Organism	Resolution (Å)	C-terminal residue	Target residue	Primary distance (Å)
1mox	Epidermal Growth Factor Receptor / Transforming Growth Factor alpha	*Homo sapiens*	2.5	Chain D (48 aa) A50	Chain B (501 aa) K465	3.3
4zgy	Ornithine Decarboxylase / Ornithine Decarboxylase Antizyme	*Homo sapiens*	2.6	Chain B (125 aa) E219	Chain A (383 aa) K92	3.5
1ory	Flagellar protein FliS / Flagellin	*Aquifex aeolicus*	2.4	Chain B (40 aa) R2518	Chain A (119 aa) K1028	3.8
2qac	Myosin A tail domain interacting protein MTIP / Myosin-A	*Plasmodium falciparum*	1.7	Chain A (144 aa) Q204	Chain T (14 aa) K813	3.9
1dml	DNA polymerase processivity factor / DNA polymerase	*Human herpesvirus 1*	2.7	Chain B (36 aa) A1235	Chain A (267 aa) K289	3.9
5yqz	Glucagon receptor / Endolysin / Glucagon analogue	*Homo sapiens, Enterobacteria phage T4*	3.0	Chain P (28 aa) T29	Chain R (558 aa) K64	4.1
1syx	Spliceosomal U5 snRNP-specific 15 kDa protein / CD2 antigen cytoplasmic tail-binding protein 2	*Homo sapiens*	2.3	Chain B (62 aa) T86	Chain A (135 aa) K125	4.3
1g0y	Interleukin-I receptor Type I / Antagonist peptide AF10847	*Homo sapiens*	3.0	Chain I (21 aa) L21	Chain R (310 aa) K95	5.5

Herein, information about the nature of a given distance (e.g. 'homomeric intermolecular' or 'heteromeric intermolecular') was derived from structural data supplemented with the dictionary information deposited with a given structure file. For instance, annotation of intramolecular or intermolecular distance was achieved by comparison of chains involved (e.g. "C-terminus of Chain A" to "a lysine residue in Chain A" is an intramolecular distance). Annotation of homomers was achieved by comparison of chains involved in a distance pair to their polymer identifier (i.e. mmCIF dictionary entry "_entity_poly.pdbx_strand_id", for example _entity_poly.pdbx_strand_id might define that both chain A and chain B consist of the same polymer, therefore the distance from C-terminus of Chain A to lysine of Chain B would be classed as an being intermolecular distance between homomers, or 'homomeric intermolecular').

3.3.3. Overview of NeissDist

The latest iteration of NeissDist was generated on 24th July 2019. 154,025 entries retrieved from wwPDB were decompressed and processed to generate 226,359 biological assemblies (with multiple biological assemblies possible per structure). For 6,926 files, biological assembly was unsuccessful, which were processed further as asymmetric units. From these 233,285 total files, 233,285 entries (100%) were deposited in the NeissDist summary table, with 224,308 (96.2%) flagged as successfully processed by disCrawl and 8,977 (3.85%) flagged as rejected. Of 8,977 rejected files, 3,422 (38.1%) were rejected due to filesize >10 MB (e.g. large biological assembly, such as a virus-like particle, or due to deposition of structure factors) to conserve computational resources, and the remaining 5,555 files were rejected due to issues in structure parsing (of which 5,280 were solution NMR structures). Target residues were defined as N-terminal residues, lysine residues (target atom "X" = Nε) and tyrosine residues (target atom "X" = OH). Structures with multiple models were analysed until the 10th model. For 3,710 of 224,308 (1.65%) files flagged as successfully processed, no distances were deposited – most corresponding to DNA, RNA or other non-protein structures with no standard amino acid residues. For 13 of these structures, C-terminal and target residues were assigned, but disCrawl failed to calculate distances between any of the defined atom pairs since neither of atom C, atom CA or atom N were resolved for the residue defined as C-terminal (PDB ID – biological assembly: 3h6d-1 & 3h6d-2, 1ivn-1 & 1ivn-2, 1j00-1 & 1j00-2, 3ad5-1, 2ynu-2, 1u8u-1 & 1u8u-2, 5lmc-1 & 5lmc-2, 5lhw-1). Corresponding to the remaining 220,598 files, representing 98.35% of files flagged as successfully processed by disCrawl or 96.15% of the total 233,285 files, 54,905,082 entries were deposited in the NeissDist distance table. These 220,598 files correspond to 146,636 entries in the PDB (current holdings: http://www.rcsb.org/pdb/statistics/holdings.do).

My primary interest was to identify structures with lysine residues on one protein in proximity to the C-terminus of a protein with which it binds (Figure 3.1). To gain insight into the distribution of structures meeting this requirement, I interrogated the NeissDist database

distance table for the number of structures featuring specific kinds of distances below a specified value, e.g. intermolecular distance between heteromers until 10 Å. These data were reduced to only include one atom to atom distance for each type of residue pair, i.e. C-terminal atom C to lysine Nε (Figure 3.4a), C-terminal atom C to tyrosine atom OH (Figure 3.4c), or C-terminal atom C to N-terminal atom N (Figure 3.4e). This reduction was intended to provide a more consistent data analysis, e.g. to exclude structures with poorly resolved residues and because the distance from atom C to lysine Nε can differ substantially to the distance from atom C to lysine atom CA (Table 3.1, Figure 3.3).

To visualize the distribution of structures in which an intermolecular distance between heteromers was the shortest, i.e. to avoid potential inhibition of the C-terminal anhydride through side-reaction (Figure 3.1), all distances of different categories were compared to each other within a given structure (Figure 3.4b,d,f). For lysine target residues, thousands of structures met both the requirement of an intermolecular distance between heteromers as the shortest distance from C-terminal atom C to target atom Nε, and that this distance was < 10 Å (< 10 Å between heteromers 10,101, of which this distance was representing the overall shortest distance in 6,584). Tyrosine residues (atom OH) showed a slightly broader distance distribution than lysine residues (atom Nε) (Figure 3.4c,b) while accounting for a similar amount of total structures (any – C to Nε: 140,954, C to OH: 139,397; intramolecular – C to Nε: 139,795, C to OH: 138,434), indicating that the nearest tyrosine residue was on average further away from the C-terminus than the nearest lysine residue. This is likely a result of reduced surface distribution of tyrosine residues compared to lysine residues (Tjong *et al.*, 2007). A similar but stronger effect was observed for the distribution of C-terminal to N-terminal residues (atom C to atom N, Figure 3.4e,f). While more structures were accounted for in total (any category – 142,434, intramolecular – 141,492), only 2,221 structures featured such a distance between heteromers < 10 Å (21.9% of atom C to lysine atom Nε).

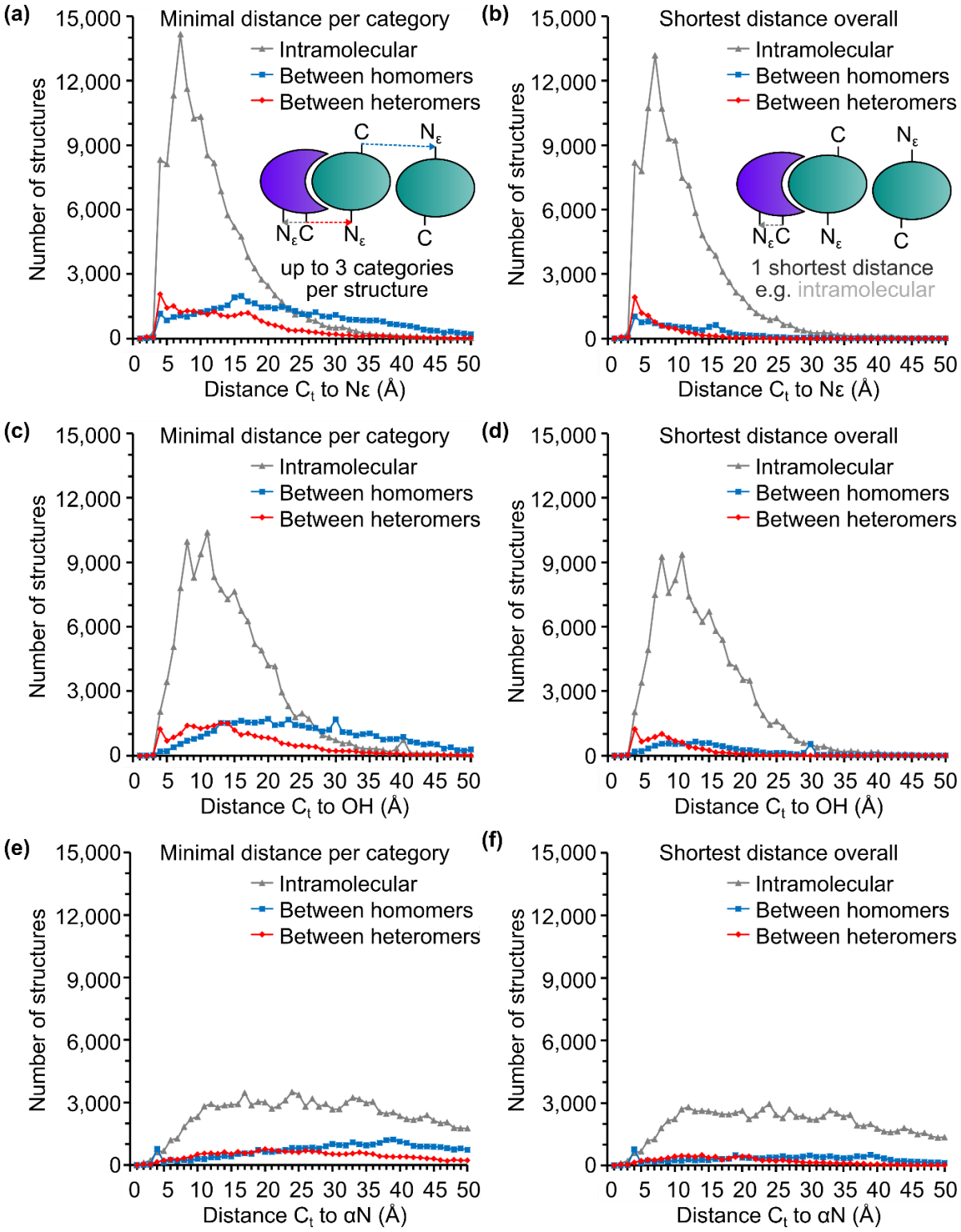


Figure 3.4. Distribution of distances in the NeissDist database. (a) Histogram of structures by their shortest distance (Å) of a given category (intramolecular, between homomers, between heteromers) from C-terminal atom C to lysine atom Nε. (b) Histogram of structures by their shortest overall distance (Å) from C-terminal atom C to lysine atom Nε, separated by category. (c,d) as (a,b) for C-terminal atom C to tyrosine atom OH. (e,f) as (a,b) for C-terminal atom C to N-terminal atom N. The NeissDist database was analysed in 0.1 Å steps, histograms are binned at 1 Å. For each PDB entry, results were pooled across biological assemblies. If no biological assemblies were generated, the asymmetric unit was selected.

3.3.4. Selection of NeissLock model complexes

Considering distances of C-terminal atom C to lysine atom Nε, from 10,101 intermolecular heteromeric distances at < 10 Å, an intermolecular heteromeric distance was the shortest distance in the entire structure for 6,584 structures. To arrive at a NeissLock candidate, I inspected structures preselected from NeissDist in PyMOL using a script for loading and rapid visualization of structures, e.g. to identify steric hindrance in the complex (see Figure 3.1). Using this tool, structures could be assessed within a few seconds, allowing for the post-screening of hundreds to thousands of structures in a matter of a few hours (Figure 3.5).

I then searched the NeissDist database for complexes which would meet the requirements for ideal NeissLock candidates. For an *in vitro* model complex, I decided on the PDB ID 4zgy (Wu *et al.*, 2015) crystal structure of the Ornithine Decarboxylase / Antizyme complex (ODC/OAZ). Furthermore, the Epidermal Growth Factor Receptor / Transforming Growth Factor α (EGFR/TGFα) complex was chosen as a candidate with interesting potential for conjugation to mammalian cells (PDB ID 1mox (Garrett *et al.*, 2002)). Rapid visualization of these complexes is shown if Figure 3.5; a highlight of these structures is shown in Figure 3.6.

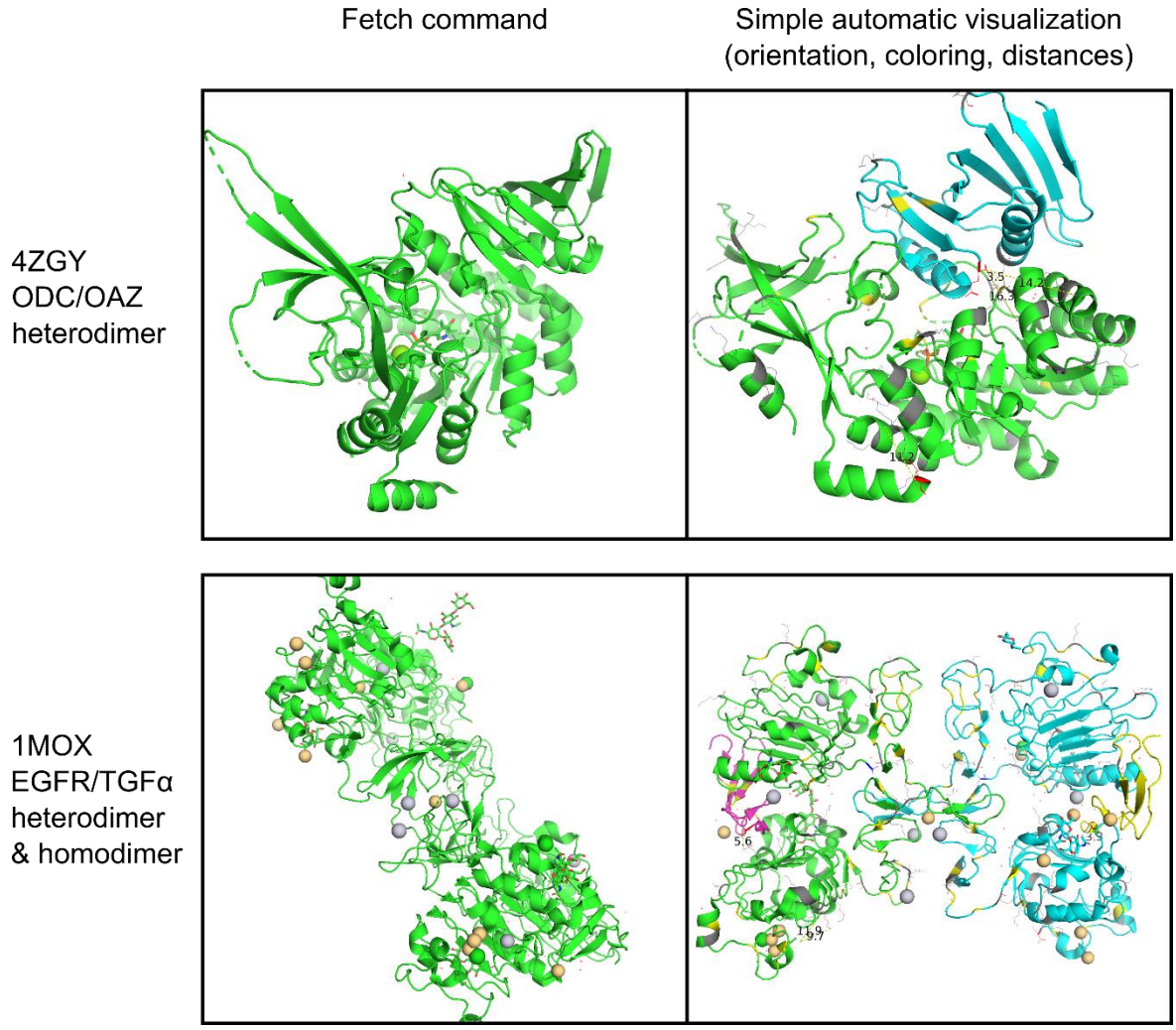


Figure 3.5. Rapid visualization of structures in PyMOL. To facilitate assessment of candidate structures, features relevant to NeissLock were automatically visualized using a dedicated PyMOL script. Examples of this visualization are shown in comparison to the PyMOL "fetch" command. Rapid visualization relied on simple commands such as orientation, colouring of chains and residues, suitable representation of residues, and labelling of distances from chain C-termini to close-by lysine residues. Structures were loaded from a queue and automatically visualized for inspection. Structures shown: PDB ID 4zgy (Wu *et al.*, 2015) and PDB ID 1mox (Garrett *et al.*, 2002), rapidly visualized in PyMOL (as seen after rapid visualization without further annotation, on white background, without ray tracing). Detail view of the same structures in Figure 3.6.

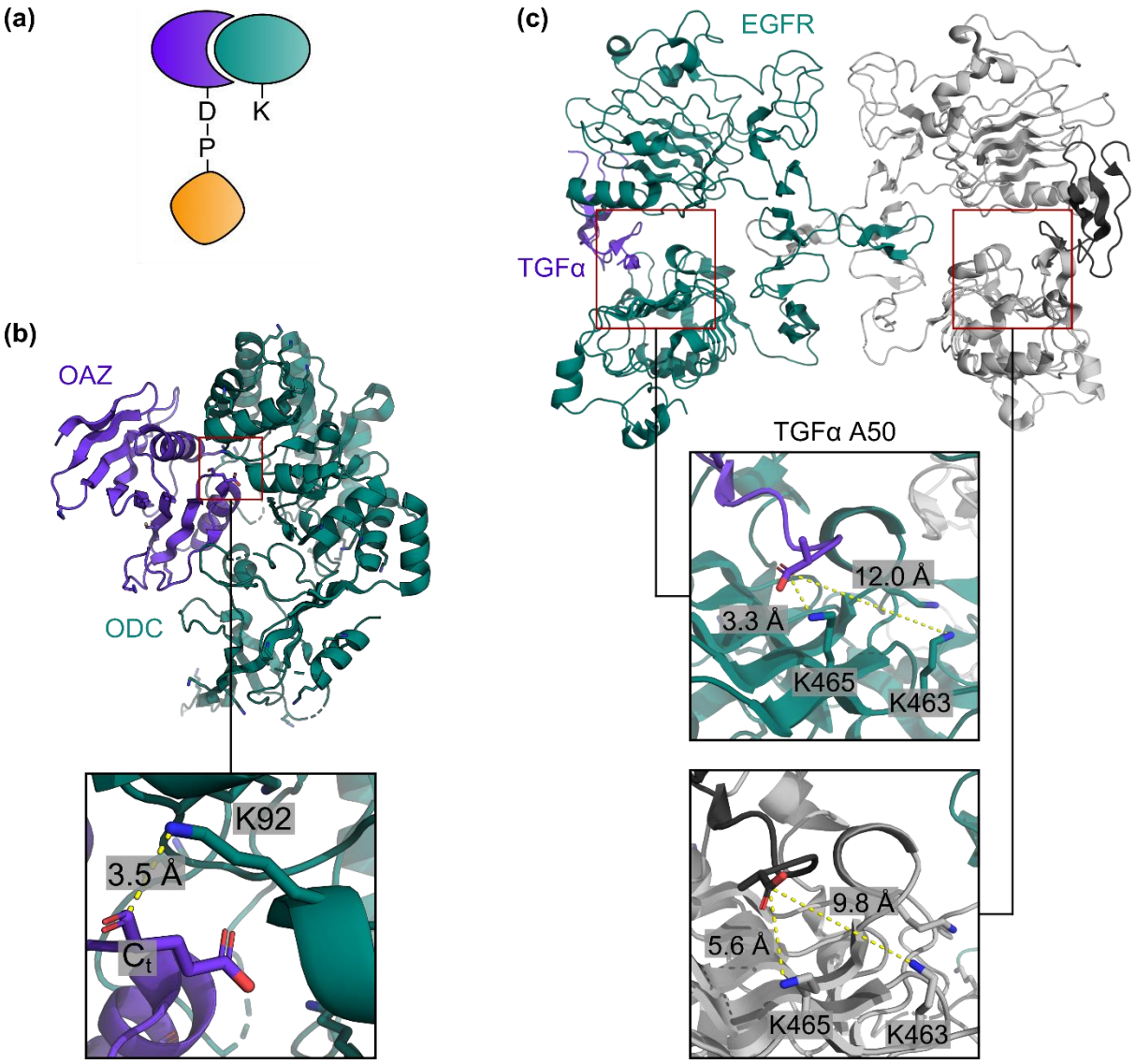


Figure 3.6. The Ornithine Decarboxylase / Antizyme complex (ODC/OAZ) and the Epidermal Growth Factor Receptor / Transforming Growth Factor α (EGFR/TGFα) complex were chosen as the model systems for this thesis. (a) Cartoon illustrating the design of a NeissLock-probe, with a binding protein (purple) bound to a target protein (green), with a lysine residue close to the SPM DP-cleavage site, enabling conjugation of the binding and target protein after anhydride formation. (b) The ODC/OAZ complex (PDB ID 4zgy (Wu *et al.*, 2015)) features a short distance from OAZ C-terminal residue E219 to K92 on ODC (3.5 Å, atom C to atom Nε) and OAZ features no lysine residues in similar proximity to E219. (c) The EGFR/TGFα complex (PDB ID 1mox (Garrett *et al.*, 2002)) shows TGFα (purple, black) bound to the ectodomain of the dimeric receptor protein EGFR (green, grey). Two lysine residues of EGFR are close to the C-terminal A50 of TGFα, with two sites resolved in the crystal structure (distances atom C to atom Nε). Structures are visualized in PyMOL.

The latest iteration of NeissDist allows for the simple filtering of structures flagged as heterodimeric, placing 4zgy at rank 169 of 1st biological assemblies (or at a lower rank with additional filters, such as filtering out ubiquitination or setting minimal and maximal chain lengths). The initial search for model complexes was conducted in a similar manner from the

second iteration of NeissDist, although with less efficient search mechanisms. Herein, 4zgy was selected from a reduced list of candidates utilizing various parameters and search terms (e.g. structures with 2 chains, 1st biological assembly, with keywords “complex”/ “bound”/ “binding”, without keywords “ubiquitin”/ “ubiquitinylated”, …). In addition to a brief distance between the C-terminal residue of the binding protein and a lysine side chain on the target protein at 3-4 Å each and no obvious competing residues on the binding protein, both complexes had additional features that led to their selection as model systems. The ODC/OAZ complex in the 4zgy crystal structure has no disulfide bonding, can readily be solubly expressed from *E. coli* and features tight binding (Hsieh *et al.*, 2011; Wu *et al.*, 2015), making it suitable for *in vitro* study. Furthermore, the C-terminus of OAZ appears accessible both for processing of SPM and conjugation to ODC K92 (Figure 3.1, Figure 3.6b). ODC/OAZ was selected as the model system to establish NeissLock conjugation *in vitro*, which is explored in detail in Chapter 4. The TGFα/EGFR complex in the 1mox crystal structure would likely present a more challenging system, as both proteins are subject to complex disulfide bonding. Nevertheless, this model system was particularly appealing as NeissLock conjugation looks very feasible (short distance from TGFα A50 to two lysines on EGFR) (Figure 3.6c) and EGFR is an important target on the cell surface, relevant to various cancer indications (Herbst, 2004). Thereby, NeissLock conjugation to EGFR would allow for the conjugation of the cell surface, a promising first step towards further applications. Therefore, the TGFα/EGFR complex was chosen as the model system for cell conjugation, which is explored in detail in Chapter 5. In brief, I successfully derived specific NeissLock-probes from OAZ, targeting ODC, and from TGFα, targeting EGFR. These were the first NeissDist-derived complexes which were tested, giving a promising indication for the rational design of NeissLock probes.

3.3.5. Exploration of additional NeissDist complexes

In addition to selection of initial model complexes, I was interested in the exploration of other structures that could enable NeissLock applications.

First, I was interested in the selection of an evolvable antibody or alternative ligand scaffold fused to SPM. In most contemporary protein scaffolds, the C-terminus is not close to the mutated binding surface (Škrlec *et al.*, 2015). Therefore, I looked for evidence of a protein complex in which the C-terminus of an evolved scaffold was close to lysines on the target protein, indicating proximity to the binding site in general. For instance, I identified the crystal structure of a monobody in complex to a MATE multidrug transporter (PDB ID 4hum (Lu *et al.*, 2013), Figure 3.7a). In this monobody, the C-terminus of the monobody is close to the binding interface and to two lysine residues on the target protein (Figure 3.7a). Another example of a monobody candidate is a monobody bound to a tyrosine kinase in PDB ID 5n7e (Reckel *et al.*, 2017) (9.5 Å from monobody T95 atom C to K622 atom Ne of Bcr-Abl). This monobody was especially interesting as it was derived with randomization of the scaffold at a β-strand (Koide *et al.*, 2012), creating a binding 'side' which was effectively closer to the C-terminus.

Secondly, I was interested in conjugation to red blood cells, which could enable stable functionalisation to enable the use of red blood cells as drug carriers (Han *et al.*, 2018). I interrogated NeissDist for a list of red blood cell surface proteins (Daniels, 2007). In this search I identified the structure of mamba toxin fasciculin II bound to acetylcholinesterase (PDB ID 1fss) (Harel *et al.*, 1995). Acetylcholinesterase is found primarily in muscles and synapses, wherein fasciculins can cause muscle twitching (Rodríguez-Ithurralde *et al.*, 1983), but a splicing variant (with an alternative C-terminal region for membrane anchoring) is also found on red blood cells (Grisaru *et al.*, 1999). The C-terminus of Fasciculin II is positioned 9.5 Å from K341 on acetylcholinesterase (Figure 3.7b). Two lysine residues of Fasciculin II are near the C-terminus of the protein itself, but this structure could provide a starting point for protein engineering of a red blood cell targeting probe.

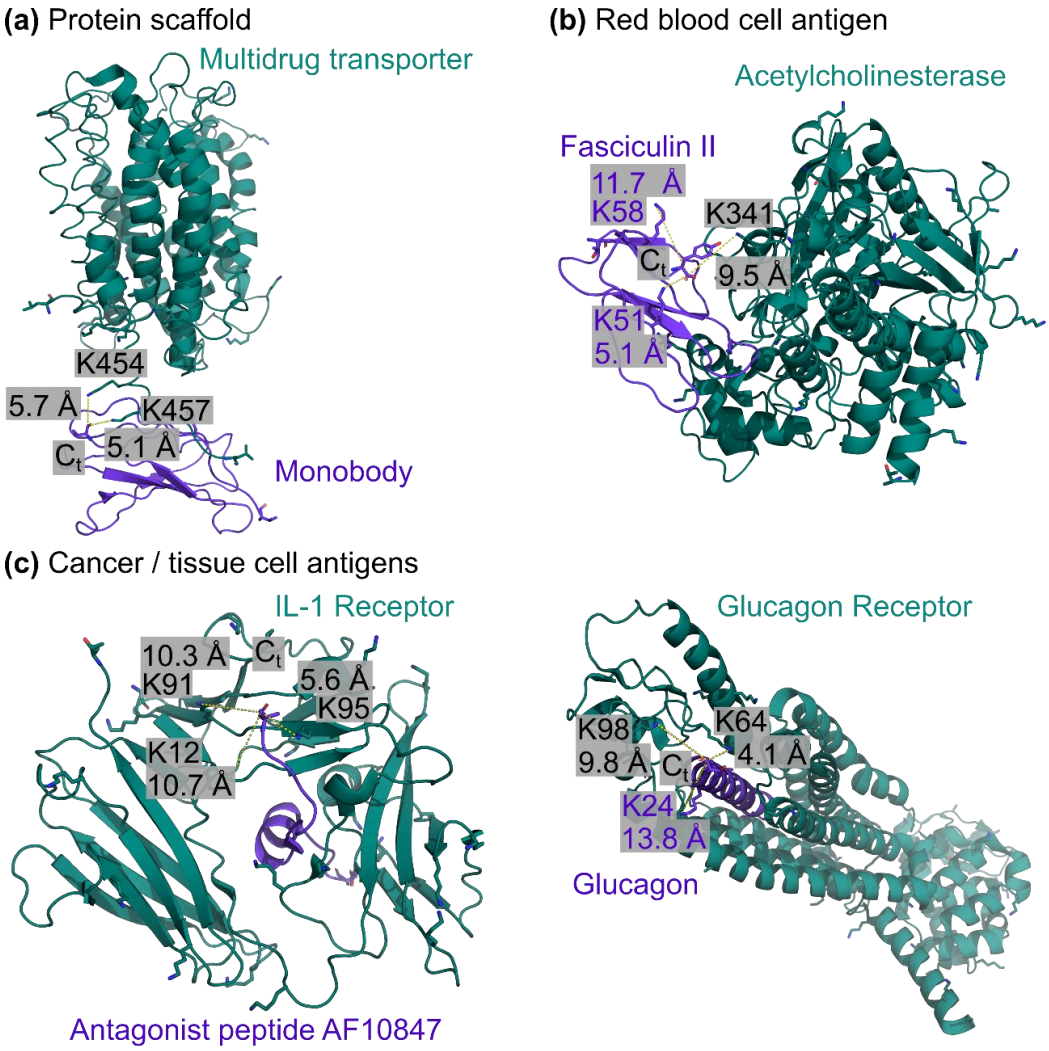


Figure 3.7. Excerpt of additional complexes which were considered for NeissLock applications. (a) Monobodies could be potential candidates for evolution of a scaffold-SPM covalent binding probe. Monobody bound to a transmembrane transporter protein featuring favourable distances from the monobody C-terminus to target lysines (PDB ID 4hum (Lu et al., 2013)). (b) Fasciculin II binds Acetylcholinesterase (PDB ID 1fss (Harel *et al.*, 1995)), which is also found in the membrane of red blood cells. Despite the potential for side-reactions, Fasciculin II shows a binding mode promising for covalent conjugation. (c) Additional proteins were identified in a broad search for cell surface conjugation. Two examples are shown, with interleukin-1 receptor bound by an antagonist peptide (PDB ID 1g0y (Vigers *et al.*, 2000)) or glucagon receptor bound by glucagon (PDB ID 5yqz (Zhang *et al.*, 2018)). Structures visualized in PyMOL, with distances from C_t atom C to target lysine (K) atom Nε.

Third, I was trying to identify additional proteins which could be utilized for conjugation to the cell surface, as alternatives to the TGFα/EGFR system. I collated a series of proteins found on the mammalian cell surface from various databases (e.g. Cell Surface Protein Atlas (Bausch-Fluck *et al.*, 2015), The Human Protein Atlas ((Thul *et al.*, 2017), https://www.proteinatlas.org/), Membrane Proteins of known 3D Structure (https://blanco.biomol.uci.edu/mpstruc/)) and used it to filter NeissDist. Herein, various

candidate complexes were identified, such as Interleukin-1 receptor bound by an antagonist peptide (Figure 3.7c) or the Glucagon receptor in complex to its native ligand (Figure 3.7c). Furthermore, I identified various other EGFR-ligands as interesting candidates for NeissLock conjugation, which are illustrated in Chapter 5 (Figure 5.1).

3.4. Alternative disCrawl implementations and uses of NeissDist

3.4.1. Search for SPM alternatives

Finally, I utilized disCrawl and the NeissDist database to answer a different question than selection of NeissLock candidates. I became interested in finding protein structures showing (potentially unidentified) Asp-Pro processing, as I wanted to gain insight into proteins with homologous activity to SPM; in addition to finding potential SPM alternatives, the NMR structure of SPM (Kuban *et al.*, 2020) was not yet available at the time of this search.

First, I used NeissDist directly. By specifically filtering for complexes in which the C-terminal residue was Asp with a short distance to an N-terminal Pro on a separate chain, I aimed to identify structures in "post-cleavage" state. Most prominently, I was able to identify SO1698, which features pH-induced (pH 7.5) cleavage at a DP bond (Osipiuk *et al.*, 2012). In the PDB ID 3n55 crystal structure, intramolecular isopeptide bond formation is observed to K98 (Osipiuk *et al.*, 2012) (Figure 3.8a). Furthermore, by filtering for C-terminal Asn and N-terminal Pro, I identified the *E. coli* Type III secretion system component EscU, which features NP cleavage activity at N262 to P263 (Zarivach *et al.*, 2008) (Figure 3.8b). Homologous components of Type III secretion systems were also identified, such as *Salmonella enterica* SpaS (NP bond, 10.3 Å post-cleavage in 4th biological assembly, PDB ID 3c01 (Zarivach *et al.*, 2008)), *Aquifex aeolicus* FlhB (NP bond, 7.8 Å post-cleavage in 3rd biological assembly, PDB ID 3b1s (Meshcheryakov *et al.*, 2013)) and *Shigella flexneri* Spa40 (NP bond, 10.1 Å post-cleavage, PDB ID 2vt1 (Deane *et al.*, 2008)). Finally, 'Repulsive Guidance Molecule B' (RGMB) in complex with Neogenin1 (Neo1) shows autoproteolytic activity at a DP bond (Bell *et al.*, 2013) (Figure 3.8c).

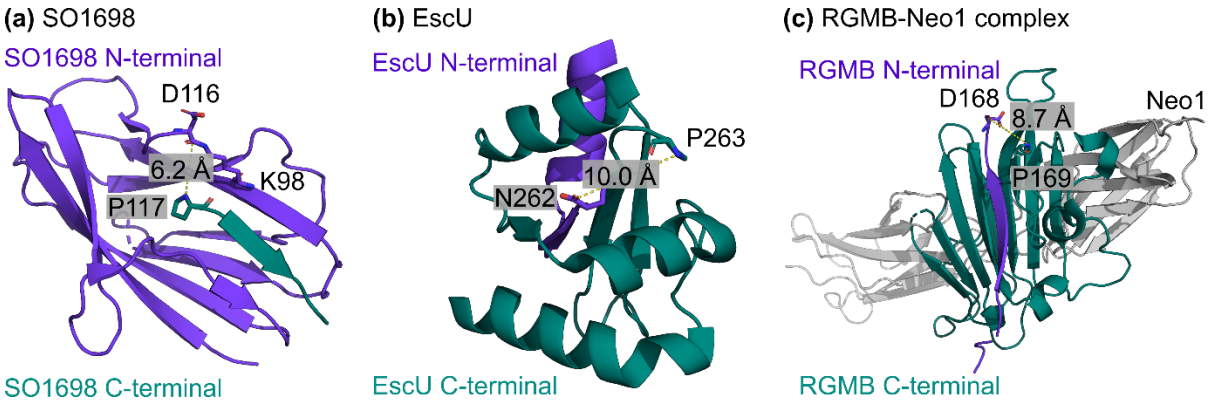


Figure 3.8. NeissDist was used to identify proteins with autoproteolytic activity at DP and NP bonds. (a) S01698 after cleavage at a DP-bond with intramolecular crosslink, PDB ID 3n55 (Osipiuk *et al.*, 2012). Wild-type structure shown, lowest ranked entry K98A mutant, PDB ID 3njg (Osipiuk *et al.*, 2012) (3.2 Å) (b) *E. coli* EscU is a part of the Type III bacterial secretion system which features self-processing at an NP-bond, biological assembly 1 of PDB ID 3bzl (Zarivach *et al.*, 2008). Wild-type structure shown, lowest ranked entry Y316D mutant, PDB ID 3bzy (7.1 Å) (c) Repulsive guidance module (RGMB) / Neogenin (Neo1) complex, in which RGMB processes at a DP-bond, biological assembly 1 of PDB ID 4bq7 (Bell *et al.*, 2013). Structures visualized in PyMOL, distances from indicated residue atom C to proline atom N.

3.4.2. PDB dipeptide analysis

Secondly, I attempted to search for Asp-Pro processing which might have escaped the authors of the structure. The first iteration of disCrawl was modified to define each residue and its preceding residue as a residue pair, calculating the bond-length of residue number n-1 atom C to residue number x atom N along the peptide backbone. Herein, subsequent residues with a bond length >1.6 Å were considered ‘overlong’ and those distances were saved in a database. This search was executed 6th March 2018 on a subset of the PDB, with asymmetric units in mmCIF format. I visualized these structures in PyMOL, using an adapted tool that automatically retrieved and visualized electron density maps and difference electron density maps. Using this method, I was able to identify an ‘overlong’ Asp-Pro bond in the 2ast crystal structure (Hao *et al.*, 2005), with 2.3 Å from S-phase kinase-associated protein 1 (Skp1) Asp1078-Pro1038. Overlay of the difference electron density map to the electron density map shows surplus electron density in the structure model (Figure 3.9). The same obsrvation is made for the 2ass crystal structure of the same complex (without p27 peptide, at 2.3 Å) (Hao *et al.*, 2005). Similarly, the region of Skp1 around the Asp1037-1038 bond is also unresolved

in other deposited crystal structures of Skp1 complexes, such as PDB ID 3wso (Skp1-FBG3) (Kumanomidou *et al.*, 2015) and PDB ID 5k35 (Skp1-AnkB) (Wong *et al.*, 2017).

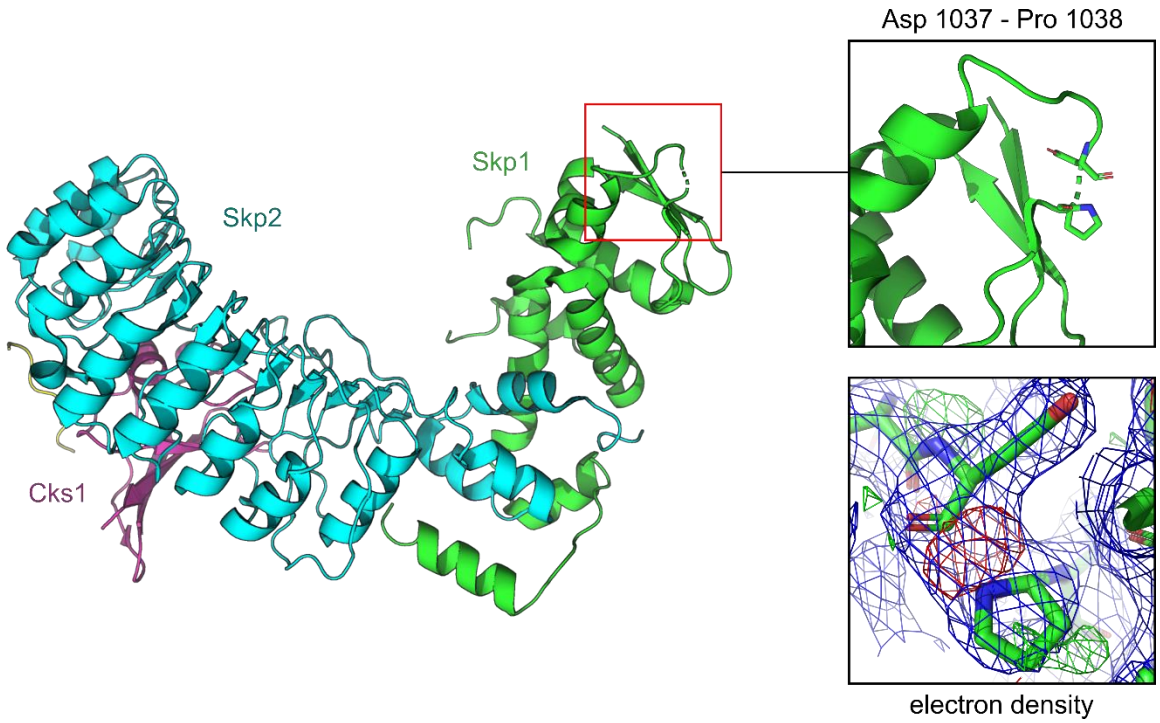


Figure 3.9. Potential overfitting in S-phase kinase-associated protein 1 (Skp1). The PDB ID 2ast crystal structure contains Skp1 in complex with Skp2, Cyclin-Dependent Kinase regulatory subunit 1 (Cks1) and a small p27 peptide (yellow) (Hao *et al.*, 2005). At 2.3 Å, the Asp1037-Pro1038 bond of Skp1 as modelled is 'overlong' (>1.6 Å), with a poor fit of the electron density map (blue) to the difference electron density map at this position (red).

3.5. Discussion

The NeissDist database proved to be a valuable tool for the identification of candidate complexes according to predetermined structural criteria. Two complexes were selected for *in vitro* study, which are explored in detail in the following chapters. Further to these, a large number of protein structures have structural features which could enable NeissLock conjugation, and examples were presented herein. The main features I considered for a NeissLock-probe were C-terminal to target residue distance, as well as intramolecular residues which could allow self-reaction. The feasibility of these parameters as a predictor of conjugation will have to be discussed after considering the experimental evidence in the following chapters. Furthermore, accessibility of the C-terminus for SPM-processing was considered as a potentially important parameter. During the course of this research, the NMR

structure of SPM has been published, supporting the partial occlusion of the DP-bond during cleavage (Kuban *et al.*, 2020).

Aside from the choice of NeissLock parameters, certain limitations exist with the design of the disCrawl program and NeissDist. First, the choice of biological assemblies for the primary analysis herein reduces the amount of structural information assessed (as not all elements of the asymmetric unit translate into the biological assembly, compare to multiple structures per chain, Figure 3.6c). By adapting the protocol for biological assembly, replicate sets of biological assemblies could capture this information. Similarly, restrictions of model count and file size could be increased or lifted altogether (since disCrawl has now been implemented on a supercomputer). Secondly, although I am mostly referring to measured 'distances', this analysis does not take into account the actual resolution of the crystal structure other than as metadata (see Table 3.2). Similarly, indicators of the quality of a protein structure were not considered (Wlodawer *et al.*, 2008). Even at low resolution of a protein structure, the position estimates of atoms are well within the considered distances (< 10 Å). Nevertheless, a reiteration of NeissDist could include B-factors of residue pairs and estimate errors in distance. Another source of inaccuracy could be flexibility in the position of side chain residues, such as lysine amines (e.g. orientation towards or away from the C-terminus). Backbone atoms can provide a more stable estimate, and I commonly considered both the distance of the C-terminus to the side chain atoms and backbone atoms when searching for NeissLock candidates. Some variation also exists independent of structural characteristics. Metadata in deposited protein structures can be incomplete (http://mmcif.wwpdb.org/) but provides a useful indication to rapidly filter and screen candidates followed by further validation. No distances were mistakenly annotated by NeissDist as intramolecular and heteromeric (which would be, by definition, impossible) and I have not yet observed wrongly annotated structures, indicating that this classification was reliable. However, ubiquitylation was a common observation for C-terminal to lysine distances close to 1.5 Å, representing covalent linkage, and I found an example of a malformed biological assembly when comparing locally

assembled structures to the protein database (3wkn). Although such limitations should be considered in the interpretation of NeissLock statistics, NeissLock was implemented to be “greedy”, that is to pre-select a broad but effective range of NeissLock-candidates from the PDB, which could then be of manageable size (hundreds to thousands) for final screening. The use of auxiliary scripts for effective filtering and suitable PyMOL visualization facilitated the identification of NeissLock candidates and could be used to effectively find candidates with various different research questions in mind (see Figure 3.6, Figure 3.7).

Finally, I also demonstrated to use of disCrawl and NeissDist for other purposes. Through assessment of ‘overlong’ Asp-Pro dipeptides, indications of poor model fit were found for a DP-bond in the PDB ID 2ast crystal structure of Skp1 (Hao *et al.*, 2005). It appears unlikely that a covalent break would have escaped various researchers, as the intact loop is modelled in other structures (e.g. NMR structure PDB 5xyl [DOI 10.2210/pdb5XYL/pdb, literature unpublished]). Nevertheless, I was able to identify an irregularity in the protein model when compared to the electron density using this method. For context, a prior analysis of relative dipeptide frequency with a non-redundant (40%) sequence dataset from UniProt did not show a reduction in Asp-Pro frequency compared to the inverse dipeptide Pro-Asp (Carugo, 2013). Alternative to the analysis of ‘overlong’ Asp-Pro bond, analysis of ‘post-cleavage’ DP-bonds or NP-bonds proved effective at identifying various proteins with self-cleaving activity.

4. Establishing the NeissLock principle

4.1. The Ornithine Decarboxylase / Antizyme in vitro model complex

To demonstrate that NeissLock-mediated protein-protein conjugation is possible and to understand its requirements, I first screened the protein database for a model system. I aimed to identify protein-protein complexes with promising steric characteristics which further appeared to be suitable for manipulation *in vitro*. Herein, the Ornithine Decarboxylase (ODC) / Ornithine Decarboxylase Antizyme 1 (OAZ) complex was identified from the NeissDist pipeline as a candidate model system (Chapter 3). The ODC/OAZ complex was selected based on its crystal structure (ODC/OAZ$^{95\text{-}228}$ PDB 4zgy (Wu *et al.*, 2015), Figure 4.1) in addition to literature information.

First, the ODC/OAZ complex features a short distance of 3.5 Å from the resolved C-terminus of OAZ (E219 atom C) to the nearest lysine on ODC (K92 atom Nε), conceivably placing an SPM-generated anhydride in proximity to the target residue (Figure 4.1a). Second, Lysine residues on the activated protein which are nearby the anhydride could be a cause of self-reaction. OAZ features no lysine residues obviously close to the C-terminus. From OAZ E219, the nearest lysine on OAZ is K153 in 19.0 Å (atom C to atom Nε) (Figure 4.1b). Third, the C-terminus of OAZ appears solvent accessible which could avoid steric clashes and facilitate SPM-processing. Fourth, the complex is of suitable strength for *in vitro* study and manipulation (expected wt K_D~0.7 µM (Hsieh *et al.*, 2011; Wu *et al.*, 2015), ~1.5 µM for OAZ$^{95\text{-}228}$ (Hsieh *et al.*, 2011)). Finally, neither ODC nor OAZ feature disulfide bonding and both ODC and OAZ can be solubly expressed in *E. coli* (Hsieh *et al.*, 2011; Wu *et al.*, 2015). Therefore, OAZ appeared to be a suitable candidate to derive a NeissLock probe for covalent conjugation, i.e. targeting ODC.

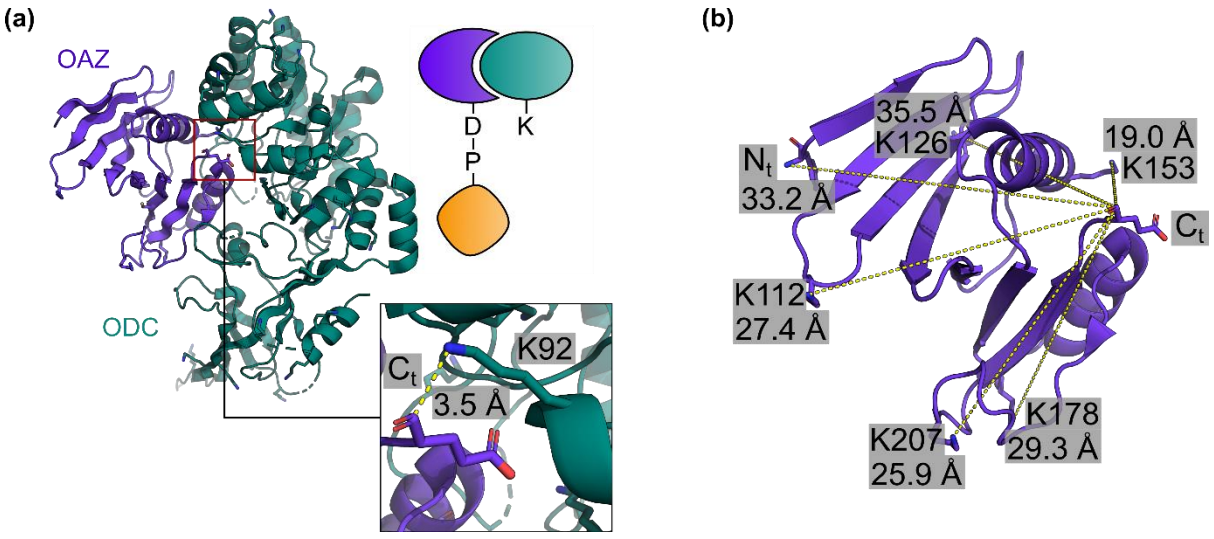


Figure 4.1. Overview of the Ornithine Decarboxylase (ODC) / Antizyme (OAZ) heterodimeric model complex for NeissLock conjugation. (a) The ODC/OAZ complex was identified from NeissDist as a candidate complex for NeissLock conjugation (PDB ID 4zgy (Wu *et al.*, 2015)). The C-terminal resolved residue of OAZ is close to a lysine on ODC. Inset: Detailed view of the resolved C-terminal residue OAZ E219 with a distance of 3.5 Å to ODC K92 (atom C to atom Nε). Cartoon: Visualizing the NeissLock binder-activator/target design compared to the crystal structure. (b) Closeup view of OAZ from the OAZ/ODC crystal structure (PDB ID 4zgy (Wu *et al.*, 2015)). At 19.0 Å, OAZ E219 is the nearest lysine of OAZ from E219 OAZ. Structures visualized in PyMOL

4.1.1. Introduction to Ornithine Decarboxylase / Antizyme biology

Ornithine Decarboxylase (ODC) and Ornithine Decarboxylase Antizyme (OAZ) are intracellular proteins promoting and regulating polyamine synthesis (Pegg, 2006). As a homodimer, ODC is a pyridoxal phosphate (PLP)-dependent enzyme that catalyzes the formation of putrescine from ornithine (Pegg, 2006) (Figure 4.2).

Putrescine is the first building block of polyamine synthesis (Pegg, 2006). After the rate-limiting decarboxylation of ornithine, putrescine is elongated by spermidine synthase to the polyamine spermidine, and spermidine is further elongated by spermine synthase to the polyamine spermine (Pegg, 2006) (Figure 4.2). Polyamine homeostasis is integral to cellular function and ODC activity is therefore subject to strict regulation (Pegg, 2006) (Figure 4.2, inset).

ODC functions as an obligate homodimer (human ODC K_D 0.1 µM (Lee *et al.*, 2014)). Its activity is regulated by Ornithine Decarboxylase Antizyme (OAZ), which binds ODC monomers with high affinity (human OAZ/ODC, K_D~0.22-0.71 µM (Hsieh *et al.*, 2011; Liu *et al.*, 2011; Wu *et al.*, 2015)). Binding of OAZ to ODC prevents ODC dimerization and recruits ODC to the

proteasome for degradation (Pegg, 2006). OAZ is generally not degraded while targeting ODC to the proteasome, causing the degradation of multiple ODC monomers (Pegg, 2006).

OAZ itself is regulated in a polyamine-concentration dependent manner (Ivanov *et al.*, 1998). The OAZ gene features a naturally occurring frameshift, which normally prevents translation of the intact protein by the ribosome (Ivanov *et al.*, 1998). However, high levels of polyamines increase the rate of ribosome slippage, thereby facilitating synthesis of intact OAZ (Matsufuji *et al.*, 1995; Ivanov *et al.*, 1998) (Figure 4.2, inset). Palanimurugan et al. have further found spermidine concentrations to positively affect the lifetime of *Saccharomyces cerevisiae* OAZ homologue (Palanimurugan *et al.*, 2004), and subsequent work showed polyamines to directly promote the OAZ-mediated degradation of ODC (Beenukumar *et al.*, 2015). Taken together, polyamine-regulated OAZ inhibition constitutes a negative feedback loop for ODC activity.

Counteracting this negative feedback, OAZ is itself subject to regulation. Antizyme Inhibitor (AzI) is an ODC homologue lacking enzymatic activity. AzI binds to OAZ at a higher affinity than ODC (in mouse: ODC/OAZ K_D 0.1 nM compared to AzI/OAZ K_D 0.017 nM (Cohavi *et al.*, 2009)). Thereby, AzI sequesters OAZ from ODC regulation (Nilsson *et al.*, 2000). Ivanov et al. reported that AzI is itself subject to translational regulation: The AzI mRNA contains an upstream coding region with a regulatory function (Ivanov *et al.*, 2008). Herein, ribosome stalling at a PPW motif requires eIF5A release, which was found to be inhibited by polyamines (Ivanov *et al.*, 2018). Therefore, as polyamine concentrations are raised and more intact OAZ protein is produced from its frameshifted mRNA, less intact AzI is produced from its mRNA. Conversely, ODC inhibition through OAZ ultimately lowers polyamine concentrations, facilitating synthesis of intact AzI and inhibition of OAZ (Figure 4.2, inset). In summary, ODC, OAZ and AzI are well studied proteins important for human metabolism.

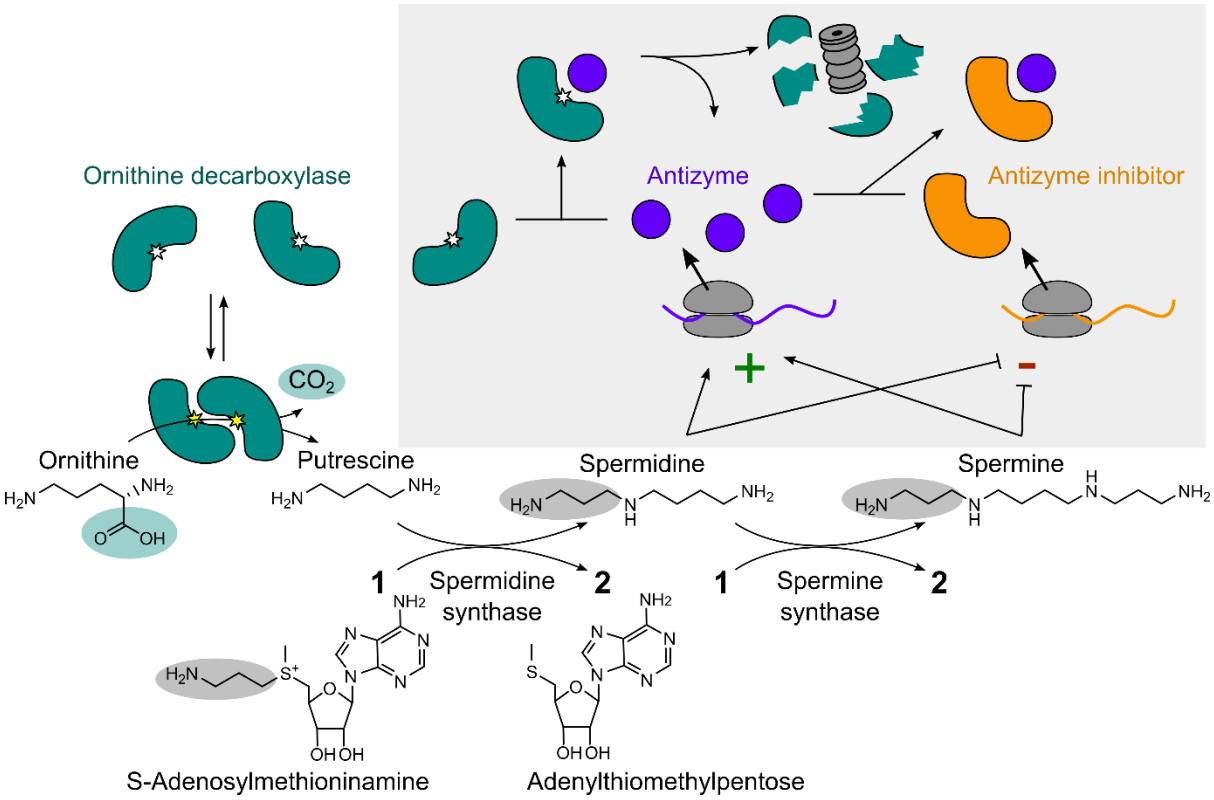


Figure 4.2. Overview of Ornithine Decarboxylase and Antizyme biology. Ornithine decarboxylase (ODC) is enzymatically active as a homodimer. Spermidine and spermine concentrations are limited by ornithine decarboxylation. Shaded inset: Translational feedback loops regulate ODC activity. High ODC activity increases polyamine concentrations. High polyamine concentrations promote frameshifting required for antizyme (OAZ) synthesis. OAZ recruits ODC to the proteasome for degradation. High polyamine concentrations also inhibit translation of mRNA for antizyme inhibitor (AzI). If polyamine concentrations decrease, AzI sequesters OAZ to permit ODC function.

4.2. Initial evidence of OAZ/ODC Neisslock-conjugation

For the first test of intermolecular NeissLock conjugation, I cloned ODC and OAZ$^{95\text{-}219}$-SPM, in which OAZ$^{95\text{-}228}$ was truncated to E219 and SPM was directly inserted thereafter (hereafter referred to as “OAZ” and “OAZ-SPM”). Initial samples of OAZ-SPM were prepared from *E. coli* according to protocols previously used in our group for SPM-fusions, with a single-stage purification through IMAC (Metzner, 2017). ODC was purified in parallel according to the same protocol. Both preparations gave moderate purity, although OAZ-SPM displayed notable precipitation during dialysis. Nevertheless, the proteins were taken forward for first processing and conjugation tests.

Upon incubation with 10 mM calcium at 37 °C, OAZ-SPM displayed self-processing activity to yield SPM and OAZ (Figure 4.3). Notably, the OAZ product resolved to two distinct species

on SDS-PAGE. When OAZ-SPM was incubated with calcium in the presence of 50 mM dithiothreitol (DTT), the OAZ product instead resolved in a single band. Thiols were previously reported to effectively quench the SPM-generated anhydride (Sadilkova *et al.*, 2008). For OAZ-SPM, thiolysis of the anhydride similarly appears to prevent the reactions leading to double-banding. Gel mobility is commonly affected by protein cyclization, as we have previously observed in our research group (Schoene *et al.*, 2014). Since OAZ formed in the presence of DTT runs parallel to the 'upper' band of double-banded OAZ, the OAZ species with decreased gel mobility were determined to correspond to linear OAZ. Correspondingly, OAZ species with increased gel mobility, forming the 'lower' of double-banded OAZ, were determined to correspond to circularized or cyclized OAZ.

In the presence of ODC, calcium-induced cleavage of OAZ-SPM led to the formation of a unique higher molecular weight species, corresponding to the expected size of an ODC:OAZ conjugate (expected 68.3 kDa). Conversely, formation of free OAZ was reduced. ODC was not visibly affected by the addition of calcium when incubated without OAZ-SPM. Taken together, these observations indicated that OAZ covalently conjugated to ODC.

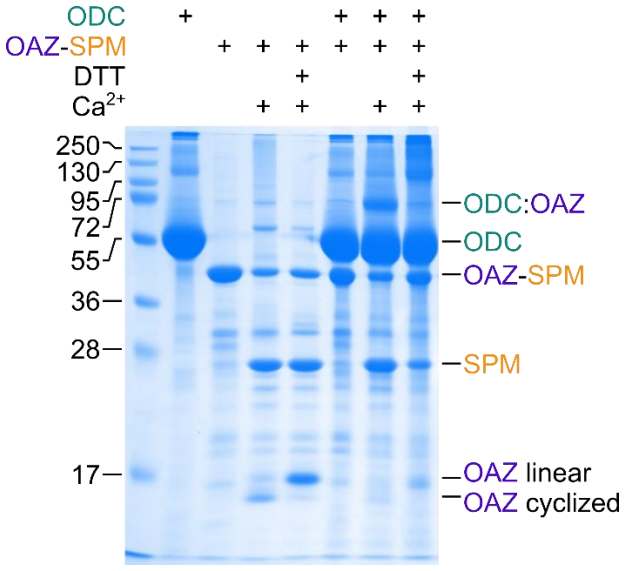


Figure 4.3. Early evidence of NeissLock conjugation with OAZ-SPM and ODC. 11.5 µM crudely purified OAZ-SPM was incubated at 37 °C for 43 h ±10 mM $CaCl_2$ and ±33 µM ODC in 50 mM HEPES, 140 mM NaCl pH 7.40. SDS-PAGE was Coomassie stained before imaging. Formation of a covalent adduct corresponding to the molecular weight of ODC:OAZ is observed in the presence of ODC. Conjugation was inhibited upon co-incubation with 50 mM dithiothreitol (DTT). Reactions were stopped with EDTA in reducing SDS-loading buffer. Samples were boiled and resolved on 16% SDS-PAGE which was stained with Coomassie before imaging.

These results were promising first evidence for the ODC:OAZ model system. However, cleavage of OAZ-SPM was slow compared to constructs reported in the literature (Osička *et al.*, 2004; Sadilkova *et al.*, 2008; Metzner, 2017). Since OAZ-SPM processing only approached completion after prolonged incubation (e.g. 1-2 days), it was difficult to effectively work with this model and estimate conjugation efficiency. Furthermore, large amounts of high molecular-weight aggregates were visible in SDS-PAGE and precipitation was commonly found in incubated samples, indicating instabilities and protein aggregation which could affect SDS-PAGE observations. Therefore, further optimization was necessary.

4.3. Optimization of conjugation rate and yield

4.3.1. Kinetics of anhydride activation

First, I aimed to test the promiscuity and lifetime of the anhydride to further explore the suitability of a reactive anhydride for covalent conjugation. Herein, I used an Affibody-SPM construct, which was easy to produce in high yield. The construct features an N-terminal His-tag, an anti-HER2 affibody with a flexible linker [$(GS)_2(GGGGS)_3GEG$] and three residues from native FrpC (HVY) before SPM, followed by a C-tag for purification. This construct had been used in our research group for early exploration experiments, for instance to confirm thiol-reactivity of the anhydride (Metzner, 2017) as had previously been reported (Sadilkova *et al.*, 2008).

To test the promiscuity of the anhydride, I incubated 20 µM Affibody-SPM with 10 mM calcium for 1 h at 37 °C in the presence of a selection of nucleophiles at 1 mM or 10 mM concentration, before stopping the SPM processing by addition of SDS-loading buffer supplemented with EDTA to sequester calcium. These nucleophiles were chosen to mimic natural amino acids and common reagents, i.e. [1] N-terminal amine, [2] Lysine side chain, [3/4] Thiols (wherein [4] can undergo stabilising S,N-acyl shift to form an amide), [5] Tyrosine side chain (Figure 4.4a). Samples were boiled and loaded on SDS-PAGE.

In the presence of calcium, Affibody-SPM readily processed to Affibody and SPM. Herein, Affibody resolved to two distinct species, as observed for OAZ-SPM, i.e. linearized (e.g.

hydrolysed or reacted with another nucleophile in solution) and cyclized (self-reacted) species. By quantifying the relative ratio of linearized to cyclized Affibody, I differentiated how readily the anhydride reacted with different nucleophiles (n=1, Figure 4.4d). Cysteine was the most effective nucleophile and substantial conjugation was still observed for dithiothreitol and Gly-Gly (as an analogue for N-terminal amine) (Fig 4.4c,d). Under the tested conditions, free lysines were much less efficient at reaction with anhydrides.

Subsequently, I wanted to estimate the lifetime of the anhydride. 7.5 µM Affibody-SPM was incubated with 10 mM calcium at 37°C. At various timepoints, the cleavage reaction was stopped by addition of EDTA, and at the same time the anhydride was quenched by addition of cysteine in the stop solution (to efficiently quench cyclization) (Figure 4.4e,f). Herein, quenching the reaction within the first 5 minutes substantially increased the formation of linear Affibody relative to cyclized Affibody (Figure 4.4e,f). This effect rapidly became less pronounced as cyclization and linear affibody formation stabilized to a ~2:1 ratio (the ratio of self-reaction to hydrolysis, compare to 4.4c,d). This indicated that most anhydride had been inactivated prior to quenching after a few minutes (Fig 4.4f).

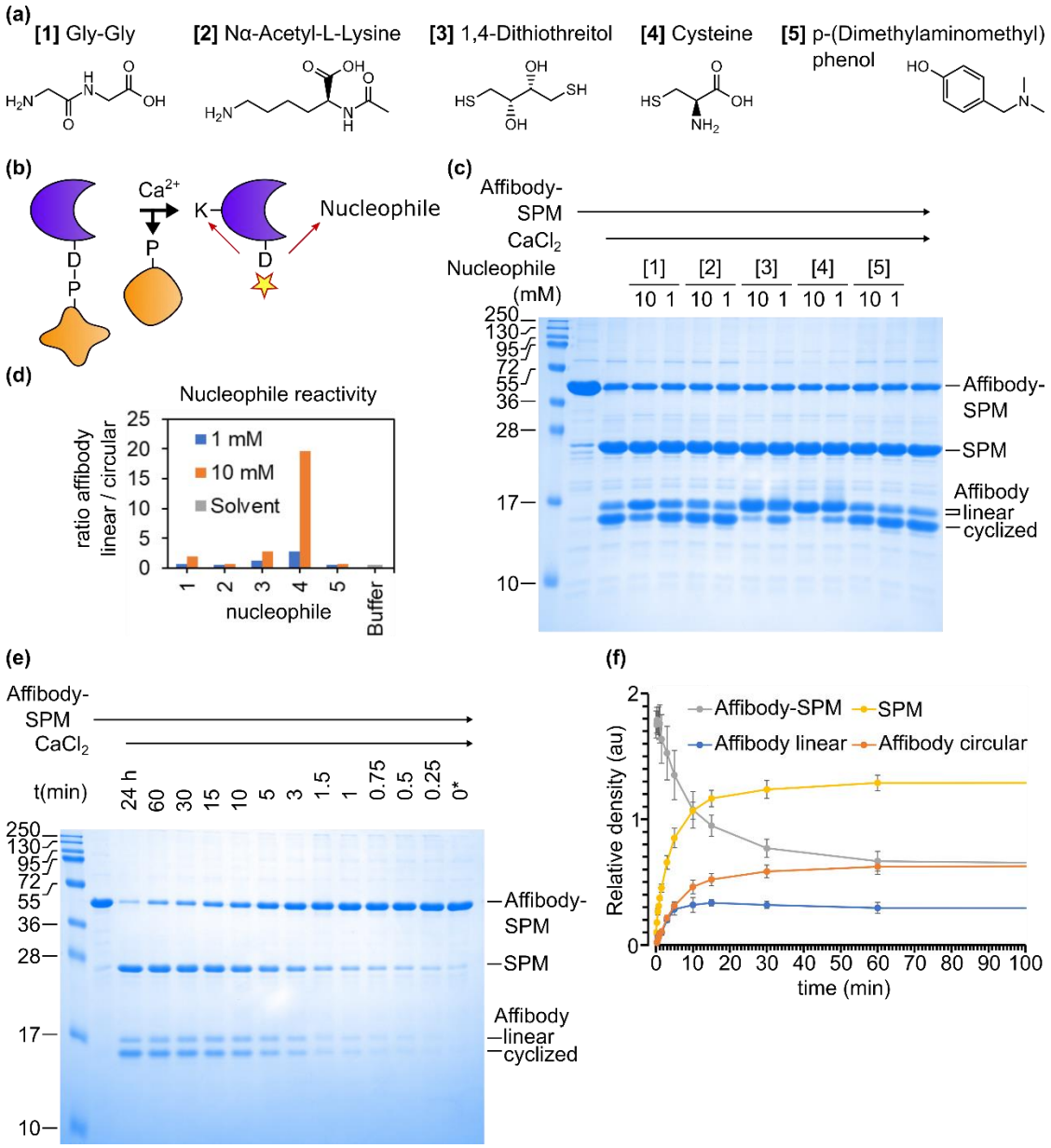


Figure 4.4. Reactivity of the protein anhydride. Affibody-SPM was used as a model to investigate reactivity of the anhydride formed during SPM-processing. (a) Different compounds used to mimic nucleophiles in natural proteins, i.e. [1] N-terminal amine, [2] Lysine side chain, [3/4] Thiols, [5] Tyrosine. [5] presents a mixture of monosubstituted (*o*- or *p*-isomer) compounds as well as twice substituted bis(dimethylaminomethyl)phenol or unsubstituted phenol (COA: 59.7% monosubstituted, 27.3% phenol). (b) Cartoon highlighting that an SPM-generated anhydride can react with nucleophiles in solution to form a linear species or react with nucleophiles on the protein itself to form a cyclized species. (c) 20 µM Affibody-SPM was incubated with 10 mM calcium for 1 h at 37°C subject to 1 mM or 10 mM of Nucleophiles [1]-[5] from (a) testing effectiveness of conjugation. Samples were boiled with SDS-loading buffer and resolved on SDS-PAGE, followed by Coomassie staining. (d) Bands corresponding to cyclized and linear affibody species in (c) were quantified and the relative ratio calculated, indicating a shift from self-reaction to reaction with nucleophiles in solution (e.g. cysteine), n=1. (e) Quenching of the reactive anhydride at various timepoints to estimate lifetime. 7.5 µM Affibody-SPM were incubated with 10 mM calcium at 37°C. At the indicated timepoints, EDTA and cysteine were added to simultaneously stop processing and quench the anhydride. 0*: EDTA and cysteine were added prior to $CaCl_2$. Samples were boiled with SDS-loading buffer and resolved on SDS-PAGE before Coomassie staining. (f) Quantification from (e) (mean ± 1 s.d.; at t>0: mean of triplicate; at t=0: mean of triplicate for Affibody-SPM, mean of duplicate for SPM, no quantification of free affibody).

4.3.2. Saturation mutagenesis of cleavage site with SpyTag-X-SPM

OAZ-SPM had markedly reduced cleavage rate compared to other constructs such as Affibody-SPM. Prior studies have reported a dependence of SPM cleavage rate on the fusion protein, potentially due to steric hindrance (Sadilkova *et al.*, 2008). Further to this, I noticed a common trend of reduced cleavage rate among unpublished constructs (Metzner, 2017) where the residue preceding the scissile bond was glycine (G).

To systematically assess if the residue preceding the scissile bond had a specific impact on cleavage rate and inform the design of an improved OAZ NeissLock probe, I used SpyTag fused to SPM as a simple model system. SpyTag is a disordered peptide (Zakeri *et al.*, 2012), which I predicted to have little specific impact on SPM cleavage rate. I cloned saturation mutants of SpyTag-X-SPM, i.e. 20 variants with X indicating any canonical amino acid.

To prepare the corresponding proteins, I used small-scale protein expression in *E. coli* in auto-induction medium for 24 h 30°C. Subsequently, cells were pelleted and lysed by lysozyme treatment supplemented with freeze-thaw cycles. Proteins were purified using small-scale IMAC in filter plate format and finally exchanged into HEPES-buffered saline during spin concentration. Samples of this purification can be found in Figure 4.5.

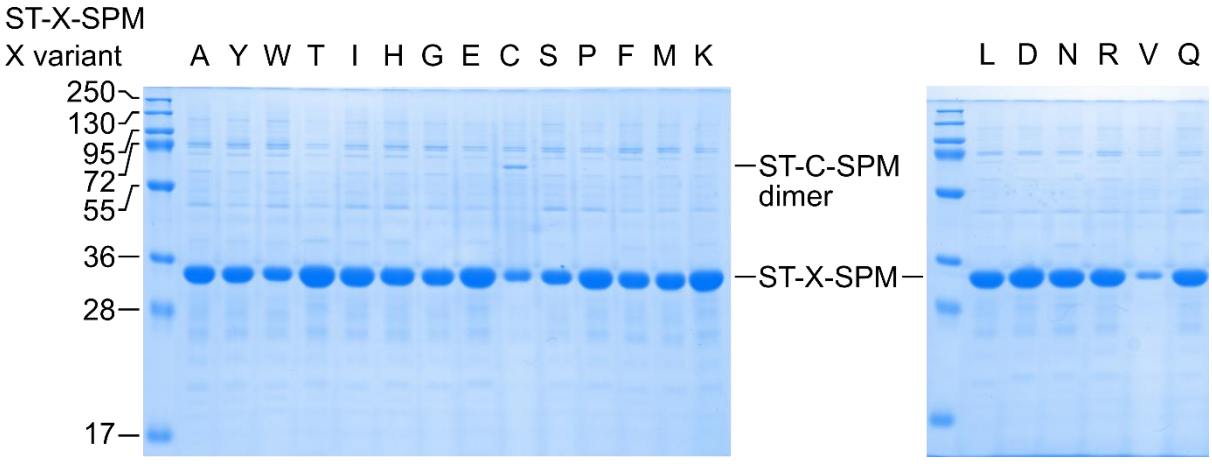


Figure 4.5. Small-scale preparations of SpyTag-X-SPM (ST-X-SPM) variants. Proteins were induced from *E. coli* in auto-induction medium and released by incubation with lysozyme and repeated freeze-thaw of cells. Proteins were eluted from Ni-Nta beads in filter plate format and buffer exchanged into HBS. A separate sample of SpyTag-V-SPM and SpyTag-A-SPM was prepared (due to low concentration and to control for batch-to-batch variation). Purification samples taken prior to buffer exchange were resolved on non-reduced SDS-PAGE and stained with Coomassie before imaging.

I then tested cleavage rate by addition of 10 mM $CaCl_2$ to 4-6 µM SpyTag-X-SPM for 5, 15, or 60 min at 37 °C in the presence of 10 mM cysteine. Cysteine served to quench side reactions (see anhydride reactivity, Figure 4.4) and reduce potential SpyTag-C-SPM adducts. Reactions were stopped by addition of EDTA in SDS loading buffer. To provide a 0 min timepoint, the stop solution was added before calcium chloride. Finally, samples were boiled and resolved on SDS-PAGE, followed by Coomassie staining.

Upon addition of calcium, all SpyTag-X-SPM variants underwent self-processing (Figure 4.6). However, there was a wide distribution in cleavage rates (relative to 0 min timepoint) (Figure 4.6c). Herein, SpyTag-G-SPM processed to <20% within an hour, whereas ST-Y-SPM reached >20% processing in 5 minutes. All aromatic residues (F, W, Y, H) showed high cleavage rate. The residue preceding the scissile Asp-Pro (DP) bond in native FrpC is Y. SpyTag-Y-SPM showed comparatively high cleavage rate at all tested timepoints (with 77%/53%/33% uncleaved protein at 5/15/60 min).

The low cleavage rate observed for SpyTag-G-SPM matched prior observations. OAZ-SPM features OAZ E219 as the residue preceding the scissile DP bond. SpyTag-E-SPM falls into the lower third of cleavage rates (with 92%/85%/56% uncleaved protein at 5/15/60 min). As SpyTag-Y-SPM featured one of the fastest cleavage rates and Y is the residue preceding SPM in native FrpC, it was chosen as a suitable residue to increase cleavage rate, i.e. to yield OAZ-Y-SPM.

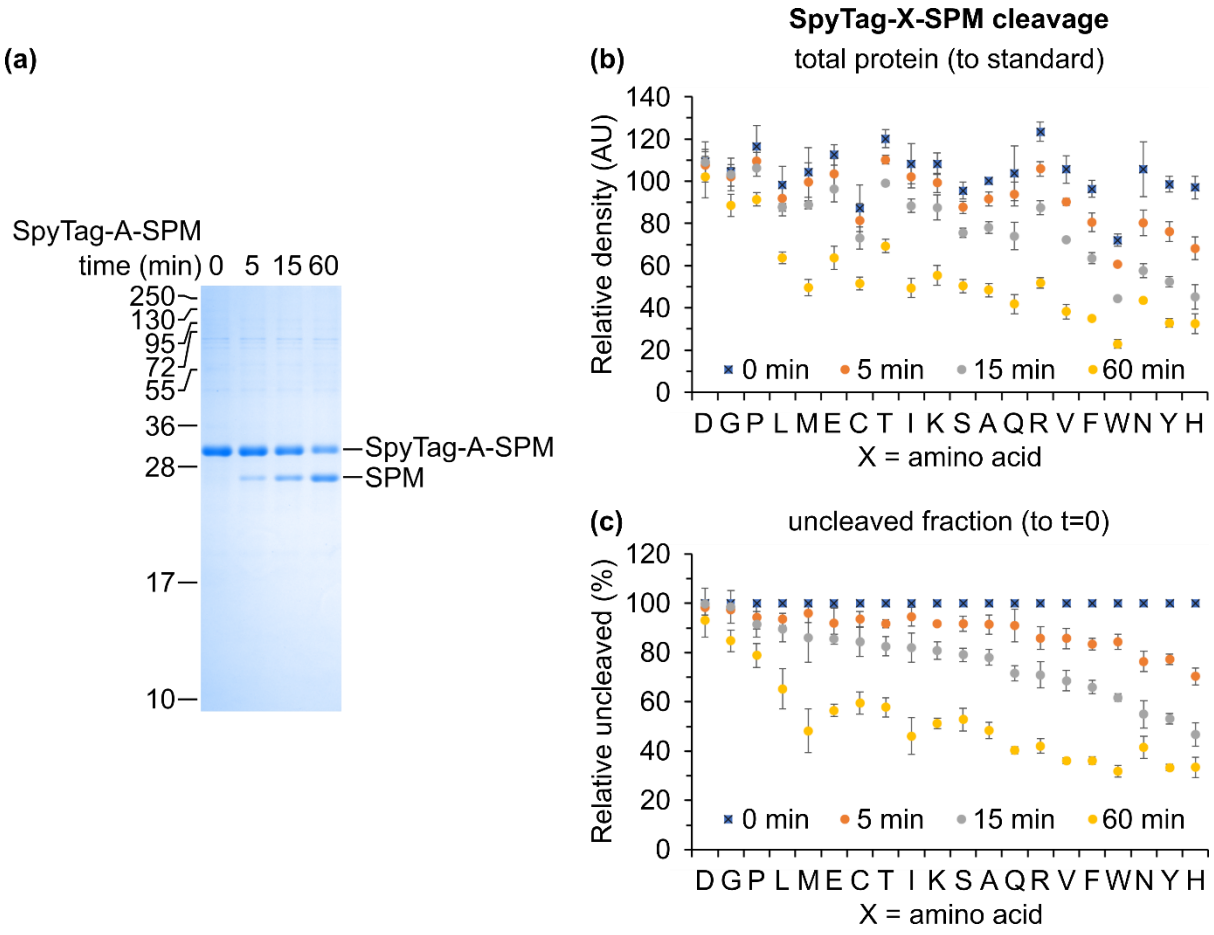


Figure 4.6. Comparison of SpyTag-X-SPM variants. (a) SDS-PAGE showing SpyTag-A-SPM as an example of processing. 4-6 µM SpyTag-X-SPM were incubated with 10 mM calcium and 10 mm cysteine at 37 °C. At the indicated timepoints, the reaction was stopped by addition of EDTA in SDS-loading buffer. Samples were boiled and resolved on SDS-PAGE before Coomassie staining. For triplicates, each replicate was resolved on a different gel (with other SpyTag-X-SPM variants). (b) Quantification of cleavage timepoints (5 min, 15 min, 60 min) and unprocessed protein (0 min) relative to protein ladder to assess variation in absolute protein concentrations. (c) Relative intensity of timepoints (0 min, 15 min, 60 min) normalised to unprocessed protein at 0 minutes shows variation in cleavage rates (mean ± s.d., n=3).

4.3.3. Purification conditions

As mentioned in 4.2, I initially purified OAZ-SPM using conditions which have previously been used in our group for SPM-fused proteins, such as Affibody-SPM (Metzner, 2017). However, for OAZ-SPM this purification procedure resulted in large amounts of precipitation upon dialysis and only yielded moderately pure protein after Ni-Nta purification. OAZ features two unpaired cysteines in its crystal structure (PDB ID 4zgy (Wu *et al.*, 2015)) which had appeared relatively obstructed but were a likely reason for protein aggregation. Notably, I wanted to avoid common reducing agent in the final buffer formulation, since the presence of free thiols in the final purification could disrupt anhydride activity (Figure 4.2, Figure 4.3).

To confirm the cause of precipitation, I used a construct encoding OAZ without fusion to SPM. Based on prior purifications of ODC and OAZ (Wu *et al.*, 2015), I adapted the purification protocol I used for OAZ-SPM to include protective reducing conditions by addition of 2-mercaptoethanol (2-ME) to purification buffers. After elution from Ni-Nta resin, I compared protein precipitation of OAZ upon dialysis into buffers with and without 2-ME. Herein, OAZ showed heavy precipitation without reducing conditions, but sustained protection of thiols in the dialysis buffer mitigated OAZ aggregation (Figure 4.7).

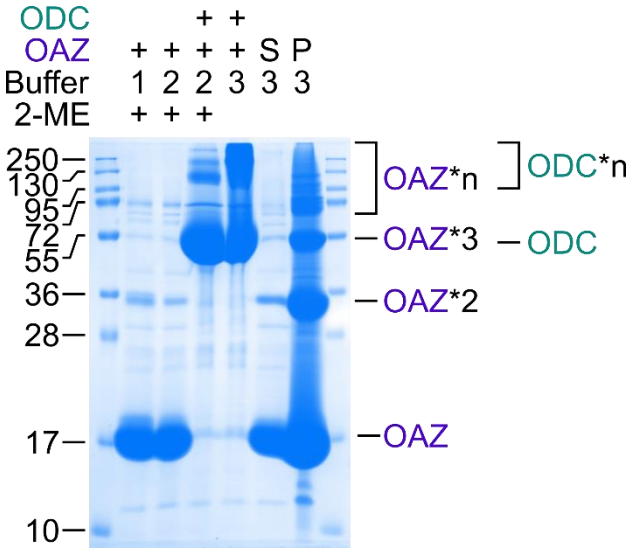


Figure 4.7. OAZ-SPM and ODC required sustained protection of thiols. OAZ and ODC samples from Ni-Nta purification were dialysed into (1) 30 mM HEPES, 250 mM NaCl, 2 mM 2-ME, ~0.02 mM pyridoxal phosphate, pH 7.4 (2) 30 mM TRIS, 250 mM NaCl, 2 mM 2-ME, ~0.02 mM pyridoxal phosphate, pH 7.4 (3) 50 mM HEPES, 150 mM NaCl, pH 7.4. Heavy aggregation was observed if samples were not protected from oxidation. Optionally, OAZ-SPM was spun at 16,900 rcf to separate aggregate from soluble protein and the aggregate was resuspended in water (S: Supernatant, P: Pellet). Samples were boiled with SDS-loading buffer and resolved on SDS-PAGE with Coomassie staining.

After confirming conditions under which OAZ was not prone to aggregation, I decided to use size exclusion chromatography for purification of OAZ-Y-SPM. Since free thiols would quench anhydride activity and NeissLock conjugation, I substituted 2-ME or the non-nucleophilic Tris-(2-carboxyethyl)-phosphine (TCEP) as the reducing agent for gel filtration. Furthermore, for subsequent purifications samples were applied to size exclusion chromatography without prior dialysis, reducing the amount of HEPES and TCEP reagents that would otherwise be required. With these conditions, both ODC and OAZ-Y-SPM could be prepared at high purity and yield (Figure 4.8).

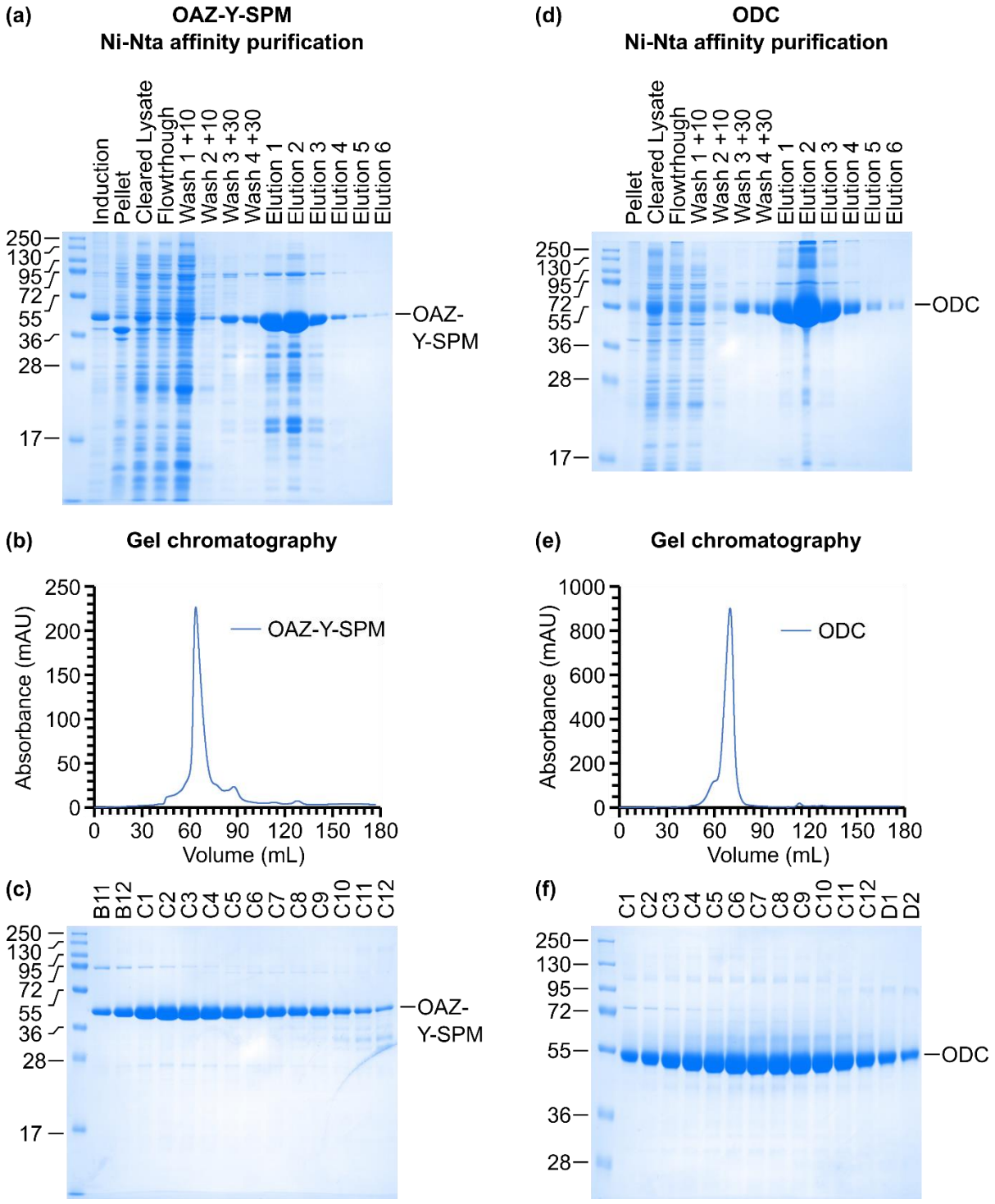


Figure 4.8. Optimized purification of ODC and OAZ-Y-SPM. (a) Ni-Nta affinity purification of OAZ-Y-SPM. OAZ-Y-SPM was solubly expressed from BL21 RIPL. After induction of expression for 16-18h at 25°C, cells were harvested, lysed and the lysate clarified by centrifugation. Cleared lysate was applied to Ni-Nta resin, washed and crudely purified OAZ-Y-SPM was eluted with 200 mM imidazole. All buffers were supplemented with reducing agent. (b) After spin concentration, the combined elution sample was directly injected into a Superdex S200 16-600 gel filtration column. Gel filtration buffer was 50 mM HEPES, 150 mM NaCl, 2 mM TCEP, pH 7.40. Shown is absorbance at 280 nM (AU) relative to run volume (mL). (c) Fractionation samples from (b) were resolved on SDS-PAGE after boiling with SDS-loading buffer. Gel was stained with Coomassie before imaging. (d,e,f) Purification procedure for ODC. For ODC, gel filtration buffer was 50 mM HEPES, 150 mM NaCl, 2 mM TCEP, 0.02 mM pyridoxal phosphate, pH 7.40.

I then reattempted conjugation of OAZ-Y-SPM to ODC. I incubated 10 µM OAZ-Y-SPM with 30 µM ODC and 10 mM calcium for 16 h at 37°C, before quenching the reaction with EDTA. Samples were boiled with SDS-loading buffer and resolving them on SDS-PAGE. Herein, conjugation of OAZ-Y-SPM to ODC was readily apparent (Figure 4.9). At a 3:1 ratio of ODC to OAZ-Y-SPM I now observed nearly quantitative conjugation of activated OAZ to ODC, with little release of free OAZ (Figure 4.9).

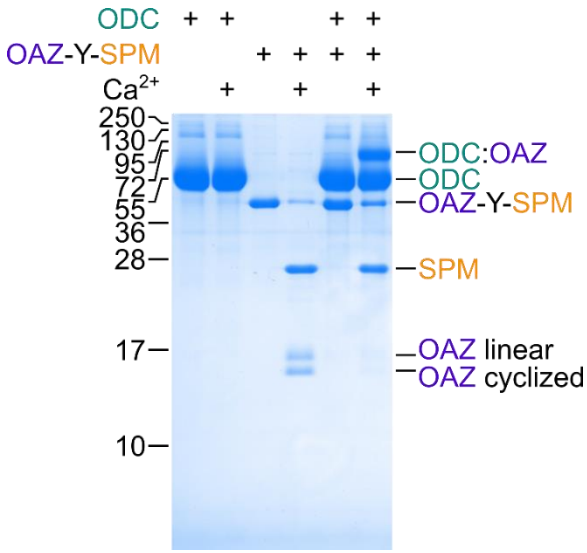


Figure 4.9. Covalent conjugation of OAZ-Y-SPM to ODC after improvement of purification conditions. 10 µM OAZ-Y-SPM was incubated at 37 °C for 16 h with or without 10 mM calcium and 30 µM ODC (at a 3:1 ratio) as indicated. Samples were resolved on reducing SDS-PAGE before staining with Coomassie and imaging.

4.3.4. Spacer insertion and steric constraints

Throughout OAZ-SPM and OAZ-Y-SPM conjugation experiments to ODC, it commonly appeared as though more OAZ-Y-SPM remained unprocessed in the presence of ODC, and therefore that the presence of ODC decreased the OAZ-(Y)-SPM cleavage rate (Figure 4.9). Therefore, I suspected that assembly of the ODC/OAZ complex might cause steric hindrance to SPM cleavage. To test this hypothesis, I incubated 10 µM OAZ-Y-SPM with increasing concentrations of ODC (2.25-30 µM) and 10 mM calcium for 18.5 h at 37°C. In all conditions, OAZ-Y-SPM showed visible processing. However, it appeared that cleavage of OAZ-Y-SPM was decreased at increased concentrations of ODC (Figure 4.10).

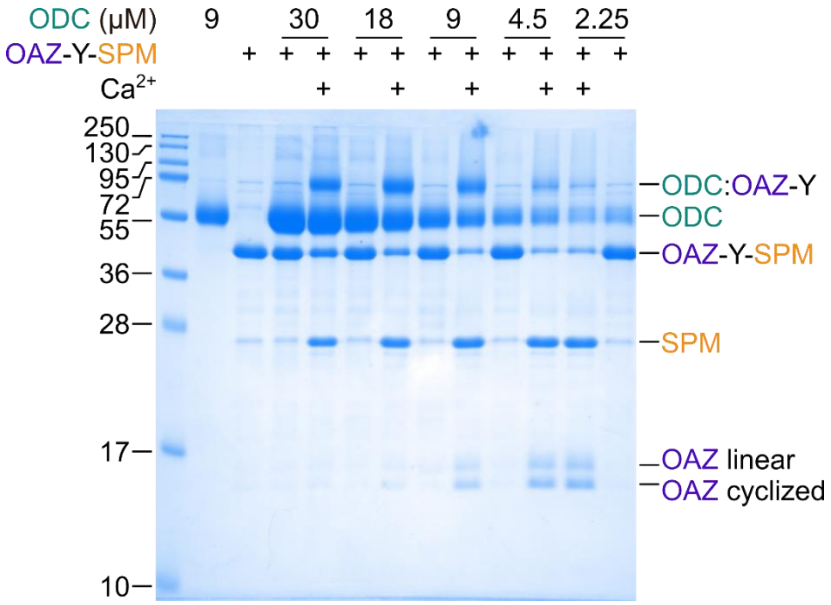


Figure 4.10. ODC inhibits OAZ-Y-SPM processing. 10 μM OAZ-Y-SPM was incubated for 18.5 h at 37°C with 10 mM calcium and with increasing concentrations of ODC (2.25 μM – 30 μM) as well as 10 mM calcium as indicated, in 50 mM HEPES, 150 mM NaCl, 1 mM TCEP, pH 7.40. Samples were boiled with SDS-loading buffer and resolved on SDS-PAGE. With higher concentrations of ODC, more unprocessed OAZ-Y-SPM was observed.

Furthermore, OAZ-(Y)-SPM displayed markedly reduced cleavage rate compared to other SPM constructs even if no ODC was present (compare to Figure 4.4, Figure 4.6, and other work (Osička *et al.*, 2004; Sadilkova *et al.*, 2008; Metzner, 2017)). Steric hindrance was previously suggested as a factor in SPM autoproteolysis rate (Sadilkova *et al.*, 2008), i.e. direct OAZ fusion could disrupt SPM folding and processing. I inserted an additional GS spacer into the OAZ-Y-SPM construct to make OAZ-GSY-SPM and purified OAZ-GSY-SPM in parallel to OAZ-Y-SPM as described in Figure 4.8.

10 μM OAZ-Y-SPM or 10 μM OAZ-GSY-SPM were incubated with 10 μM ODC and 10 mM calcium at 37°C, and the reaction was allowed to proceed for increasing periods of time. In the presence of an equimolar ratio of ODC, OAZ-GSY-SPM processed faster than OAZ-Y-SPM, reaching ~50% autoproteolysis after 2-5 minutes, as opposed to 30-60 minutes for OAZ-Y-SPM (Figure 4.11).

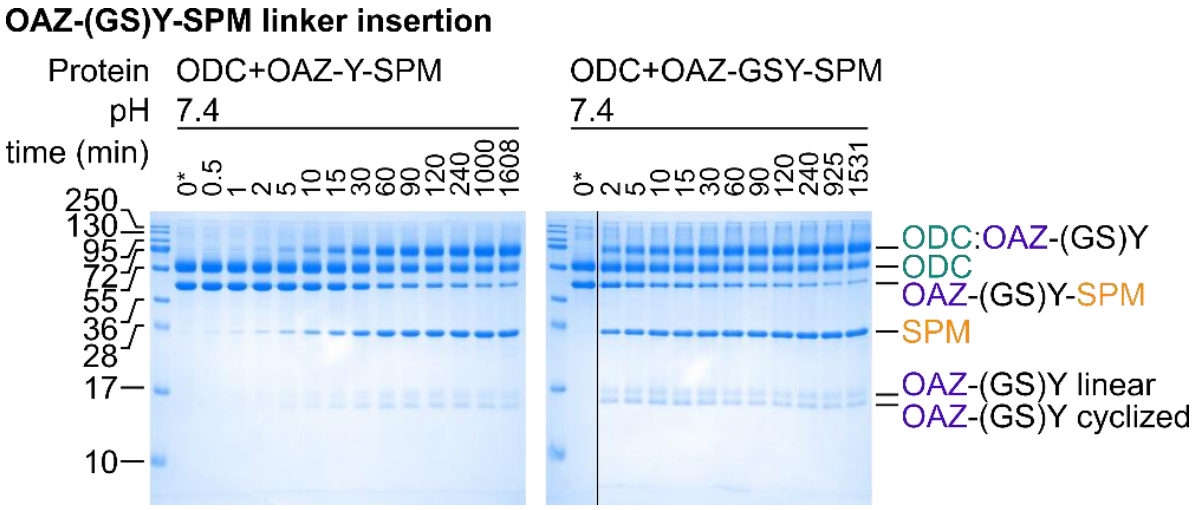


Figure 4.11. GS linker insertion increase OAZ-Y-SPM cleavage rate. 10 µM OAZ-Y-SPM or 10 µM OAZ-GSY-SPM were diluted together with 10 µM ODC in 50 mM HEPES, 150 mM NaCl, 2 mM TCEP, pH 7.4. Samples were incubated with 10 mM $CaCl_2$ at 37 °C. After the indicated time, the reaction was stopped by addition of EDTA in SDS-loading buffer. 0*: EDTA in SDS-loading buffer was added before $CaCl_2$ to provide a timepoint 0. Samples were resolved on SDS-PAGE before Coomassie-staining and imaging. Black line: Gel was digitally cut (see methods).

Taken together, these observations suggest that steric effects can interfere with NeissLock-probe processing, through steric hindrance of SPM-processing due to formation of the binder/target OAZ/ODC complex or due to direct fusion of SPM to OAZ. Steric hindrance could be alleviated for OAZ-Y-SPM through insertion of a GS spacer.

4.3.5. pH-dependence

With an average pK_a of 10.7 in proteins (Pahari *et al.*, 2019), lysine residues are generally protonated at physiological pH of 7.4. Nevertheless, I observed high amounts of conjugation at pH 7.4 in HEPES-buffered saline. Upon co-incubation of *N. meningitidis* with A549 human cells, Sviridova et al. observed evidence for conjugation of N-terminal FrpC to plasma membrane samples of A549 cells (Sviridova *et al.*, 2017), mimicking physiological conditions, and Osička et al. previously showed SPM cleavage to be effective around pH 6-8 (Osička *et al.*, 2004). Similarly, I aimed to investigate the impact of pH on both cleavage rate and conjugation efficiency to determine optimized conjugation conditions.

I therefore prepared 50 mM HEPES, 150 mM NaCl, 2 mM TCEP buffers adjusted to varying pH values (pH 7.4, pH 8.0 or pH 8.4). In parallel to testing reaction of OAZ-Y-SPM or OAZ-GSY-SPM with ODC (Figure 4.11), I also tested reaction rate of OAZ-Y-SPM and ODC diluted

in buffer with pH 8.0 or pH 8.4, as well as reaction of OAZ-Y-SPM without ODC (Figure 4.12, OAZ-Y-SPM with ODC at pH 7.4 also in Figure 4.11).

For OAZ-Y-SPM, I observed decreased cleavage rate with increasing pH, with a strong decrease at pH 8.4, both with and without ODC. With ODC, conjugation rate showed an inverse correlation to pH, in parallel to decreased cleavage rate. Adjusted for amount of cleaved protein, overall conjugation efficiency was apparently unaffected by pH (maximum ~1:1). Therefore, I concluded that the apparent dependence of conjugation rate on pH was mainly driven by the effect of pH on cleavage rate.

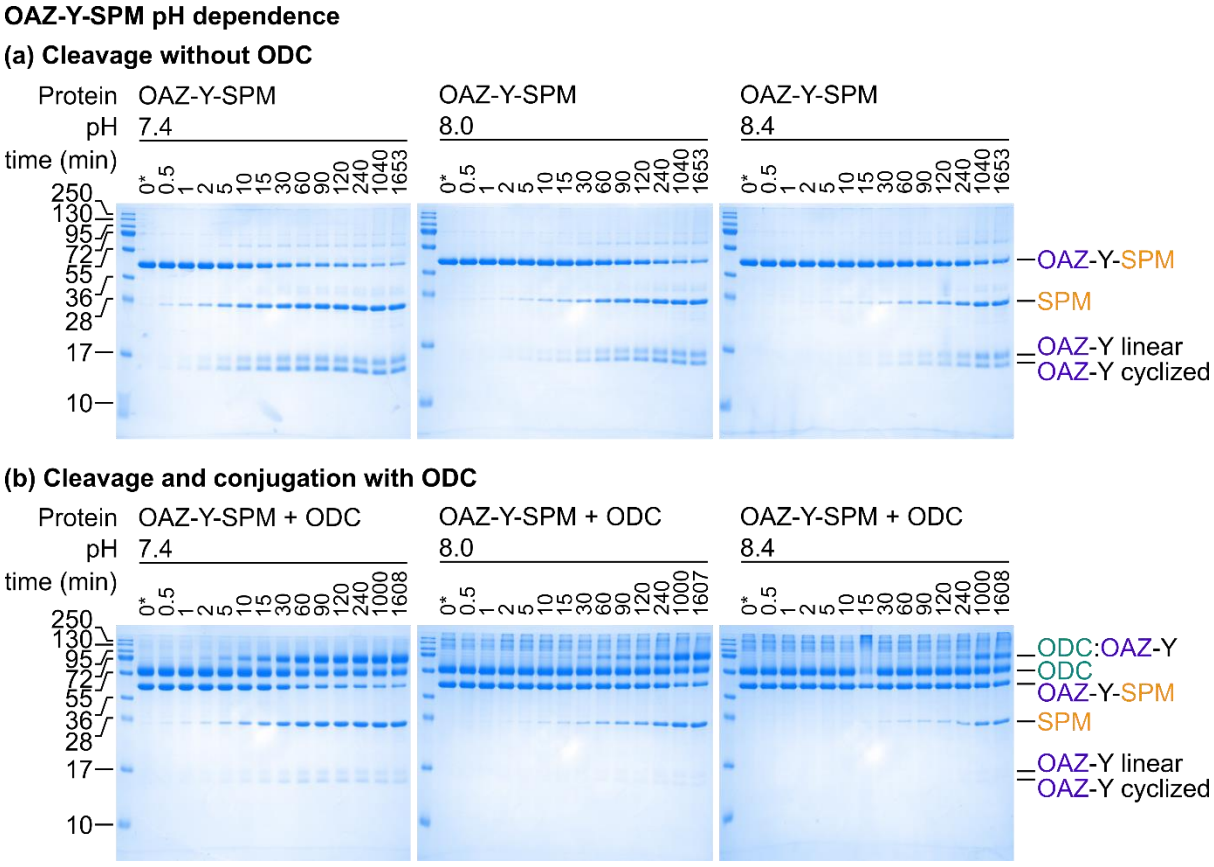


Figure 4.12. Screen of conditions affecting NeissLock conjugation to ODC. General reaction conditions as in Figure 4.11. 10 µM OAZ-Y-SPM and 10 µM ODC as indicated were diluted in 50 mM HEPES, 150 mM NaCl, 2 mM TCEP adjusted to the pH 7.4, pH 8.0 or pH 8.4. Samples were incubated with 10 mM calcium (diluted in the corresponding buffer) at 37 °C. After the indicated time before the reaction was stopped by addition of EDTA in SDS-loading buffer. 0*: EDTA in SDS-loading buffer was added before $CaCl_2$ to provide a timepoint 0. Samples were resolved on SDS-PAGE before Coomassie-staining and imaging. (a) Cleavage of OAZ-Y-SPM without ODC. (b) Cleavage and conjugation with ODC. Cleavage and conjugation rate steadily decreased from pH 7.4 to 8.4. OAZ-Y-SPM + ODC (in b) also shown in Figure 4.11.

With pKa_1 of 3 and pKa_2 of 7.5, pH 8.4 is at the end of the buffering range for HEPES. Based on my initial observations, Sheryl Lim in the same research group conducted replication experiments at a larger pH scale by including MES in the buffer formulation (150 mM NaCl, 50 mM HEPES, 50 mM MES, 2 mM TCEP), using OAZ-GSY-SPM as the construct (Figure 4.13b,c, adapted from Scheu *et al.* (Scheu *et al.*, 2021)). In a range from 6.5 to 8.5, the cleavage rate of OAZ-GSY-SPM gradually decreased with increasing pH, with a distinct reduction from pH 8.0 to pH 8.5. These observations agree with my initial observations on cleavage rate, and similarly compare to previous reports (Osička *et al.*, 2004).

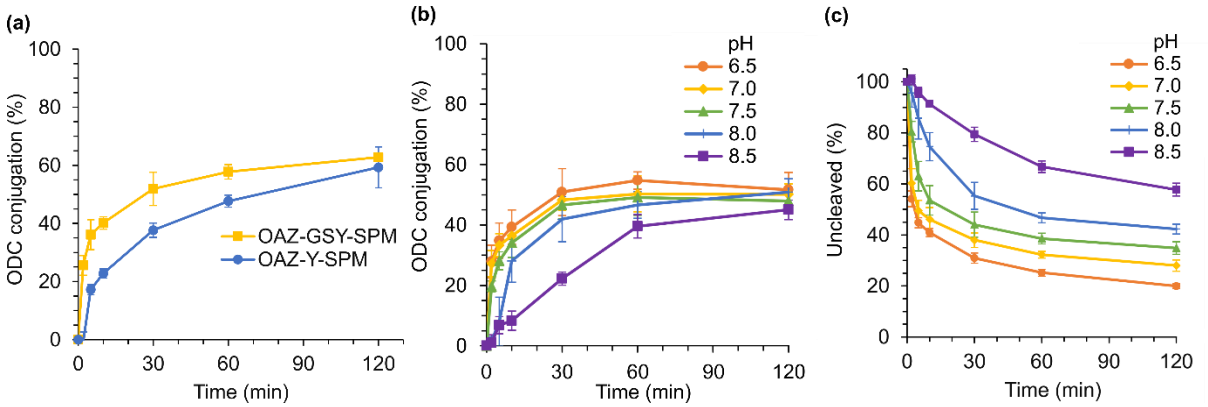


Figure 4.13. Quantitative analysis of NeissLock-conjugation to ODC. (a) Comparison of conjugation rate for 10 µM OAZ-Y-SPM or 10 µM OAZ-GSY-SPM incubated with 10 µM ODC in the presence of 10 mM calcium in 50 mM HEPES, 150 mM NaCl, 2 mM TCEP, pH 7.40 (b) Conjugation rate of 10 µM OAZ-GSY-SPM to 10 µM ODC with 10 mM CaCl2 in 150 mM NaCl, 50 mM HEPES, 50 mM MES, 2 mM TCEP dependent on buffer pH. (c) As in (b), but for cleavage rate. Quantification from Coomassie-stained SDS-PAGE. Error bars: standard deviation, n=3. Adapted from Sheryl Lim, with permission (Scheu *et al.*, 2021).

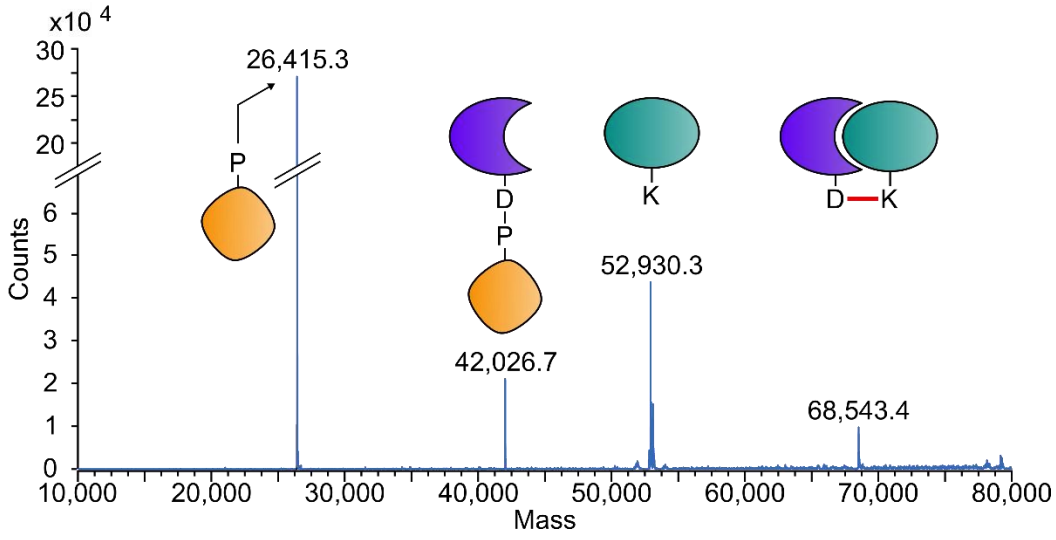
4.4. Intact mass spectrometry of OAZ:ODC conjugation

To test the chemical identity of the crosslink, I assessed the conjugation of OAZ-Y-SPM and OAZ-GSY-SPM to ODC via intact mass spectrometry (MS). 10 µM of OAZ-Y-SPM or OAZ-GSY-SPM were mixed with 10 mM $CaCl_2$ and 30 µM ODC and the reaction was allowed to proceed for 16 h at 37 °C before stopping SPM processing by addition of EDTA. Subsequently, samples were acidified with formic acid and injected onto electrospray ionization mass spectrometry (ESI-MS) using RapidFire automated sample processing (RF-MS).

Theoretical masses were calculated using ExPASY ProtParam according to the corresponding linear protein sequences without an initiating methionine, i.e. for full-length ODC, full-length OAZ-Y-SPM as well as OAZ-YD *(OAZ-Y retaining Asp from DP cleavage, usually excluded from nomenclature herein)* and P-SPM *(SPM retaining Pro from DP cleavage, usually excluded from nomenclature herein)* protein fragments. For the conjugate ODC:OAZ-Y, the mass was estimated as the combined mass of ODC and OAZ-YD, with loss of water.

After processing of OAZ-Y-SPM in the presence of ODC, I observed masses corresponding to ODC (observed 52,930.3 Da, predicted 52,929.9 Da), OAZ-Y-SPM (observed 42,026.7 Da, predicted 42,024.7 Da) and the conjugation product ODC:OAZ-Y (observed 68,543.4 Da, predicted 68,539.4 Da) as well as SPM (observed 26,415.3 Da, predicted 26,415.1 Da) (Figure 4.14a). The conjugation product ODC:OAZ-Y featured a loss of ~18.0 Da compared to the calculated mass for ODC+OAZ-Y, as would be expected for loss of water. For OAZ-GSY-SPM, all masses were increased by the corresponding amount (GS: 144.13 Da) compared to OAZ-Y-SPM (Figure 4.14b).

a)



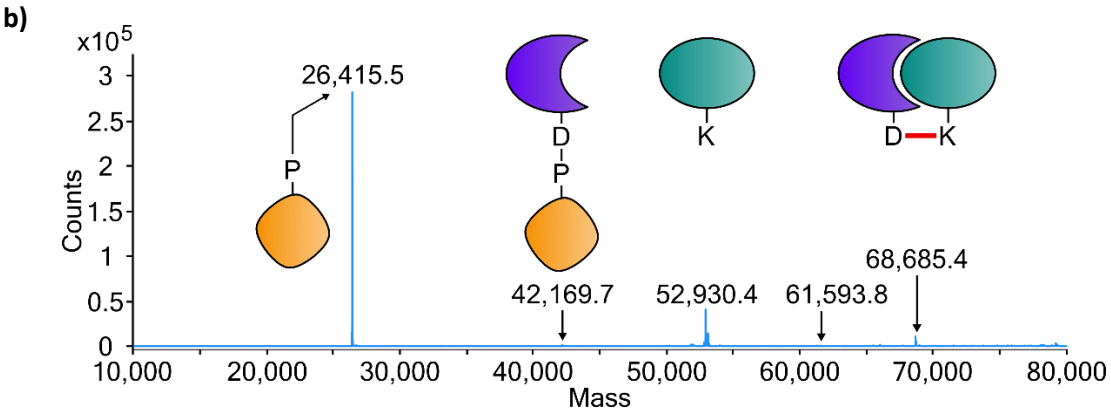


Figure 4.14. RF-MS of OAZ-Y-SPM and OAZ-GSY-SPM after conjugation with ODC. OAZ-Y-SPM or OAZ-GSY-SPM were incubated with calcium in the presence of ODC. After stopping self-processing with EDTA, samples were injected into electrospray ionization mass spectrometry. (a) The covalent adduct of ODC and OAZ-Y was observed with loss of water (predicted: 68,539.4 kDa), as were ODC (predicted 52,929.9 Da), OAZ-Y-SPM (predicted 42,024.7 Da), and SPM (predicted 26,415.1 Da) (b) The corresponding conjugate for OAZ-GSY-SPM showed an increase in mass as expected for insertion of the GS spacer (GS predicted: +144.13 Da; ODC:OAZ-GSYD predicted: 68,683.6 Da), as did OAZ-GSY-SPM (predicted 42,168.8) compared to OAZ-Y-SPM. Samples from these reactions were also resolved on SDS-PAGE as Figure 4.9 after denaturation with SDS-loading buffer.

Free SPM gave a strong signal compared to other species, making it difficult to assess other proteins (Figure 4.14). I utilized liquid chromatography - mass spectrometry (LC-MS) to gain more insight into all product species. 20 µM OAZ-Y-SPM with 20 µM ODC and 10 mM $CaCl_2$ were incubated at 37°C for 1 h before the reaction was stopped by addition of EDTA. The sample was diluted with water and passed onto LC for separation after which it was injected into MS. Herein, the same MS system was used for LC-MS as for RF-MS. On the LC-MS chromatogram, the sample separated into broad peaks (Figure 4.15a). As for RF-MS, SPM gave a strong signal. However, the separation in LC allowed for clear distinction of the expected protein species. Upon deconvolution of spectra obtained at peak positions (2.993 min / 3.970 min / 5.792 min), I identified species corresponding to predominantly free OAZ-Y (2.993 min), SPM (3.970 min), and a mix of ODC and ODC:OAZ-Y (5.792 min) (Figure 4.15b).

Using LC-MS, unconjugated OAZ was readily detected in the presence of ODC (Figure 4.15b). The predicted mass for the linear fragment of OAZ-Y is 15627.56. However, this mass was only visible upon closer inspection of the deconvoluted spectrum (at 15,628.24 Da, Figure 4.15c). In addition to the mass for linear OAZ-Y, a stronger neighbouring peak was observed

at 15,609.59, corresponding to a dehydrated form (-18 Da). As outlined in Chapter 4.2 and Chapter 4.3 (Figure 4.3, Figure 4.9), released OAZ resolves as two distinct species on SDS-PAGE. Thiolysis of the anhydride (Figure 4.3) suggested that one band corresponds to a linear (e.g. hydrolysed or thiolysed) species of OAZ and one to a cyclized species of OAZ (e.g. C_t to Nα or C_t to Nε). Therefore, the observations of LC-MS and SDS-PAGE both match this hypothesis.

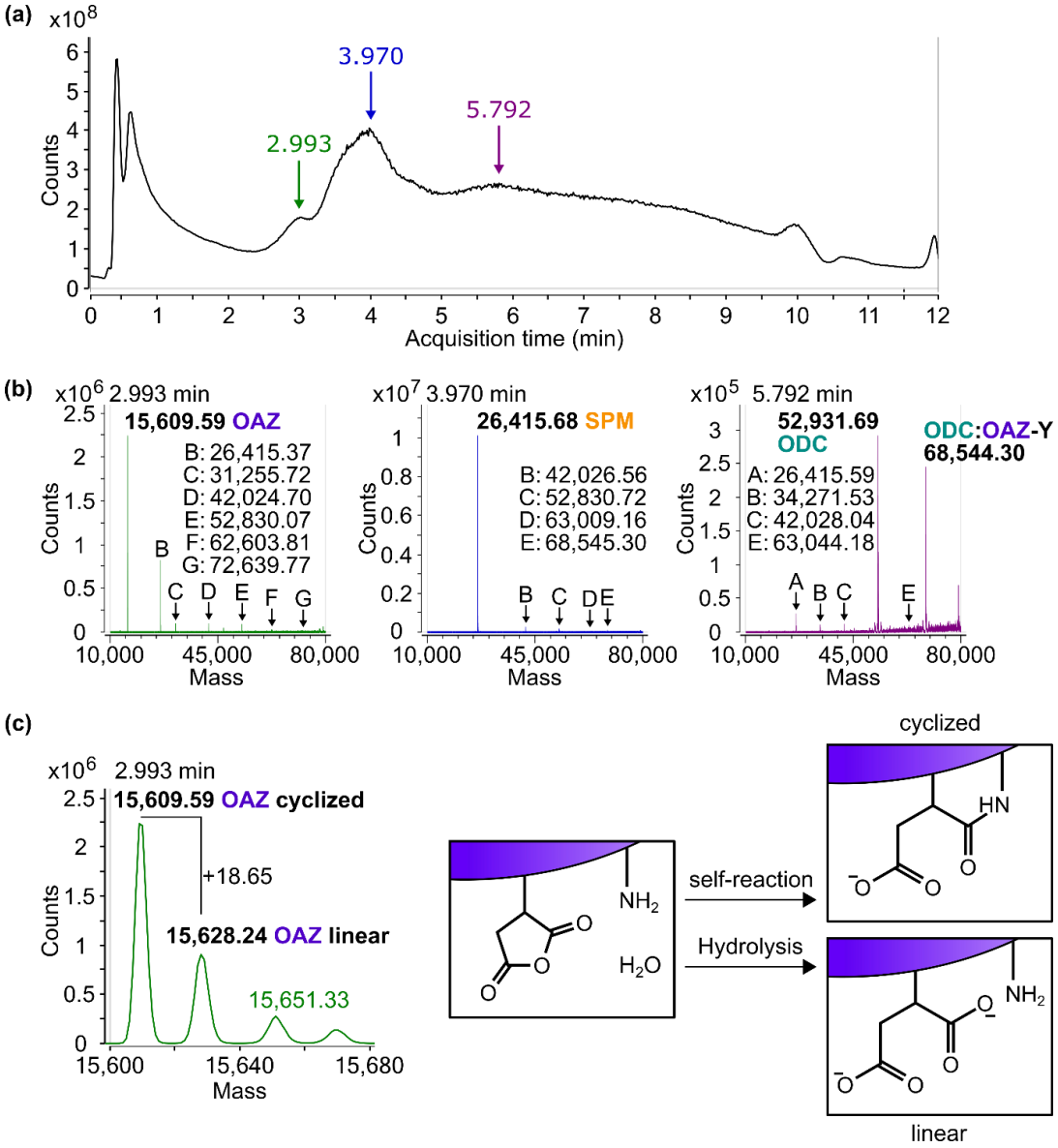


Figure 4.15. Liquid chromatography with coupled mass spectrometry of OAZ-Y-SPM conjugated to ODC. (a) liquid chromatography profile after incubation of OAZ-Y-SPM with calcium in the presence of ODC, showing separation of the major protein species. (b) Coupled MS spectrum corresponding to peak positions in (a), showing individual masses for identified proteins (compare to Figure 4.14). (c) Detail view of coupled MS spectrum corresponding to liquid chromatography at 2.993 min. Closeup around 15,640 Da, showing masses corresponding to self-reacted and hydrolysed (+18 Da) OAZ-Y-SPM cleavage products (OAZ-YD) and cartoon thereof.

4.5. Mapping of OAZ:ODC crosslinking sites

4.5.1. Mutation of the target site

After validating chemical conjugation of ODC:OAZ-Y and ODC:OAZ-GSY via SDS-PAGE and intact MS, I aimed to identify the position(s) at which crosslink occurs. First, I mutated K92 in ODC, which I hypothesized to be the main target residue based on its proximity to the C-terminus of OAZ (3.5 Å in 4zgy, i.e. its selection criterion as the NeissLock model complex). K92R was chosen as a mutation to retain a basic residue while removing nucleophilicity. After induction of processing, OAZ-Y-SPM or OAZ-GSY-SPM retained high amounts of conjugation to ODC K92R (Figure 4.16a,b). However, a long runtime at high gel density (1:40 h, 18%, 180 V) resolved two distinct bands for ODC K92R:OAZ-Y or ODC K92R:OAZ-GSY conjugation products, indicating at least two distinct conjugate species. For OAZ-GSY-SPM, a second band could already be resolved for conjugation to wt ODC. These observations suggest that K92 is a target site, but alternative residues can be conjugated.

4.5.2. Introduction of alternative sites

To further explore the spatial requirements for crosslinking, I attempted to rescue wild type-like conjugation of OAZ-Y-SPM or OAZ-GSY-SPM to ODC K92R by reintroducing K residues in proximity to the mutation site K92. Along the α-helix on which K92 is positioned, I introduced T93K, Q96K or S100K into the ODC K92R background, to act as a 'ruler' facing away from K92 (Figure 4.16c). Measured from atom C of OAZ E219 to atom CA (backbone) of the target residue, the distances were 7.6 Å to K92, 8.8 Å to T93, 12.0 Å to Q96, and 17.5 Å to S100 (Figure 4.16c). Furthermore, I tested ODC K92R T396K (with 14.6 Å from E219 atom C to T396 atom CA), as I hypothesized this region to be accessible for the C-terminus of OAZ based on the OAZ NMR structure PDB ID 1zo0 [DOI 10.2210/pdb1ZO0/pdb, literature unpublished].

I observed formation of a predominant product band for conjugation of OAZ-Y-SPM or OAZ-GSY-SPM to ODC K92R Q96K (Figure 4.16a,b), similar to the respective wt ODC conjugate (Figure 4.16a,b, compare to Figure 4.17a). ODC K92R T93K or ODC K92R S100K

did not show such a banding pattern, but instead appeared like conjugation to ODC K92R with two strong product bands (Figure 4.16a,b, compare to Figure 4.17).

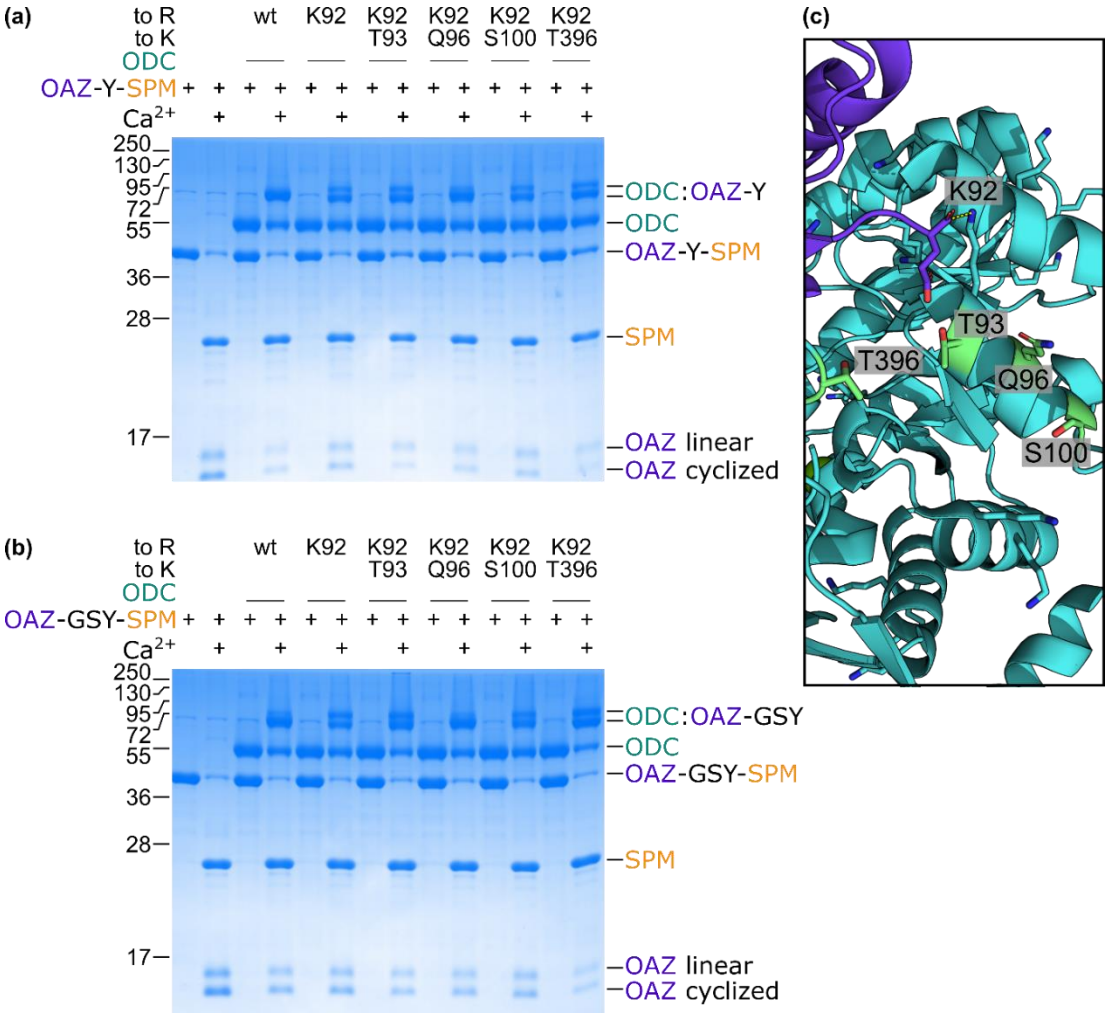


Figure 4.16. Mutation of the target residue K92 and introduction of alternate target sites. ODC K92 is a target site, but OAZ-(GS)Y-SPM can conjugate to alternate target sites. (a) 10 µM OAZ-Y-SPM was incubated at 37°C with 10 µM wt ODC, ODC K92R, or mutants of ODC K92R with introduction of an additional lysine in proximity to test specificity around the target site. Samples were incubated in the presence of 10 mM calcium as indicated. Reactions were stopped with EDTA in SDS-loading buffer. Samples were boiled and resolved on SDS-PAGE before Coomassie staining. Conjugation to ODC K92R causes distinct double banding, which is not apparent for ODC K92R Q96 (b) As (a), with OAZ-GSY-SPM. (c) Cartoon of positions which were mutated on ODC (PDB ID 4zgy (Wu *et al.*, 2015)).

I only observed the distinct banding pattern for ODC K92R conjugates concurrent to setting up this analysis of double mutants. As overall yield did not appear affected in ODC K92R, yield would likely be a weak indicator of wild-type like conjugation. However, a covalent adduct at a site close to K92 on the primary sequence (i.e. T93K/Q96K/S100K) comparably to conjugates at K92 in (denaturing) SDS-PAGE. Therefore, I concluded that ODC K92R Q96K

but not ODC K92R T93K or ODC K92R S100K rescued wt-like conjugation from ODC K92R. Similarly, conjugation at ODC K92R T396K appeared like conjugation at ODC K92R, although it is unclear if conjugation at T396K would rescue the original banding pattern. In summary, conjugation in ODC K92R Q96K confirmed that conjugation of ODC K92R can be rescued by introduction of nearby K residues, while estimated proximity is not the only parameter in determining conjugation behaviour.

4.5.3. Enzymatic digest and LC-MS/MS

I utilized tryptic liquid chromatography tandem mass spec (LC-MS/MS) of conjugation products to identify crosslinking sites. After conjugation of OAZ-Y-SPM with ODC wt (Figure 4.17a), I submitted the single conjugated product band for further processing with the Biochemistry proteomics team. After conjugation of OAZ-Y-SPM to ODC K92R, I separated the two distinct bands with a blade and submitted them as independent conjugation products. Data were processed by Dr. Shabaz Mohammed. For conjugation of OAZ-Y-SPM to wt ODC, adduct peptides corresponding to conjugation of OAZ-YD (including the C-terminal aspartate) and ODC wt at K92 were detected (Figure 4.17c). No other crosslinking site was confirmed in this sample. For conjugation of OAZ-Y-SPM to ODC K92R, K121 was identified as a crosslinking site in the higher running band (Figure 4.17d). Compared to K92, K121 is positioned on the neighbouring α-helix, with 14.2 Å from E219 C to K121 Nε (3.5 Å for K92) or 11.0 Å E219 C to K121 CA (7.6 Å for K92) (Figure 4.17b). No crosslinking site was confirmed for the lower running band.

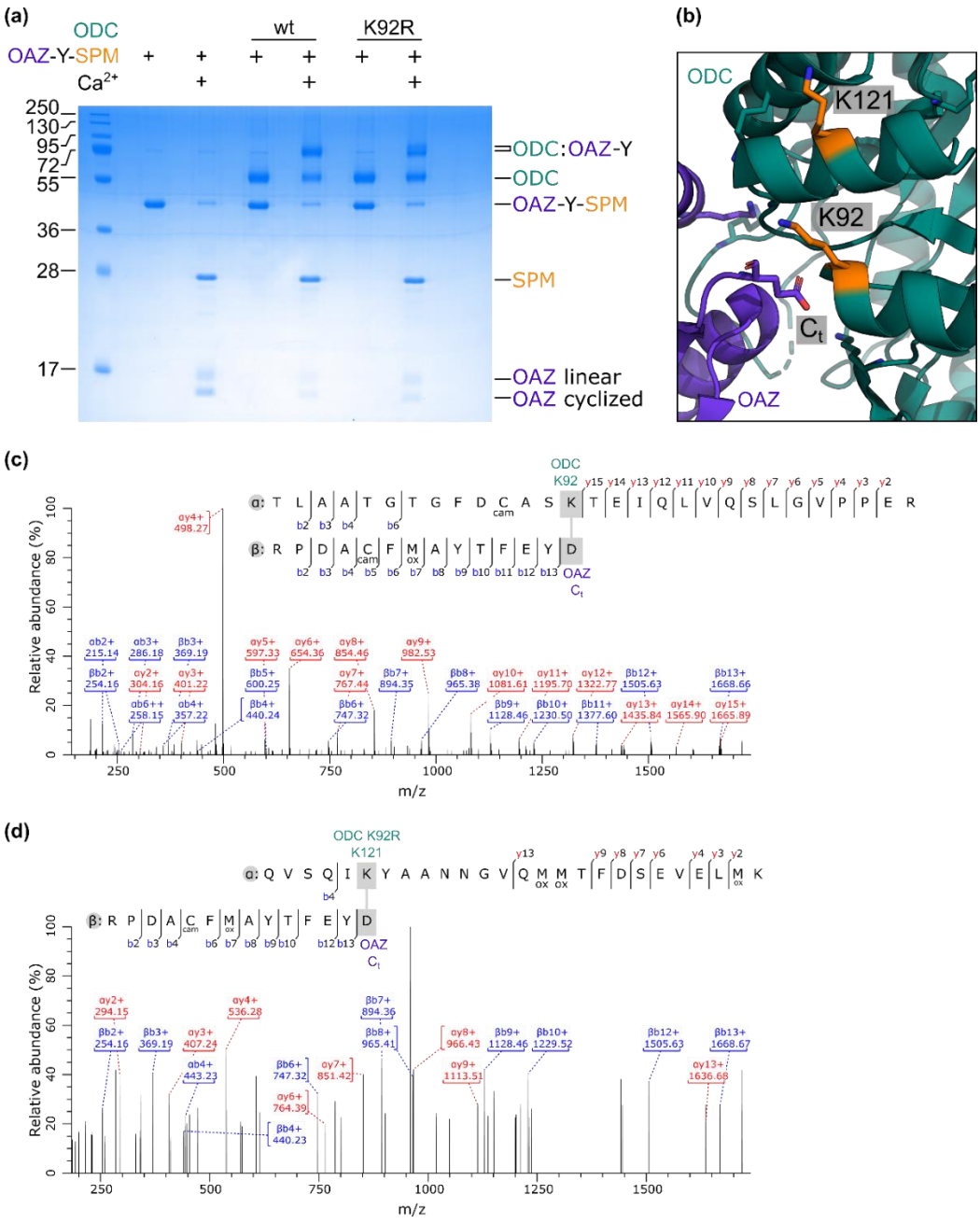


Figure 4.17. Tryptic digestion and liquid chromatography with tandem mass spectrometry (tryptic LC-MS/MS) of OAZ-Y-SPM conjugation to ODC or ODC K92R. (a) The product of OAZ-Y-SPM conjugation to ODC K92R can be resolved into two distinct bands, from which samples were cut for submission to LC-MS/MS. (b) ODC K92 annotated on the crystal structure PDB ID 4zgy (Wu *et al.*, 2015), as well as the nearby residue K121. (c) LC-MS/MS spectrum as well as annotated structure for crosslinked peptide confirming crosslinking of the OAZ-Y-SPM C-terminus post-cleavage (OAZ-YD) to wt ODC at ODC K92. (d) LC-MS/MS spectrum and annotated structure for crosslinked peptide confirming crosslinking of the OAZ-Y-SPM C-terminus post-cleavage (OAZ-YD) to ODC K92R at ODC K121. No crosslinking to K92 was detected for conjugation to ODC K92R, and no crosslinking to K121 was detected for conjugation to wt ODC.

4.6. Assessment of specificity

4.6.1. Design of binding mutants

To gain insight into the required binding affinity to facilitate conjugation, I next introduced mutations into OAZ affecting OAZ binding to ODC. Cohavi *et al.* previously screened a series of alanine scan point mutants of mouse OAZ for disrupted binding to ODC (Cohavi *et al.*, 2009), with the most notable disruption compared to wt (k_{off}: $2*10^{-4}$/s; K_D: 0.1 nM) in K153A (k_{off}: $4.5*10^{-4}$/s; K_D: 0.88 nM, K_A 12% of wt), V198A (k_{off}: $6*10^{-4}$/s; K_D: 1.05 nM, K_A 10% of wt) and F213A (k_{off}: $4.4*10^{-4}$/s; K_D: 0.92 nM, K_A 11% of wt) (Cohavi *et al.*, 2009). The PDB ID 4zgy (Wu *et al.*, 2015) crystal structure contains the human OAZ/ODC complex, which was used in this study. Judging by this structure, F213 does not directly interface with ODC residues, whereas K153 is involved in a polar and V198 in an apolar binding pocket (Figure 4.18).

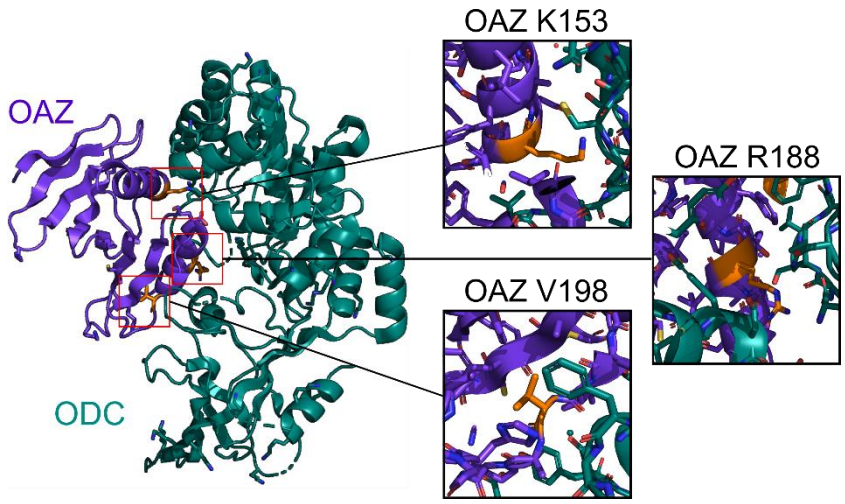


Figure 4.18. Residues at the OAZ and ODC interface which were mutated to reduce binding affinity. Three distinct binding pockets were chosen for mutation along the ODC/OAZ interface. For mouse orthologues of ODC/OAZ, OAZ K153 and OAZ V198 have previously been identified as critical to the interaction (Cohavi *et al.*, 2009). According to the crystal structure of human ODC/OAZ (PDB ID 4zgy (Wu *et al.*, 2015)), OAZ R188 appeared to interact with ODC in an extensive polar interface which could be disrupted by charge inversion.

To avoid potential perturbations of OAZ structure, I opted for mutation of residues directly in contact with ODC, i.e. K153A and V198A were chosen for mutation. Furthermore, I introduced A215R into OAZ, substituting a small to a bulky residue at the interface to ODC F397. As V198 and A215 are in proximity at the same binding pocket, I decided not to simultaneously mutate

V198A and A215R to avoid unpredictable disruption of OAZ. I therefore cloned OAZ variants including K153A or K153A/V198A or K153A/A215R.

These constructs were based on OAZ-SPM and prepared prior to optimization of purification conditions (subject to partial precipitation upon dialysis, without protection of thiols by reducing agent). Upon induction of SPM-processing with calcium, I observed conjugation for all tested OAZ-SPM variants in the presence of ODC (Figure 4.19).

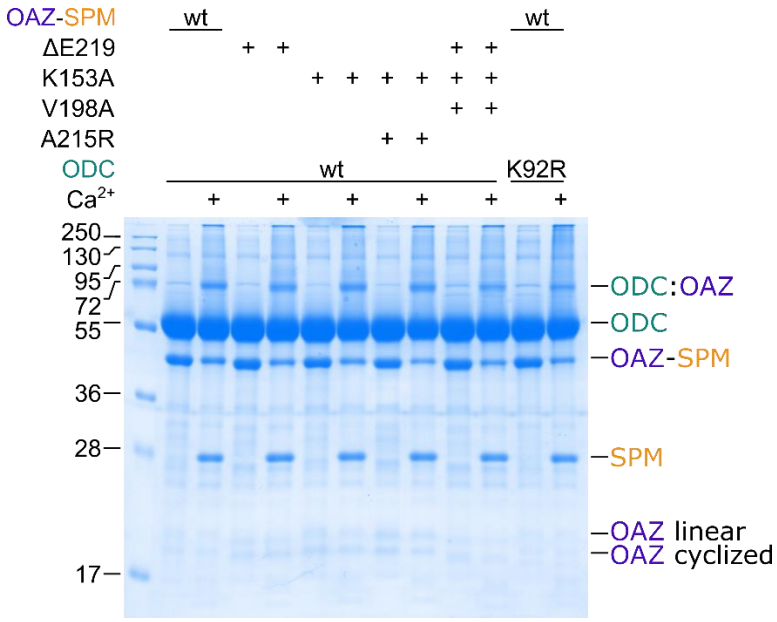


Figure 4.19. First test of OAZ-SPM binding mutations on ODC conjugation. 10 µM OAZ-SPM or binding mutants thereof were incubated at 37 °C overnight with 30 µM ODC in the presence of 12.5 mM calcium in 50 mM HEPES, 140 mM NaCl, pH 7.40 as indicated. Samples were boiled with reducing SDS-loading buffer and resolved on SDS-PAGE before Coomassie staining.

Subsequently, Sheryl Lim and I decided on introduction of R188E as a charge-inversing residue at a third interface to further reduce binding (Figure 4.18). Furthermore, she substituted K153A to K153E as another charge-inversing mutation. At low component concentrations (0.5 µM), she observed strong reduction of conjugation efficiency (Figure 4.20 (Scheu *et al.*, 2021)). Further details are provided in Scheu *et al.* (Scheu *et al.*, 2021).

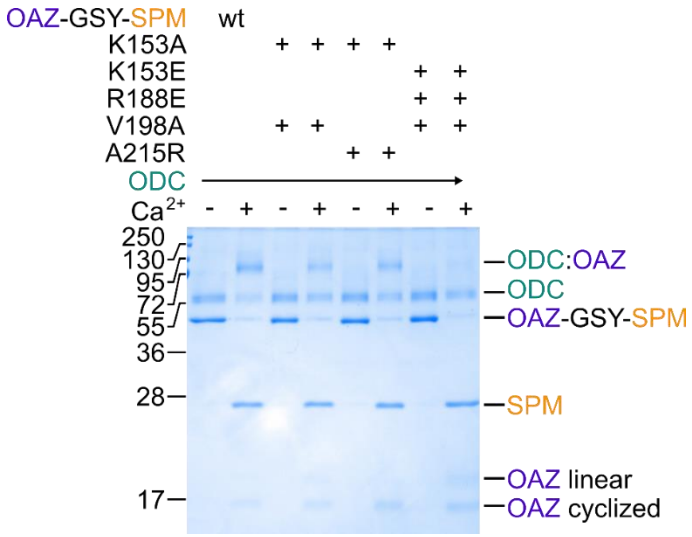


Figure 4.20. OAZ[K153E,R188E,V198A]-GSY-SPM shows reduced conjugation efficiency to ODC. Mutants of OAZ-GSY-SPM at 0.5 µM were incubated with 0.5 µM ODC and 10 mM calcium as indicated, for 0 min or 60 min at 37°C. Samples were boiled with SDS-loading buffer and resolved on SDS-PAGE before Coomassie staining. Adapted from Sheryl Lim, with permission (Scheu *et al.*, 2021).

4.6.2. Nonspecific protein conjugation

To test that conjugation was specific to ODC/OAZ complex formation, I wanted to test conjugation of OAZ-GSY-SPM to a non-cognate protein. For an initial test, I chose bovine serum albumin (BSA) as a protein with no specific interaction with OAZ or SPM. BSA is a protein which is very soluble, commonly used and readily available for purchase; furthermore, BSA is commonly used to stabilize cellular assays (see Chapter 5).

I incubated 6.7 µM OAZ-GSY-SPM with increasing concentrations of BSA with or without the presence of 10 mM calcium overnight at 37 °C (Figure 4.21). With calcium, I observed formation of a new band close to the size predicted for a BSA:OAZ-GSY conjugate as well as a new band close to the size predicted for BSA featuring two OAZ-GSY adducts [semi-log estimation: OAZ-GSY-SPM 45 kDa (predicted: 42.2 kDa), BSA 67 kDa (predicted: 66.4 kDa), BSA:OAZ-GSY 87 kDa (predicted: 82.2 kDa), BSA:OAZ-GSY*2 111 kDa (predicted: 98.0 kDa)]. With increasing concentration of BSA, both bands became only slightly stronger (~2-fold increase from 4 µM to 33 µM BSA). Substantial amounts of free OAZ formation were observed even at high BSA concentrations (Compare to Figure 4.10). Therefore, I concluded that there was initial evidence for non-specific conjugation of OAZ-GSY-SPM to BSA, but that this conjugation was not effective even at concentrations relevant to cell biology (e.g. 1% (w/v)

BSA = 150 µM, see Chapter 5) and that BSA would be unlikely to interfere with relevant applications. While the heterogeneity of commercial BSA hindered a conclusive analysis, subsequent reactions to further non-binding, recombinantly produced proteins were tested by Sheryl Lim (Scheu *et al.*, 2021).

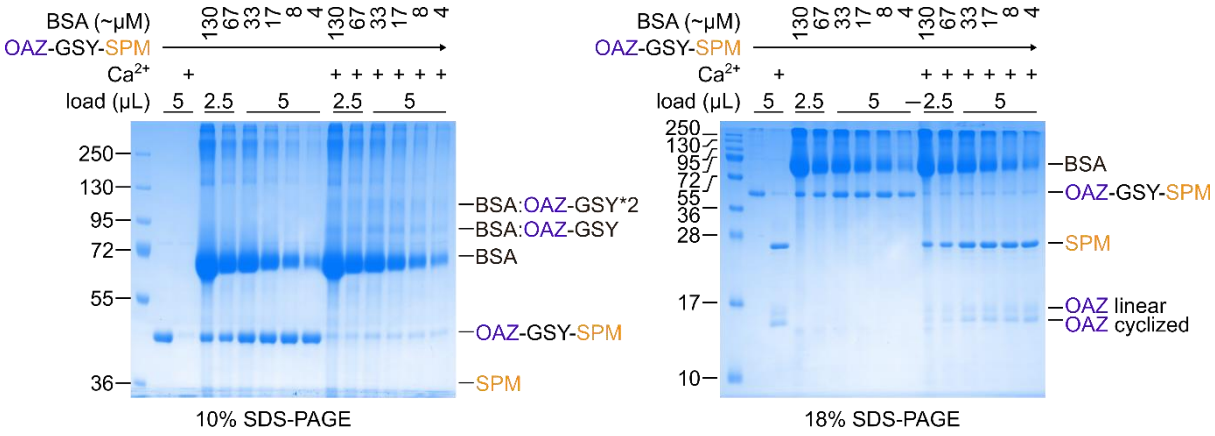


Figure 4.21. OAZ-GSY-SPM conjugation test to unspecific protein. 6.7 µM OAZ-GSY-SPM was incubated with increasing concentrations of bovine serum albumin (BSA) as a control protein and 10 mM calcium in 50 mM HEPES, 150 mM NaCl, 2 mM TCEP, pH 7.4 as indicated. After overnight incubation at 37 °C, the samples were boiled with SDS-loading buffer and resolved on 10% SDS-PAGE or 18% SDS-PAGE before Coomassie staining. For high concentrations of BSA, half the sample was loaded to reduce oversaturation. BSA concentration estimated from dissolved dry mass at 66.4 kDa molecular weight.

4.7. Discussion

Although protein-protein crosslinking has previously been observed after cleavage of full-length or partially truncated FrpC (Osička *et al.*, 2004; Sadilkova *et al.*, 2008) and our research group had previously observed evidence for protein cyclization in SPM-fusion proteins (also see (Metzner, 2017)), prior attempts of our research group to derive targeted intermolecular crosslinking were unsuccessful (Metzner, 2017). ODC/OAZ was the first model complex identified from NeissDist that was tested experimentally and OAZ-SPM became a valuable NeissLock model system. The initial conjugation between OAZ-SPM and ODC gave promising yields, but crosslinking efficiency could vary substantially in early preparations.

Screening of SpyTag-X-SPM revealed that SPM-processing rate depended on the residue preceding the cleavage site (XDP). Sadilkova *et al.* previously reported efficient cleavage for constructs preceding the DP bond with “alanine, glutamine, histidine, serine, tryptophan,

tyrosine, or valine" (Sadilkova *et al.*, 2008). For SpyTag-X-SPM all these residues similarly showed intermediate to very good cleavage rates. However, I observed low cleavage rates SpyTag-X-SPM where X was glycine, proline or aspartic acid. Two NMR structures for the FrpC self-processing module were recently published, one after processing (PDB ID 6sjw) and one of a non-processing P415A variant (PDB ID 6sjx) (Kuban *et al.*, 2020). Herein, it appears that the residue preceding the cleavage site could contact the C-terminal region of SPM as well as a loop from T430 to G450. Kubáň *et al.* were working with a shortened version of SPM (FrpC 414-591 (Kubáň *et al.*, 2015; Matyska Liskova *et al.*, 2016; Lišková *et al.*, 2019; Kuban *et al.*, 2020) compared to FrpC 414-657). FrpC 414-657 is a rough delineation of the domain determined by testing of cleavage activity in various truncations of FrpC (Osička *et al.*, 2004; Sadilkova *et al.*, 2008). A truncation to FrpC 414-591 agrees closely with domain predictions I conducted to validate the design of Kubáň et al., either by alignment of homologues or automated analysis (Kim *et al.*, 2004), to residues 414-586. However, our research group previously observed reduced cleavage rates for FrpC 414-591 (Metzner, 2017). In a comparison of OAZ-GSY-SPM with SPM as FrpC 414-591, FrpC 414-613 FprC 414-635 or FrpC 414-657, Sheryl Lim observed reduced cleavage rate for FrpC 414-591 (Scheu *et al.*, 2021). Since Kubáň *et al.* also state that the loop from T430 to G450 could be deleted without adverse effect (Kuban *et al.*, 2020), it is possible that these loops and the cleavage site interact in a way not apparent from the published structures.

After optimizing the purification conditions, OAZ-(GS)Y-SPM consistently showed efficient conjugation to ODC. The strong reactivity of the anhydride towards thiols agrees with previous studies of SPM-activation (Sadilkova *et al.*, 2008; Liu *et al.*, 2014; Metzner, 2017). In particular, quenching the anhydride and supressing protein-protein conjugation had been utilized for protein purification purposes (Sadilkova *et al.*, 2008; Liu *et al.*, 2014). Conversely, although thiols would quench the protein anhydride, TCEP was compatible with NeissLock-conjugation. For Affibody-SPM processing, I have observed that most anhydride-activated affibody can be quenched with cysteine within the first few minutes of inducing cleavage. If the reaction was

stopped and quenched at 10-15 minutes, I observed reduced impact on the ratio of cyclized to linear affibody, suggesting that most free anhydride had already dissipated. For comparison, reports on succinic anhydride hydrolysis at neutral pH suggest a half-life of 4-5 minutes at 25.1°C (k_1 = 2,69*10^{-2}/s at 25.1°C (Bunton *et al.*, 1963), i.e. $t_{1/2}$ = 4.29 min), wherein resonance stabilisation can substantially alter reactivity of an anhydride (e.g. maleic anhydride k1 = 31,40*10^{-2}/s at 25°C (Bunton *et al.*, 1963), i.e. $t_{1/2}$ = 2.20 min)).

Although most lysines are expected to be protonated around neutral pH, I observed tolerance of conjugation relative to self-processing at pH 7.4, 8.0 and 8.4. An explanation for this pH tolerance could be that an increase in anhydride hydrolysis counteracts increases in lysine nucleophilicity. Notably, the charge-inverting modification of lysine with similar carboxylic anhydrides (e.g. succinic anhydride, citraconic anhydride) has long precedent for protein modification (Lundblad, 2004), commonly involving modification at pH 7.0 to pH 8.5. Increased pH reduced SPM cleavage rate, as was previously reported for truncated FrpC (Osička *et al.*, 2004). Compared to Osička *et al.* (Osička *et al.*, 2004), the drop in cleavage rate of OAZ-Y-SPM was observed slightly earlier than they observed for truncated FrpC (pH 8.4 vs pH 9.0). A potential cause could be differences between the tested proteins or the use of TRIS-buffered system in their experiments (which markedly decreases in pH upon transition from room temperature to 37 °C, whereas the pH of HEPES buffers is less sensitive to temperature changes (Good *et al.*, 1966)). Nevertheless, NeissLock appeared suitable for applications around pH 7.4 to pH 8.0 (e.g. pH 7.4 as physiological pH of blood). While the lower bound for efficient conjugation has not yet been determined, we have also reported OAZ-GSY-SPM to be tolerant to and even increasing in cleavage rate at pH 6.5 (Scheu *et al.*, 2021).

Both the decrease of processing for OAZ-Y-SPM at high concentrations of ODC (occupying an interface close to the OAZ C-terminus) and the stark difference in cleavage rate between OAZ-Y-SPM and OAZ-GSY-SPM suggested that steric clashes affect processing. The NMR structure revealed the cleavage site to be at least partially enveloped by the SPM protein (Kuban *et al.*, 2020), again making steric hindrance seem likely – a common problem in

recombinant protein fusions. In addition to its effect on cleavage rate, the comparison of OAZ-Y-SPM and OAZ-GSY-SPM was also interesting for the understanding of NeissLock conjugation itself. Although NeissDist primarily considers the distance of the C-terminal resolved residue to target amines (i.e. 3.5 Å for ODC/OAZ), the actual anhydride becomes C-terminally extended according to the binder protein: First, even direct fusion of SPM to the C-terminus leaves the Asp anhydride after processing (OAZ-YD). This extension can potentially be avoided if the C-terminus is already Asp, or if other residues are identified as suitable for cleavage and conjugation (e.g. EP instead of DP, see (Osička *et al.*, 2004)). Second, if the residue preceding the cleavage site negatively impacts cleavage rate (e.g. GDP), another residue might be added (e.g. tyrosine, see 4.3.2 Saturation mutagenesis of cleavage site with SpyTag-X-SPM). Third, a spacer might be incorporated to mitigate steric hindrance. Alternatively, a spacer might be introduced to intentionally extend the flexible 'range' of the aspartic anhydride for target protein conjugation. Therefore, the C-termini of both OAZ-Y-SPM and OAZ-GSY-SPM are extended after cleavage compared to OAZ in the crystal structure, i.e. OAZ95-219-YD or OAZ95-219-GSYD. Both OAZ-Y-SPM and OAZ-GSY-SPM constructs were able to conjugate with various amines on ODC which could compensate for mutation of the target residue K92 (e.g. ODC K92R at K121, or ODC K92R Q96K). For OAZ-GSY-SPM but not OAZ-Y-SPM, distinct double banding was already apparent for conjugation to wt ODC (Figure 4.16, Figure 4.17), suggesting increased promiscuity of target sites compared to OAZ-Y-SPM, potentially due to additional flexibility and range of the C-terminal anhydride. For comparison, Hamachi *et al.* report crosslinking of a proximity-enabled small molecule probe to a lysine residues 11.4 Å from the primary binding site (Tamura *et al.*, 2018).

It proved difficult to identify all conjugation sites of crosslinked OAZ/ODC by tryptic LC-MS/MS alone. Although crosslinking of OAZ-Y-SPM to K92 (with wt ODC) and K121 (with ODC K92R) were confirmed, double-banding of ODC K92R suggested at least one additional crosslinking site. Here, it would have been interesting to further investigate samples with LC-MS/MS, e.g. following digestion with different enzymes or using different experimental settings. Through

site-directed mutagenesis and truncation, Sheryl Lim identified the N-terminus as another major conjugation site (Scheu *et al.*, 2021), which was not identified in tryptic LC-MS/MS. Another aspect of steric requirements became apparent when attempting to rescue wild type-like conjugation to ODC K92R. ODC K92 (wt), ODC K92R T93K, ODC K92R Q96K and ODC K92R S100 were originally intended to act as a 'ruler' with increasing distance of the resolved C-terminal residue OAZ E219 to a target lysine, but conjugation was not rescued according to this principle (i.e. K92>T93>Q96>S100) (Figure 4.16). First, this order is less apparent when accounting for the additional residues introduced after E219, i.e. -YD or -GSYD. Second, although ODC T93 is closer to OAZ E219 than ODC Q96 in 4zgy, the orientation of the residues alternates along the α-helix axis and T93/S100 are partially opposing K92/Q96 (Figure 4.22). Compared to ODC K92R T93K and ODC K92R S100, ODC K92R Q96K could therefore be a preferred conjugation site due to an orientation similar to ODC K92 (wt), i.e. bringing the lysine in proximity to an area in which the C-terminal anhydride of OAZ-YD or OAZ-GSYD preferentially conjugates (close to K121). Alternatively, nearby residues could influence crosslinking, e.g. by increasing reactivity of lysine nucleophiles at K92, Q96 and K121 (Baeza *et al.*, 2015) or by more effectively competing with lysine nucleophiles at T93 and S100.

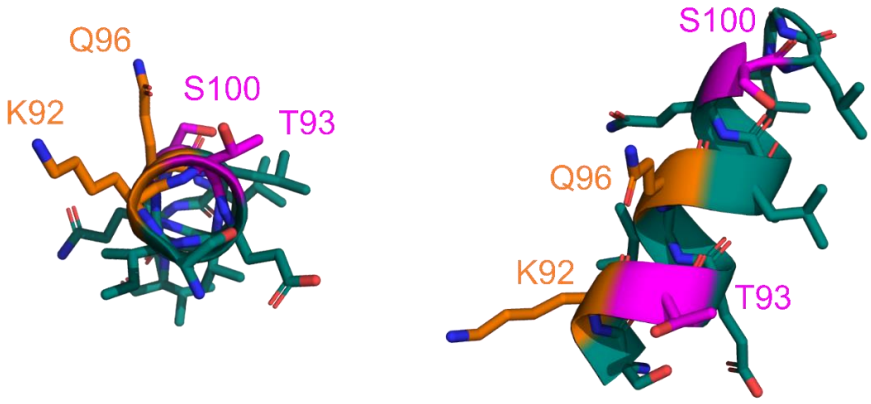


Figure 4.22. ODC K92, T93, Q96 and S100 are in alternating positions across an α-helix. ODC residues 91-103 from the 4zgy crystal structure are annotated on a view through the helical axis and a side-view. K92 and Q96 are highlighted in orange (ODC K92R Q96K shows similar conjugation to wt ODC at K92), T93 and S100 in magenta.

It is promising that the tolerance for positioning of the C-terminus could allow a broad range of complexes to be suitable for conjugation (compare to Chapter 3). Nevertheless, questions on the precise steric requirements and their optimisation remain. Linker design is an important consideration for electrophilic unnatural amino acids (Xiang *et al.*, 2014) and chemical warhead (Tamura *et al.*, 2012) development, and similarly linker design appears to be an important consideration for optimization for NeissLock probes.

Initial exploration of OAZ-(GS)Y-SPM specificity suggests that off-target conjugation would be much less efficient than NeissLock-targeted conjugation (Figure 4.21, further work by Sheryl Lim (Scheu *et al.*, 2021)). Sadilkova *et al.* captured glutathione-S-transferase (GST) fused to SPM with a polyhistidine-tag (GST-SPM-His) on a Ni-Nta column, as well as maltose-binding protein (MalE) fused to SPM with a chitin-binding domain (MalE-SPM-CBD) on a chitin column (Sadilkova *et al.*, 2008). For both proteins, activation of SPM processing allowed purification after washing. For GST-SPM-His and MalE-SPM-CBD they did not observe spontaneous protein-protein crosslinking *in vitro,* unlike their observations for truncated FrpC (Sadilkova *et al.*, 2008). Nevertheless, they report that more protein was released from the columns for cleavage in the presence of the thiol DTT. Therefore, they suggest that the increased concentration of proteins on the columns could increase protein-protein crosslinking (Sadilkova *et al.*, 2008). In addition, anhydride quenching after processing of SPM-fusion proteins could prevent reattachment of the N-terminal cleavage fragment with the C-terminal fragment. It is also conceivable that such reattachment would not always result in native SPM-fusion proteins still capable of cleavage, e.g. via attachment at a different site or through conjugation via an isopeptide bond, and could contribute a subspecies of “unprocessed” protein.

What is the lower end of binding strength required to drive conjugation? At the tested concentrations (30 µM ODC / 10 µM OAZ-SPM variant), the reduction in binding for OAZ-SPM mutants was not sufficient to ablate conjugation. From data on mouse ODC / OAZ, I expected introduction of K153A or V198A to OAZ to each reduce affinity by one order of magnitude,

(Cohavi *et al.*, 2009), potentially reducing affinity of a double mutant to about 1% of the original value. Sheryl Lim was able to demonstrate a reduction in affinity of OAZ-GSY-SPM to ODC from K_D 0.12 µM to 15 µM (0.8%) for K153, A215R or 25 µM (0.48%) for K153, V198 (Scheu *et al.*, 2021). Herein, observed k_{off} values each increased by about one order of magnitude (wt: 0.0028/s; K153A, A215R: 0.017/s; K153A, V198A: 0.034/s) (Scheu *et al.*, 2021). Covalent inhibitors depend on the rate of irreversible conjugation (k_{irr}) relative to the rate of dissociation (k_{off}), i.e. at sufficiently fast k_{irr} relative to k_{off}, a compound would never dissociate after binding, effectively achieving “infinite affinity” (Butlin and Meares, 2006). At the same time, NeissLock is distinct from typical covalent drugs because NeissLock utilizes a reactive species with limited lifetime that is generated *in situ*. While K_D values give an orientation for protein occupancy at equilibrium state, slow association rates (k_{on}) could therefore limit endpoint conjugation yields of NeissLock; most anhydride could simply react with water before target binding. SPR data suggest that the ODC / OAZ-GSY-SPM system (at relevant *in vitro* concentrations) reaches thermodynamic equilibrium within seconds to minutes (Scheu *et al.*, 2021); it can therefore be assumed that the time-frame for *in vitro* experimental setup of ODC / OAZ-GSY-SPM NeissLock reactions is sufficient to reach equilibrium, even before considering SPM processing rate. Therefore, in a K_D range of 15-25 µM, most OAZ would still be expected to be bound to ODC at the tested concentration of 10 µM OAZ and 30 µM ODC. Using subsequent mutations (OAZ[K153E,V198A,R188E]-GSY-SPM), Sheryl Lim was able to reduce the affinity of OAZ-GSY-SPM to ODC to be undetectable by SPR (expected K_D >100 µM) (Scheu *et al.*, 2021). At 0.5 µM concentration, OAZ[K153E,V198A,R188E]-GSY-SPM only conjugated to ODC in trace amounts (Scheu *et al.*, 2021). For comparison, specific low affinity protein-interactions can be in the 0.1-1 mM range, such as for ubiquitin-binding domains (Hurley *et al.*, 2006). Therefore, I expect that NeissLock could meet specificity requirements for many applications without further optimization.

5. Cellular NeissLock conjugation

5.1. Targeting the Epidermal Growth Factor Receptor with NeissLock

In parallel to identification of a model complex to validate NeissLock conjugation in vitro (Chapter 4), I sought to find a NeissLock system that could allow interesting therapeutic or diagnostic application. One such application would be NeissLock conjugation to the cell surface, for which I searched NeissDist for cellular surface proteins (Chapter 3). Here, I identified the Transforming Growth Factor α (TGFα) / Epidermal Growth Factor Receptor (EGFR) complex as a promising candidate.

EGFR is a transmembrane receptor and a well-studied member of the receptor-tyrosine kinase (RTK) family (Herbst, 2004). EGFR activation regulates cell growth and survival and the receptor is commonly overexpressed in cancer, such as in breast cancer (Herbst, 2004). Therefore, EGFR could make a valuable target for covalent modification, e.g. to effectively modify aberrant cells, potentially to deliver therapeutic agents. EGFR is activated by soluble ligands, most prominently Epidermal Growth Factor (EGF) and Transforming Growth Factor α (TGFα) (Singh and Coffey, 2014).

Ligand binding stabilises activating confirmations of the receptor, facilitating its homodimerization and subsequent transphosphorylation of the intracellular domains of EGFR (Freed *et al.*, 2017). Although structurally homologous, the various ligands binding to EGFR differ in properties such as affinity, dissociation and geometry of EGFR activation, leading to different receptor fates (Freed *et al.*, 2017). EGF and TGFα are ligands causing a 'strongly activating' conformational change of EGFR which encourages the formation of active homodimers at the cell surface (Freed *et al.*, 2017) (Figure 5.1a,b). Upon subsequent endocytosis of the EGFR dimer, TGFα is more likely to dissociate upon acidification, whereas most EGF remains bound (Ebner and Derynck, 1991; Roepstorff *et al.*, 2009). *In vitro*, artificial pH-change induces dissociation from EGFR at the cell surface at pH 6.5 for TGFα and pH 5.5 for EGF (Ebner and Derynck, 1991; Roepstorff *et al.*, 2009). Subsequently, EGF promotes degradation of the receptor and thereby desensitization to EGF due to decreased surface

levels of EGFR (Ebner and Derynck, 1991; Roepstorff *et al.*, 2009). Other ligands, such as Epigen or Epiregulin, can cause reduced but prolonged activation of the receptor through stabilisation of alternate receptor confirmations (Freed *et al.*, 2017) (Figure 5.1c,d).

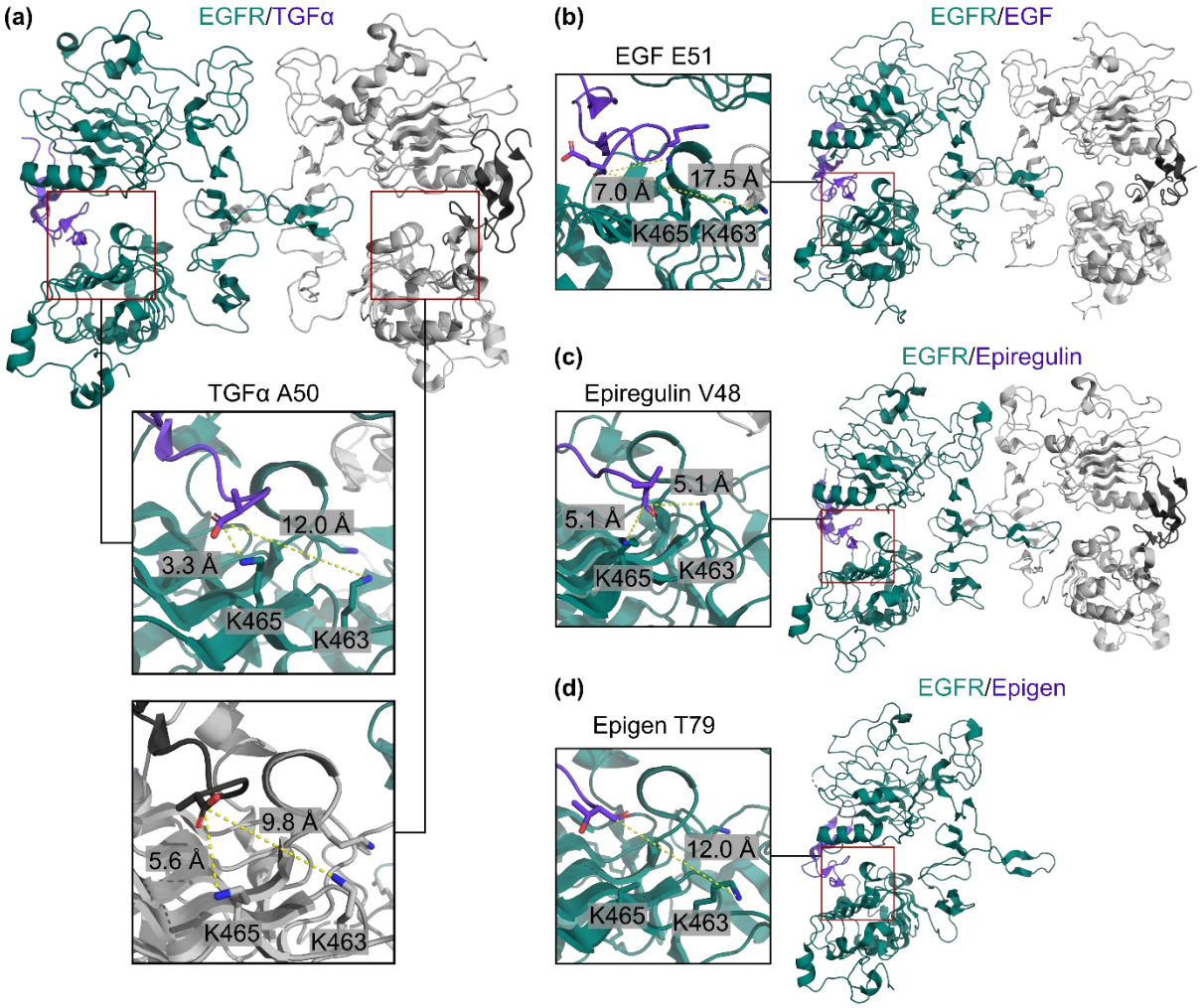


Figure 5.1. Overview of Epidermal Growth Factor receptor (EGFR) / Ligand complexes and potential for NeissLock conjugation. (a) The EGFR/TGFα complex was identified from NeissDist as a candidate complex for cellular NeissLock conjugation (crystal structure 1mox (Garrett *et al.*, 2002)). TGFα binding stabilizes formation of an activated EGFR homodimer. For both EGFR chains, the C-terminal resolved residue of TGFα is close to K465 as well as K463 on EGFR. Green: EGFR chain D, Purple: TGFα chain B, Grey: EGFR chain C, Black: TGFα chain A. (b,c,d) Crystal structures of the EGFR/EGF complex (1IVO (Ogiso *et al.*, 2002)), the EGFR/Epiregulin complex (5WB7 (Freed *et al.*, 2017)) and EGFR/Epigen complex (5WB8 (Freed *et al.*, 2017)) with ligand C-termini in inset. (b) Inset showing dimer with shorter distance from EGF E51 to EGFR K465, at 7.0 Å (chain C to A) instead of 7.2 Å (chain D to B), atom C to atom Nε. (c) Inset showing Epiregulin chain with more resolved residues, i.e. chain E (Epiregulin S2-V48) instead of chain H (Epiregulin S2-F45). (d) In the biological assembly of EGFR/Epigen, EGFR is not dimerized. In the inset, EGFR K465 atom Nε is not resolved.

TGFα was identified as a promising NeissLock candidate from NeissDist (Chapter 3, Figure 5.1a). According to the crystal structure of the TGFα/EGFR ectodomain complex (1mox

(Garrett *et al.*, 2002)), two lysine residues on EGFR are close to the C-terminus of TGFα, with two resolved chains each. For TGFα chain D to EGFR chain B, the distance between the C-terminal residue A50 (atom C) of TGFα to EGFR K465 (atom Nε) is calculated as 3.3 Å and the distance to EGFR K463 (atom Nε) is calculated as 12.0 Å. (Figure 5.1a). Similarly, for TGFα chain C to EGFR chain A, the distance between the C-terminal residue A50 (atom C) of TGFα to EGFR K465 (atom Nε) is 5.6 Å and the distance to EGFR K463 (atom Nε) is 9.8 Å. (Figure 5.1a). The structure is resolved at 2.50 Å. Therefore, EGFR K465 appeared ideally positioned for conjugation with a TGFα C-terminal anhydride (compare to ODC 3.5 Å, Chapter 4). In addition to TGFα, multiple other EGFR-ligands were identified as promising NeissLock candidates from NeissDist, most notably EGF (Figure 5.1b-d). TGFα and EGFR contain complex disulfide bonding, a disadvantage not observed for the OAZ/ODC system. However, the TGFα/EGFR complex is of striking biological relevance and was therefore exciting for validation of NeissLock-applications.

5.2. *In vitro* conjugation of EGFR/TGFα

5.2.1. Initial validation of TGFα-SPM cleavage activity

As the first construct I cloned TGFα-SPM, featuring an N-terminal His-tag, then TGFα, with fusion of SPM directly after the C-terminal A50 of TGFα. Soluble protein was produced either from BL21 RIPL or from RosettaGami-2 (featuring thioredoxin/glutaredoxin reductase mutation for folding of disulfide-bonded proteins (Seras-Franzoso *et al.*, 2012)). After induction at 18°C for 16-18h (Figure 5.2a), the protein was dialysed into 50 mM HEPES, 140 mM NaCl, pH 7.40 before spin concentration.

I incubated 10 μM of TGFα-SPM prepared from BL21 RIPL or RosettaGami-2 with or without 10 mM calcium at 37°C. After 1 h or 18 h, the reactions were stopped by addition of EDTA in reducing SDS-loading buffer. After boiling, the samples were resolved on SDS-PAGE and Coomassie stained. With the addition of calcium, TGFα-SPM underwent self-processing, both after purification from BL21 RIPL and Rosetta-Gami2 strains. After overnight-processing, essentially all TGFα-SPM was cleaved (Figure 5.2b).

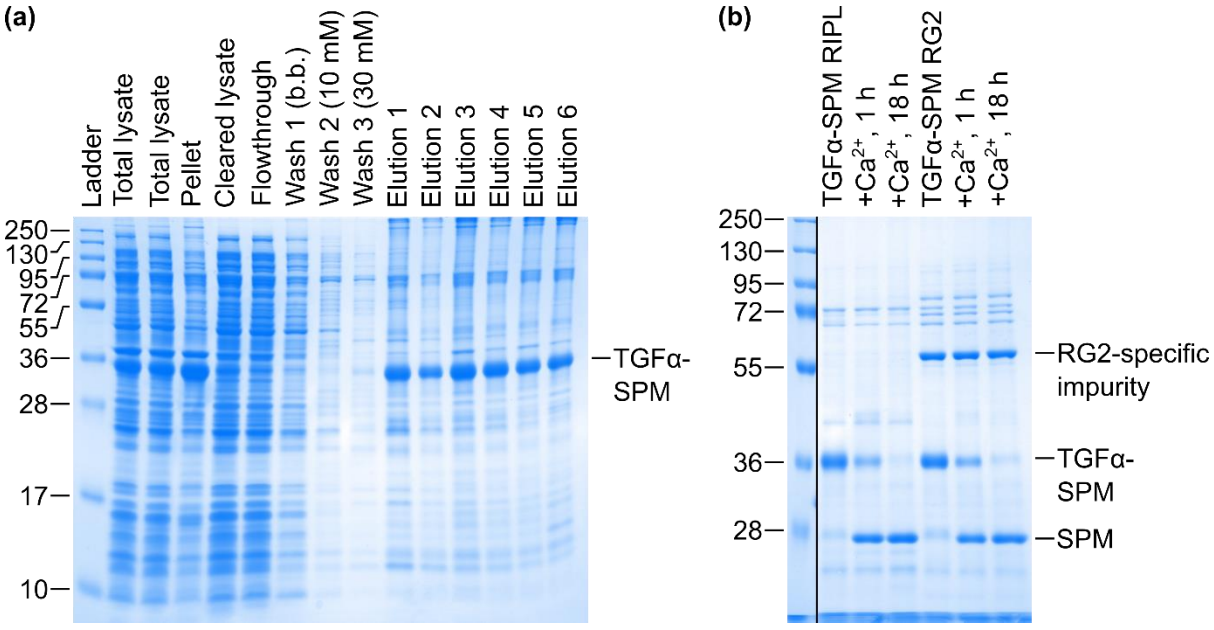


Figure 5.2. Purification and initial validation of TGFα-SPM activity. (a) Samples from TGFα-SPM Ni-Nta affinity chromatography, induced from BL21-RIPL for 16 h at 18°C, resolved on reducing SDS-PAGE. b.b.: binding buffer; wash buffers supplemented with indicated amounts of imidazole. (b) TGFα-SPM from RIPL and RosettaGami2 (RG2) undergoes self-processing upon addition of calcium. 10 mM calcium was added to 10 µM TGFα-SPM and samples were incubated for 1 h to 18 h at 37°C before stopping the reaction by addition of EDTA in reducing SDS-loading buffer. Samples were resolved on SDS-PAGE and Coomassie stained.

5.2.2. Preparation of soluble EGFR for *in vitro* study

To test if TGFα-GSY-SPM is suitable for NeissLock-conjugation to EGFR *in vitro*, I decided to use sEGFR501, a truncation from the extracellular soluble domain of EGFR, sEGFR621. sEGFR501 was previously shown to be highly effective in TGFα binding, with about 2-fold lower K_D compared to sEGFR621 (Elleman *et al.*, 2001). I cloned sEGFR501 with a tissue plasminogen activator (tPA) signal sequence for mammalian secretion. Following transient transfection of Expi293 cells, the cells were incubated for 4 days at 37°C, 125 rpm shaking with 8% CO_2. Optionally, I added Kifunensine in parallel to transfection, which is an inhibitor of mannosidase I (Elbein *et al.*, 1990). Secreted sEGFR501 was purified from the cell supernatant using Ni-Nta purification (Figure 5.3). Herein, sEGFR501 purified from Kifunensine-treated cells showed a sharpened band compared to sEGFR501 purified from untreated cells, indicating simplified glycosylation (Figure 5.3b). sEGFR501 preparations

made with or without Kifunensine (sEGFR501 ±Kifunensine) were then utilized for NeissLock conjugation tests.

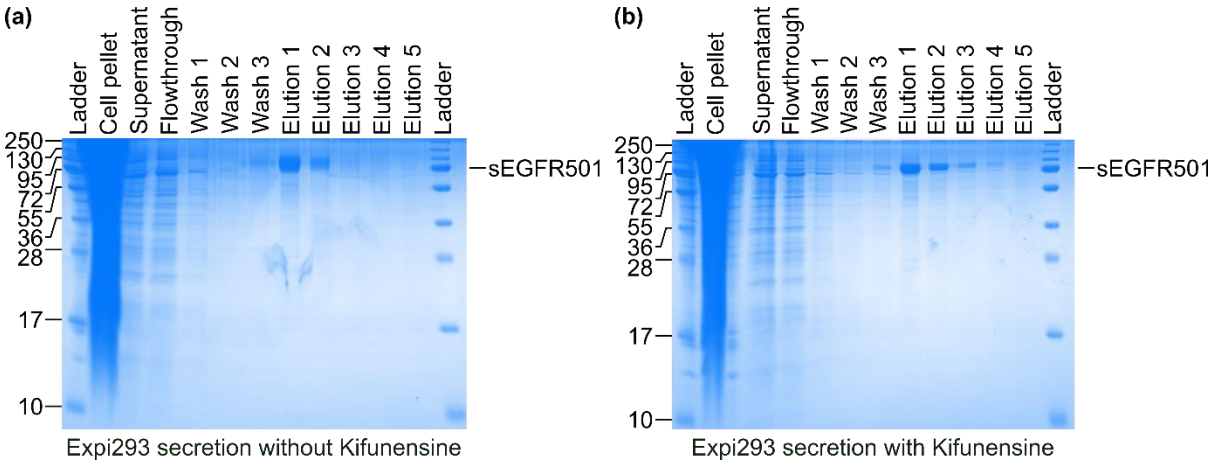


Figure 5.3. Purification of sEGFR501 from Expi293. Ni-Nta affinity chromatography of sEGFR501 secreted from Expi293 cells Optionally, cells were treated with the mannosidase-inhibitor Kifunensine. Samples were boiled with SDS-loading buffer and resolved on SDS-PAGE before Coomassie staining. After Kifunensine treatment, sEGFR501 became more focused on SDS-PAGE, indicating reduced glycosylation.

5.2.3. Initial evidence of EGFR:TGFα conjugation

To facilitate detection of an EGFR/TGFα conjugate, I first decided to incorporate SpyTag003 into TGFα-SPM (to make 'ST3-TGFα-SPM'). The peptide tag SpyTag003 rapidly and covalently reacts with the protein SpyCatcher003 and thereby serves as a convenient handle for further functionalisation (Keeble *et al.*, 2019). ST3-TGFα-SPM was purified by Ni-Nta purification followed by C-tag purification. I incubated 5 µM sEGFR501 -Kifunensine or 7 µM sEGFR501 +Kifunensine with 20 µM ST3-TGFα-SPM ±1 mM calcium overnight at 37°C. Subsequently, SpyTag003 was labelled by addition of SpyCatcher003:DyLight conjugate for 1 h at 23°C.

After overnight incubation with calcium, a large fraction of ST3-TGFα-SPM had self-processed (Figure 5.4, Coomassie). On Coomassie, both sEGFR501 ±Kifunensine showed initial evidence of sEGFR501:ST3-TGFα conjugation. After incubation of sEGFR501 ±Kifunensine in the presence of ST3-TGFα-SPM and calcium, the electrophoretic mobility of sEGFR501 -Kifunensine appeared reduced, and a new faint band was observed for

sEGFR501 +Kifunensine (Figure 5.4, Coomassie). However, the closely stacked sEGFR501 species made it difficult to validate conjugation.

Conjugation of SpyTag003 with sub-stoichiometric amounts of SpyCatcher003:DyLight conjugate allowed the direct comparison of fluorescence imaging to Coomassie-staining. Herein, I clearly observed the appearance of conjugate bands (Figure 5.4, Fluorescence) as well as a distinct shift in molecular weight of tagged species upon SpyCatcher003:DyLight-conjugation (Figure 5.4, Overlay). This high-molecular weight band was sharper for sEGFR501 -Kifunensine treatment than the band for sEGFR501 derived without Kifunensine treatment (Figure 5). As this band was both specifically labelled by SpyCatcher003:DyLight (indicating the presence of ST3-TGFα) and sensitive to Kifunensine (indicating the presence of sEGFR501), this experiment gave the first compelling evidence of specific TGFα/EGFR NeissLock-conjugation using ST3-TGFα-SPM to sEGFR501.

Finally, unlabelled ST3-TGFα had migrated off the gel upon incubation of ST3-TGFα-SPM with calcium (Figure 5.4, Coomassie). However, ST3-TGFα labelled with SpyCatcher003:DyLight was retained and now visible in the fluorescent image (Figure 5.4, Fluorescence). Labelled ST3-TGFα portrayed distinct double banding, indicative of self-reaction after processing, as was previously observed for Affibody-SPM and OAZ-SPM (Chapter 4). After optimization of OAZ-SPM to OAZ-GSY-SPM (Chapter 4.3.3.), and with evidence of self-reaction in ST3-TGFα-SPM (Figure 5.4), I decided to adjust the future design of TGFα-SPM to include a GSY spacer. Furthermore, I decided to move the N-terminal His-tag to the C-terminus of SPM to make TGFα-GSY-SPM-H_6 (as “TGFα-GSY-SPM”). To avoid impurities introduced in purification from RosettaGami2 (Figure 5.2), BL21-RIPL was used for further study.

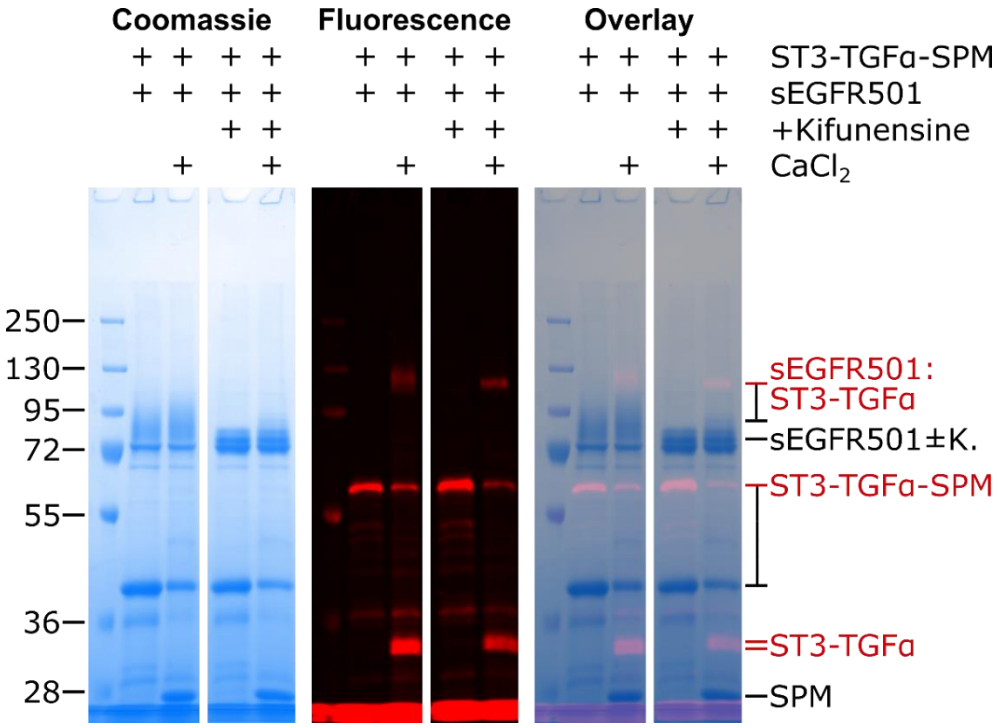


Figure 5.4. SpyCatcher003:DyLight labelling provides first evidence of TGFα as an EGFR-targeting NeissLock probe. 20 µM ST3-TGFα-SPM was incubated overnight at 37°C with 5-7 µM sEGFR501 (purified from cells with or without Kifunensine-treatment) and 1 mM $CaCl_2$ as indicated. SpyTag003 was labelled for 1 h at 23°C with sub-stoichiometric quantities of SpyCatcher003:DyLight. Samples were reduced, heat denatured (3 min 99°C) and resolved on SDS-PAGE. Samples were fluorescence-imaged before Coomassie staining. (a) Coomassie staining of SDS-PAGE. (b) Fluorescence imaging of SDS-PAGE. (c) Overlay of Coomassie staining and fluorescence imaging by alignment of the protein ladder.

5.2.4. sEGFR501 deglycosylation and *in vitro* conjugation yield

ST3-TGFα-SPM gave initial evidence of sEGFR501 conjugation but did not provide a system for convenient estimation of conjugation efficiency. Since Kifunensine had proved helpful to simplify sEGFR501 glycosylation (Figure 5.3, Figure 5.4), I used Peptide:N-glycosidase F (PNGase F) to further unify sEGFR501. Initial samples of TGFα-GSY-SPM were obtained from BL21-RIPL by Ni-Nta purification.

5 µM sEGFR+Kifunensine were incubated with 50 µM TGFα-GSY-SPM and 2 mM $CaCl_2$ at 37°C for 1 h. Then, samples were reduced and denatured using Glycoprotein Denaturing Buffer (NEB) with incubation for 10 min at 100 °C. Additional PNGase F buffer components were added, with or without addition of PNGase F enzyme. The digestion was incubated for 1 h at 37 °C before addition of SDS-loading buffer. Samples were boiled and resolved on SDS-PAGE followed by Coomassie staining.

Multiple bands were observed for sEGFR501 even from cells treated with Kifunensine (Figure 5.5a,b). After PNGase F treatment, sEGFR501 collapsed to a single band. Upon incubation of TGFα-GSY-SPM with sEGFR501 and calcium, I observed clear evidence of sEGFR501:TGFα-GSY conjugation (Figure 5.5a,b). At 10-fold excess of TGFα-GSY-SPM, over half of sEGFR501 was conjugated to form a new band corresponding to sEGFR501/TGFα-GSY.

Next, I wanted to test if conjugation was limited by the ratio of TGFα-GSY-SPM to sEGFR501 or absolute concentration (e.g. due to dissociation / low binding). 5 µM sEGFR+Kifunensine were incubated with calcium and decreasing amounts of TGFα-GSY-SPM as indicated, before reduction, denaturation and PNGase F deglycosylation (Figure 5.5c). Alternatively, sEGFR501 and TGFα-GSY-SPM were combined at a 1:10 ratio and then diluted to change the total concentration (Figure 5.5c).

With decreasing excess of TGFα-GSY-SPM compared to sEGFR501, there was a visible reduction in conjugation to sEGFR501. At a 10-fold amount of TGFα-GSY-SPM (50 µM), over half of sEGFR501 was conjugated, at a 5-fold amount of TGFα-GSY-SPM (25 µM) more than a quarter of sEGFR501 was conjugated, and at a 2-fold amount (10 µM) or equimolar (5 µM) amount of TGFα-GSY-SPM only small amounts of conjugation were observed (Figure 5.5c). For samples where the same relative rate between sEGFR501 and TGFα-GSY-SPM was maintained, decreasing the concentration did not have an obvious effect on conjugation ratio (Figure 5.5c). These observations suggest that conjugation is not limited by dissociation at the given concentrations, but by the fraction of TGFα-GSY-SPM conjugating to sEGFR501 (compare to ODC/OAZ, Figure 4.10).

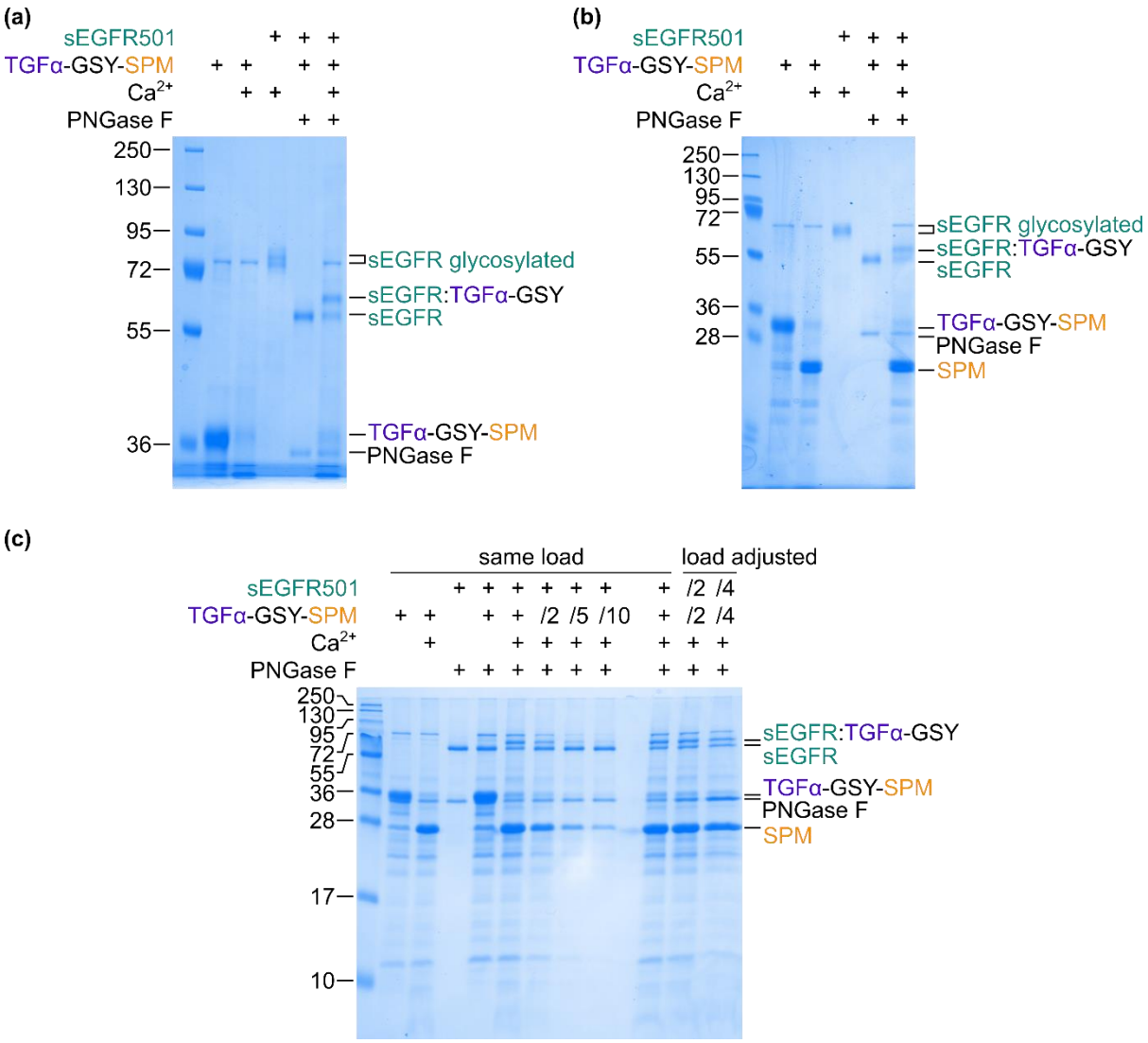


Figure 5.5. PNGase F digestion facilitates sEGFR501/TGFα conjugate detection. (a,b) 5 µM sEGFR501 +Kifunensine was incubated with 50 µM of TGFα-GSY-SPM for 1 h at 37 °C. Subsequently, the samples were reduced and denatured. As indicated, PNGase F was added for digestion of glycosylation. Herein, sEGFR501 collapsed to a single band. This digestion allowed for the easy detection of sEGFR501/TGFα-GSY conjugate. Samples were boiled with SDS-loading buffer and resolved on SDS-PAGE of different density before Coomassie staining. (c) TGFα-GSY-SPM partially conjugates to sEGFR501. 5 µM sEGFR501 +Kifunensine were incubated with decreasing concentrations of TGFα-GSY-SPM (50 µM at "+", concentration halved (25 µM) at "/2", 10 µM at "/5", 5 µM at "/10") and 2 mM $CaCl_2$ for 1.5 h at 37°C. With decreasing excess of TGFα-GSY-SPM to sEGFR501, conjugation efficiency dropped to trace amounts. Alternatively, the concentration of both TGFα-GSY-SPM and sEGFR501 was decreased while maintaining their stoichiometric ratio (5 µM sEGFR501 + 50 µM TGFα-GSY-SPM, then undiluted, diluted 1:2 or diluted 1:4). Decreasing absolute concentrations showed no appreciable effect on conjugation efficiency. Samples were reduced and denatured before PNGase F treatment, then boiled with SDS-loading buffer before SDS-PAGE and Coomassie staining.

5.2.5. Size exclusion chromatography and TGFα-GSY-SPM folding

Next, I used size exclusion chromatography to assess the structural integrity and purity of TGFα-GSY-SPM. Herein, I subjected TGFα-GSY-SPM to size exclusion chromatography in

50 mM HEPES, 150 mM NaCl, pH 7.40 immediately following Ni-Nta purification. TGFα-GSY-SPM was induced from BL21-RIPL. The cells were harvested, fractured by sonication and TGFα-GSY-SPM was isolated using Ni-Nta affinity purification (Figure 5.6a). After spin concentration, samples were loaded onto a Superdex S200 16-600 column in an ÄKTA pure system.

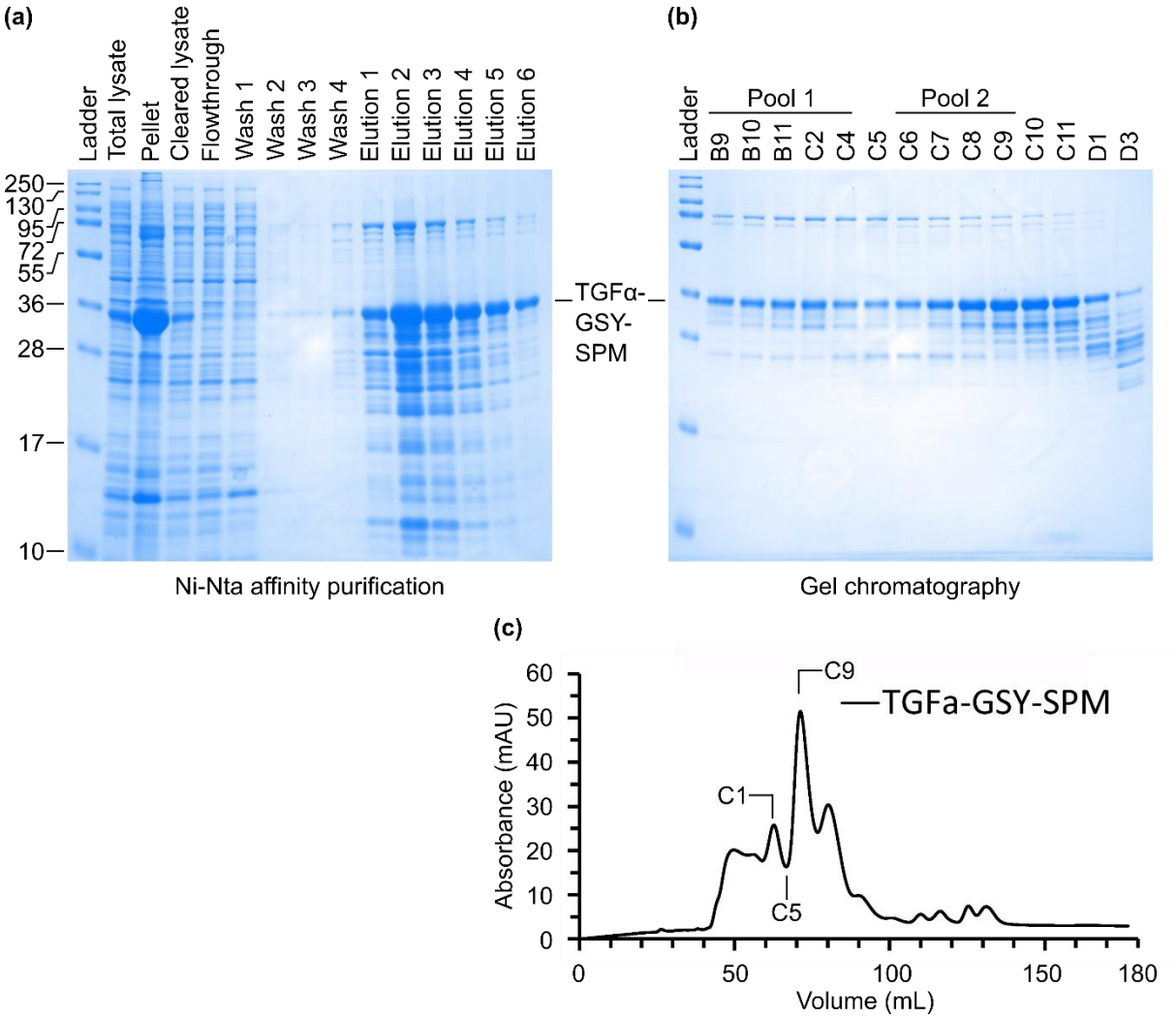


Figure 5.6. Gel chromatography of TGFα-GSY-SPM. (a) Ni-Nta affinity purification of TGFα-GSY-SPM induced from BL21-RIPL. Wash 1: Ni-Nta buffer + 10 mM Imidazole. Wash 2: Ni-Nta buffer + 30 mM Imidazole. Elution: Ni-Nta buffer + 200 mM Imidazole. (b,c) Eluates from (a) were spin concentrated and resolved by gel chromatography on a Superdex S200 16-600 column. Samples from fractions surrounding main peak positions in (b), with chromatogram at A_{280} in (c). In (a,b), SDS-loading buffer was added to samples before boiling. Samples were resolved on SDS-PAGE and Coomassie stained.

The size exclusion chromatogram resolved two peaks indicating high protein concentrations (Figure 5.6b,c). According to the fractions resolved on SDS-PAGE, both peaks consisted primarily of TGFα-GSY-SPM (Figure 5.6b,c). However, I observed a reduction in concentration

between peak positions, indicating that TGFα-GSY-SPM in fact resolved in two different peaks during gel chromatography. This was an indication that the preparation was subject to heterogeneity (e.g. multimerization or partial misfolding). I separately pooled samples from both peaks and concentrated them in Vivaspin columns (‘pool 1’ or ‘peak 1’ at lower retention volume, ‘pool 2’ or ‘peak 2’ at higher retention volume).

I then tested conjugation of sEGFR501 with TGFα-GSY-SPM purified in this manner. After PNGase F treatment, sEGFR501 collapsed to a single band, both for sEGFR501 purified from cells with or without Kifunensine treatment (Figure 5.7). As prior, TGFα-GSY-SPM cleaved to near completion in the presence of calcium. Upon co-incubation of ‘peak 2’ TGFα-GSY-SPM with sEGFR501 ±Kifunensine, I observed the appearance of a distinctly resolved band corresponding to EGFR/TGFα-GSY conjugation. Herein, both sEGFR501 obtained from cells treated with Kifunensine or untreated cells displayed similar conjugation efficiency. There was no evidence of unspecific protein-protein conjugation.

I also tested conjugation of the fraction of ‘peak 1’ TGFα-GSY-SPM resolved at a lower retention volume in gel chromatography (Figure 5.6). Similar amounts of ‘peak 1’ TGFα-GSY-SPM protein appeared to be resolved on SDS-PAGE (Figure 5.7). However, conjugation efficiency in the presence of sEGFR501 was reduced for TGFα-GSY-SPM corresponding to ‘peak 1’ compared to ‘peak 2’, both for sEGFR501 ±Kifunensine (Figure 5.7). Therefore, TGFα-GSY-SPM obtained from ‘peak 2’ was chosen for further experiments unless stated otherwise.

Finally, TGFα R42A has been reported to effectively reduce TGFα activity and binding to EGFR (Defeo-Jones *et al.*, 1989; Lazar *et al.*, 1989). To provide additional controls, I cloned TGFα[R42A]-GSY-SPM in addition to TGFα-GSY-[DA]SPM (i.e. featuring an inactivating DP to AP mutation at the SPM cleavage site) and purified them as described. During gel purification of these variants, I observed slight changes in retention volume and adjusted the peak position accordingly (± 1 Fraction). TGFα[R42A]-GSY-SPM showed strongly reduced conjugation to sEGFR501, both with and without Kifunensine (Figure 5.7, compare to Figure

4.20). As expected, TGFα-GSY-[DA]SPM was not capable of self-processing or conjugation to sEGFR501 (Figure 5.7).

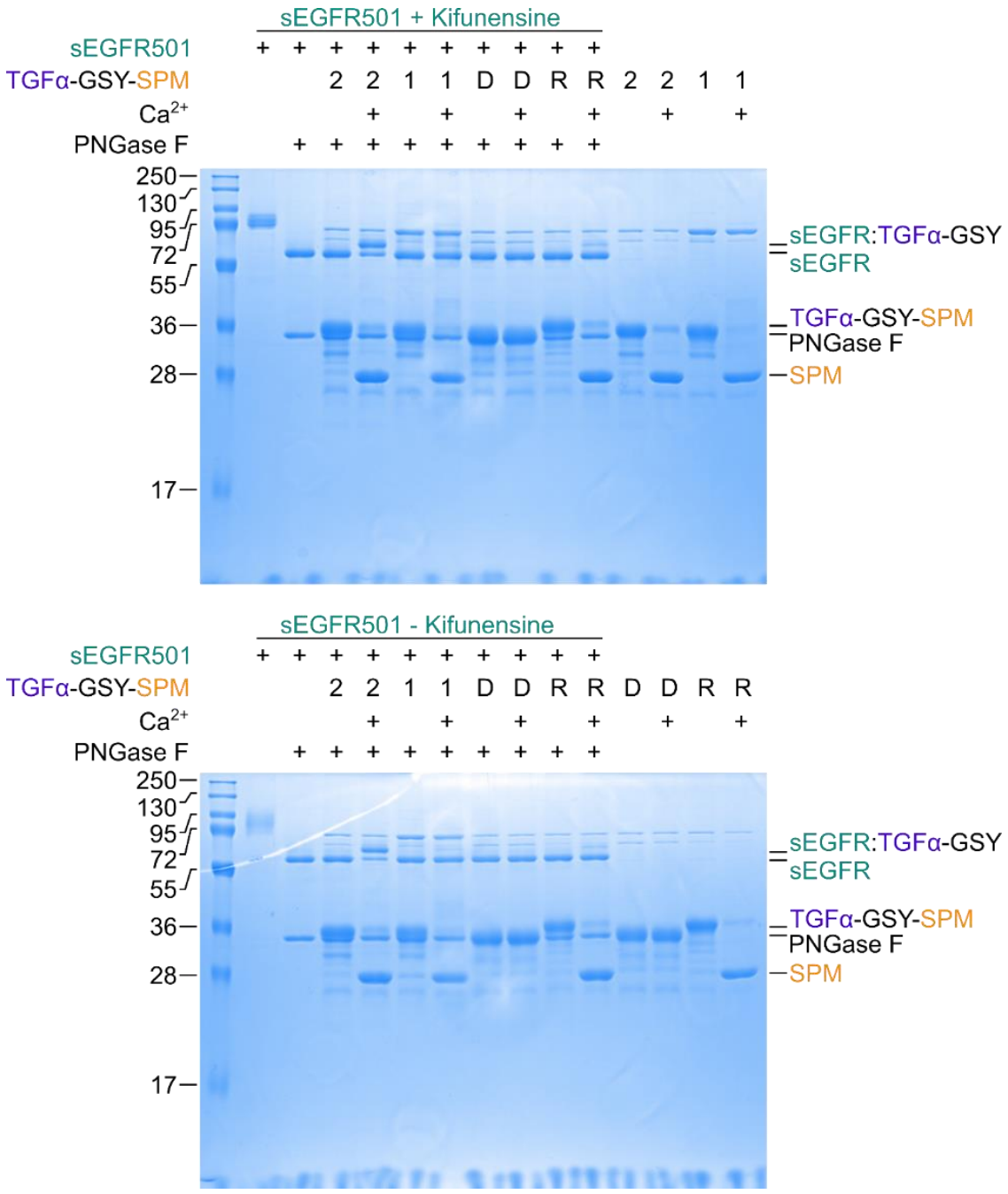


Figure 5.7. TGFα-GSY-SPM, resolved at distinct gel chromatography retention volumes, differentially conjugates to sEGFR501. TGFα-GSY-SPM resolved into two peaks on Superdex S200. TGFα-GSY-SPM samples from a peak with lower retention volume (1) were compared to samples from a peak with higher retention volume (2), corresponding to two separate purification batches. D: TGFα-GSY-[DA]SPM corresponding to (2), R: TGFα[R42A]-GSY-SPM corresponding to (2). 12.5 µM per TGFα-GSY-SPM sample were incubated with 2.5 µM sEGFR501 ±Kifunensine for 5 h at 37 °C. Samples were reduced and denatured before incubation with PNGase F or water as indicated. Samples were boiled with SDS-loading buffer and resolved on SDS-PAGE before Coomassie staining.

5.3. Characterisation of sEGFR501:TGFα-GSY conjugate

5.3.1. Mapping of crosslinking sites

I wanted to identify crosslinking sites of TGFα-GSY-SPM on sEGFR. In the crystal structure, EGFR K463 and K465 were identified as lysine residues in proximity to the C-terminus of TGFα (Figure 5.1, Figure 5.8b). I attempted to identify evidence of site-specific conjugation by tryptic LC-MS/MS. For this, 2.5 µM sEGFR501+Kifunensine were incubated with 12.5 µM TGFα-GSY-SPM and 2 mM calcium as indicated for 5 h at 37°C. Samples were reduced and denatured before deglycosylation with PNGase F and resolved on SDS-PAGE (Figure 5.8a). The conjugate product band was cut from the gel and prepared for in-gel tryptic digest and analysis by LC-MS/MS. Herein, crosslinking of the C-terminus of TGFα-GSY to sEGFR501 K465 was confirmed (Figure 5.8c), representing a residue close to the TGFα C-terminus residue (Figure 5.8b). No peptides were confirmed for conjugation to K463.

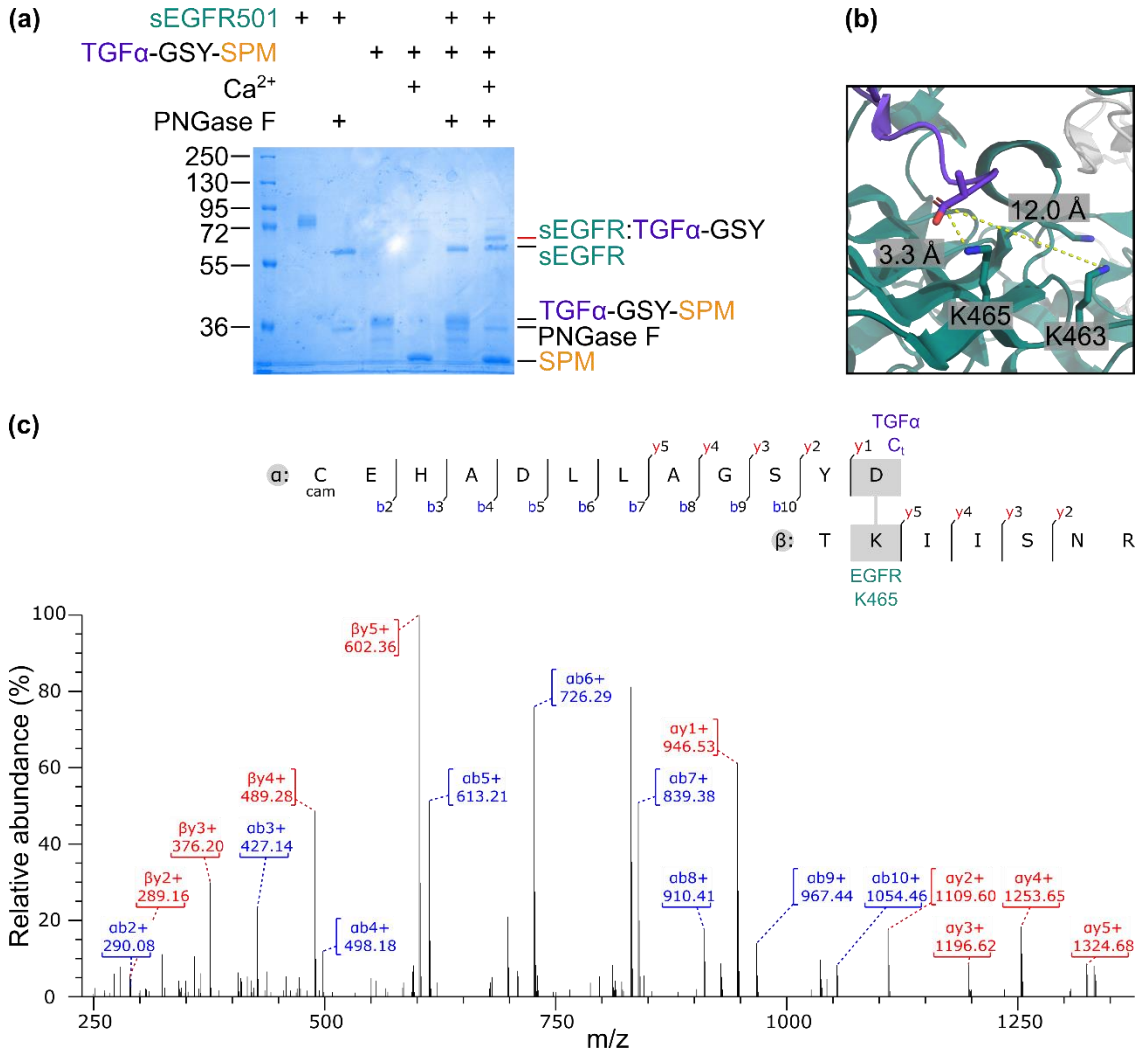


Figure 5.8. Tryptic digestion and liquid chromatography with tandem mass spectrometry (tryptic LC-MS/MS) of TGFα-GSY-SPM conjugation to EGFR. (a) 2.5 µM sEGFR501 were incubated with 12.5 µM TGFα-GSY-SPM and 2 mM $CaCl_2$ or buffer as indicated. sEGFR501 samples were deglycosylated with PNGase F before boiling with SDS-loading buffer. Samples were resolved on SDS-PAGE with lane spacing, and the sEGFR501/TGFα conjugate (indicated by red line) was cut from gel after brief Coomassie staining. (b) TGFα A50 annotated on the crystal structure PDB ID 1mox (Garrett *et al.*, 2002), as well as the nearby residues EGFR K465 and EGFR K463. (c) LC-MS/MS spectra as well as annotated structure of crosslinked peptide identified from LC-MS/MS, indicating conjugation of TGFα-GSY-SPM C-terminus post-cleavage (TGFα-GSYD) to EGFR K465.

5.3.2. *In vitro* Western blot

Before analysis of TGFα-GSY-SPM conjugation to EGFR on cells, I wanted to establish Western blotting of TGFα-GSY-SPM conjugation to sEGFR501 *in vitro*. I used samples of TGFα-GSY-SPM conjugation to sEGFR501 from the same experiment as shown in Figure 5.5c for *in vitro* Western blot (first five samples of Figure 5.5c). Herein, 50 µM TGFα-GSY-SPM had been incubated with 5 µM EGFR +Kifunensine with 2 mM calcium for 1.5 h at 37°C before reduction, denaturation and PNGase F deglycosylation. Samples were resolved on two

separate 18% SDS-PAGE gels (for Coomassie staining and Western blotting). The samples were transferred to methanol-activated polyvinylidene fluoride (PVDF) membrane and stained with mouse anti-EGFR antibody, as well as anti-mouse antibody:horseradish peroxidase (HRP) conjugate. Luminescence was accumulated upon incubation with HRP substrate, the membrane was transilluminated to detect the ladder position. The gel from which protein had been transferred to the PVDF membrane was also Coomassie stained and imaged (Figure 5.9).

Western blot with anti-EGFR showed a sharp band corresponding to sEGFR501, as previously resolved on Coomassie (Figure 5.9). In addition, a distinct band was observed for sEGFR501/TGFα-GSY conjugate, confirming band identity (Figure 5.9, compare to Figure 5.4). Coomassie-staining of the gel from which protein was transferred showed that migration of high-molecular weight species was poor, and that sEGFR501 was partially retained in the gel (Figure 5.9). Therefore, transfer conditions were altered to 16 h at 30 V, 4 °C from lower density SDS-PAGE (e.g. 10%) for cellular experiments with higher-molecular weight of EGFR.

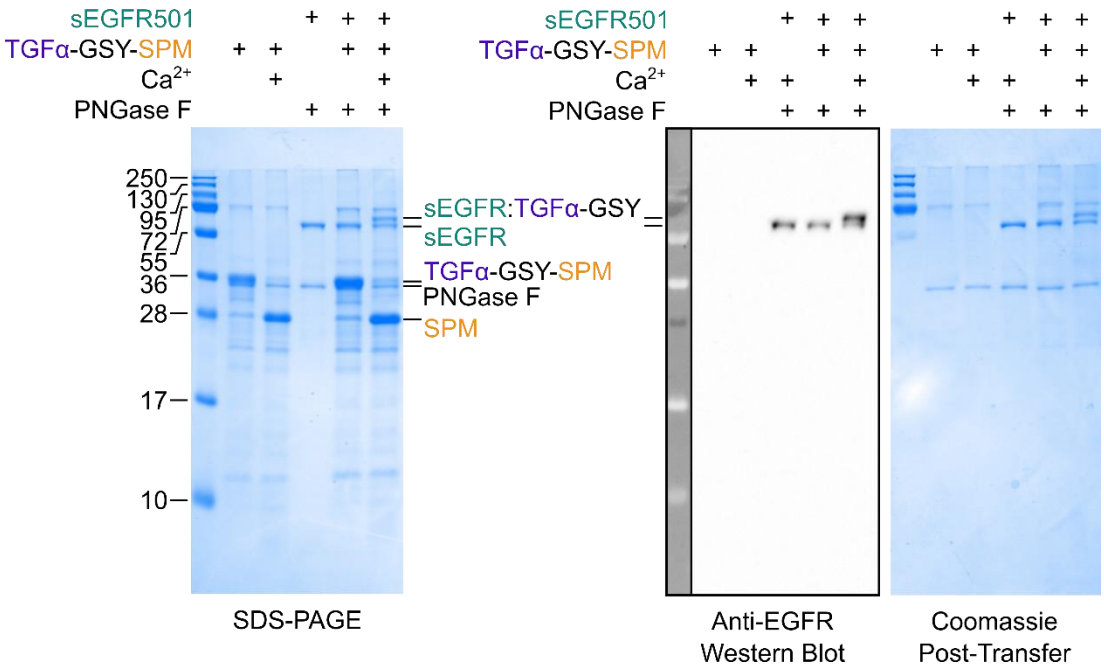


Figure 5.9. Western blot of *in-vitro* sEGFR501:TGFα-GSY conjugation. 50 μM TGFα-GSY-SPM was incubated with 5 μM sEGFR501 and 2 mM calcium as indicated for 1.5 h at 37 °C. sEGFR501 was deglycosylated with PNGase F before boiling with SDS-loading buffer and resolving on SDS-PAGE. Two gels were prepared, one for immediate Coomassie staining and one for Western blot transfer. Anti-EGFR detected both sEGFR501 and sEGFR501:TGFα-GSY.

5.4. Cellular NeissLock-conjugation

5.4.1. TGFα-GSY-SPM cell binding

Finally, following *in vitro* conjugation of TGFα-GSY-SPM to sEGFR501, I wanted to establish if TGFα-GSY-SPM could be used for conjugation to EGFR displayed on cells. I used A431 as a cell line displaying high levels of EGFR (Davidson *et al.*, 1987) for tests of TGFα-GSY-SPM conjugation. First, I confirmed specific binding of TGFα-GSY-SPM to A431 by comparison to TGFα[R42A]-GSY-SPM (wherein TGFα R42A reduces receptor binding (Defeo-Jones *et al.*, 1989), also see Figure 5.7). A431 cells were seeded on glass-bottom dishes and then incubated with 5 µM TGFα-GSY-SPM, TGFα-GSY-[DA]SPM or TGFα[R42A]-GSY-SPM as indicated for 1 h at 4°C. Subsequently, cells were washed and proteins detected with anti-His-antibody:Phycoerythrin (PE) conjugate.

Fluorescence imaging shows that both TGFα-GSY-SPM and TGFα-GSY-[DA]SPM effectively bind A431 cell membranes (Figure 5.10). However, TGFα[R42A]-GSY-SPM strongly reduced cell binding, indicating that the interaction is specific to TGFα (Figure 5.10).

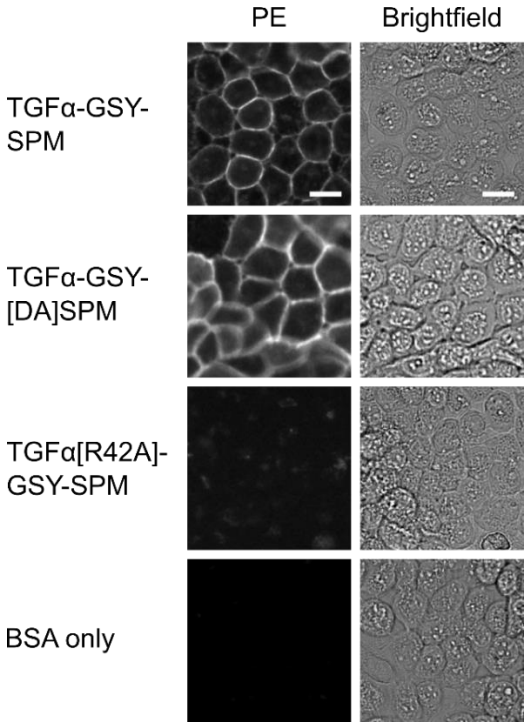


Figure 5.10. TGFα-GSY-SPM specifically binds to EGFR-presenting A431 cells. A431 cells were stained with 5 µM TGFα-GSY-SPM, TGFα-GSY-[DA]SPM, or TGFα[R42A]-GSY-SPM followed by anti-His-Phycoerythrin (PE) labelling. Fluorescence imaging revealed that TGFα[R42A]-GSY-SPM reduces specific binding to A431 cells. TGFα-GSY-SPM and TGFα-GSY-[DA]SPM both effectively bound A431 cell membranes. Scale bar: 20 µm.

5.4.2. Screening of cell conjugation conditions

I screened multiple conjugation conditions for conjugation of TGFα-GSY-SPM to arrive at an efficient method for cell conjugation. A431 cells were serum starved in DMEM before the experiment. Two dishes were treated with 80 µM Dynasore for 30 minutes to inhibit receptor endocytosis (Macia *et al.*, 2006). I further tested conjugation at different temperatures (4°C or 37°C) and different regimens of calcium induction: cells were treated with 1 µM TGFα-GSY-SPM in 50 mM HEPES, 150 mM NaCl, pH 7.40 supplemented with 5 mM $MgCl_2$ (HBS-M), at 4 °C or 37 °C. After 5 minutes at 37 °C or 35 min at 4 °C, the cells were washed, followed by incubation with 2 mM calcium as indicated. Alternatively, calcium at 2x concentration was added immediately with the protein solution or added to the cells without washing in between. After lysis, samples were processed in anti-TGFα and anti-EGFR Western blots.

For cell treated with Dynasore prior to incubation with TGFα-GSY-SPM and then calcium at 37 °C, a prominent high molecular weight band was detected with anti-TGFα antibody. Similarly, a prominent band was detected for cells which were simultaneously incubated with TGFα-GSY-SPM and calcium at 37 °C. At a high molecular weight, anti-EGFR detected EGFR in all samples, although with varying signal strength. The anti-TGFα and anti-EGFR detected high molecular weight bands both run concurrent to the 250 kDa ladder, suggesting that these represent EGFR (for anti-EGFR) as well as EGFR/TGFα-GSY conjugate (for anti-EGFR, anti-TGFα). Both co-incubation and Dynasore treatment appeared effective, although Dynasore-treated conjugation (1^{st} sample) showed comparatively low anti-EGFR signal. As co-incubation was deemed less likely to interfere with cellular processes, it was chosen for subsequent cellular experiments.

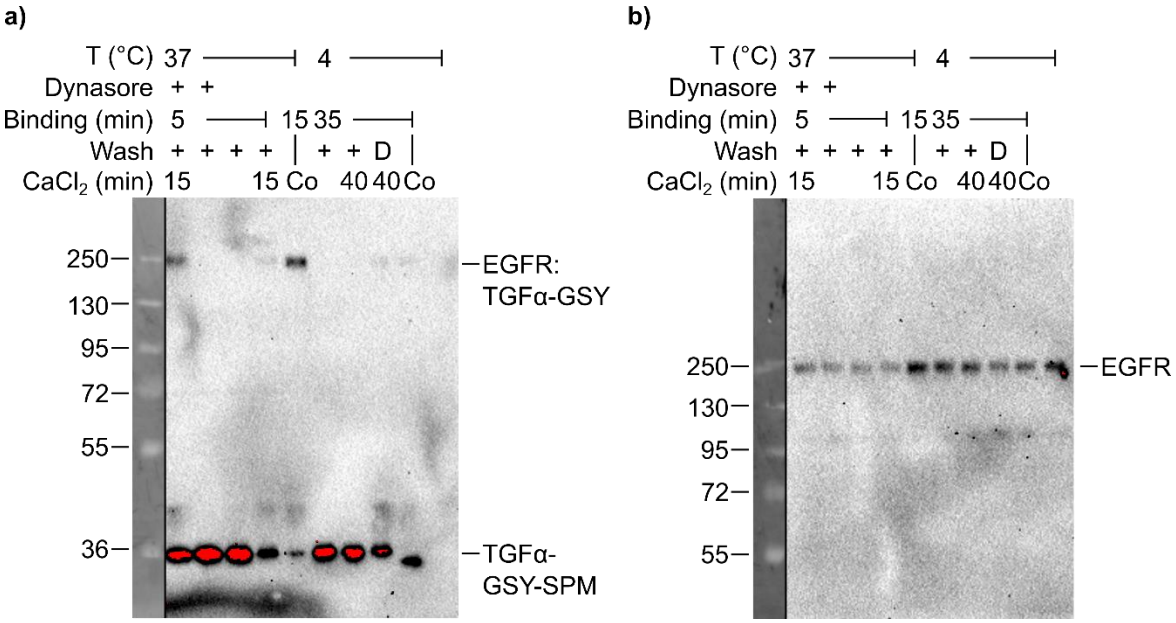


Figure 5.11. Screen of conditions for TGFα-GSY-SPM conjugation to EGFR on the cell surface. A431-cells were incubated with 1 µM TGFα-GSY-SPM at 37 °C for 5 min or 4 °C for 35 min, before washing and incubation with 2 mM calcium for additional time as indicated. Optionally, cells were treated with 80 µM Dynasore before conjugation. Optionally, cells were not washed before calcium addition ("D"). Optionally, calcium was added to 2 mM immediately following TGFα-GSY-SPM and cells were co-incubated ("Co"). Cells were harvested by hot SDS lysis and sonication. Western blots with anti-EGFR or anti-TGFα shown. Red: Overexposure. Inhibition of endocytosis with Dynasore as well as co-incubation of TGFα-GSY-SPM with cells at 37 °C provided effective conjugation conditions.

5.4.3. Sensitivity of cellular conjugation to TGFα-GSY-SPM mutant design

Next, I wanted to confirm that this conjugation was dependent on specific TGFα-GSY-SPM binding to EGFR. A431 cells were incubated for 15 min with 0.5 µM TGFα-GSY-SPM, TGFα-GSY-[DA]SPM or TGFα[R42A]-GSY-SPM in HBS-M, with or without 2 mM calcium as indicated. Anti-TGFα staining showed conjugation for TGFα-GSY-SPM, with a single band at the previously observed molecular weight for EGFR:TGFα-GSY (Figure 5.12a, compare to Figure 5.11). No conjugation was observed with either TGFα-GSY-[DA]SPM or TGFα[R42A]-GSY-SPM, indicating that conjugation was dependent on both cleavage and binding. For both TGFα-GSY-[DA]SPM or TGFα[R42A]-GSY-SPM, it appeared as though the amount of unprocessed protein decreased with calcium incubation.

No TGFα staining was detected for cells incubated with only buffer (1st lane left to TGFα-GSY-SPM wt). Furthermore, immediately after incubation with TGFα-GSY-SPM variants and calcium, the supernatant was recovered from cells and the reaction was mixed with EDTA.

The samples were boiled with SDS-loading buffer and resolved on SDS-PAGE followed by Coomassie staining. Coomassie staining of these supernatant samples confirmed that TGFα-GSY-[DA]SPM did not undergo calcium-dependent processing (Figure 5.12b). TGFα-GSY-[DA]SPM and TGFα[R42A]-GSY-SPM samples portrayed the same changes in gel mobility as observed in Figure 5.7 and on Western blot (Figure 5.12a). Coomassie staining also showed that most TGFα-GSY-SPM and TGFα[R42A]-GSY-SPM in the supernatant had processed during the experiment (Figure 5.12a,b).

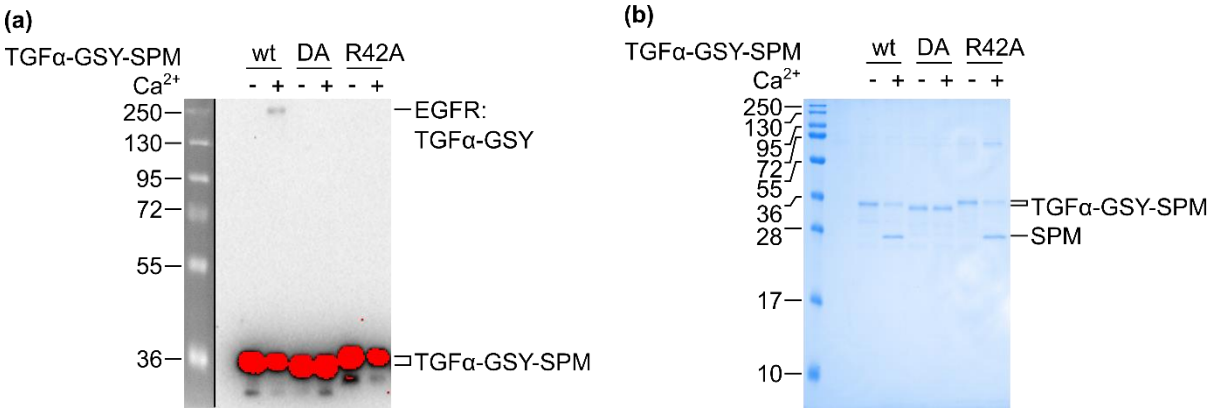


Figure 5.12. TGFα-GSY-SPM conjugation to EGFR-presenting cells is dependent on both TGFα-mediated binding and SPM-mediated processing. (a) A431 cells were incubated with 0.5 µM TGFα-GSY-SPM variant and 2 mM calcium as indicated for 15 min. Cells were harvested with hot SDS lysis and sonication, samples were analysed by Western blot using anti-TGFα primary antibody. Conjugation to EGFR was observed for TGFα-GSY-SPM after calcium processing, but not for TGFα[R42A]-GSY-SPM or non-cleaving TGFα-GSY-[DA]SPM. Transillumination of ladder was spliced with luminescence signal. 1st lane after ladder: Cells incubated with only buffer (not visible). Red: Overexposure. (b) Supernatant from (a) boiled with SDS-loading buffer, resolved on SDS-PAGE, with Coomassie staining.

5.4.4. Initial evidence of TGFα/EGFR conjugate fate

Finally, I wanted to obtain initial evidence to understand the processing of conjugated TGFα. Herein, A431 was grown in a 24-well plate and then starved for 44 h in DMEM medium. At varying points prior to cell lysis, cells were conjugated for 15 min at 37°C with 1 µM TGFα-GSY-SPM and 2 mM $CaCl_2$ in HBS-M, after which cells were placed back in DMEM medium. Optionally, TGFα-GSY-SPM was diluted in DMEM (which already contains 1.8 mM calcium ions). After conjugation of the final samples, all cells were washed and stored at -80 °C. Cells were lysed with hot SDS lysis-buffer; due to the small volumes, samples could not be subjected to sonication (increasing viscosity). Nevertheless, samples were boiled with SDS-

loading buffer and resolved on SDS-PAGE before transfer. Samples were resolved in Western blot with anti-EGFR or anti-TGFα.

As previously, TGFα-GSY-SPM conjugated to EGFR, whereas no such evidence was observed for TGFα-GSY-[DA]SPM. Herein, the strongest conjugate signal was observed for the sample 1 h post-conjugation. Over time, I saw a decrease in EGFR/TGFα-GSY conjugate, indicating degradation of the conjugated species (Figure 5.13a). For anti-EGFR staining, I observed variation in EGFR concentrations indicative of uneven lysis, loading, or transfer, with almost complete loss for TGFα-GSY-[DA]SPM in DMEM (Figure 5.13b). However, EGFR signal was relatively constant for TGFα-GSY-SPM conjugation in HBS.

Finally, for the sample corresponding to prolonged incubation of cells with TGFα-GSY-SPM in DMEM, I observed evidence of efficient conjugation. This could provide an even easier method for specific cell conjugation since it requires no change of medium.

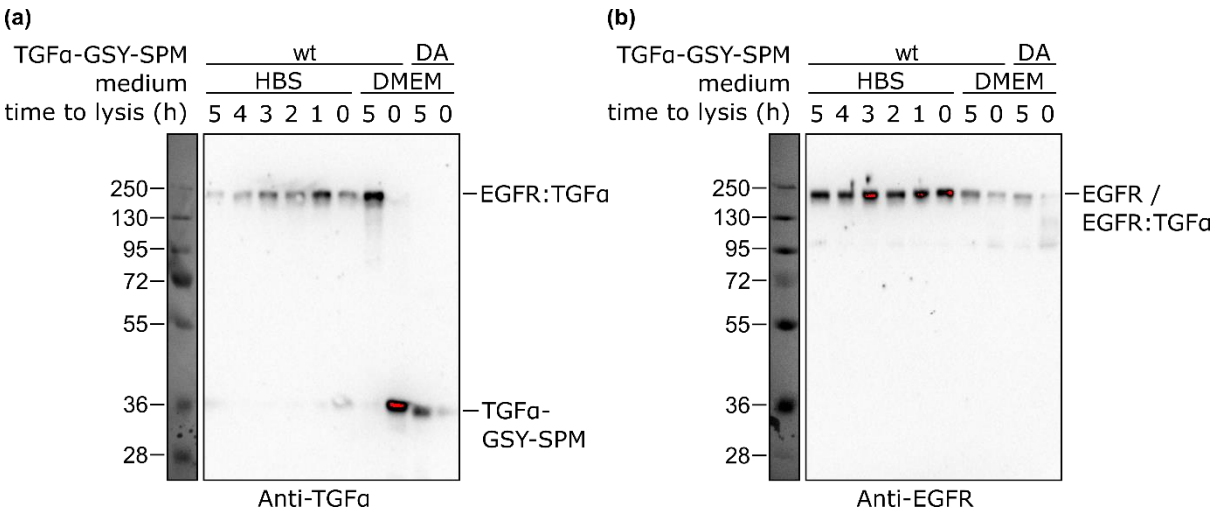


Figure 5.13. Initial evidence of EGFR:TGFα-GSY degradation. Western blots of EGFR:TGFα-GSY conjugation. A431 seeded in a 24-well plate were incubated with 1 μM TGFα-GSY-SPM or TGFα-GSY-[DA]SPM and 2 mM calcium in HBS-M for 15 min. Optionally, protein was added to DMEM instead of HBS-M, which contains calcium. Cells were lysed after indicated time. a) Anti-TGFα shows degradation of EGFR:TGFα-GSY conjugate over time. Red: Overexposed. b) EGFR and EGFR:TGFα-GSY conjugate are labelled with anti-EGFR.

5.5. Discussion

With TGFα-GSY-SPM, I have demonstrated the first NeissLock-probe for cellular conjugation, with initial evidence of receptor fate after conjugation. As for the OAZ/ODC *in vitro* model system, various insights from the TGFα/EGFR cellular model guide future *in vitro* NeissLock design.

I obtained multiple lines of evidence that TGFα-GSY-SPM is capable of specifically binding and conjugating to EGFR, both when presented on cells or as sEGFR501 *in vitro*. Conjugation was reliant on this interaction and could be reduced by introduction of an TGFα R42A mutation. This also confirmed specificity of NeissLock in general and could be compared to investigation of OAZ/ODC mutants (Chapter 4.6). Overall, TGFα-GSY-SPM conjugation *in vitro* was less efficient than conjugation of OAZ-Y-SPM / OAZ-GSY-SPM to ODC (Chapter 4). To reach comparable quantities of TGFα-GSY conjugation to sEGFR501, a substantial excess of TGFα-GSY-SPM to sEGFR501 was required (Figure 5.5, Figure 5.7). For OAZ-Y-SPM / OAZ-GSY-SPM conjugation to ODC, free OAZ-Y or OAZ-GSY was effectively consumed even at small molar excess (Chapter 4, Figure 4.10, Figure 4.11). However, separation of TGFα-GSY-SPM by gel chromatography and subsequent conjugation tests suggested that conjugation efficiency can be improved by removing heterogenous TGFα-GSY-SPM from the preparation (Figure 5.7).

For OAZ-Y-SPM and OAZ-GSY-SPM conjugation to ODC, a major hurdle to overcome had been issues with the quality of purified protein if OAZ-derived proteins had not been consistently reduced. For TGFα-GSY-SPM, I expected the protein to be only partially active, as TGFα is prone to misfolding: The three disulfide bonds which must correctly form in the 50 aa protein allow 75 different configurations, commonly resulting in partial activity of TGFα samples (Groenen *et al.*, 1994). Size exclusion chromatography suggested differential folding of TGFα-GSY-SPM, as the protein eluted around two separate peak position (Figure 5.6). The conjugation efficiency appeared to differ substantially between samples purified from either peak (Figure 5.7). To improve the conjugation efficiency of TGFα-GSY-SPM, it therefore

appears promising to further optimise the purification process. As SPM does not appear to be compatible with the mammalian secretory pathway without further engineering, TGFα-GSY-SPM cannot currently be produced in a more native fashion. Nevertheless, alternative purification methods for TGFα have been described. Winkler *et al.* purified TGFα as a TrpLE fusion protein (TrpLE: tryptophan operon tag for insoluble expression (Hwang *et al.*, 2014)) from inclusion bodies (Winkler *et al.*, 1986). Herein, refolding has been shown to increase the activity of fusTGFα by a factor of 20-50 (Winkler *et al.*, 1986). Similarly, after cleavage of the tag and HPLC purification, they found that the activity of TGFα corresponding to two peaks in HPLC differed by >99% (Winkler *et al.*, 1986).

Further to purification conditions, a second question is the design of the TGFα-GSY-SPM probe itself, which for instance could influence self-reaction of the SPM-activated NeissLock probe (Figure 5.4, compare to OAZ/ODC, Figure 4.9). Self-reaction of the activated NeissLock probe could inactivate the protein before conjugation and can be faster than hydrolysis alone (see Figure 4.4, Figure 4.9). One consideration in the design of TGFα-GSY-SPM was the removal of N-terminal tags (i.e. TGFα-GSY-SPM-**H$_6$** instead of **H$_6$**-TGFα-SPM), as I suspected that the N-terminus could be a cause of self-reaction (Figure 5.14). Although the effect of the position of the His-tag on TGFα-GSY-SPM-**H$_6$** (instead of **H$_6$**-TGFα-GSY-SPM) was not specifically validated, Sheryl Lim later observed that disordered ODC termini can be an unexpected conjugation site for OAZ-GSY-SPM (Scheu *et al.*, 2021).

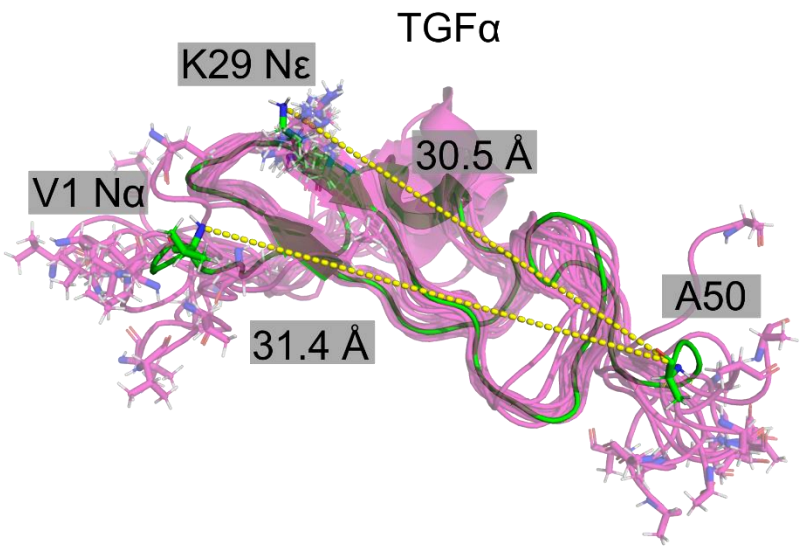


Figure 5.14. Overlay of TGFα NMR-states highlighting distance from the TGFα C-terminus to K29 Nε and the N-terminal Nα. TGFα solution structure PDB ID 1yuf (Moy *et al.*, 1993) is represented. The first NMR state is shown in green with no transparency, and the distance from A50 atom C to K29 atom Nε and V1 atom Nα is annotated for this state, with these residues shown in stick representation. 15 additional states in 1yuf are overlayed in magenta with transparency. A50, K29 and V1 are again represented as sticks, highlighting the flexibility of N-terminus and C-terminus.

Furthermore, conjugation of TGFα-GSY-SPM to EGFR could be less effective than conjugation of OAZ-Y-SPM or OAZ-GSY-SPM to ODC due to factors inherent to the protein complex. In the crystal structure PDB ID 1mox (Garrett *et al.*, 2002), the C-terminus of TGFα-GSY-SPM bound to EGFR appears more obstructed than the C-terminus of OAZ in the OAZ/ODC complex, which could be alleviated by adequate linker design. Alternatively, the local environment could be more activated for ODC/OAZ conjugation than for TGFα-GSY-SPM conjugation to EGFR, e.g. due to local pKa modulation or other factors affecting lysine-regioselectivity (Matos *et al.*, 2018; R. Liu *et al.*, 2019). Here, studies with additional protein complexes will provide additional insight, as well as further variations of TGFα-Neisslock probes and similar EGFR-targeting proteins. Nevertheless, crosslinking was identified at the predicted site K465 using tryptic LC-MS/MS, highlighting that the NeissLock principle worked as intended from NeissDist. Like for OAZ-Y-SPM, it would be interesting to see if other crosslinking sites could be identified with further MS study, e.g. using digestion with different enzymes. As K465 is a close site, K463 and K465 could also be subjected to point mutation.

Other than optimisation, what are potential directions in which to develop this technology? Despite the aforementioned limitations, TGFα-GSY-SPM purified from *E. coli* proved to be exceptionally useful for *in vitro* and cellular studies, if deployed at a high ratio to mitigate potentially reduced activity (herein 0.5-5 μM instead of typical concentrations e.g. 1-10 nM (Reddy *et al.*, 1996), MCF-7 16 nM EGF / CHO 50 nM EGF (Freed *et al.*, 2017), Hep2 ~10 nM TGFα or EGF (Roepstorff *et al.*, 2009)). TGFα-GSY-SPM was able to specifically bind to EGFR-presenting A431 cells and specifically conjugate to both sEGFR501 *in vitro* and EGFR on the cell surface. Importantly, this binding and conjugation could be abrogated with a single point mutation. After activation by both EGF and TGFα, EGFR is rapidly endocytosed (around 50% after 15 minutes incubation with saturating concentrations of TGFα at 37°C) (Ebner and Derynck, 1991; Roepstorff *et al.*, 2009). In a screen for efficient NeissLock conjugations (Figure 5.11), co-incubation of cells with TGFα-GSY-SPM at the same time as calcium at 37°C was an effective strategy, as was incubation of Dynasore-treated cells with TGFα-GSY-SPM followed by a wash and calcium treatment. If cells were not Dynasore-treated, conjugation with TGFα-GSY-SPM was less effective if cells had been washed between protein and calcium addition. This suggests that TGFα-GSY-SPM becomes insensitive to calcium-incubation in a Dynasore-sensitive manner, potentially involving TGFα-GSY-SPM endocytosis before calcium application.

After conjugation, I have observed initial evidence that covalently linked EGFR:TGFα-GSY is degraded within a few hours, despite limitations of the lysis and Western blot protocols used herein (e.g. in detection of anti-EGFR in Western blot). Although hot SDS lysis is commonly recommended for membrane lysis, RIPA buffer proved more successful (Scheu *et al.*, 2021). After additional stabilization of growth factor starvation and lysis conditions using RIPA buffer, Sheryl Lim in the Howarth group has confirmed my initial observations on degradation (Scheu *et al.*, 2021). Furthermore, Sheryl Lim has been able to show effective modulation of cellular signalling, wherein EGFR-downstream signalling was modulated in a covalent-conjugation dependent manner (Scheu *et al.*, 2021).

Non-covalently bound TGFα commonly dissociates from EGFR upon acidification of the endosome, allowing for the recycling of EGFR to the cell surface (Ebner and Derynck, 1991; Roepstorff *et al.*, 2009). To give a potential explanation of differential signalling, covalent conjugation of TGFα to EGFR could retain TGFα in an inactive bound state, encouraging immediate rebinding after recycling, or lead to an EGF-like continued anchoring, preventing dissociation altogether. Figure 5.15 summarizes this hypothesis of how complex stabilization via NeissLock-conjugation could affect TGFα/EGFR trafficking compared to normal ligand binding. Furthermore, Figure 5.16 presents an overview of how the different conjugation conditions described herein could affect efficiency of ligand/receptor conjugation and ultimately influence receptor fate.

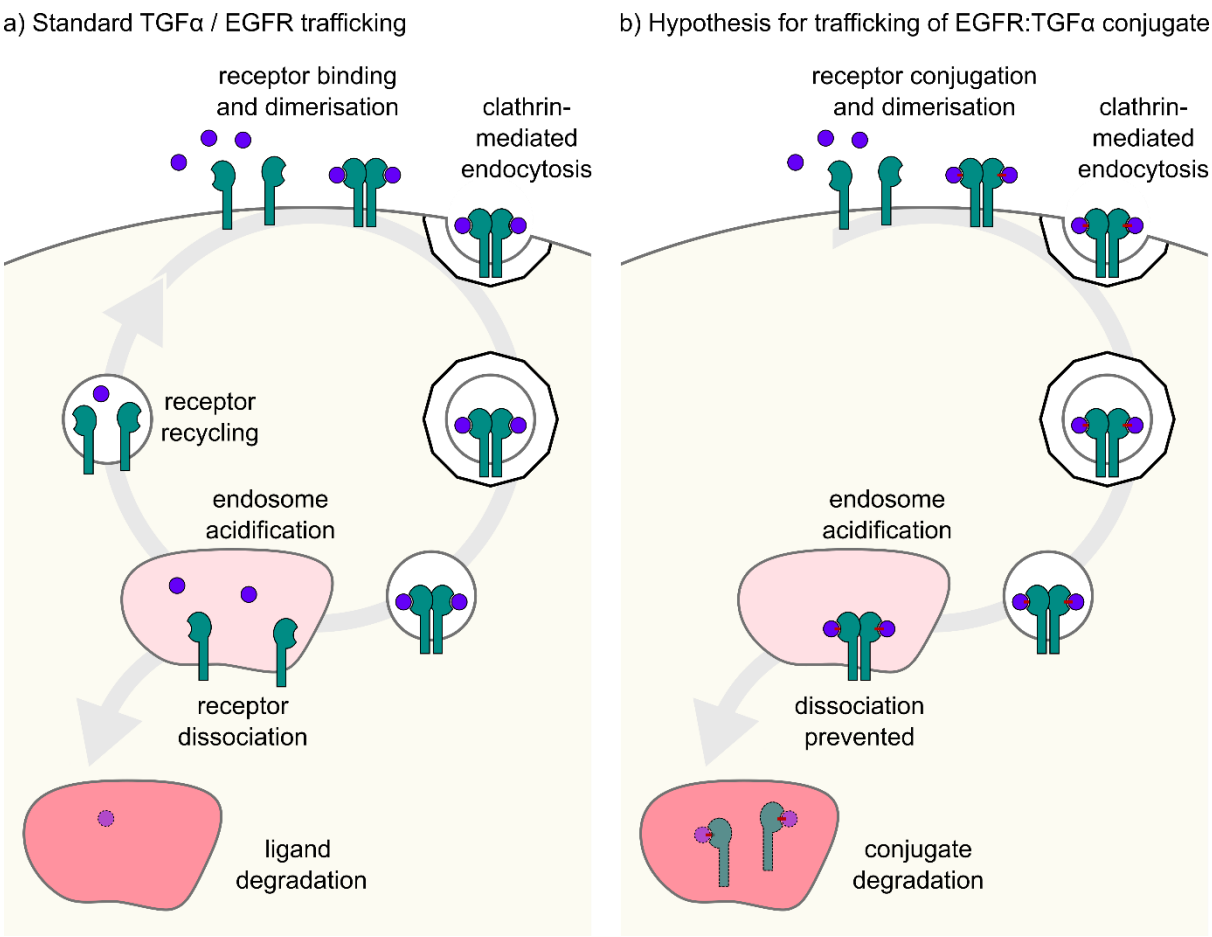


Figure 5.15. Hypothesis for the influence of covalent NeissLock-conjugation on TGFα/EGFR complex fate. a) In standard TGFα/EGFR trafficking, TGFα (purple) dissociates from EGFR (green) upon acidification of the endosome. Inactivation of the receptor promotes recycling of receptor and ligand (Ebner and Derynck, 1991; Roepstorff *et al.*, 2009). Herein, work by Roepstorff *et al.* indicates that TGFα does not stimulate degradation of EGFR, although partial degradation of TGFα has been reported by Ebner & Derynck (Ebner and Derynck, 1991; Roepstorff *et al.*, 2009). b) Covalently conjugated TGFα is unable to dissociate from EGFR, potentially causing its continued activation. Similar to EGF, covalently conjugated TGFα could cause sustained signalling and ultimately promote degradation of EGFR. This effect could even exceed EGF, which is more pH resistant than TGFα (dissociation from EGFR at pH 5.5 instead of pH 6.5), but for which partial recycling has been observed (Ebner and Derynck, 1991; Roepstorff *et al.*, 2009). Figure based on evidence and schematics of differential TGFα and EGF trafficking by Ebner & Derynck and Roepstorff *et al.* (Ebner and Derynck, 1991; Roepstorff *et al.*, 2009). Also compare to data in Figure 5.13 and (Scheu *et al.*, 2021).

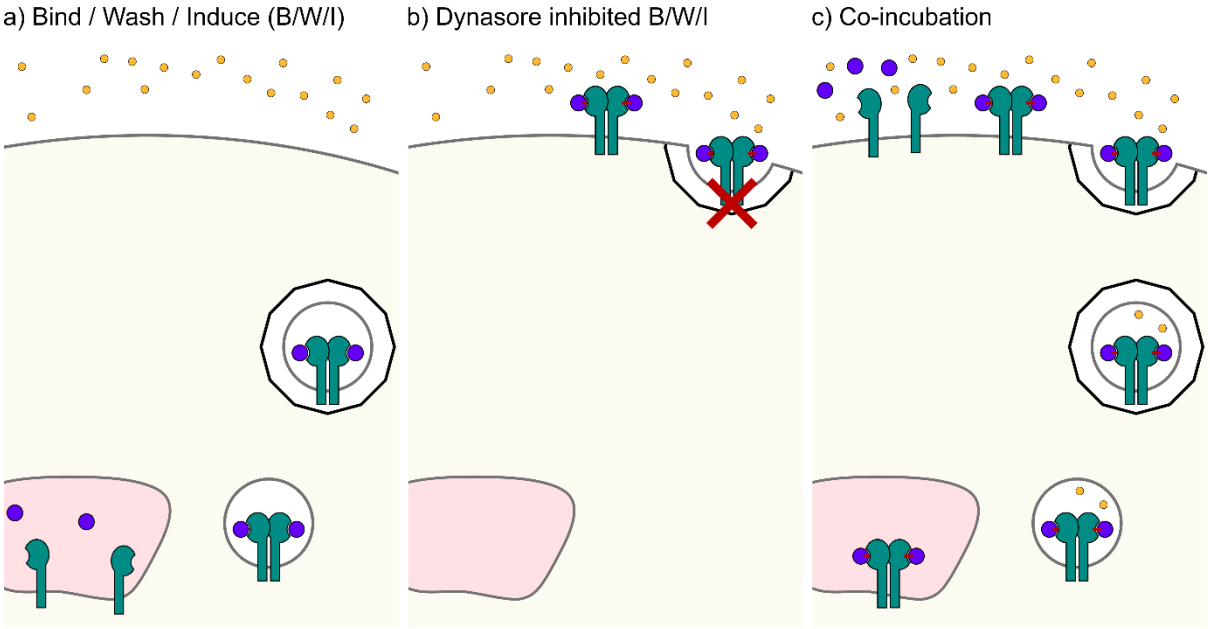


Figure 5.16. Potential causes for differences in NeissLock efficiency between cellular conjugation conditions. a) In the "Bind/Wash/Induce" (B/W/I) condition, cells are first incubated with TGFα-GSY-SPM (TGFα in purple). After a given period, cells are washed to remove excess ligand. Then, SPM processing is induced with calcium (yellow). Rapid endocytosis of EGFR (green) potentially limits the access of calcium to TGFα-GSY-SPM, preventing SPM activation and covalent conjugation of TGFα to EGFR. Note that this effect would likely be temperature dependent (e.g. at 37°C, Figure 5.11). Prior work on EGFR endocytosis involved equilibration of cells with TGFα at 4°C before internalisation was observed at 37°C (Ebner and Derynck, 1991; Roepstorff *et al.*, 2009). This suggests that NeissLock-mediated conjugation at 4°C could be affected or compounded by a different effect, such as reduction in cleavage or conjugation rates (compare co-incubation at 37°C to co-incubation at 4°C, Figure 5.11). b) If cells are treated with Dynasore before ligand binding, clathrin-mediated endocytosis of the receptor is inhibited (red cross). Bound ligand/receptor complexes would thereby remain surface exposed, and calcium could effectively induce NeissLock conjugation (red lines). c) Alternatively, cells are incubated with TGFα-SPM and calcium at the same time. Herein, calcium could promote NeissLock-conjugation of TGFα-SPM to EGFR from the time of complex formation at the cell surface, and potentially even sustain SPM activity during endocytosis. Note that calcium concentrations decrease substantially in the early endosome due to H^+/Ca^{2+} antiport before calcium import during further maturation (Lloyd-Evans *et al.*, 2010). Also compare to Figure 5.11 (supporting data) and Figure 5.15 (supporting schematic).

It will be interesting to further investigate the subcellular localisation of the covalent TGFα/EGFR complex and its degradation, for instance how the conjugation affects EGFR trafficking. Although fluorescent microscopy was used herein to validate cellular recognition by TGFα, it is not sufficient to study EGFR trafficking. In a comprehensive study of the effect of different mitogens on EGFR fate, Roepstorff *et al.* effectively used both flow cytometry and fluorescent imaging (Roepstorff *et al.*, 2009). Herein, flow cytometry provided quantitative data of surface-available EGFR via anti-EGFR antibody labelling, while confocal microscopy was used to observe and quantify co-localisation of EGFR with the endosomal marker EEA1

(Roepstorff *et al.*, 2009). Similar analyses could be used to improve understanding of EGFR trafficking after NeissLock-conjugation with TGFα-GSY-SPM. In addition, acid wash of mitogens was previously used to quantify ligand release from EGFR, either directly (for release of radiolabelled EGF) or by subsequent probing of receptor availability with (commercially available) radiolabelled EGF (Ebner and Derynck, 1991; Roepstorff *et al.*, 2009). Similarly, study of TGFα-GSY-SPM (either suitably photolabelled or radiolabelled, or by probing via labelled EGF) after acid wash could provide additional insight into the degree of (covalent) TGFα NeissLock-conjugation as well as TGFα fate, as could the release of undegraded ligand after recycling (e.g. via differential solubility of degraded peptides) (Ebner and Derynck, 1991).

Depending on the precise fate of the covalent TGFα/EGFR complex, N-terminal fusion of therapeutic agents to TGFα-GSY-SPM could become an exciting mechanism for cell therapy. Enzyme-antibody conjugates have been used to facilitate localised conversion of a pro-drug to an active drug, directly at the surface of a target cell or in the cell endosome, which could similarly be implemented as an enzyme-TGFα covalent NeissLock drug (Tietze and Krewer, 2009). A convenient conjugation handle can be introduced to TGFα-GSY-SPM to rapidly iterate on multiple such assays, e.g. by SpyCatcher-mediated conjugation (Zakeri *et al.*, 2012).

In summary, I have hereby provided a promising starting point for conjugation of cells using NeissLock. EGFR provides a promising target, and EGFR-specific NeissLock is already an avenue to exciting applications. Moreover, with TGFα/EGFR and OAZ/ODC I successfully derived two out of two NeissLock systems selected from NeissDist (Chapter 3, Chapter 4). Therefore, the success of the TGFα/EGFR cellular conjugation system provides an encouraging foundation for the exploration of further routes to cellular conjugation (such as described in Figure 3.7, Figure 5.1).

6. Summary and Future Work

In this work, I have validated SPM-mediated chemical activation of recombinant proteins for conjugation to native proteins, termed ‘NeissLock’. NeissLock relies solely on the standard genetic code, and high conjugation yields can be achieved (Chapter 4, Chapter 5). Key conjugation parameters have been explored, such as complex formation, linker design, pH, and mutagenesis of target residues (Chapter 4, Chapter 5). With disCrawl and NeissDist, I have provided an initial tool for evaluation of NeissLock candidates and a database of hundreds to thousands of protein complexes with potential for application to NeissLock (Chapter 3). *In vitro* conjugation was shown for two out of two pre-selected model complexes, and tryptic LC-MS/MS and mutagenesis confirmed the fundamental prediction of target residues by NeissDist (Chapter 4, Chapter 5). Finally, I have shown that NeissLock can facilitate conjugation to live cells (Chapter 5).

With a promising scope of evidence, this work constitutes just a start for the investigation of NeissLock-mediated protein conjugation technologies. Additional experimental work will improve our understanding on the mechanisms of NeissLock conjugation: can SPM folding actively interfere with target protein complex formation, i.e. does SPM cleavage always or partially occur within a bound complex? What is the precise requirement for anhydride positioning and are there localized effects in target conjugation? With additional *in vitro* study of NeissDist-selected complexes and the SPM NMR structure now available (Kuban *et al.*, 2020), it will likely be possible to increase the predictive power of NeissDist and optimize the design of NeissLock conjugation in future complexes. The study of thermodynamic and kinetic parameters of ODC / OAZ-GSY-SPM by Sheryl Lim gave a brief introduction to the reliance of NeissLock specificity and yield on target complex formation, wherein she effectively increased K_D and k_{off} by site-directed mutagenesis of OAZ to abolish covalent conjugation (Figure 4.20, Chapter 4.7, published data (Scheu *et al.*, 2021)). Under the assumption of sufficient pre-equilibration (i.e. co-incubation of target and NeissLock probe until equilibrium is reached, prior to calcium-induced probe activation), K_D could constitute a useful indicator

for NeissLock design, that is to derive protein occupancy at time of anhydride generation. Herein, the NeissLock process differs from traditional covalent probes in that it generates a reactive intermediate *in situ*, potentially uncoupling initial binding kinetics from conjugation kinetics (especially considering the addition of calcium as 'timepoint 0' as opposed to the addition of the NeissLock probe). Furthermore, NeissLock uses a relatively unstable intermediate for conjugation. Although comparison to the short lifetimes of photochemical probes does not seem appropriate, a comparison to pharmacokinetic parameters regarding the clearance of unconjugated covalent drugs might provide useful insight into *in vitro* kinetic properties of a labile reactive species with half-life in the minutes (*Note that this comparison would not account for potential reversible inhibitory effects of inactivated NeissLock probes*). Strelow *et al.* modeled the effects of drug clearance on target occupancy for two different covalent drugs (Strelow, 2017); herein, it becomes apparent that effective binding and conjugation are a requirement to achieving high target occupancy when considering performance of a drug subject to degradation (Strelow, 2017). Pre-equilibration should be especially relevant for low K_D values arising at exceptionally low k_{off} values, assuming high enough conjugation rate (k_{irr}) (i.e. featuring low k_{on} compensated by k_{off} far below k_{irr}, rendering the actual value of k_{off} kinetically irrelevant). Conversely, if both k_{on} and k_{off} values are high, pre-equilibration might not be applicable or yield appreciable benefit despite low K_D (considering fast equilibration times). Adequate equilibration conditions might thereby provide means to control NeissLock-specificity *either* kinetically *or* thermodynamically (while considering effects on equilibration and occupancy between *in vitro* and *in vivo* assays, such as actual concentrations (Tonge, 2018; Jarmoskaite *et al.*, 2020), and in comparison to the potential for "*kinetic* selectivity […] in the absence of *thermodynamic* selectivity" for classical covalent drugs (Tonge, 2018)). Considering a classical covalent drug, Pettinger *et al.* impressively execute on general kinetic considerations to achieve rapid covalent conjugation, while highlighting the importance of probe design on k_{irr}, e.g. in precise positioning of the reactive group relative to the target protein, and its interplay with k_{on} and k_{off} (Pettinger *et al.*, 2019). For NeissLock, cellular experiments indicate the importance of further assessing

potential pre-equilibration effects on target proteins, such as target internalisation (see Figure 5.16), and potential SPM-mediated effects, such as target dissociation and rebinding, for fine-tuning of future NeissLock designs.

NeissLock then has the potential to open a broad range of applications. Crosslinking of proteins via SPM could be used for the assembly of biological complexes and nanoassembly; the covalent crosslinking of oligomeric complexes can be used to increase the stability of such assemblies (Banerjee and Howarth, 2018). For example, the heptameric IMX313 scaffold has been used as a platform for antigen decoration via SpyTag/SpyCatcher; wherein the subunits in this scaffold are interconnected via disulfide bonds (Brune *et al.*, 2017). Another application could be the formation of covalent protein-based gels with defined interaction sites, as has been achieved using the SpyTag/SpyCatcher system (Gao *et al.*, 2016). Potential benefits of NeissLock could include ease of multiplexing by choice of distinct binder-target pairs, including many-to-one targeting by engaging multiple epitopes on a single target protein, as well as simple control of crosslinking by calcium induction.

NeissLock could also benefit applications which are essentially limited by target dissociation, for instance in magnetic cell capture (which was improved by enhancing complex stability and valency (Jain *et al.*, 2013; Fierer *et al.*, 2014)). In another application, an SPM-activated protein ligand could be used to bestow new characteristics to an endogenous protein, e.g. to introduce a permanent fluorescent label, for delivery of drugs using endogenous carriers (such as via red blood cells (Villa *et al.*, 2016)), or to deliver enzymes for pro-drug conversion (Tietze and Krewer, 2009) (Figure 6.1a). In addition to cell conjugation described in Chapter 5, immediate follow-up work by Sheryl Lim suggested that NeissLock conjugation of TGFα to EGFR on the cell surface (via TGFα-GSY-SPM) can differentially modify cell biology, i.e. to achieve pronounced stimulation of pSTAT1 not found with non-covalent binding (Scheu *et al.*, 2021).

In parallel to direct applications, the accelerated investigation of NeissLock could also advance other uses of SPM. Aside from the use as a purification strategy, another potential application

of SPM could involve the calcium-triggered cleavage of the polypeptide chain for functional purposes, e.g. to relieve inhibition in a protein domain. Compared to similar self-cleaving activity in other proteins, FrpC SPM offers a short minimal N-terminal fragment (D) and simple mode of activation (Sadilkova *et al.*, 2008). Furthermore, under the assumption of an S,N-acyl shift (Sadilkova *et al.*, 2008) a plethora of coupling agents for native chemical ligation would become available, for example summarized by Conibear *et al.* (Conibear *et al.*, 2018). Upon activation, these could then be coupled to any protein genetically fused to SPM (Metzner, 2017). Such an approach has successfully been described using inteins, termed expressed protein ligation (Muir *et al.*, 1998). For instance, SPM could be used in the covalent modification of proteins for drug deployment, such as antibody-drug conjugates, which likewise has been achieved using inteins (Möhlmann *et al.*, 2011). Another intriguing application of SPM could be the production of cyclic peptides. Aside from native chemical ligation by intramolecular conjugation to N-terminal Cysteine after chemical synthesis (Clark and Craik, 2010), split-inteins have previously been utilized to circularise peptides (Camarero *et al.*, 2007). Similarly, it could be possible to achieve peptide circularisation via native chemical ligation using SPM, or peptides might be circularised by reaction with the N-terminus or internal lysines.

In such applications, adequate regioselectivity for reaction with the aspartic anhydride should be considered (due to the availability of two reactive sites, i.e. to form a [native] peptide or [non-native] isopeptide bond). Prior analyses of protected aspartic anhydrides suggest that solvent effects are a prime factor in determining regioselectivity for reaction with small chemicals (Huang *et al.*, 1997; Ibatullin and Selivanov, 2009). For instance, Huang *et al.* report the reaction of aniline with Fmoc-protected Asp yielding product in a 3:1 ratio of a peptide bond over an isopeptide bond in benzene, but to a 1:6 ratio in DMSO; the α-carbonyl would be chemically activated in a solvent-sensitive manner, potentially via “an intramolecular hydrogen bond […] between the hydrogen on the α-amino nitrogen and the oxygen of the α-carbonyl” that is abrogated in aprotic polar solvent (Huang et al., 1997). They also conferred

(weaker) steric effects towards the accessible β-carbonyl by comparison of protective groups (i.e. Cbz/Fmoc/Boc), although noting that “this effect is more pronounced in the case of glumatic acid anhydrides” (Huang et al., 1997). Regioselectivity should likewise be assessed in SPM-activated reactions, especially for applications in which it is a requirement to obtain uniform conjugation products. In aging proteins, iso-aspartate can convert from asparagine or aspartate via a succinimide intermediate (Johnson *et al.*, 1989; Aswad *et al.*, 2000). Herein, products of specific reaction of L-isoaspartyl methyltransferase with iso-aspartate containing proteins can be quantified via HPLC (Johnson *et al.*, 1989; Johnson and Aswad, 1991), and this method could similarly be used to assess isopeptide formation in NeissLock reactions.

Finally, NeissLock itself could be made more potent by protein engineering, e.g. to increase SPM-mediated cleavage rate, to implement an SPM with reduced length or altered ion activation, or to entirely alter aspects of the NeissLock concept. For instance, a ‘3-part’ NeissLock approach could be envisioned, in which a third protein non-covalently mediates the interaction between an SPM-fused peptide or protein and a target protein (Figure 6.1b). Sophisticated linker design could enable precise positioning of the anhydride, away from the primary binding site (Figure 6.1c). Protection of the reactive species by occlusion could protect reaction in a pre-activated state, similar to complement proteins (Janssen *et al.*, 2005, 2006; Wiesmann *et al.*, 2006), e.g. before protein-protein binding. A “split” variant of SPM (unpublished work by Sheryl Lim) could be derived to self-process only upon pre-mediated reconstitution, e.g. during binding of two co-localized protein domains, similar to the Co-LOCKR system (Lajoie *et al.*, 2020). Alternatively, quenching with small molecules could potentially reduce unspecific conjugation.

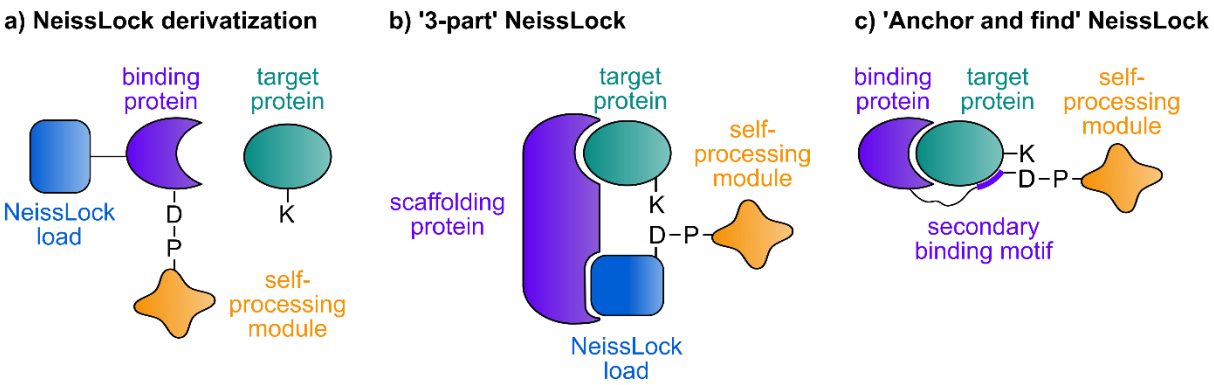


Figure 6.1. Complex NeissLock designs. a) NeissLock could be derivatized to include additional functionality, e.g. to anchor enzymes to target cells for localized pre-drug conversion (as for antibody-enzyme conjugates (Tietze and Krewer, 2009)). b) A scaffolding protein (or non-covalently attached binding protein) could facilitate direct conjugation of a NeissLock load, precluding conjugation of a binding protein in the final assembly. c) A secondary binding motif could be used for precise positioning of the reactive anhydride, facilitating NeissLock conjugation distant from the primary protein complex. Such a motif could also be introduced to a NeissLock load described in b).

In summary, SPM-mediated NeissLock conjugation provides the first example of rationally designed integration of natural domains to facilitate generation of reactive protein activation for protein-protein conjugation. Herein, in situ generation of a protein anhydride enables the use of a higher reactivity than traditional proximity-enabled probes and UAAs (Chapter 1.2) for protein conjugation. While other recent work trying to accomplish this goal relies on complex processes, such as ‘Genetically Encoded Chemical COnversion’ (GECCO) (Yang *et al.*, 2019), SPM-mediated activation has an advantage in its simplicity. In this work, I suggest avenues to new applications of NeissLock, SPM and SPM-like protein domains. Finally, the demonstration of *in vitro* and cellular conjugation opens the exploration of NeissLock’s therapeutic and diagnostic applications.

7. References

Abbot, E.L., Smith, W.D., Siou, G.P.S., Chiriboga, C., Smith, R.J., Wilson, J.A., Hirst, B.H. and Kehoe, M.A. (2007) 'Pili mediate specific adhesion of Streptococcus pyogenes to human tonsil and skin', *Cellular Microbiology*, 9:1822–1833. doi:10.1111/j.1462-5822.2007.00918.x.

Adams, P.D., Afonine, P. V., Baskaran, K., Berman, H.M., Berrisford, J., Bricogne, G., Brown, D.G., Burley, S.K., Chen, M., Feng, Z., *et al.* (2019) 'Announcing mandatory submission of PDBx/mmCIF format files for crystallographic depositions to the protein data bank (PDB)', *Acta Crystallographica Section D: Structural Biology*, 75:451–454. doi:10.1107/S2059798319004522.

Alfonta, L., Zhang, Z., Uryu, S., Loo, J.A. and Schultz, P.G. (2003) 'Site-Specific Incorporation of a Redox-Active Amino Acid into Proteins', *Journal of the American Chemical Society*, 125:14662–14663. doi:10.1021/ja038242x.

Amitai, G., Dassa, B. and Pietrokovski, S. (2004) 'Protein Splicing of Inteins with Atypical Glutamine and Aspartate C-terminal Residues', *Journal of Biological Chemistry*, 279:3121–3131. doi:10.1074/jbc.M311343200.

Aswad, D.W., Paranandi, M. V. and Schurter, B.T. (2000) 'Isoaspartate in peptides and proteins: Formation, significance, and analysis', *Journal of Pharmaceutical and Biomedical Analysis*, 21:1129–1136. doi:10.1016/S0731-7085(99)00230-7.

Baeza, J., Smallegan, M.J. and Denu, J.M. (2015) 'Site-specific reactivity of nonenzymatic lysine acetylation', *ACS Chemical Biology*, 10:122–128. doi:10.1021/cb500848p.

Ball, H.R. and Winn, S.E. (1982) 'Acylation of Egg White Proteins with Acetic Anhydride and Succinic Anhydride', *Poultry Science*, 61:1041–1046. doi:10.3382/ps.0611041.

Banerjee, A. and Howarth, M. (2018) 'Nanoteamwork: covalent protein assembly beyond duets towards protein ensembles and orchestras', *Current Opinion in Biotechnology*, 51:16–23. doi:10.1016/j.copbio.2017.10.006.

Baruah, H., Puthenveetil, S., Choi, Y.A., Shah, S. and Ting, A.Y. (2008) 'An engineered aryl azide ligase for site-specific mapping of protein-protein interactions through photo-cross-linking', *Angewandte Chemie - International Edition*, 47:7018–7021. doi:10.1002/anie.200802088.

Batra, P.P., Roebuck, M.A. and Uetrecht, D. (1990) 'Effect of lysine modification on the secondary structure of ovalbumin', *Journal of Protein Chemistry*, 9:37–44. doi:10.1007/BF01024982.

Bausch-Fluck, D., Hofmann, A., Bock, T., Frei, A.P., Cerciello, F., Jacobs, A., Moest, H., Omasits, U.,

Gundry, R.L., Yoon, C., *et al.* (2015) 'A mass spectrometric-derived cell surface protein atlas', *PLoS ONE*, 10:e0121314. doi:10.1371/journal.pone.0121314.

Beenukumar, R.R., Gödderz, D., Palanimurugan, R. and Dohmen, R.J. (2015) 'Polyamines directly promote antizyme-mediated degradation of ornithine decarboxylase by the proteasome', *Microbial Cell*, 2:197–205. doi:10.15698/mic2015.06.206.

Bell, C.H., Healey, E., Van Erp, S., Bishop, B., Tang, C., Gilbert, R.J.C., Radu Aricescu, A., Jeroen Pasterkamp, R. and Siebold, C. (2013) 'Structure of the Repulsive Guidance Molecule (RGM)-neogenin signaling hub', *Science*, 341:77–80. doi:10.1126/science.1232322.

Berman, H., Henrick, K. and Nakamura, H. (2003) 'Announcing the worldwide Protein Data Bank', *Nature Structural Biology*, 10:980. doi:10.1038/nsb1203-980.

Bischoff, R. and Schlüter, H. (2012) 'Amino acids: Chemistry, functionality and selected non-enzymatic post-translational modifications', *Journal of Proteomics*, 75:2275–2296. doi:10.1016/j.jprot.2012.01.041.

Böttke, T., Ernicke, S., Serfling, R., Ihling, C., Burda, E., Gurevich, V. V, Sinz, A. and Coin, I. (2020) 'Exploring GPCR-arrestin interfaces with genetically encoded crosslinkers', *EMBO reports*, 21:1–11. doi:10.15252/embr.202050437.

Brune, K.D., Buldun, C.M., Li, Y., Taylor, I.J., Brod, F., Biswas, S. and Howarth, M. (2017) 'Dual Plug-and-Display Synthetic Assembly Using Orthogonal Reactive Proteins for Twin Antigen Immunization', *Bioconjugate Chemistry*, 28:1544–1551. doi:10.1021/acs.bioconjchem.7b00174.

Bumba, L., Masin, J., Macek, P., Wald, T., Motlova, L., Bibova, I., Klimova, N., Bednarova, L., Veverka, V., Kachala, M., *et al.* (2016) 'Calcium-Driven Folding of RTX Domain β-Rolls Ratchets Translocation of RTX Proteins through Type I Secretion Ducts', *Molecular Cell*, 62:47–62. doi:10.1016/j.molcel.2016.03.018.

Bunton, C.A., Fuller, N.A., Perry, S.G. and Shiner, V.J. (1963) 'The hydrolysis of carboxylic anhydrides. Part III.* Reactions in initially neutral solution', *Journal of the Chemical Society (Resumed)*, 3028–3036. doi:10.1039/jr9630002918.

Burdine, L., Gillette, T.G., Lin, H.J. and Kodadek, T. (2004) 'Periodate-triggered cross-linking of DOPA-containing peptide-protein complexes', *Journal of the American Chemical Society*, 126:11442–11443. doi:10.1021/ja045982c.

Butlin, N.G. and Meares, C.F. (2006) 'Antibodies with infinite affinity: Origins and applications', *Accounts of Chemical Research*, 39:780–787. doi:10.1021/ar020275e.

Camarero, J.A., Kimura, R.H., Woo, Y.H., Shekhtman, A. and Cantor, J. (2007) 'Biosynthesis of a fully functional cyclotide inside living bacterial cells', *ChemBioChem*, 8:1363–1366. doi:10.1002/cbic.200700183.

Carugo, O. (2013) 'Frequency of dipeptides and antidipeptides', *Computational and Structural Biotechnology Journal*, 8:e201308001. doi:10.5936/csbj.201308001.

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 Chemical Biology*, 9:1956–1961. doi:10.1021/cb500453a.

Chin, J.W., Martin, A.B., King, D.S., Wang, L. and Schultz, P.G. (2002) 'Addition of a photocrosslinking amino acid to the genetic code of Escherichia coli', *Proceedings of the National Academy of Sciences of the United States of America*, 99:11020–11024. doi:10.1073/pnas.172226299.

Chin, J.W., Santoro, S.W., Martin, A.B., King, D.S., Wang, L. and Schultz, P.G. (2002) 'Addition of p-azido-L-phenylalanine to the genetic code of Escherichia coli', *Journal of the American Chemical Society*, 124:9026–9027. doi:10.1021/ja027007w.

Chivers, C.E., Koner, A.L., Lowe, E.D. and Howarth, M. (2011) 'How the biotin-streptavidin interaction was made even stronger: Investigation via crystallography and a chimaeric tetramer', *Biochemical Journal*, 435:55–63. doi:10.1042/BJ20101593.

Chmura, A.J., Orton, M.S. and Meares, C.F. (2001) 'Antibodies with infinite affinity', *Proceedings of the National Academy of Sciences of the United States of America*, 98:8480–8484. doi:10.1073/pnas.151260298.

Cigler, M., Müller, T.G., Horn-Ghetko, D., von Wrisberg, M.K., Fottner, M., Goody, R.S., Itzen, A., Müller, M.P. and Lang, K. (2017) 'Proximity-Triggered Covalent Stabilization of Low-Affinity Protein Complexes In Vitro and In Vivo', *Angewandte Chemie - International Edition*, 56:15737–15741. doi:10.1002/anie.201706927.

Clark, R.J. and Craik, D.J. (2010) 'Native chemical ligation applied to the synthesis and bioengineering of circular peptides and proteins.', *Biopolymers*, 94:414–422. doi:10.1002/bip.21372.

Cock, P.J.A., Antao, T., Chang, J.T., Chapman, B.A., Cox, C.J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B., *et al.* (2009) 'Biopython: Freely available Python tools for computational molecular biology and bioinformatics', *Bioinformatics*, 25:1422–1423. doi:10.1093/bioinformatics/btp163.

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', *Journal of Molecular Biology*, 390:503–515. doi:10.1016/j.jmb.2009.05.029.

Conibear, A.C., Watson, E.E., Payne, R.J. and Becker, C.F.W. (2018) 'Native chemical ligation in protein synthesis and semi-synthesis', *Chemical Society Reviews*, 47:9046–9068. doi:10.1039/c8cs00573g.

Dadová, J., Galan, S.R. and Davis, B.G. (2018) 'Synthesis of modified proteins via functionalization of dehydroalanine', *Current Opinion in Chemical Biology*, 46:71–81. doi:10.1016/j.cbpa.2018.05.022.

Daniels, G. (2007) 'Functions of red cell surface proteins', *Vox Sanguinis*, 93:331–340. doi:10.1111/j.1423-0410.2007.00970.x.

Davidson, N.E., Gelmann, E.P., Lippman, M.E. and Dickson, R.B. (1987) 'Epidermal growth factor receptor gene expression in estrogen receptor-positive and negative human breast cancer cell lines', *Molecular Endocrinology*, 1:216–223. doi:10.1210/mend-1-3-216.

Deane, J.E., Graham, S.C., Mitchell, E.P., Flot, D., Johnson, S. and Lea, S.M. (2008) 'Crystal structure of Spa40, the specificity switch for the Shigella flexneri type III secretion system', *Molecular Microbiology*, 69:267–276. doi:10.1111/j.1365-2958.2008.06293.x.

Defeo-Jones, D., Tai, J.Y., Vuocolo, G.A., Wegrzyn, R.J., Schofield, T.L., Riemen, M.W. and Oliff, A. (1989) 'Substitution of lysine for arginine at position 42 of human transforming growth factor-alpha eliminates biological activity without changing internal disulfide bonds.', *Molecular and Cellular Biology*, 9:4083–4086. doi:10.1128/mcb.9.9.4083.

Dodds, A.W., Ren, X.D., Willis, A.C. and Law, S.K.A. (1996) 'The reaction mechanism of the internal thioester in the human complement component C4', *Nature*, 379:177–179. doi:10.1038/379177a0.

Durán, D., Imperial, J., Palacios, J., Ruiz-Argüeso, T., Göttfert, M., Zehner, S. and Rey, L. (2018) 'Characterization of a novel MIIA domain-containing protein (MdcE) in Bradyrhizobium spp.', *FEMS Microbiology Letters*, 365:1–8. doi:10.1093/femsle/fnx276.

Ebner, R. and Derynck, R. (1991) 'Epidermal growth factor and transforming growth factor-α: Differential intracellular routing and processing of ligand-receptor complexes', *Molecular Biology of the Cell*, 2:599–612. doi:10.1091/mbc.2.8.599.

Elbein, A.D., Tropea, J.E., Mitchell, M. and Kaushal, G.P. (1990) 'Kifunensine, a potent inhibitor of the glycoprotein processing mannosidase I', *Journal of Biological Chemistry*, 265:15599–15605.

Elleman, T.C., Domagala, T., McKern, N.M., Nerrie, M., Lönnqvist, B., Adams, T.E., Lewis, J., Lovrecz, G.O., Hoyne, P.A., Richards, K.M., *et al.* (2001) 'Identification of a determinant of epidermal growth

factor receptor ligand-binding specificity using a truncated, high-affinity form of the ectodomain', *Biochemistry*, 40:8930–8939. doi:10.1021/bi010037b.

Fierer, J.O., Veggiani, G. and Howarth, M. (2014) 'SpyLigase peptide-peptide ligation polymerizes affibodies to enhance magnetic cancer cell capture', *Proceedings of the National Academy of Sciences of the United States of America*, 111:E1176–E1181. doi:10.1073/pnas.1315776111.

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. doi:10.1016/j.cell.2017.09.017.

Friedrich, M.G., Wang, Z., Schey, K.L. and Truscott, R.J.W. (2018) 'Spontaneous cross-linking of proteins at aspartate and asparagine residues is mediated via a succinimide intermediate', *Biochemical Journal*, 475:3189–3200. doi:10.1042/BCJ20180529.

Furman, J.L., Kang, M., Choi, S., Cao, Y., Wold, E.D., Sun, S.B., Smider, V. V., Schultz, P.G. and Kim, C.H. (2014) 'A genetically encoded aza-michael acceptor for covalent cross-linking of protein-receptor complexes', *Journal of the American Chemical Society*, 136:8411–8417. doi:10.1021/ja502851h.

Gadjeva, M., Dodds, A.W., Taniguchi-Sidle, A., Willis, A.C., Isenman, D.E. and Law, S.K.A. (1998) 'The covalent binding reaction of C3', *Molecular Immunology*, 35:376. doi:10.1016/s0161-5890(98)90716-7.

Gao, X., Fang, J., Xue, B., Fu, L. and Li, H. (2016) 'Engineering Protein Hydrogels Using SpyCatcher-SpyTag Chemistry', *Biomacromolecules*, 17:2812–2819. doi:10.1021/acs.biomac.6b00566.

Garrett, T.P.J., McKern, N.M., Lou, M., Elleman, T.C., Adams, T.E., Lovrecz, G.O., Zhu, H.J., Walker, F., Frenkel, M.J., Hoyne, P.A., *et al.* (2002) 'Crystal structure of a truncated epidermal growth factor receptor extracellular domain bound to transforming growth factor α', *Cell*, 110:763–773. doi:10.1016/S0092-8674(02)00940-6.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D. and Bairoch, A. (2005) 'Protein Identification and Analysis Tools on the ExPASy Server', *The Proteomics Protocols Handbook*, 50:571–607. doi:10.1385/1-59259-890-0:571.

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', *Nature Methods*, 6:343–345. doi:10.1038/nmeth.1318.

Gold, A.M. and Fahrney, D. (1964) 'Sulfonyl Fluorides as Inhibitors of Esterases. II. Formation and Reactions of Phenylmethanesulfonyl α-Chymotrypsin', *Biochemistry*, 3:783–791.

doi:10.1021/bi00894a009.

Good, N.E., Winget, G.D., Winter, W., Connolly, T.N., Izawa, S. and Singh, R.M. (1966) 'Hydrogen ion buffers for biological research.', *Biochemistry*, 5:467–477. doi:10.1021/bi00866a011.

Griffin, B.A., Adams, S.R. and Tsien, R.Y. (1998) 'Specific covalent labeling of recombinant protein molecules inside live cells', *Science*, 281:269–272. doi:10.1126/science.281.5374.269.

Grisaru, D., Sternfeld, M., Eldor, A., Glick, D. and Soreq, H. (1999) 'Structural roles of acetylcholinesterase variants in biology and pathology', *European Journal of Biochemistry*, 264:672–686. doi:10.1046/j.1432-1327.1999.00693.x.

Groenen, L.C., Nice, E.C. and Burgess, A.W. (1994) 'Structure-function relationships for the EGF/TGF-α family of mitogens', *Growth Factors*, 11:235–257. doi:10.3109/08977199409010997.

Hamelryck, T. and Manderick, B. (2003) 'PDB file parser and structure class implemented in Python', *Bioinformatics*, 19:2308–2310. doi:10.1093/bioinformatics/btg299.

Han, X., Wang, C. and Liu, Z. (2018) 'Red Blood Cells as Smart Delivery Systems', *Bioconjugate Chemistry*, 29:852–860. doi:10.1021/acs.bioconjchem.7b00758.

Hao, B., Zheng, N., Schulman, B.A., Wu, G., Miller, J.J., Pagano, M. and Pavletich, N.P. (2005) 'Structural basis of the Cks1-dependent recognition of p27 Kip1 by the SCF Skp2 ubiquitin ligase', *Molecular Cell*, 20:9–19. doi:10.1016/j.molcel.2005.09.003.

Harel, M., Kleywegt, G.J., Ravelli, R.B., Silman, I. and Sussman, J.L. (1995) 'Crystal structure of an acetylcholinesterase-fasciculin complex: interaction of a three-fingered toxin from snake venom with its target', *Structure*, 3:1355–1366. doi:10.1016/S0969-2126(01)00273-8.

Hayashi, T. and Hamachi, I. (2012) 'Traceless affinity labeling of endogenous proteins for functional analysis in living cells', *Accounts of Chemical Research*, 45:1460–1469. doi:10.1021/ar200334r.

Herbst, R.S. (2004) 'Review of epidermal growth factor receptor biology', *International Journal of Radiation Oncology Biology Physics*, 59:S21–S26. doi:10.1016/j.ijrobp.2003.11.041.

Hermanson, G.T. (2013) 'Functional Targets for Bioconjugation', in *Bioconjugate Techniques*, 127–228. doi:10.1016/b978-0-12-382239-0.00002-9.

Högbom, M., Eklund, M., Nygren, P.Å. and Nordlund, P. (2003) 'Structural basis for recognition by an in vitro evolved affibody', *Proceedings of the National Academy of Sciences of the United States of America*, 100:3191–3196. doi:10.1073/pnas.0436100100.

Holm, L., Moody, P. and Howarth, M. (2009) 'Electrophilic affibodies forming covalent bonds to protein targets', *Journal of Biological Chemistry*, 284:32906–32913. doi:10.1074/jbc.M109.034322.

Hoppmann, C., Lacey, V.K., Louie, G. V., Wei, J., Noel, J.P. and Wang, L. (2014) 'Genetically encoding photoswitchable click amino acids in Escherichia coli and mammalian cells', *Angewandte Chemie - International Edition*, 53:3932–3936. doi:10.1002/anie.201400001.

Hoppmann, C., Maslennikov, I., Choe, S. and Wang, L. (2015) 'In Situ Formation of an Azo Bridge on Proteins Controllable by Visible Light', *Journal of the American Chemical Society*, 137:11218–11221. doi:10.1021/jacs.5b06234.

Hoyt, E.A., Cal, P.M.S.D., Oliveira, B.L. and Bernardes, G.J.L. (2019) 'Contemporary approaches to site-selective protein modification', *Nature Reviews Chemistry*, 3:147–171. doi:10.1038/s41570-019-0079-1.

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*, 6:26–28. doi:10.1371/journal.pone.0024366.

Hsu, S.C. and Hung, M.C. (2007) 'Characterization of a novel tripartite nuclear localization sequence in the EGFR family', *Journal of Biological Chemistry*, 282:10432–10440. doi:10.1074/jbc.M610014200.

Huang, X., Luo, X., Roupioz, Y. and Keillor, J.W. (1997) 'Controlled Regioselective Anilide Formation from Aspartic and Glutamic Acid Anhydrides', *Journal of Organic Chemistry*, 62:8821–8825. doi:10.1021/jo971375e.

Hurley, J.H., Lee, S. and Prag, G. (2006) 'Ubiquitin-binding domains', *Biochemical Journal*, 399:361–372. doi:10.1042/BJ20061138.

Hwang, P.M., Pan, J.S. and Sykes, B.D. (2014) 'Targeted expression, purification, and cleavage of fusion proteins from inclusion bodies in Escherichia coli', *FEBS Letters*, 588:247–252. doi:10.1016/j.febslet.2013.09.028.

Ibatullin, F.M. and Selivanov, S.I. (2009) 'Reaction of N-Fmoc aspartic anhydride with glycosylamines: a simple entry to N-glycosyl asparagines', *Tetrahedron Letters*, 50:6351–6354. doi:10.1016/j.tetlet.2009.08.106.

Isenegger, P.G. and Davis, B.G. (2020) 'Concepts of catalysis in site-selective protein modifications', *Journal of the American Chemical Society*, 141:8005–8013. doi:10.1021/jacs.8b13187.

Isidro-Llobet, A., Álvarez, M. and Albericio, F. (2009) 'Amino acid-protecting groups', *Chemical*

Reviews, 109:2455–2504. doi:10.1021/cr800323s.

Ivanov, I.P., Simin, K., Letsou, A., Atkins, J.F. and Gesteland, R.F. (1998) ' The Drosophila Gene for Antizyme Requires Ribosomal Frameshifting for Expression and Contains an Intronic Gene for snRNP Sm D3 on the Opposite Strand ', *Molecular and Cellular Biology*, 18:1553–1561. doi:10.1128/mcb.18.3.1553.

Ivanov, I.P., Shin, B.S., Loughran, G., Tzani, I., Young-Baird, S.K., Cao, C., Atkins, J.F. and Dever, T.E. (2018) 'Polyamine Control of Translation Elongation Regulates Start Site Selection on Antizyme Inhibitor mRNA via Ribosome Queuing', *Molecular Cell*, 70:254-264.e6. doi:10.1016/j.molcel.2018.03.015.

Ivanov, I.P., Loughran, G. and Atkins, J.F. (2008) 'uORFs with unusual translational start codons autoregulate expression of eukaryotic ornithine decarboxylase homologs', *Proceedings of the National Academy of Sciences of the United States of America*, 105:10079–10084. doi:10.1073/pnas.0801590105.

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 Research*, 73:2310–2321. doi:10.1158/0008-5472.CAN-12-2956.

Janssen, B.J.C., Huizinga, E.G., Raaijmakers, H.C.A., Roos, A., Daha, M.R., Nilsson-Ekdahl, K., Nilsson, B. and Gros, P. (2005) 'Structures of complement component C3 provide insights into the function and evolution of immunity', *Nature*, 437:505–511. doi:10.1038/nature04005.

Janssen, B.J.C., Christodoulidou, A., McCarthy, A., Lambris, J.D. and Gros, P. (2006) 'Structure of C3b reveals conformational changes that underlie complement activity', *Nature*, 444:213–216. doi:10.1038/nature05172.

Jarmoskaite, I., Alsadhan, I., Vaidyanathan, P.P. and Herschlag, D. (2020) 'How to measure and evaluate binding affinities', *eLife*, 9:1–34. doi:10.7554/ELIFE.57264.

Johnson, B.A., Shirokawa, J.M., Hancock, W.S., Spellman, M.W., Basa, L.J. and Aswad, D.W. (1989) 'Formation of isoaspartate at two distinct sites during in vitro aging of human growth hormone', *Journal of Biological Chemistry*, 264:14262–14271. doi:10.1016/s0021-9258(18)71672-4.

Johnson, B.A. and Aswad, D.W. (1991) 'Optimal conditions for the use of protein l-isoaspartyl methyltransferase in assessing the isoaspartate content of peptides and proteins', *Analytical Biochemistry*, 192:384–391. doi:10.1016/0003-2697(91)90553-6.

Keeble, A.H., Turkki, P., Stokes, S., Anuar, I.N.A.K., Rahikainen, R., Hytönen, V.P. and Howarth, M.

(2019) ‘Approaching infinite affinity through engineering of peptide-protein interaction’, *Proceedings of the National Academy of Sciences of the United States of America*, 116:26523–26533. doi:10.1073/pnas.1909653116.

Kim, D.E., Chivian, D. and Baker, D. (2004) ‘Protein structure prediction and analysis using the Robetta server’, *Nucleic Acids Research*, 32:W526-31. doi:10.1093/nar/gkh468.

Koide, A., Wojcik, J., Gilbreth, R.N., Hoey, R.J. and Koide, S. (2012) ‘Teaching an old scaffold new tricks: Monobodies constructed using alternative surfaces of the FN3 scaffold’, *Journal of Molecular Biology*, 415:393–405. doi:10.1016/j.jmb.2011.12.019.

Krissinel, E. and Henrick, K. (2007) ‘Inference of Macromolecular Assemblies from Crystalline State’, *Journal of Molecular Biology*, 372:774–797. doi:10.1016/j.jmb.2007.05.022.

Kuban, V., Macek, P., Hritz, J., Nechvatalova, K., Nedbalcova, K., Faldyna, M., Sebo, P., Zidek, L. and Bumba, L. (2020) ‘Structural basis of Ca2+-dependent self-processing activity of repeat-in-toxin proteins’, *mBio*. Edited by C. Buchrieser, 11:1–18. doi:10.1128/mBio.00226-20.

Kubáň, V., Nováček, J., Bumba, L. and Žídek, L. (2015) ‘NMR assignment of intrinsically disordered self-processing module of the FrpC protein of Neisseria meningitidis’, *Biomolecular NMR Assignments*, 9:435–440. doi:10.1007/s12104-015-9625-z.

Kumanomidou, T., Nishio, K., Takagi, K., Nakagawa, T., Suzuki, A., Yamane, T., Tokunaga, F., Iwai, K., Murakami, A., Yoshida, Y., *et al.* (2015) ‘The structural differences between α glycoprotein specific F-box protein Fbs1 and its homologous protein FBG3’, *PLoS ONE*, 10:e0140366. doi:10.1371/journal.pone.0140366.

Lajoie, M.J., Boyken, S.E., Salter, A.I., Bruffey, J., Rajan, A., Langan, R.A., Olshefsky, A., Muhunthan, V., Bick, M.J., Gewe, M., *et al.* (2020) ‘Designed protein logic to target cells with precise combinations of surface antigens’, *Science*, 369:1–13. doi:10.1126/science.aba6527.

Lam, S.S., Martell, J.D., Kamer, K.J., Deerinck, T.J., Ellisman, M.H., Mootha, V.K. and Ting, A.Y. (2014) ‘Directed evolution of APEX2 for electron microscopy and proximity labeling’, *Nature Methods*, 12:51–54. doi:10.1038/nmeth.3179.

Law, S.K.A. and Dodds, A.W. (1997) ‘The internal thioester and the covalent binding properties of the complement proteins C3 and C4’, *Protein Science*, 6:263–274. doi:10.1002/pro.5560060201.

Lazar, E., Vicenzi, E., Van Obberghen-Schilling, E., Wolff, B., Dalton, S., Watanabe, S. and Sporn, M.B. (1989) ‘Transforming growth factor alpha: an aromatic side chain at position 38 is essential for biological activity.’, *Molecular and Cellular Biology*, 9:860–864. doi:10.1128/mcb.9.2.860.

Lee, C.Y., Liu, Y.L., Lin, C.L., Liu, G.Y. and Hung, H.C. (2014) 'Functional roles of the dimer-interface residues in human ornithine decarboxylase', *PLoS ONE*, 9:1–11. doi:10.1371/journal.pone.0104865.

Lee, H.S., Dimla, R.D. and Schultz, P.G. (2009) 'Protein-DNA photo-crosslinking with a genetically encoded benzophenone-containing amino acid', *Bioorganic and Medicinal Chemistry Letters*, 19:5222–5224. doi:10.1016/j.bmcl.2009.07.011.

Lee, Y.J., Wu, B., Raymond, J.E., Zeng, Y., Fang, X., Wooley, K.L. and Liu, W.R. (2013) 'A genetically encoded acrylamide functionality', *ACS Chemical Biology*, 8:1664–1670. doi:10.1021/cb400267m.

Li, J.C., Liu, T., Wang, Y., Mehta, A.P. and Schultz, P.G. (2018) 'Enhancing Protein Stability with Genetically Encoded Noncanonical Amino Acids', *Journal of the American Chemical Society*, 140:15997–16000. doi:10.1021/jacs.8b07157.

Li, Q., Chen, Q., Klauser, P.C., Li, M., Zheng, F., Wang, N., Li, X., Zhang, Q., Fu, X., Wang, Q., *et al.* (2020) 'Developing Covalent Protein Drugs via Proximity-Enabled Reactive Therapeutics', *Cell*, 182:85-97.e16. doi:10.1016/j.cell.2020.05.028.

Li, X. and Liu, D.R. (2004) 'DNA-templated organic synthesis: Nature's strategy for controlling chemical reactivity applied to synthetic molecules', *Angewandte Chemie - International Edition*, 43:4848–4870. doi:10.1002/anie.200400656.

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', *Biochemical Journal*, 399:121–129. doi:10.1042/BJ20060443.

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', *Journal of Biological Chemistry*, 278:13944–13951. doi:10.1074/jbc.M210069200.

Lin, C.W. and Ting, A.Y. (2006) 'Transglutaminase-catalyzed site-specific conjugation of small-molecule probes to proteins in vitro and on the surface of living cells', *Journal of the American Chemical Society*, 128:4542–4543. doi:10.1021/ja0604111.

Lin, S., He, D., Long, T., Zhang, S., Meng, R. and Chen, P.R. (2014) 'Genetically encoded cleavable protein photo-cross-linker', *Journal of the American Chemical Society*, 136:11860–11863. doi:10.1021/ja504371w.

Linhartová, I., Bumba, L., Mašn, J., Basler, M., Osička, R., Kamanová, J., Procházková, K., Adkins, I., HejnováHolubová, J., Sadílková, L., *et al.* (2010) 'RTX proteins: A highly diverse family secreted bya common mechanism', *FEMS Microbiology Reviews*, 34:1076–1112. doi:10.1111/j.1574-6976.2010.00231.x.

Linke-Winnebeck, C., Paterson, N.G., Young, P.G., Middleditch, M.J., Greenwood, D.R., Witte, G. and Baker, E.N. (2014) 'Structural model for covalent adhesion of the Streptococcus pyogenes pilus through a thioester bond', *Journal of Biological Chemistry*, 289:177–189. doi:10.1074/jbc.M113.523761.

Lišková, P., Konopásek, I. and Fišer, R. (2019) 'Simple Way to Detect Trp to Tb 3+ Resonance Energy Transfer in Calcium-Binding Peptides Using Excitation Spectrum', *Journal of Fluorescence*, 29:9–14. doi:10.1007/s10895-018-2326-0.

Liu, B., Archer, C.T., Burdine, L., Gillette, T.G. and Kodadek, T. (2007) 'Label transfer chemistry for the characterization of protein-protein interactions', *Journal of the American Chemical Society*, 129:12348–12349. doi:10.1021/ja072904r.

Liu, B., Burdine, L. and Kodadek, T. (2006) 'Chemistry of periodate-mediated cross-linking of 3,4-dihydroxylphenylalanine-containing molecules to proteins', *Journal of the American Chemical Society*, 128:15228–15235. doi:10.1021/ja065794h.

Liu, C.C. and Schultz, P.G. (2010) 'Adding new chemistries to the genetic code', *Annual Review of Biochemistry*, 79:413–444. doi:10.1146/annurev.biochem.052308.105824.

Liu, J., Li, S., Aslam, N.A., Zheng, F., Yang, B., Cheng, R., Wang, N., Rozovsky, S., Wang, P.G., Wang, Q., *et al.* (2019) 'Genetically Encoding Photocaged Quinone Methide to Multitarget Protein Residues Covalently in Vivo', *Journal of the American Chemical Society*, 141:9458–9462. doi:10.1021/jacs.9b01738.

Liu, J., Cai, L., Sun, W., Cheng, R., Wang, N., Jin, L., Rozovsky, S., Seiple, I.B. and Wang, L. (2019) 'Photocaged Quinone Methide Crosslinkers for Light-Controlled Chemical Crosslinking of Protein–Protein and Protein–DNA Complexes', *Angewandte Chemie - International Edition*, 58:18839–18843. doi:10.1002/anie.201910135.

Liu, M., Ji, Z., Zhang, M. and Xia, J. (2017) 'Versatile Site-Selective Protein Reaction Guided by WW Domain-Peptide Motif Interaction', *Bioconjugate Chemistry*, 28:2199–2205. doi:10.1021/acs.bioconjchem.7b00334.

Liu, R., Yue, Z., Tsai, C.C. and Shen, J. (2019) 'Assessing Lysine and Cysteine Reactivities for Designing Targeted Covalent Kinase Inhibitors', *Journal of the American Chemical Society*, 141:6553–6560. doi:10.1021/jacs.8b13248.

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 Expression and Purification*, 98:18–24. doi:10.1016/j.pep.2014.02.016.

Liu, Y.C., Hsu, D.H., Huang, C.L., Liu, Y.L., Liu, G.Y. and Hung, H.C. (2011) 'Determinants of the differential antizyme-binding affinity of ornithine decarboxylase', *PLoS ONE*, 6:e26835. doi:10.1371/journal.pone.0026835.

Lloyd-Evans, E., Waller-Evans, H., Peterneva, K. and Platt, F.M. (2010) 'Endolysosomal calcium regulation and disease', *Biochemical Society Transactions*, 38:1458–1464. doi:10.1042/BST0381458.

Lu, M., Symersky, J., Radchenko, M., Koide, A., Guo, Y., Nie, R. and Koide, S. (2013) 'Structures of a Na+-coupled, substrate-bound MATE multidrug transporter', *Proceedings of the National Academy of Sciences of the United States of America*, 110:2099–2104. doi:10.1073/pnas.1219901110.

Lu, Y., Huang, F., Wang, J. and Xia, J. (2014) 'Affinity-guided covalent conjugation reactions based on PDZ-peptide and sh3-peptide interactions', *Bioconjugate Chemistry*, 25:989–999. doi:10.1021/bc500134w.

Lundblad, R. (2004) 'The Modification of Amino Groups', in *Chemical Reagents for Protein Modification, Third Edition*. 3rd edn. CRC Press, 31–66. doi:10.1201/9781420039511.ch2.

Macia, E., Ehrlich, M., Massol, R., Boucrot, E., Brunner, C. and Kirchhausen, T. (2006) 'Dynasore, a Cell-Permeable Inhibitor of Dynamin', *Developmental Cell*, 10:839–850. doi:10.1016/j.devcel.2006.04.002.

Mák, M., Mezö, G., Skribanek, Z. and Hudecz, F. (1998) 'Stability of Asp-Pro bond under high and low energy collision induced dissociation conditions in the immunodominant epitope region of herpes simplex virion glycoprotein D', *Rapid Communications in Mass Spectrometry*, 12:837–842. doi:10.1002/(SICI)1097-0231(19980715)12:13<837::AID-RCM248>3.0.CO;2-Z.

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', *Nature Biotechnology*, 30:1143–1148. doi:10.1038/nbt.2375.

Matos, M.J., Oliveira, B.L., Martínez-Sáez, N., Guerreiro, A., Cal, P.M.S.D., Bertoldo, J., Maneiro, M., Perkins, E., Howard, J., Deery, M.J., *et al.* (2018) 'Chemo- and Regioselective Lysine Modification on Native Proteins', *Journal of the American Chemical Society*, 140:4004–4017. doi:10.1021/jacs.7b12874.

Matsufuji, S., Matsufuji, T., Miyazaki, Y., Murakami, Y., Atkins, J.F., Gesteland, R.F. and Hayashi, S. ichi (1995) 'Autoregulatory frameshifting in decoding mammalian ornithine decarboxylase antizyme', *Cell*, 80:51–60. doi:10.1016/0092-8674(95)90450-6.

Matyska Liskova, P., Fiser, R., MacEk, P., Chmelik, J., Sykora, J., Bednarova, L., Konopasek, I. and Bumba, L. (2016) 'Probing the Ca2+-assisted π-π Interaction during Ca2+-dependent protein folding', *Soft Matter*, 12:531–541. doi:10.1039/c5sm01796c.

Meshcheryakov, V.A., Kitao, A., Matsunami, H. and Samatey, F.A. (2013) 'Inhibition of a type III secretion system by the deletion of a short loop in one of its membrane proteins', *Acta Crystallographica Section D: Biological Crystallography*, 69:812–820. doi:10.1107/S0907444913002102.

Metzner, F.J. (2017) '*Design of a Ca2+-cleavable and reactive protein tag based on a protein from Neisseria meningitidis*', Master thesis, University of Oxford.

Mills, K. V., Connor, K.R., Dorval, D.M. and Lewandowski, K.T. (2006) 'Protein purification via temperature-dependent, intein-mediated cleavage from an immobilized metal affinity resin', *Analytical Biochemistry*, 356:86–93. doi:10.1016/j.ab.2006.04.055.

Minteer, C.J., Siegart, N.M., Colelli, K.M., Liu, X., Linhardt, R.J., Wang, C., Gomez, A. V., Reitter, J.N. and Mills, K. V. (2017) 'Intein-Promoted Cyclization of Aspartic Acid Flanking the Intein Leads to Atypical N-Terminal Cleavage', *Biochemistry*, 56:1042–1050. doi:10.1021/acs.biochem.6b00894.

Möhlmann, S., Bringmann, P., Greven, S. and Harrenga, A. (2011) 'Site-specific modification of ED-B-targeting antibody using intein-fusion technology', *BMC Biotechnology*, 11:76. doi:10.1186/1472-6750-11-76.

Mora, M., Bensi, G., Capo, S., Falugi, F., Zingaretti, C., Manetti, A.G.O., Maggi, T., Taddei, A.R., Grandi, G. and Telford, J.L. (2005) 'Group A Streptococcus produce pilus-like structures containing protective antigens and Lancefield T antigens', *Proceedings of the National Academy of Sciences of the United States of America*, 102:15641–15646. doi:10.1073/pnas.0507808102.

Moy, F.J., Li, Y.C., Rauenbuehler, P., Winkler, M.E., Scheraga, H.A. and Montelione, G.T. (1993) 'Solution Structure of Human Type-α Transforming Growth Factor Determined by Heteronuclear NMR Spectroscopy and Refined by Energy Minimization with Restraints', *Biochemistry*, 32:7334–7353. doi:10.1021/bi00080a003.

Muir, T.W., Sondhi, D. and Cole, P.A. (1998) 'Expressed protein ligation: A general method for protein engineering', *Proceedings of the National Academy of Sciences of the United States of America*, 95:6705–6710. doi:10.1073/pnas.95.12.6705.

Murale, D.P., Hong, S.C., Haque, M.M. and Lee, J.S. (2017) 'Photo-affinity labeling (PAL) in chemical proteomics: A handy tool to investigate protein-protein interactions (PPIs)', *Proteome Science*, 15:14.

doi:10.1186/s12953-017-0123-3.

Nilsson, J., GRAHN, B. and HEBY, O. (2000) 'Antizyme inhibitor is rapidly induced in growth-stimulated mouse fibroblasts and releases ornithine decarboxylase from antizyme suppression', *Biochemical Journal*, 346:699–704. doi:10.1042/0264-6021:3460699.

Nödling, A.R., Spear, L.A., Williams, T.L., Luk, L.Y.P. and Tsai, Y.H. (2019) 'Using genetically incorporated unnatural amino acids to control protein functions in mammalian cells', *Essays in Biochemistry*, 63:237–266. doi:10.1042/EBC20180042.

Oehler, S. and Müller-Hill, B. (2010) 'High Local Concentration: A Fundamental Strategy of Life', *Journal of Molecular Biology*, 395:242–253. doi:10.1016/j.jmb.2009.10.056.

Ogiso, H., Ishitani, R., Nureki, O., Fukai, S., Yamanaka, M., Kim, J.H., Saito, K., Sakamoto, A., Inoue, M., Shirouzu, M., *et al.* (2002) 'Crystal structure of the complex of human epidermal growth factor and receptor extracellular domains', *Cell*, 110:775–787. doi:10.1016/S0092-8674(02)00963-7.

Osička, R., Kalmusová, J., Křížová, P. and Šebo, 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', *Infection and Immunity*, 69:5509–5519. doi:10.1128/IAI.69.9.5509-5519.2001.

Osička, R., Procházková, K., Šulc, M., Linhartová, I., Havlíček, V. and Šebo, 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', *Journal of Biological Chemistry*, 279:24944–24956. doi:10.1074/jbc.M314013200.

Osipiuk, J., Mulligan, R., Bargassa, M., Hamilton, J.E., Cunningham, M.A. and Joachimiak, A. (2012) 'Characterization of member of DUF1888 protein family, self-cleaving and self-assembling endopeptidase', *Journal of Biological Chemistry*, 287:19452–19461. doi:10.1074/jbc.M112.358069.

Pahari, S., Sun, L. and Alexov, E. (2019) 'PKAD: A database of experimentally measured pKa values of ionizable groups in proteins', *Database*, 2019:1–7. doi:10.1093/database/baz024.

Palanimurugan, R., Scheel, H., Hofmann, K. and Dohmen, R.J. (2004) 'Polyamines regulate their synthesis by inducing expression and blocking degradation of ODC antizyme', *EMBO Journal*, 23:4857–4867. doi:10.1038/sj.emboj.7600473.

Pegg, A.E. (2006) 'Regulation of ornithine decarboxylase', *Journal of Biological Chemistry*, 281:14529–14532. doi:10.1074/jbc.R500031200.

Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D.J., Inuganti, A., Griss, J., Mayer, G., Eisenacher, M., *et al.* (2019) 'The PRIDE database and related tools and resources in 2019: Improving support for quantification data', *Nucleic Acids Research*, 47:D442–D450. doi:10.1093/nar/gky1106.

Pettinger, J., Carter, M., Jones, K. and Cheeseman, M.D. (2019) 'Kinetic Optimization of Lysine-Targeting Covalent Inhibitors of HSP72', *Journal of Medicinal Chemistry*, 62:11383–11398. doi:10.1021/acs.jmedchem.9b01709.

Piszkiewicz, D., Landon, M. and Smith, E.L. (1970) 'Anomalous cleavage of aspartyl-proline peptide bonds during amino acid sequence determinations', *Biochemical and Biophysical Research Communications*, 40:1173–1178. doi:10.1016/0006-291X(70)90918-6.

Pointon, J.A., Smith, W.D., Saalbach, G., Crow, A., Kehoe, M.A. and Banfield, M.J. (2010) 'A highly unusual thioester bond in a pilus adhesin is required for efficient host cell interaction', *Journal of Biological Chemistry*, 285:33858–33866. doi:10.1074/jbc.M110.149385.

Preston, G.W. and Wilson, A.J. (2013) 'Photo-induced covalent cross-linking for the analysis of biomolecular interactions', *Chemical Society Reviews*, 42:3289–3301. doi:10.1039/c3cs35459h.

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', *Journal of Biological Chemistry*, 280:3251–3258. doi:10.1074/jbc.M411232200.

Proft, T. (2010) 'Sortase-mediated protein ligation: An emerging biotechnology tool for protein modification and immobilisation', *Biotechnology Letters*, 32:1–10. doi:10.1007/s10529-009-0116-0.

Rawlings, N.D., Barrett, A.J. and Bateman, A. (2011) 'Asparagine peptide lyases: A seventh catalytic type of proteolytic enzymes', *Journal of Biological Chemistry*, 286:38321–38328. doi:10.1074/jbc.M111.260026.

Reckel, S., Gehin, C., Tardivon, D., Georgeon, S., Kükenshöner, T., Löhr, F., Koide, A., Buchner, L., Panjkovich, A., Reynaud, A., *et al.* (2017) 'Structural and functional dissection of the DH and PH domains of oncogenic Bcr-Abl tyrosine kinase', *Nature Communications*, 8:2101. doi:10.1038/s41467-017-02313-6.

Recktenwald, C. V. and Hansson, G.C. (2016) 'The reduction-insensitive bonds of the MUC2 mucin are isopeptide bonds', *Journal of Biological Chemistry*, 291:13580–13590. doi:10.1074/jbc.M116.726406.

Reddy, C.C., Niyog, S.K., Wells, A., Wiley, H.S. and Lauffenburger, D.A. (1996) 'Engineering epidermal growth factor for enhanced mitogenic potency', *Nature Biotechnology*, 14:1696–1699.

doi:10.1038/nbt1296-1696.

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. doi:10.1126/science.1230593.

Rodríguez-Ithurralde, D., Silveira, R., Barbeito, L. and Dajas, F. (1983) 'Fasciculin, a powerful anticholinesterase polypeptide from Dendroaspis angusticeps venom', *Neurochemistry International*, 5:267–274. doi:10.1016/0197-0186(83)90028-1.

Roepstorff, K., Grandal, M.V., Henriksen, L., Knudsen, S.L.J., Lerdrup, M., Grøvdal, L., Willumsen, B.M. and Van Deurs, B. (2009) 'Differential effects of EGFR ligands on endocytic sorting of the receptor', *Traffic*, 10:1115–1127. doi:10.1111/j.1600-0854.2009.00943.x.

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', *Journal of Cell Biology*, 196:801–810. doi:10.1083/jcb.201112098.

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 Science*, 17:1834–1843. doi:10.1110/ps.035733.108.

Saito, G., Swanson, J.A. and Lee, K.D. (2003) 'Drug delivery strategy utilizing conjugation via reversible disulfide linkages: Role and site of cellular reducing activities', *Advanced Drug Delivery Reviews*, 55:199–215. doi:10.1016/S0169-409X(02)00179-5.

Scheu, A.H.A. (2020) 'NeissDist: NeissLock Distance Database (NeissLock: an inducible protein anhydride for covalent targeting of endogenous proteins)', *Zenodo*. doi:10.5281/zenodo.4322640.

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', *Nature Communications*, 12:717. doi:10.1038/s41467-021-20963-5.

Schirrmeister, J., Friedrich, L., Wenzel, M., Hoppe, M., Wolf, C., Göttfert, M. and Zehner, S. (2011) 'Characterization of the self-cleaving effector protein NopE1 of Bradyrhizobium japonicum', *Journal of Bacteriology*, 193:3733–3739. doi:10.1128/JB.00437-11.

Schirrmeister, J., Zocher, S., Flor, L., Göttfert, M. and Zehner, S. (2013) 'The domain of unknown function DUF1521 exhibits metal ion-inducible autocleavage activity - a novel example from a putative effector protein of Vibrio coralliilyticus ATCC BAA-450', *FEMS Microbiology Letters*, 343:177–182. doi:10.1111/1574-6968.12145.

Schoene, C., Fierer, J.O., Bennett, S.P. and Howarth, M. (2014) 'SpyTag/Spycatcher cyclization confers resilience to boiling on a mesophilic enzyme', *Angewandte Chemie - International Edition*, 53:6101–6104. doi:10.1002/anie.201402519.

Ségalas, I., Thai, R. and Claudio Vita, R.M. (1995) 'A particularly labile Asp-Pro bond in the green mamba muscarinic toxin MTX2. Effect of protein conformation on the rate of cleavage', *FEBS Letters*, 371:171–175. doi:10.1016/0014-5793(95)00844-Y.

Seras-Franzoso, J., Affentranger, R., Ferrer-Navarro, M., Daura, X., Villaverde, A. and García-Fruitósa, E. (2012) 'Disulfide bond formation and activation of Escherichia coli β-galactosidase under oxidizing conditions', *Applied and Environmental Microbiology*, 78:2376–2385. doi:10.1128/AEM.06923-11.

Shah, N.H. and Muir, T.W. (2014) 'Inteins: Nature's gift to protein chemists', *Chemical Science*, 5:446–461. doi:10.1039/c3sc52951g.

Shiao, D.D.F., Lumry, R. and Rajender, S. (1972) 'Modification of Protein Properties by Change in Charge: Succinylated Chymotrypsinogen', *European Journal of Biochemistry*, 29:377–385. doi:10.1111/j.1432-1033.1972.tb01999.x.

Singh, B. and Coffey, R.J. (2014) 'From wavy hair to naked proteins: The role of transforming growth factor alpha in health and disease', *Seminars in Cell and Developmental Biology*, 28:12–21. doi:10.1016/j.semcdb.2014.03.003.

Singh, J., Petter, R.C., Baillie, T.A. and Whitty, A. (2011) 'The resurgence of covalent drugs', *Nature Reviews Drug Discovery*, 10:307–317. doi:10.1038/nrd3410.

Škrlec, K., Štrukelj, B. and Berlec, A. (2015) 'Non-immunoglobulin scaffolds: A focus on their targets', *Trends in Biotechnology*, 33:408–418. doi:10.1016/j.tibtech.2015.03.012.

Stones, D.H. and Krachler, A.M. (2016) 'Against the tide: The role of bacterial Adhesion in host colonization', *Biochemical Society Transactions*, 44:1571–1580. doi:10.1042/BST20160186.

Strelow, J.M. (2017) 'A Perspective on the Kinetics of Covalent and Irreversible Inhibition', *Journal of Biomolecular Screening*, 22:3–20. doi:10.1177/1087057116671509.

Suchanek, M., Radzikowska, A. and Thiele, C. (2005) 'Photo-leucine and photo-methionine allow identification of protein-protein interactions in living cells', *Nature Methods*, 2:261–267. doi:10.1038/nmeth752.

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', *Scientific Reports*, 7:40408. doi:10.1038/srep40408.

Takaoka, Y., Ojida, A. and Hamachi, I. (2013) 'Protein organic chemistry and applications for labeling and engineering in live-cell systems', *Angewandte Chemie - International Edition*, 4088–4106. doi:10.1002/anie.201207089.

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', *Nature Communications*, 9:1–12. doi:10.1038/s41467-018-04343-0.

Tamura, T. and Hamachi, I. (2019) 'Chemistry for Covalent Modification of Endogenous/Native Proteins: From Test Tubes to Complex Biological Systems', *Journal of the American Chemical Society*, 141:2782–2799. doi:10.1021/jacs.8b11747.

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', *Journal of the American Chemical Society*, 134:2216–2226. doi:10.1021/ja209641t.

Telford, J.L., Barocchi, M.A., Margarit, I., Rappuoli, R. and Grandi, G. (2006) 'Pili in Gram-positive pathogens', *Nature Reviews Microbiology*, 4:509–519. doi:10.1038/nrmicro1443.

Thompson, S.A., Wang, L.L., West, A. and Sparling, P.F. (1993) 'Neisseria meningitidis produces iron-regulated proteins related to the RTX family of exoproteins', *Journal of Bacteriology*, 175:811–818. doi:10.1128/jb.175.3.811-818.1993.

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', *Infection and Immunity*, 61:2906–2911. doi:10.1128/iai.61.7.2906-2911.1993.

Thul, P.J., Akesson, L., Wiking, M., Mahdessian, D., Geladaki, A., Ait Blal, H., Alm, T., Asplund, A., Björk, L., Breckels, L.M., *et al.* (2017) 'A subcellular map of the human proteome', *Science*, 356:eaal3321. doi:10.1126/science.aal3321.

Thuveson, M. and Fries, E. (1999) 'Intracellular proteolytic processing of the heavy chain of rat pre-α-inhibitor. The COOH-terminal propeptide is required for coupling to bikunin', *Journal of Biological Chemistry*, 274:6741–6746. doi:10.1074/jbc.274.10.6741.

Thuveson, M. and Fries, E. (2000) 'The low pH in Trans-Golgi triggers autocatalytic cleavage of pre-α-inhibitor hearty chain precursor', *Journal of Biological Chemistry*, 275:30996–31000. doi:10.1074/jbc.M002399200.

Tietze, L.F. and Krewer, B. (2009) 'Antibody-directed enzyme prodrug therapy: A promising approach for a selective treatment of cancer based on prodrugs and monoclonal antibodies', *Chemical Biology and Drug Design*, 74:205–211. doi:10.1111/j.1747-0285.2009.00856.x.

Tippmann, E.M., Liu, W., Summerer, D., Mack, A. V. and Schultz, P.G. (2007) 'A genetically encoded diazirine photocrosslinker in Escherichia coli', *ChemBioChem*, 8:2210–2214. doi:10.1002/cbic.200700460.

Tjong, H., Qin, S. and Zhou, H.X. (2007) 'PI 2PE: Protein interface/interior prediction engine', *Nucleic Acids Research*, 35:357–362. doi:10.1093/nar/gkm231.

Tonge, P.J. (2018) 'Drug-Target Kinetics in Drug Discovery', *ACS Chemical Neuroscience*, 9:29–39. doi:10.1021/acschemneuro.7b00185.

Tsukiji, S., Miyagawa, M., Takaoka, Y., Tamura, T. and Hamachi, I. (2009) 'Ligand-directed tosyl chemistry for protein labeling in vivo', *Nature Chemical Biology*, 5:341–343. doi:10.1038/nchembio.157.

Tsukiji, S. and Hamachi, I. (2014) 'Ligand-directed tosyl chemistry for in situ native protein labeling and engineering in living systems: From basic properties to applications', *Current Opinion in Chemical Biology*, 21:136–143. doi:10.1016/j.cbpa.2014.07.012.

Umeda, A., Thibodeaux, G.N., Zhu, J., Lee, Y.A. and Zhang, Z.J. (2009) 'Site-specific protein cross-linking with genetically incorporated 3,4-dihydroxy-l-phenylalanine', *ChemBioChem*, 10:1302–1304. doi:10.1002/cbic.200900127.

Vane, J.R. and Botting, R.M. (2003) 'The mechanism of action of aspirin', *Thrombosis Research*, 110:255–258. doi:10.1016/S0049-3848(03)00379-7.

Vigers, G.P.A., Dripps, D.J., Edwards, C.K. and Brandhuber, B.J. (2000) 'X-ray crystal structure of a small antagonist peptide bound to interleukin-1 receptor type 1', *Journal of Biological Chemistry*, 275:36927–36933. doi:10.1074/jbc.M006071200.

Villa, C.H., Anselmo, A.C., Mitragotri, S. and Muzykantov, V. (2016) 'Red blood cells: Supercarriers for drugs, biologicals, and nanoparticles and inspiration for advanced delivery systems', *Advanced Drug Delivery Reviews*, 88–103. doi:10.1016/j.addr.2016.02.007.

Virji, M. (2009) 'Pathogenic neisseriae: Surface modulation, pathogenesis and infection control', *Nature Reviews Microbiology*, 7:274–286. doi:10.1038/nrmicro2097.

Walden, M., Crow, A., Nelson, M.D. and Banfield, M.J. (2014) 'Intramolecular isopeptide but not

internal thioester bonds confer proteolytic and significant thermal stability to the S. pyogenes pilus adhesin Spy0125', *Proteins: Structure, Function and Bioinformatics*, 82:517–527. doi:10.1002/prot.24420.

Walden, M., Edwards, J.M., Dziewulska, A.M., Bergmann, R., Saalbach, G., Kan, S.Y., Miller, O.K., Weckener, M., Jackson, R.J., Shirran, S.L., *et al.* (2015) 'An internal thioester in a pathogen surface protein mediates covalent host binding', *eLife*, 4:1–24. doi:10.7554/eLife.06638.

Wang, H., Koshi, Y., Minato, D., Nonaka, H., Kiyonaka, S., Mori, Y., Tsukiji, S. and Hamachi, I. (2011) 'Chemical cell-surface receptor engineering using affinity-guided, multivalent organocatalysts', *Journal of the American Chemical Society*, 133:12220–12228. doi:10.1021/ja204422r.

Wang, J., Schiller, S.M. and Schultz, P.G. (2007) 'A biosynthetic route to dehydroalanine-containing proteins', *Angewandte Chemie - International Edition*, 46:6849–6851. doi:10.1002/anie.200702305.

Wang, J., Yu, Y. and Xia, J. (2014) 'Short peptide tag for covalent protein labeling based on coiled coils', *Bioconjugate Chemistry*, 25:178–187. doi:10.1021/bc400498p.

Wang, L. (2017) 'Genetically encoding new bioreactivity', *New Biotechnology*, 38:16–25. doi:10.1016/j.nbt.2016.10.003.

Wang, N., Yang, B., Fu, C., Zhu, H., Zheng, F., Kobayashi, T., Liu, J., Li, S., Ma, C., Wang, P.G., *et al.* (2018) 'Genetically encoding fluorosulfate- l -tyrosine to react with lysine, histidine, and tyrosine via SuFEx in proteins in vivo', *Journal of the American Chemical Society*, 140:4995–4999. doi:10.1021/jacs.8b01087.

Wang, R., Lu, D., Bai, H., Jin, C., Yan, G., Ye, M., Qiu, L., Chang, R., Cui, C., Liang, H., *et al.* (2016) 'Using modified aptamers for site specific protein-aptamer conjugations', *Chemical Science*, 7:2157–2161. doi:10.1039/c5sc02631h.

Wang, Z., Lyons, B., Truscott, R.J.W. and Schey, K.L. (2014) 'Human protein aging: Modification and crosslinking through dehydroalanine and dehydrobutyrine intermediates', *Aging Cell*, 13:226–234. doi:10.1111/acel.12164.

Wiesmann, C., Katschke, K.J., Yin, J.P., Helmy, K.Y., Steffek, M., Fairbrother, W.J., McCallum, S.A., Embuscado, L., DeForge, L., Hass, P.E., *et al.* (2006) 'Structure of C3b in complex with CRIg gives insights into regulation of complement activation', *Nature*, 444:217–220. doi:10.1038/nature05263.

Winkler, M.E., Bringman, T. and Marks, B.J. (1986) 'The purification of fully active recombinant transforming growth factor α produced in Escherichia coli', *Journal of Biological Chemistry*, 261:13838–13843. doi:10.1016/s0021-9258(18)67096-6.

Wlodawer, A., Minor, W., Dauter, Z. and Jaskolski, M. (2008) 'Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures', *FEBS Journal*, 275:1–21. doi:10.1111/j.1742-4658.2007.06178.x.

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. doi:10.1021/bi00912a013.

Wong, K., Perpich, J.D., Kozlov, G., Cygler, M., Abu Kwaik, Y. and Gehring, K. (2017) 'Structural Mimicry by a Bacterial F Box Effector Hijacks the Host Ubiquitin-Proteasome System', *Structure*, 25:376–383. doi:10.1016/j.str.2016.12.015.

Wright, T.H., Bower, B.J., Chalker, J.M., Bernardes, G.J.L., Wiewiora, R., Ng, W.L., Raj, R., Faulkner, S., Vallée, M.R.J., Phanumartwiwath, A., *et al.* (2016) 'Posttranslational mutagenesis: A chemical strategy for exploring protein side-chain diversity', *Science*, 354. doi:10.1126/science.aag1465.

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', *Proceedings of the National Academy of Sciences of the United States of America*, 112:11229–11234. doi:10.1073/pnas.1508187112.

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', *Nature Methods*, 10:885–888. doi:10.1038/nmeth.2595.

Xiang, Z., Lacey, V.K., Ren, H., Xu, J., Burban, D.J., Jennings, P.A. and Wang, L. (2014) 'Proximity-enabled protein crosslinking through genetically encoding haloalkane unnatural amino acids', *Angewandte Chemie - International Edition*, 53:2190–2193. doi:10.1002/anie.201308794.

Xuan, W., Li, J., Luo, X. and Schultz, P.G. (2016) 'Genetic Incorporation of a Reactive Isothiocyanate Group into Proteins', *Angewandte Chemie - International Edition*, 55:10065–10068. doi:10.1002/anie.201604891.

Xuan, W., Shao, S. and Schultz, P.G. (2017) 'Protein Crosslinking by Genetically Encoded Noncanonical Amino Acids with Reactive Aryl Carbamate Side Chains', *Angewandte Chemie - International Edition*, 56:5096–5100. doi:10.1002/anie.201611841.

Yang, B., Tang, S., Ma, C., Li, S.T., Shao, G.C., Dang, B., DeGrado, W.F., Dong, M.Q., Wang, P.G., Ding, S., *et al.* (2017) 'Spontaneous and specific chemical cross-linking in live cells to capture and identify protein interactions', *Nature Communications*, 8:2240. doi:10.1038/s41467-017-02409-z.

Yang, B., Wang, N., Schnier, P.D., Zheng, F., Zhu, H., Polizzi, N.F., Ittuveetil, A., Saikam, V., Degrado, W.F., Wang, Q., *et al.* (2019) 'Genetically Introducing Biochemically Reactive Amino Acids Dehydroalanine and Dehydrobutyrine in Proteins', *Journal of the American Chemical Society*, 141:7698–7703. doi:10.1021/jacs.9b02611.

Yang, Y., Song, H., He, D., Zhang, S., Dai, S., Lin, S., Meng, R., Wang, C. and Chen, P.R. (2016) 'Genetically encoded protein photocrosslinker with a transferable mass spectrometry-identifiable label', *Nature Communications*, 7:12299. doi:10.1038/ncomms12299.

Yang, Y., Song, H., He, D., Zhang, S., Dai, S., Xie, X., Lin, S., Hao, Z., Zheng, H. and Chen, P.R. (2017) 'Genetically encoded releasable photo-cross-linking strategies for studying protein-protein interactions in living cells', *Nature Protocols*, 12:2147–2168. doi:10.1038/nprot.2017.090.

Yu, Y., Nie, Y., Feng, Q., Qu, J., Wang, R., Bian, L. and Xia, J. (2017) 'Targeted Covalent Inhibition of Grb2-Sos1 Interaction through Proximity-Induced Conjugation in Breast Cancer Cells', *Molecular Pharmaceutics*, 14:1548–1557. doi:10.1021/acs.molpharmaceut.6b00952.

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', *Proceedings of the National Academy of Sciences of the United States of America*, 109:E690–E697. doi:10.1073/pnas.1115485109.

Zarivach, R., Deng, W., Vuckovic, M., Felise, H.B., Nguyen, H. V., Miller, S.I., Finlay, B.B. and Strynadka, N.C.J. (2008) 'Structural analysis of the essential self-cleaving type III secretion proteins EscU and SpaS', *Nature*, 453:124–127. doi:10.1038/nature06832.

Zhang, Haonan, Qiao, A., Yang, L., Van Eps, N., Frederiksen, K.S., Yang, D., Dai, A., Cai, X., Zhang, Hui, Yi, C., *et al.* (2018) 'Structure of the glucagon receptor in complex with a glucagon analogue', *Nature*, 553:106–110. doi:10.1038/nature25153.

8. Appendix

Table 8.1. Amino acid sequences of proteins used in this study.

Affibody-SPM from Felix Metzner, annotation adapted (Metzner, 2017)

MGSSHHHHHHSSGLVPRGSHMGLNDIFEAQKIEWHEGSGASMTGGQQMGRDPGVDNKFNKEMRNAYWEIALLPNLNNQQKRAF
IRSLYDDPSQSANLLAEAKKLNDAQAPKGLEGSGSGGGGSGGGGSGGGGSGEGHVYDPLALDLDGDGIETVATKGFAGSLFDH
TNNGIRTATGWVSADDGLLVRDLNGNGIIDNGAELFGDNTKLADGSFAKHGYAALAELDSNGDNIINAADAAFQTLRVWQDLN
QDGISQANELRTLEELGIQSLDLAYKDVNKNLGNGNTLAQQGSYTKTDGTTAKMGDLLLAADNLHSRFKDKVELTAEQAKAAN
LAGIGRLRDLREAAALSGDLANMLKAYSAAETKEAQLALLDNLIHKWAETDGSGEPEA-

H_6, thrombin cleavage site, AviTag, anti-HER2 Affibody, $FrpC^{411-413}$ spacer (residues preceding cleavage site), SPM ($FrpC^{414-647}$), C-tag

SpyTag-X-SPM [X: canonical amino acid, here X=A]

MGSSHHHHHHSSGLVPRGSHMGAHIVMVDAYKPTKADPLALDLDGDGIETVATKGFAGSLFDHTNNGIRTATGWVSADDGLLV
RDLNGNGIIDNGAELFGDNTKLADGSFAKHGYAALAELDSNGDNIINAADAAFQTLRVWQDLNQDGISQANELRTLEELGIQS
LDLAYKDVNKNLGNGNTLAQQGSYTKTDGTTAKMGDLLLAADNLHSRFKDKVELTAEQAKAANLAGIGRLRDLREAAALSGDL
ANMLKAYSAAETKEAQLALLDNLIHKWAETDGSGEPEA-

H_6, thrombin cleavage site, SpyTag, Alanine, SPM ($FrpC^{414-647}$), C-tag

OAZ-SPM

MGSSHHHHHHFYSDDRLNVTEELTSNDKTRILNVQSRLTDAKRINWRTVLSGGSLYIEIPGGALPEGSKDSFAVLLEFAEEQL
RADHVFICFHKNREDRAALLRTFSFLGFEIVRPGHPLVPKRPDACFMAYTFEDPLALDLDGDGIETVATKGFAGSLFDHTNNG
IRTATGWVSADDGLLVRDLNGNGIIDNGAELFGDNTKLADGSFAKHGYAALAELDSNGDNIINAADAAFQTLRVWQDLNQDGI
SQANELRTLEELGIQSLDLAYKDVNKNLGNGNTLAQQGSYTKTDGTTAKMGDLLLAADNLHSRFKDKVELTAEQAKAANLAGI
GRLRDLREAAALSGDLANMLKAYSAAETKEAQLALLDNLIHKWAETDGSGEPEA-

H_6, OAZ^{95-219}, SPM ($FrpC^{414-647}$), C-tag

OAZ[ΔE219]-SPM

MGSSHHHHHHFYSDDRLNVTEELTSNDKTRILNVQSRLTDAKRINWRTVLSGGSLYIEIPGGALPEGSKDSFAVLLEFAEEQL
RADHVFICFHKNREDRAALLRTFSFLGFEIVRPGHPLVPKRPDACFMAYTFDPLALDLDGDGIETVATKGFAGSLFDHTNNGI
RTATGWVSADDGLLVRDLNGNGIIDNGAELFGDNTKLADGSFAKHGYAALAELDSNGDNIINAADAAFQTLRVWQDLNQDGIS
QANELRTLEELGIQSLDLAYKDVNKNLGNGNTLAQQGSYTKTDGTTAKMGDLLLAADNLHSRFKDKVELTAEQAKAANLAGIG
RLRDLREAAALSGDLANMLKAYSAAETKEAQLALLDNLIHKWAETDGSGEPEA-

H_6, OAZ^{95-218}, SPM ($FrpC^{414-647}$), C-tag

OAZ[K153A]-SPM

MGSSHHHHHHFYSDDRLNVTEELTSNDKTRILNVQSRLTDAKRINWRTVLSGGSLYIEIPGGALPEGS**A**DSFAVLLEFAEEQL
RADHVFICFHKNREDRAALLRTFSFLGFEIVRPGHPLVPKRPDACFMAYTFEDPLALDLDGDGIETVATKGFAGSLFDHTNNG
IRTATGWVSADDGLLVRDLNGNGIIDNGAELFGDNTKLADGSFAKHGYAALAELDSNGDNIINAADAAFQTLRVWQDLNQDGI
SQANELRTLEELGIQSLDLAYKDVNKNLGNGNTLAQQGSYTKTDGTTAKMGDLLLAADNLHSRFKDKVELTAEQAKAANLAGI
GRLRDLREAAALSGDLANMLKAYSAAETKEAQLALLDNLIHKWAETDGSGEPEA-

H_6, OAZ^{95-219}[**K153A**], SPM ($FrpC^{414-647}$), C-tag

OAZ[K153A, A215R]-SPM

MGSSHHHHHHFYSDDRLNVTEELTSNDKTRILNVQSRLTDAKRINWRTVLSGGSLYIEIPGGALPEGSADSFAVLLEFAEEQL
RADHVFICFHKNREDRAALLRTFSFLGFEIVRPGHPLVPKRPDACFMRYTFEDPLALDLDGDGIETVATKGFAGSLFDHTNNG
IRTATGWVSADDGLLVRDLNGNGIIDNGAELFGDNTKLADGSFAKHGYAALAELDSNGDNIINAADAAFQTLRVWQDLNQDGI
SQANELRTLEELGIQSLDLAYKDVNKNLGNGNTLAQQGSYTKTDGTTAKMGDLLLAADNLHSRFKDKVELTAEQAKAANLAGI
GRLRDLREAAALSGDLANMLKAYSAAETKEAQLALLDNLIHKWAETDGSGEPEA-

H_6, $OAZ^{95\text{-}219}$[**K153A**,**A215R**], SPM ($FrpC^{414\text{-}647}$), C-tag

OAZ1[K153A,V198A,ΔE219]-SPM

MGSSHHHHHHFYSDDRLNVTEELTSNDKTRILNVQSRLTDAKRINWRTVLSGGSLYIEIPGGALPEGSADSFAVLLEFAEEQL
RADHVFICFHKNREDRAALLRTFSFLGFEIARPGHPLVPKRPDACFMAYTFDPLALDLDGDGIETVATKGFAGSLFDHTNNGI
RTATGWVSADDGLLVRDLNGNGIIDNGAELFGDNTKLADGSFAKHGYAALAELDSNGDNIINAADAAFQTLRVWQDLNQDGIS
QANELRTLEELGIQSLDLAYKDVNKNLGNGNTLAQQGSYTKTDGTTAKMGDLLLAADNLHSRFKDKVELTAEQAKAANLAGIG
RLRDLREAAALSGDLANMLKAYSAAETKEAQLALLDNLIHKWAETDGSGEPEA-

H_6, $OAZ^{95\text{-}218}$[**K153A**,**V198A**], SPM ($FrpC^{414\text{-}647}$), C-tag

OAZ-Y-SPM

MGSSHHHHHHFYSDDRLNVTEELTSNDKTRILNVQSRLTDAKRINWRTVLSGGSLYIEIPGGALPEGSKDSFAVLLEFAEEQL
RADHVFICFHKNREDRAALLRTFSFLGFEIVRPGHPLVPKRPDACFMAYTFEYDPLALDLDGDGIETVATKGFAGSLFDHTNN
GIRTATGWVSADDGLLVRDLNGNGIIDNGAELFGDNTKLADGSFAKHGYAALAELDSNGDNIINAADAAFQTLRVWQDLNQDG
ISQANELRTLEELGIQSLDLAYKDVNKNLGNGNTLAQQGSYTKTDGTTAKMGDLLLAADNLHSRFKDKVELTAEQAKAANLAG
IGRLRDLREAAALSGDLANMLKAYSAAETKEAQLALLDNLIHKWAETDGSGEPEA-

H_6, $OAZ^{95\text{-}219}$, Y spacer, SPM ($FrpC^{414\text{-}647}$), C-tag

OAZ-GSY-SPM

MGSSHHHHHHFYSDDRLNVTEELTSNDKTRILNVQSRLTDAKRINWRTVLSGGSLYIEIPGGALPEGSKDSFAVLLEFAEEQL
RADHVFICFHKNREDRAALLRTFSFLGFEIVRPGHPLVPKRPDACFMAYTFEGSYDPLALDLDGDGIETVATKGFAGSLFDHT
NNGIRTATGWVSADDGLLVRDLNGNGIIDNGAELFGDNTKLADGSFAKHGYAALAELDSNGDNIINAADAAFQTLRVWQDLNQ
DGISQANELRTLEELGIQSLDLAYKDVNKNLGNGNTLAQQGSYTKTDGTTAKMGDLLLAADNLHSRFKDKVELTAEQAKAANL
AGIGRLRDLREAAALSGDLANMLKAYSAAETKEAQLALLDNLIHKWAETDGSGEPEA-

H_6, $OAZ^{95\text{-}219}$, GSY spacer, SPM ($FrpC^{414\text{-}647}$), C-tag

OAZ

MGSSHHHHHHFYSDDRLNVTEELTSNDKTRILNVQSRLTDAKRINWRTVLSGGSLYIEIPGGALPEGSKDSFAVLLEFAEEQL
RADHVFICFHKNREDRAALLRTFSFLGFEIVRPGHPLVPKRPDACFMAYTFEGSGEPEA-

H_6, $OAZ^{95\text{-}219}$, C-Tag

ODC

MGSSHHHHHHSSGNNFGNEEFDCHFLDEGFTAKDILDQKINEVSSSDDKDAFYVADLGDILKKHLRWLKALPRVTPFYAVKCN
DSKAIVKTLAATGTGFDCASKTEIQLVQSLGVPPERIIYANPCKQVSQIKYAANNGVQMMTFDSEVELMKVARAHPKAKLVLR
IATDDSKAVCRLSVKFGATLRTSRLLLERAKELNIDVVGVSFHVGSGCTDPETFVQAISDARCVFDMGAEVGFSMYLLDIGGG
FPGSEDVKLKFEEITGVINPALDKYFPSDSGVRIIAEPGRYYVASAFTLAVNIIAKKIVLKEQTGSDDEDESSEQTFMYYVND
GVYGSFNCILYDHAHVKPLLQKRPKPDEKYYSSSIWGPTCDGLDRIVERCDLPEMHVGDWMLFENMGAYTVAAASTFNGFQRP
TIYYVMSGPAWQLMQQFQNPDFPPEVEEQDASTLPVSCAWESGMKRHRAACASASINVGSGEPEA-

H_6, ODC, C-Tag

ODC K92R

MGSSHHHHHHSSGNNFGNEEFDCHFLDEGFTAKDILDQKINEVSSSDDKDAFYVADLGDILKKHLRWLKALPRVTPFYAVKCN
DSKAIVKTLAATGTGFDCASRTEIQLVQSLGVPPERIIYANPCKQVSQIKYAANNGVQMMTFDSEVELMKVARAHPKAKLVLR
IATDDSKAVCRLSVKFGATLRTSRLLLERAKELNIDVVGVSFHVGSGCTDPETFVQAISDARCVFDMGAEVGFSMYLLDIGGG
FPGSEDVKLKFEEITGVINPALDKYFPSDSGVRIIAEPGRYYVASAFTLAVNIIAKKIVLKEQTGSDDEDESSEQTFMYYVND
GVYGSFNCILYDHAHVKPLLQKRPKPDEKYYSSSIWGPTCDGLDRIVERCDLPEMHVGDWMLFENMGAYTVAAASTFNGFQRP
TIYYVMSGPAWQLMQQFQNPDFPPEVEEQDASTLPVSCAWESGMKRHRAACASASINVGSGEPEA-

H6, ODC[K92R], C-tag

ODC K92R T93K

MGSSHHHHHHSSGNNFGNEEFDCHFLDEGFTAKDILDQKINEVSSSDDKDAFYVADLGDILKKHLRWLKALPRVTPFYAVKCN
DSKAIVKTLAATGTGFDCASRKEIQLVQSLGVPPERIIYANPCKQVSQIKYAANNGVQMMTFDSEVELMKVARAHPKAKLVLR
IATDDSKAVCRLSVKFGATLRTSRLLLERAKELNIDVVGVSFHVGSGCTDPETFVQAISDARCVFDMGAEVGFSMYLLDIGGG
FPGSEDVKLKFEEITGVINPALDKYFPSDSGVRIIAEPGRYYVASAFTLAVNIIAKKIVLKEQTGSDDEDESSEQTFMYYVND
GVYGSFNCILYDHAHVKPLLQKRPKPDEKYYSSSIWGPTCDGLDRIVERCDLPEMHVGDWMLFENMGAYTVAAASTFNGFQRP
TIYYVMSGPAWQLMQQFQNPDFPPEVEEQDASTLPVSCAWESGMKRHRAACASASINVGSGEPEA-

H6, ODC[K92R,T93K], C-tag

ODC K92R Q96K

MGSSHHHHHHSSGNNFGNEEFDCHFLDEGFTAKDILDQKINEVSSSDDKDAFYVADLGDILKKHLRWLKALPRVTPFYAVKCN
DSKAIVKTLAATGTGFDCASRTEIKLVQSLGVPPERIIYANPCKQVSQIKYAANNGVQMMTFDSEVELMKVARAHPKAKLVLR
IATDDSKAVCRLSVKFGATLRTSRLLLERAKELNIDVVGVSFHVGSGCTDPETFVQAISDARCVFDMGAEVGFSMYLLDIGGG
FPGSEDVKLKFEEITGVINPALDKYFPSDSGVRIIAEPGRYYVASAFTLAVNIIAKKIVLKEQTGSDDEDESSEQTFMYYVND
GVYGSFNCILYDHAHVKPLLQKRPKPDEKYYSSSIWGPTCDGLDRIVERCDLPEMHVGDWMLFENMGAYTVAAASTFNGFQRP
TIYYVMSGPAWQLMQQFQNPDFPPEVEEQDASTLPVSCAWESGMKRHRAACASASINVGSGEPEA-

H6, ODC1 [K92R,Q96K], C-tag

ODC K92R S100K

MGSSHHHHHHSSGNNFGNEEFDCHFLDEGFTAKDILDQKINEVSSSDDKDAFYVADLGDILKKHLRWLKALPRVTPFYAVKCN
DSKAIVKTLAATGTGFDCASRTEIQLVQKLGVPPERIIYANPCKQVSQIKYAANNGVQMMTFDSEVELMKVARAHPKAKLVLR
IATDDSKAVCRLSVKFGATLRTSRLLLERAKELNIDVVGVSFHVGSGCTDPETFVQAISDARCVFDMGAEVGFSMYLLDIGGG
FPGSEDVKLKFEEITGVINPALDKYFPSDSGVRIIAEPGRYYVASAFTLAVNIIAKKIVLKEQTGSDDEDESSEQTFMYYVND
GVYGSFNCILYDHAHVKPLLQKRPKPDEKYYSSSIWGPTCDGLDRIVERCDLPEMHVGDWMLFENMGAYTVAAASTFNGFQRP
TIYYVMSGPAWQLMQQFQNPDFPPEVEEQDASTLPVSCAWESGMKRHRAACASASINVGSGEPEA-

H6, ODC1 [K92R,S100], C-tag

ODC K92R T396K

MGSSHHHHHHSSGNNFGNEEFDCHFLDEGFTAKDILDQKINEVSSSDDKDAFYVADLGDILKKHLRWLKALPRVTPFYAVKCN
DSKAIVKTLAATGTGFDCASRTEIQLVQSLGVPPERIIYANPCKQVSQIKYAANNGVQMMTFDSEVELMKVARAHPKAKLVLR
IATDDSKAVCRLSVKFGATLRTSRLLLERAKELNIDVVGVSFHVGSGCTDPETFVQAISDARCVFDMGAEVGFSMYLLDIGGG
FPGSEDVKLKFEEITGVINPALDKYFPSDSGVRIIAEPGRYYVASAFTLAVNIIAKKIVLKEQTGSDDEDESSEQTFMYYVND
GVYGSFNCILYDHAHVKPLLQKRPKPDEKYYSSSIWGPTCDGLDRIVERCDLPEMHVGDWMLFENMGAYTVAAASKFNGFQRP
TIYYVMSGPAWQLMQQFQNPDFPPEVEEQDASTLPVSCAWESGMKRHRAACASASINVGSGEPEA-

H6, ODC1 [K92R,T396], C-tag

TGFα-SPM

MGSSHHHHHHSSGVVSHFNDCPDSHTQFCFHGTCRFLVQEDKPACVCHSGYVGARCEHADLLADPLALDLDGDGIETVATKGF
AGSLFDHTNNGIRTATGWVSADDGLLVRDLNGNGIIDNGAELFGDNTKLADGSFAKHGYAALAELDSNGDNIINAADAAFQTL
RVWQDLNQDGISQANELRTLEELGIQSLDLAYKDVNKNLGNGNTLAQQGSYTKTDGTTAKMGDLLLAADNLHSRFKDKVELTA
EQAKAANLAGIGRLRDLREAAALSGDLANMLKAYSAAETKEAQLALLDNLIHKWAETDGSGEPEA-

H_6, TGFα, SPM (FrpC$^{414-647}$), C-tag

TGFα-GSY-SPM

MVVSHFNDCPDSHTQFCFHGTCRFLVQEDKPACVCHSGYVGARCEHADLLAGSYDPLALDLDGDGIETVATKGFAGSLFDHTN
NGIRTATGWVSADDGLLVRDLNGNGIIDNGAELFGDNTKLADGSFAKHGYAALAELDSNGDNIINAADAAFQTLRVWQDLNQD
GISQANELRTLEELGIQSLDLAYKDVNKNLGNGNTLAQQGSYTKTDGTTAKMGDLLLAADNLHSRFKDKVELTAEQAKAANLA
GIGRLRDLREAAALSGDLANMLKAYSAAETKEAQLALLDNLIHKWAETDGSSHHHHHHSSGEPEA-

TGFα, GSY spacer, SPM (FrpC$^{414-647}$), H_6, C-tag

TGFα[R42A]-GSY-SPM

MVVSHFNDCPDSHTQFCFHGTCRFLVQEDKPACVCHSGYVGA**A**CEHADLLAGSYDPLALDLDGDGIETVATKGFAGSLFDHTN
NGIRTATGWVSADDGLLVRDLNGNGIIDNGAELFGDNTKLADGSFAKHGYAALAELDSNGDNIINAADAAFQTLRVWQDLNQD
GISQANELRTLEELGIQSLDLAYKDVNKNLGNGNTLAQQGSYTKTDGTTAKMGDLLLAADNLHSRFKDKVELTAEQAKAANLA
GIGRLRDLREAAALSGDLANMLKAYSAAETKEAQLALLDNLIHKWAETDGSSHHHHHHSSGEPEA-

TGFα[R42A], GSY spacer, SPM (FrpC$^{414-647}$), H_6, C-tag

TGFα-GSY-[DA]SPM

MVVSHFNDCPDSHTQFCFHGTCRFLVQEDKPACVCHSGYVGARCEHADLLAGSY**A**PLALDLDGDGIETVATKGFAGSLFDHTN
NGIRTATGWVSADDGLLVRDLNGNGIIDNGAELFGDNTKLADGSFAKHGYAALAELDSNGDNIINAADAAFQTLRVWQDLNQD
GISQANELRTLEELGIQSLDLAYKDVNKNLGNGNTLAQQGSYTKTDGTTAKMGDLLLAADNLHSRFKDKVELTAEQAKAANLA
GIGRLRDLREAAALSGDLANMLKAYSAAETKEAQLALLDNLIHKWAETDGSSHHHHHHSSGEPEA-

TGFα, GSY spacer, [**DA**]SPM (FrpC$^{414-647}$[**D414A**]), H_6, C-tag

ST3-TGFα-SPM

MGSSHHHHHHSSGLVPRGSRGVPHIVMVDAYKRYKGSGESGVVSHFNDCPDSHTQFCFHGTCRFLVQEDKPACVCHSGYVGAR
CEHADLLADPLALDLDGDGIETVATKGFAGSLFDHTNNGIRTATGWVSADDGLLVRDLNGNGIIDNGAELFGDNTKLADGSFA
KHGYAALAELDSNGDNIINAADAAFQTLRVWQDLNQDGISQANELRTLEELGIQSLDLAYKDVNKNLGNGNTLAQQGSYTKTD
GTTAKMGDLLLAADNLHSRFKDKVELTAEQAKAANLAGIGRLRDLREAAALSGDLANMLKAYSAAETKEAQLALLDNLIHKWA
ETDGSGEPEA-

H_6, thrombin cleavage site, SpyTag003, TGFα, SPM (FrpC$^{414-647}$), C-tag

sEGFR501

MDAMKRGLCCVLLLCGAVFVSPSQEIHARFRRLEEKKVCQGTSNKLTQLGTFEDHFLSLQRMFNNCEVVLGNLEITYVQRNYD
LSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNMYYENSYALAVLSNYDANKTGLKELPMRNLQEILHGAVRFSNNPALCN
VESIQWRDIVSSDFLSNMSMDFQNHLGSCQKCDPSCPNGSCWGAGEENCQKLTKIICAQQCSGRCRGKSPSDCCHNQCAAGCT
GPRESDCLVCRKFRDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNYVVTDHGSCVRACGADSYEMEEDGVR
KCKKCEGPCRKVCNGIGIGEFKDSLSINATNIKHFKNCTSISGDLHILPVAFRGDSFTHTPPLDPQELDILKTVKEITGFLLI
QAWPENRTDLHAFENLEIIRGRTKQHGQFSLAVVSLNITSLGLRSLKEISDGDVIISGNKNLCYANTINWKKLFGTSGQKTKI
ISNRGENSCKATGQVCHALCSPEGCWGPEPRDCVSGSGESGHHHHHH-

tPA secretion leader, sEGFR501, H_6