Synthetic Biology Engineering to Catalyse Unbreakable Linkage Between Peptide Building Blocks

Can Martin Buldun

Department of Biochemistry
Corpus Christi College
University of Oxford

Trinity 2017

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

Abstract

Protein conjugation is an important tool for research and the development of novel therapeutics. Peptide tags are ideal handles for bioconjugation, as they are minimally disruptive to protein structure and function and are weakly immunogenic. In this thesis, I have developed SnoopLigase, a protein which is able to covalently link the two peptide tags SnoopTagJr and DogTag by site-specific transamidation. The three modules were obtained by splitting an isopeptide bond-containing domain of the *Streptococcus pneumoniae* adhesin RrgA. Using rational and computational protein design, rate and yield of the reaction were improved, while allowing efficient reaction in different buffers and temperatures.

The separation of conjugated proteins from unreacted substrate can be challenging when using established bioconjugation techniques. SnoopLigase bound strongly to its reaction product, which enabled the purification of conjugated SnoopTagJr-DogTag by SnoopLigase solid-phase immobilisation and subsequent product elution. Hence, SnoopLigase could be useful to produce high quality bioconjugate therapeutics.

Vaccines are important medical interventions, saving millions of lives each year. Recombinantly produced antigens are cost-effective and have an unmatched safety profile, yet often show poor immunogenicity. Protein multimerisation can strongly enhance the immune response against the antigen. Based on the SnoopLigase system and the heptameric protein IMX313, I developed a highly stable, modular vaccine platform. Covalent conjugation of SnoopTagJr-linked proteins to the platform nanoparticle IMX313-DogTag allowed simple, efficient and modular protein multimerisation. The efficient coupling by SnoopLigase and the small size of the tags provide potential to accelerate development of modular vaccines against malaria and other diseases.

Authorship Declaration

I declare that this thesis is entirely my own work, except for cases described below. This thesis has not been submitted for any other degree at the University of Oxford or another institution.

Experiments performed by others:

Chapter 5

Mass spectrometry was performed by Dr David Staunton

Proteins obtained by others:

Chapter 3

SnoopTag-4E6-SpyTag was produced by Dr Gianluca Veggiani

Chapter 5

SnoopTagJr-CIDR and CyRPA-SnoopTagJr were produced by Dr Anne-Marie Andersson

Acknowledgements

I would like to thank Professor Mark Howarth for his guidance, patience, teaching me the importance of efficiently streamlining projects to reach a set goal rather than getting lost in minor details, and allowing me to get an insight into academic entrepreneurship.

I would also like to express my gratitude to my funding bodies, Corpus Christi College and the Engineering and Physical Sciences Research Council (EPSRC), for their financial support and enabling me to study at the University of Oxford.

Thanks to the members of the Sherratt and Whitby laboratories for creating a wonderful and supportive atmosphere.

To all current and former members of the Howarth laboratory – I am grateful not only for your technical advice, but also your infinite source of positivity and entertainment. I would like to express special thanks to Chris and Sam who never ceased to surprise me with incredibly entertaining and sometimes inappropriate anecdotes, Matteo for introducing me to the world of programming; and Theo for being an amazing friend inside and outside the lab. For my development as a scientist and as a person I would like to especially thank Gianluca, who was an important mentor and used to call me his “Padawan”, and Karl, whose tranquillity, infectious ambitiousness and permanent innovativeness have inspired me.

I would also like to thank my friends Eric, Nadiya and Tianyi for making my academic and non-academic life in Oxford an enjoyable experience and reminding me that there is a life outside the laboratory.

Finally, very special thanks to Isabel and my mother for supporting me through the ups and downs of my studies.

Abbreviations

AP	Acceptor peptide
BCA	Bicinchoninic acid
Bio	Biotin
CD-HIT	Cluster Database at High Identity with Tolerance
CHAPS	3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate hydrate
CyRPA	Cysteine-rich protective antigen
DDM	*n*-Dodecyl β-D-maltoside
DHFR	Dihydrofolate reductase
DogT	DogTag
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
ESI-MS	Electrospray Ionisation Mass Spectrometry
FACS	Fluorescence-activated cell sorting
FPLC	Fast protein liquid chromatography
GST	Glutathione S-transferase
HIV	Human immunodeficiency virus
IPTG	Isopropyl β-D-1-thiogalactopyranoside
LB	Luria Bertani broth
MBP	Maltose binding protein
MSA	Multiple sequence alignment
MUSCLE	Multiple Sequence Comparison by Log-Expectation
MWCO	Molecular weight cut-off
Ni-NTA	Nickel-nitrilotriacetic acid
NMR	Nuclear magnetic resonance
OD	Optical density
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PEI	Polyethyleneimine
PMSF	Phenylmethylsulfonylfluoride
PROSS	Protein Repair One Stop Shop

PSI-BLAST	Position-Specific Iterated Basic Local Alignment Search Tool
RrgAL	RrgALigase
SDS	Sodium dodecyl sulfate
SEC	Size exclusion chromatography
SnL	SnoopLigase
SnT	SnoopTag
SnT2	SnoopTag2
SnTJr	SnoopTagJr
SUMO	Small Ubiquitin Modifier
TLR	Toll-like receptor
TMAO	Trimethylamine N-oxide
UAA	Unnatural amino acid
VLP	Virus-like particle
YESS	Yeast endoplasmic reticulum sequestration screening

Table of Contents

List of Figures and Tables

1. Introduction

1.1. Peptide-Based Systems for Covalent Protein Conjugation

Protein conjugation is an important tool for research and the development of novel therapeutics, allowing the incorporation of desired features of two distinct proteins into one molecule. This general principle enables a broad application spectrum for protein conjugates, such as optical tracking of fluorescently labelled proteins (Kaether and Gerdes, 1995), targeting toxins to specific cell types (McCluskey and Collier, 2013) or improving the half-life of protein drugs in blood circulation (Schellenberger et al., 2009).

Protein conjugation can be achieved by several genetic, chemical or enzymatic strategies. Genetic engineering has the advantage of being inherently site-specific. However, genetic fusions can interfere with protein expression, folding and function (Giepmans et al., 2006).

A wide variety of chemical conjugation techniques target functional groups of the naturally occurring amino acids, such as Lys (Chen et al., 2012), Trp (Antos and Francis, 2004), Tyr (Joshi et al., 2004) or Cys (Chalker et al., 2009). However, such modifications often occur at all accessible reactive sites of a target protein, yielding a combination of products, each with different properties. Precise control over location and stoichiometry of protein modifications for homogenous conjugates is a highly sought after target (Hu et al., 2016; Junutula et al., 2008; Kennedy et al., 2017; Meldal and Schoffelen, 2016; Shen et al., 2017; Sochaj et al., 2015; 2016). Site-specificity can be improved by incorporation of unnatural amino acids (UAAs) into the protein of interest. There is a rapidly expanding repertoire of UAAs to choose from, many of which allow the use of coupling techniques that do not target the natural amino acids (Brofelth et al., 2017; Wallace and Chin, 2014; Wang et al.,

2003a). However, the establishment and incorporation of UAAs can be challenging, as it requires the simultaneous expression of i) an engineered tRNA, ii) an engineered tRNA synthetase with high specificity and iii) the protein of interest with altered codon recognized by the tRNA. Furthermore, sufficient cytoplasmic delivery of the UAA needs to be achieved for efficient incorporation into the protein (Hallam et al., 2015; Xie and Schultz, 2005).

Peptides offer a great handle for bioconjugation, due to their small size and the possibility to genetically encode them with the protein of interest using the standard set of amino acids. Furthermore, the confinement of the conjugation reaction to a specific peptide sequence generally allows high site-specificity. Some of the most commonly used peptide-based bioconjugation systems are discussed below.

1.1.1. Transglutaminases

Transglutaminases are enzymes found in species across all kingdoms that catalyse the formation of isopeptide bonds between the *γ*-carboxamide group of a glutamine and the *ε*-amino group of a lysine residue in proteins (Figure 1.1) (Griffin et al., 2002). While not strictly speaking a peptide-based conjugation system, transglutaminases show substrate preference for certain amino acid sequences, rather than pure Gln and Lys specificity (Kuramoto et al., 2013; Strop et al., 2012; Sugimura et al., 2006, 2008).

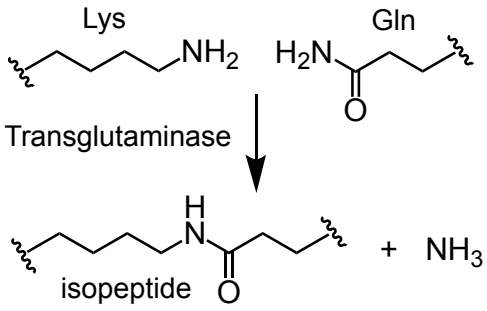


Figure 1.1 Transglutaminase reaction. Molecular basis of isopeptide bond formation mediated by transglutaminases. The transamidation reaction between Lys and Gln releases a molecule of ammonia.

As naturally occurring enzymes, transglutaminases are involved in processes like endocytosis, extracellular matrix construction, apoptosis or blood clotting (Odii and Coussons, 2014). The mild reaction conditions as well as broad substrate specificity have enabled the exploitation of transglutaminases for research and biotechnological applications (Rachel and Pelletier, 2013). These include tissue engineering (Ito et al., 2003), antibody-drug conjugation (Jeger et al., 2010), polyethylene glycol conjugation of pharmaceutical proteins (Fontana et al., 2008), formation of hydrogels (Teixeira et al., 2012), leather and fabric processing for the textile industry (Cortez et al., 2004; Tesfaw et al., 2014) or as biological glue in the food industry to modify the texture and mechanical strength of meat, soy, dairy and wheat products (Kieliszek and Misiewicz, 2014). The mild reaction conditions allow the use of transglutaminases in cellular environments (Lin and Ting, 2006).

Transglutaminases generally show poor substrate specificity – most Gln and Lys that are accessible for the enzyme can act as substrates (Groenen et al., 1994; Gundersen et al., 2014). The acyl-acceptor amine substrate can, apart from the lysine side chain, include a variety of other primary amines (Folk et al., 1980). A major advantage of transglutaminases as bioconjugation tool is that, due to the

molecular basis of the reaction, it is often possible to conjugate naturally occurring proteins – without the need of any protein engineering. While this characteristic is useful for macro-scale cross-linking of protein-rich products like meat, the poor specificity can be problematic for highly precise applications like drug development, where off-target conjugations of as low as 1.3% are considered problematic (Farias et al., 2014).

Although transglutaminases can in theory use all Gln and Lys of a protein as substrates, conjugation occurs only at a small number of sites (Mero et al., 2009). Since the molecular basis for the substrate preference of transglutaminases is not entirely clear, it is difficult to predict the conjugation behaviour for proteins of interest. Therefore, with the current state of knowledge it is required to empirically determine the number and exact location of conjugations when using new substrate proteins (Spolaore et al., 2012). Undesirable results require substrate or enzyme engineering (Farias et al., 2014; Zhao et al., 2010).

1.1.2. Inteins

Inteins are protein sequences that catalyse the post-translational modification process of protein splicing, in which an internal intein sequence excises itself from a protein, linking the N- and C-terminally flanking extein regions together (Kane et al., 1990). They exist in many different organisms and over 1,000 inteins have been found (Novikova et al., 2016). Most of these catalytic elements perform catalysis in cis (Figure 1.2a), but some inteins exist as separately expressed modules to perform trans-splicing (Figure 1.2b) (Saleh and Perler, 2006).

Additionally, artificial trans-splicing has been achieved by splitting cis-splicing inteins (Aranko et al., 2014a; Wu et al., 1998).

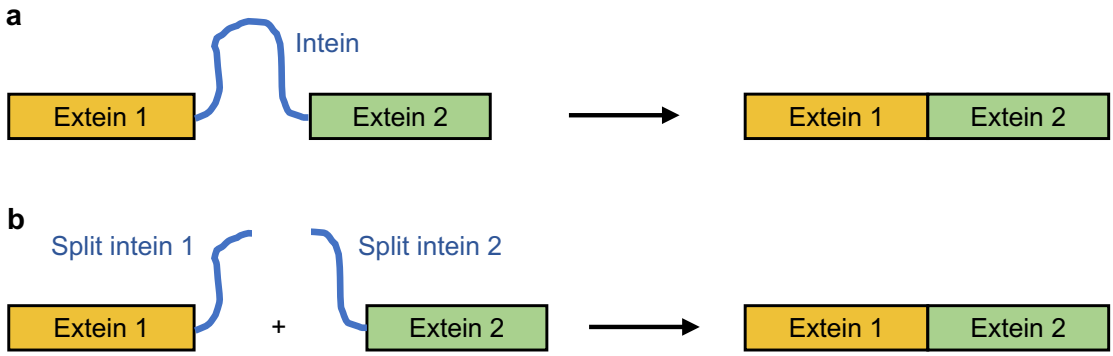


Figure 1.2 Inteins and split inteins for bioconjugation. a) An internal intein sequence excises itself from a polypeptide, linking the N- and C-terminally flanking exteins together. b) Split inteins enable trans-splicing, linking the N- and C-terminally flanking exteins together when expressed on separate polypeptides.

Problems that early described intein systems showed were slow kinetics and strong dependence on the immediately flanking extein sequences (Southworth et al., 1999). Inteins with improved properties have both been found in nature (Zettler et al., 2009) and developed through engineering (Aranko et al., 2014b; Stevens et al., 2016). Furthermore, while naturally occurring inteins have sizes of hundreds to over a thousand amino acids (Yang et al., 2004), engineering allowed the development of small split inteins, as short as 6 amino acid residues (Lin et al., 2013).

Inteins have been used for a wide range of applications, including drug discovery (Lennard and Tavassoli, 2014), protein purification (Chong et al., 1997; Coolbaugh et al., 2017), protein biotinylation (Lesaicherre et al., 2002), materials development (Ramirez and Chen, 2017) control of protein activity (Yuen et al., 2006) and enzyme cyclisation (Kanno et al., 2007). Intein-based synthetic protein components able to detect defined DNA sequences and subsequently elicit a

programmable intracellular response added to the repertoire of inducible parts for synthetic biology. The response was triggered through intein-mediated splicing of a trans-activator signalling to a genetic circuit that includes a response or reporter gene (Slomovic and Collins, 2015).

Despite their widespread use, inteins have several drawbacks. Although improvements have been achieved, the preference of inteins for specific flanking extein sequences remain a common theme (Lin et al., 2013), which could require screening different inteins systems when using new exteins. Furthermore, many naturally occurring and engineered inteins show solubility problems (Oeemig et al., 2009; Sydor et al., 2002; Zheng et al., 2012), rendering them unsuitable for working with poorly soluble exteins. Also, many fast-reacting inteins are derived from the same intein family and therefore show substantial cross-reactivity with homologous sequences (Aranko et al., 2014b), which makes the ligation of multiple fragments challenging. Another important limitation inherent to the intein reaction is that only N- or C-terminal, but not internal conjugation of exteins can be facilitated.

1.1.3. Sortases

Sortases are a family of membrane-associated enzymes in Gram-positive bacteria that mediate the covalent anchorage of proteins to the cell wall. Different sortases vary in their recognition motifs and specificities. The most extensively studied and utilised version is *Staphylococcus aureus* sortase A, which mediates the transpeptidation between an N-terminal oligoglycine and a C-terminal LPXTG sequence, where X represents any amino acid. Cleavage occurs between Thr and Gly of LPXTG, such that the carboxyl terminus of Thr gets joined with the amino terminus of oligoglycine (Figure 1.3) (Mao et al., 2004).

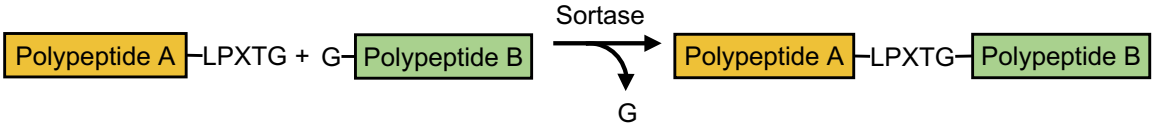


Figure 1.3 Sortase A for bioconjugation. Sortase A mediates the transpeptidation between an LPXTG- and an oligoglycine-tagged polypeptide.

Due to the small size of the oligoglycine and LPXTG moieties, the conjugation handles for sortagging minimally affect protein structure. Furthermore, orthogonal sortase systems allow construction of conjugates with precise architectures (Dorr et al., 2014; Raeeszadeh-Sarmazdeh et al., 2015).

Sortases have been used for a variety of applications, including triggerable conjugation of metabolic enzymes for substrate channelling (Matsumoto et al., 2016), fusion of antibody-like molecules with toxins for targeted toxicity (McCluskey and Collier, 2013), development of a modular hydrogel system (Valdez et al., 2017), or treatment of autoimmune diseases using modified erythrocytes (Pishesha et al., 2017). Another powerful use of sortases is segmental isotope-labelling for nuclear magnetic resonance (NMR) spectroscopy studies of proteins (Freiburger et al., 2015). NMR can provide valuable insight into the structure, dynamics and interactions of proteins (Marion, 2013). However, the analysis of large proteins represents a major challenge for this method (Marion, 2013). Uniformly enriching proteins with ^{13}C and ^{15}N isotopes increased the size of proteins that could be analysed by NMR (Kay et al., 1990; Shan et al., 1996). Further improvements could be achieved by segmental isotope-labelling, in which only a specific region of the protein is labelled (Volkmann and Iwaï, 2010; Xue et al., 2012). In order to achieve segmental isotope labelling, the different protein domains have to be expressed separately and subsequently assembled together, as achieved with sortase

(Freiburger et al., 2015). NMR spectroscopy requires ≥ 250 µM protein concentration, prohibiting the study of many biologically interesting but poorly soluble proteins (Amer et al., 2016). Using sortase-mediated attachment of a non-labelled solubilising protein to isotope-labelled proteins of interest, solubility problems of proteins to be used for NMR analyses could be overcome (Amer et al., 2016).

Despite their widespread use, sortase-mediated ligations have several disadvantages. Most sortases require millimolar concentrations of Ca^{2+}, limiting the utility of the system *in vivo*. However, engineering of *S. aureus* sortase A yielded a Ca^{2+}-independent sortase, which had inferior kinetic properties compared to the wildtype, but could be applied in cellular environments (Hirakawa et al., 2012; Pishesha et al., 2017; Swee et al., 2015). With K_m values of 7.6 mM for the LPXTG and 140 µM for the oligoglycine substrate (Chen et al., 2011), another drawback of sortagging is the high amount of substrate required (up to millimolar concentration) to achieve efficient product formation (Guimaraes et al., 2013), although an engineering trick enabled improvement of this aspect for *in vitro* applications (Wang et al., 2017). Furthermore, only N- and C-terminal, but not internal conjugations can be facilitated by sortase. Although the LPXTG recognition sequence can be placed at an internal site, due to the nature of the transpeptidation reaction only the N-terminal polypeptide sequence will be included in the conjugate (Guimaraes et al., 2013). Other problems with sortase-mediated ligation can be hydrolytic activity, formation of unwanted side-products, or strongly varying reaction rates dependent on the substrate (Antos et al., 2016; Heck et al., 2014). Sortase ligations often yield substrate conversion below 75% and only rarely over 90% (David Row et al., 2015;

Levary et al., 2011; Pritz et al., 2007; Wu and Guo, 2012). Using depsipeptides as substrate, nearly 100% conversion was achieved (Williamson et al., 2012).

1.1.4. Butelase

Butelase 1 is a Asn/Asp-specific cysteine ligase found in the plant *Clitoria ternatea* (Nguyen et al., 2014) which catalyses the transpeptidation between a C-terminal Asx-His-Val sequence (where Asx is Asn or Asp) and th N-terminus of a polypeptide (Figure 1.4). The specificity of butelase for the N-terminal sequence is relatively broad, accepting most residues at the first position and Ile, Leu, Val or Cys at the second position (Nguyen et al., 2016a). Its natural function is the cyclisation of peptides (Nguyen et al., 2016a). The enzyme shows high catalytic efficiency (K_{cat}/K_m 5.4 × 10^5 M^{-1} s^{-1}) and is able to reach more than 95% maximum substrate conversion (Nguyen et al., 2014).

Butelase has been exploited for the cyclisation of D-amino acid-containing peptides (Nguyen et al., 2016b), the synthesis of long peptides (Hemu et al., 2016), protein cyclisation (Nguyen et al., 2015) and bacterial cell surface labelling (Bi et al., 2017).

The catalytic profile and the small recognition motif make butelase a useful tool for bioconjugation. Compared to sortase, reactions typically require 100 – 1000 fold less input of enzyme, an important aspect for industrial applications (Rosen and Francis, 2017). However, similar to sortase, major limitations of butelase are the excess of one substrate required over the other (> 5-fold) to drive efficient intermolecular bioconjugation and the limitation of the reaction to the termini of peptides and proteins (Nguyen et al., 2016a).

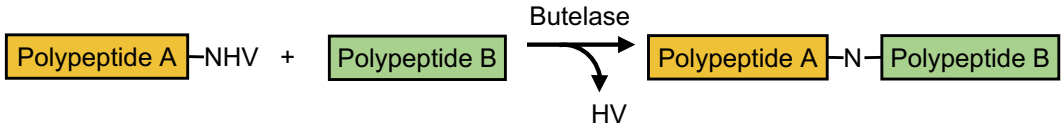


Figure 1.4 Butelase for bioconjugation. Butelase mediates the transpeptidation between a polypeptide with C-terminal NHV and the N-terminus of a second polypeptide.

Recently, an enzyme named OaAEP1 has been identified (Harris et al., 2015) and engineered (Yang et al., 2017). The protein has 66% sequence identity to butelase and shares similar reaction characteristics in its engineered form, but uses the C-terminal recognition sequence Asn-Gly-Leu . (Yang et al., 2017).

1.2. Spontaneous Isopeptide Bond Formation

Isopeptide bonds are amide bonds that are not part of the polypeptide backbone. They form between the ε-amino group of a lysine residue and the carboxyl or carboxamide group of Asn, Gln, Asp, Glu or a protein's C-terminus (Asquith et al., 1974). Several enzymes catalyse the formation of isopeptide bonds in nature, such as ubiquitin ligases (Pickart, 2001) or the family of transglutaminases described in section 1.1.1. Autocatalytic formation of intermolecular isopeptide bonds in HK97 phage has also been described (Wikoff et al., 2000). Intramolecular isopeptide bonds have first been identified in the major pilin protein Spy0128 of *Streptococcus pyogenes* (Figure 1.5a) (Kang et al., 2007) and were soon found in other surface proteins of Gram-positive bacteria (Forsgren et al., 2010; Izoré et al., 2010; Paterson and Baker, 2011; Vengadesan et al., 2011; Walden et al., 2014; Zähner et al., 2011).

Spontaneous isopeptide bond formation is enabled by a catalytic Glu or Asp residue and the reactive triad is located in the hydrophobic protein core to favour

the autocatalytic isopeptide bond formation (Hu et al., 2011). Based on this knowledge, an isopeptide bond was recently engineered *de novo* into an immunoglobulin-like protein (Kwon et al., 2017).

Intramolecular isopeptide bonds confer increased mechanical and thermostability, as well as resistance to proteolytic degradation (Alegre-Cebollada et al., 2010; Kang and Baker, 2009). These covalent bonds are resilient to extreme conditions (Kang and Baker, 2011) and dynamic force spectroscopy of a split protein suggested that isopeptide bonds have similar mechanical stability as other covalent bonds (Zakeri et al., 2012).

Splitting proteins into two or more fragments allows controlled reassembly of the functional parent domain upon mixing of the protein fragments. Successfully split proteins include fluorescent proteins (Zhang et al., 2004), dihydrofolate reductase (Remy et al., 2007), tobacco etch virus protease (Wehr et al., 2006) and ubiquitin (Johnsson and Varshavsky, 1994). Most split protein systems have been used as reporter systems for protein-protein interaction (Shekhawat and Ghosh, 2011).

The isopeptide bond-containing protein Spy0128 was split into two parts, such that the reactive lysine and asparagine were located on separate fragments (Figure 1.5b). Upon mixing, the two modules named fragment pilin-C and isopeptag formed a covalent bond (Zakeri and Howarth, 2010). Despite slow reaction and incomplete re-constitution, this approach was a promising tool for protein bioconjugation. Following this example, other isopeptide bond-containing protein domains have been split, some of which are described below.

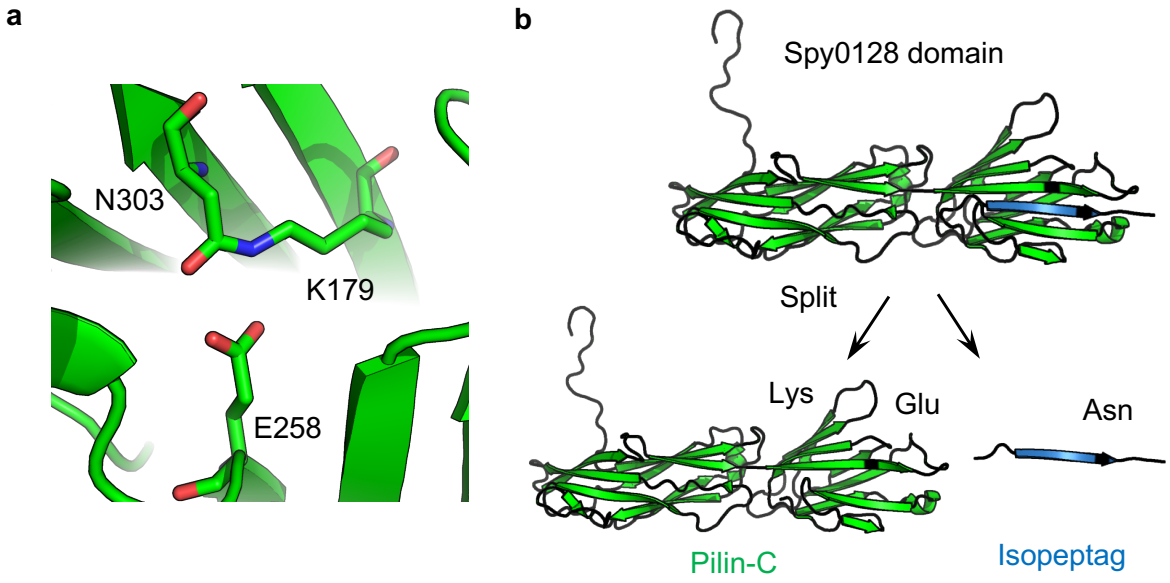


Figure 1.5 Principle of split Spy0128. a) Spy0128 (Protein Data Bank 3B2M) contains an isopeptide bond, formed between Lys and Asn, catalysed by Glu. b) Cartoon of Spy0128 splitting. The domain was split into two modules, with pilin-C (green) containing the catalytic Glu and reactive Lys and isopeptag (blue) the reactive Asn. Key residues for isopeptide bond formation are highlighted in black.

1.2.1. SpyCatcher

Splitting the CnaB2 domain from the fibronectin binding protein FbaB into two parts, the covalent bioconjugation system SpyCatcher-SpyTag was developed (Zakeri et al., 2012). The name of the system refers to *Streptococcus pyogenes*, the organism the protein sequence was derived from. SpyCatcher has a size of ca. 15 kDa and contains the catalytic Glu and the reactive Lys, while SpyTag consists of 13 amino acids including the reactive Asp residue (Figure 1.6a).

Compared to the split Spy0128 system, SpyCatcher-SpyTag reconstitution is significantly faster, with a reaction half time of ca. 70 s at 10 µM concentration of both reaction partners and reaches a higher maximum fragment reconstitution. The reaction is efficient under a wide range of conditions, including temperatures, pH values, buffers, reducing agent and non-ionic detergents. Although the isopeptide

bond formation between SpyCatcher and SpyTag releases a molecule of water (Figure 1.6b), no reverse reaction was detected in aqueous solution, where water concentration is ca. 55 M (Zakeri et al., 2012). SpyTag can be placed not only at the N- or C-terminus of a protein, but also between two proteins (Zakeri et al., 2012). Furthermore, the lack of requirement for cofactors or reducing agent makes SpyCatcher-SpyTag an advantageous system for use in cellular environments.

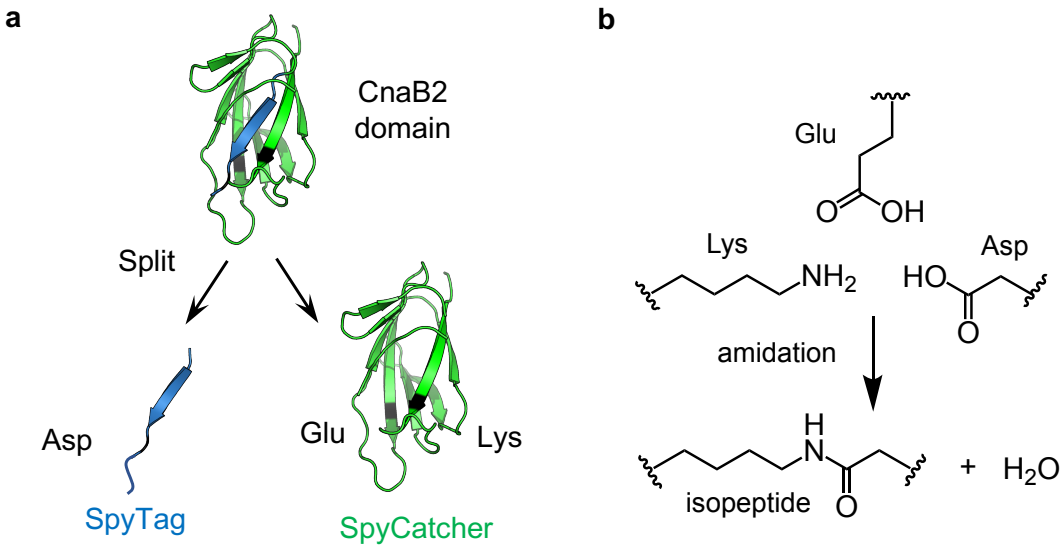


Figure 1.6 Principle of SpyCatcher design. a) Cartoon of CnaB2 splitting (Protein Data Bank 2X5P). The domain was split into two modules, with SpyCatcher (green) containing the catalytic Glu and reactive Lys and SpyTag (blue) the reactive Asp. Key residues for isopeptide bond formation are highlighted in black. b) Molecular basis of isopeptide bond formation in the SpyCatcher system. The amidation reaction between Lys and Asp is catalysed by Glu, eliminating a molecule of water.

Since its development in 2012, the SpyCatcher technology has been used for a variety of applications such as extracellular self-assembly of protein-protein conjugates with tunable composition (Gilbert et al., 2017), whole-cell biocatalysis (Peschke et al., 2017), modular vaccine-enhancing technologies (Brune et al., 2016; Liu et al., 2014), generation of bispecific antibodies (Yumura et al., 2017), enzyme

cyclisation for improved thermoresistance (Schoene et al., 2014) and cost-effective single-molecule DNA sequencing (Stranges et al., 2016).

Albeit the SpyCatcher technology allows fast and robust reaction, the primary disadvantage is the size of the SpyCatcher module, which can influence expression, folding and function of the protein it gets fused to. Furthermore, the size of SpyCatcher makes its insertion into proteins difficult, limiting its utility to the N- or C-terminus of fusion proteins.

1.2.2. SpyLigase

To overcome the problems connected to heterologous SpyCatcher insertion, Jacob Fierer in the Howarth laboratory split the CnaB2 domain into three instead of two modules. While leaving the previously developed SpyTag unchanged, he split the SpyCatcher protein into two modules, generating the ca. 11 kDa protein SpyLigase containing the catalytic Glu residue, and KTag, consisting of 10 amino acids including the reactive Lys residue (Figure 1.7). After further engineering of SpyLigase by circular permutation, the enzyme was able to catalyse the covalent conjugation of SpyTag and KTag (Fierer et al., 2014)

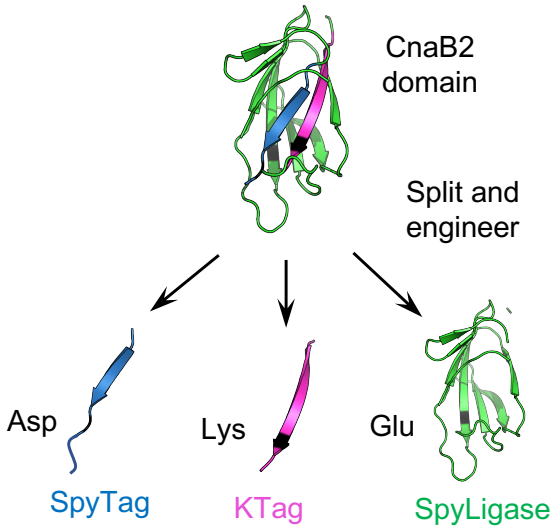


Figure 1.7 Principle of SpyLigase design. Cartoon of CnaB2 splitting (Protein Data Bank 2X5P). The domain was split into three modules, with SpyLigase (green) containing the catalytic Glu, SpyTag (blue) the reactive Asp, and KTag (pink) the reactive Lys. Key residues for isopeptide bond formation are highlighted in black.

Besides the polymer formation for enhanced cancer cell capture reported in the original publication (Fierer et al., 2014), SpyLigase has been used to produce antibody-drug conjugates (Siegmund et al., 2016). However, the general utility of the system is limited by its reaction characteristics. Ideally the reaction is performed in non-physiological conditions such as the presence of 1.5 M of the chemical chaperone trimethylamine N-oxide (TMAO) and a temperature of 4 °C. SpyLigase does not show substrate turnover and needs to be in four-fold molar excess over the substrate concentration for efficient reaction. The reaction half time is ca. 24 h and even after 72 h reaction time, the maximum substrate conversion only reaches slightly over 60% (Fierer, 2013; Fierer et al., 2014).

Despite its limited utility, SpyLigase was a useful proof of concept for the feasibility of tripartite splits of isopeptide bond-containing proteins.

1.2.3. SnoopCatcher

Following the principle of split Spy0128 and SpyCatcher-SpyTag, the C-terminal domain of the pilus-associated adhesin RrgA from *Streptococcus pneumoniae* was split by the Howarth laboratory to yield a novel Catcher-Tag pair. The initial split was performed by Raphael Gayet, yielding a peptide tag with the reactive Lys residue and a Catcher protein containing the reactive Asn and the catalytic Glu residues. Tomohiko Nakamura and Gianluca Veggiani further optimised the system, by shortening the tag's C-terminus by three residues and introducing the two point mutations G842T and D848G to stabilise secondary structure elements. The modules were named SnoopCatcher and SnoopTag (Veggiani et al., 2016).

The system reacts over a pH range of 6 – 9 and optimal temperatures of 12 to 37 °C. The reaction was not affected by high salt concentration, non-ionic detergents or reducing agent. The addition of 1.5 M of the chemical chaperone trimethylamine N-oxide (TMAO) improved the reaction slightly. The reaction can be pushed to completion by using a two-fold molar excess of one of the reaction partners, which is a major advantage over the maximal yield of sortases (Veggiani et al., 2016). Overall, its reaction characteristics make the SnoopCatcher system a widely applicable bioconjugation tool.

With a reaction half time of about 2.5 min SnoopCatcher reacts slightly slower than SpyCatcher. The Snoop- and SpyCatcher system do not show cross-reactivity. This orthogonality has been harnessed to build protein polymers with defined subunit composition and arrangement (Veggiani et al., 2016). Furthermore, a highly stable, multimeric and modular vaccine platform has been developed for simultaneous immunisation against two different antigens (Brune et al., 2017).

Similar to SpyCatcher, the major limitation of the SnoopCatcher system is the size of the Catcher protein, which may influence fusion protein expression, folding and function. Additionally, the reaction is slower than SpyCatcher, even in the presence of 1.5 M TMAO.

1.3. Vaccine Technologies

Vaccines are one of the most important medical interventions to prevent infectious diseases and have even allowed the complete and near eradication of new smallpox and polio incidences around the world (Behbehani, 1983; Delany et al., 2014). Furthermore, veterinary vaccines can have considerable indirect impact on human health and economy (Meeusen et al., 2007; Rich and Perry, 2011). A prominent example is the eradication of Rinderpest in 2011 – a cattle virus with nearly 100% morbidity rate which had caused severe economic and political crises throughout human history (Youde, 2013).

National immunisation programs have enabled the containment of most historically widespread infections in first world countries (Greenwood, 2014). Global interventions led by the World Health Organization and other alliances are working towards the same goal in low- and middle-income countries (Feikin et al., 2016). However, many diseases still lack effective vaccines. These include HIV, tuberculosis and malaria (Bourzac, 2014; Rappuoli and Aderem, 2011).

Historically, vaccines were live, attenuated or inactivated pathogens or toxins and have yielded effective vaccines against smallpox, influenza and hepatitis A, among others (Marmion, 1980; Salk and Salk, 1977; Ulmer et al., 2006). However, the use of vaccines derived from whole pathogens has important drawbacks. First, it can be challenging to establish a feasible and safe production procedure for large

quantities of the vaccine (van Aarle, 2006; Thomassen et al., 2013; Ulmer et al., 2006). Second, albeit live vaccines are cultured for the pathogen to lose its virulence, an important safety concern is the pathogen's reversion to virulence once administered to the patient (Hanley, 2011). Third, live, attenuated and inactivated vaccines can cause severe side-effects, especially the ones derived from bacteria as they contain a complex mixture of highly immunogenic proteins and polysaccharides (Jacobson, 2003; Lambert et al., 2013; Stratton et al., 1994). Fourth, vaccine efficacy often depends on the generation of neutralising antibodies (Siegrist, 2013). Complex vaccines like those derived from intact pathogens provide a wide range of different epitopes. Many of these epitopes do not yield neutralising antibodies, rendering the vaccines relatively ineffective (McComb and Martchenko, 2016; Reason et al., 2008; Zinkernagel et al., 2001).

Advances in biotechnology such as the design and production of recombinant proteins have helped to circumvent some of the problems described above.

1.3.1. Recombinant Subunit Vaccines

Producing vaccines based on recombinantly expressed and purified proteins rather than on pathogen-derived complex mixtures can be a safer, cheaper and more effective alternative (Moyle and Toth, 2013; Powers et al., 1995). One of the first successfully employed recombinant antigens targeted Hepatitis B Virus (Michel and Tiollais, 2010; Valenzuela et al., 1982). However, proteins purely based on the amino acid sequences occurring in pathogens often induce a weaker immune response, with occasional exceptions (Egan et al., 1987; Heinz et al., 1995; Moyle and Toth, 2013; Rota et al., 1990). Performance often needs to be enhanced by

addition of immunostimulatory adjuvants (Leroux-Roels, 2015; Reed et al., 2013). The formulation of antigen-adjuvant mixtures plays a key role in the development of modern vaccines (Reed et al., 2013). Furthermore, suitable delivery systems such as emulsions, liposomes or polymers can significantly improve immunogenicity (Sahdev et al., 2014).

Technological and computational advances in genomic and structural biology have allowed the design of improved vaccine antigens, by identifying protein regions that elicit the production of neutralising antibodies (Liljeroos et al., 2015). Structure-related strategies include binding studies of purified neutralising antibodies from previous vaccination studies, analysis of toxin crystal structures to identify functional protein regions and protein mutational studies (McComb and Martchenko, 2016; Zuckerman, 1985). This so-called “epitope mapping” allowed for example the identification of a loop structure important in Anthrax toxin action (Singh et al., 2001). Vaccination with based on this loop sequence enabled complete protection of rabbits against Anthrax (Oscherwitz et al., 2013). Genomic analysis of type B meningococcus, a bacterium with strong surface antigen variation, allowed the identification of five representative antigens that constituted the first universal vaccine against this infection (Giuliani et al., 2006).

Besides epitope identification, structural and computational approaches have helped to produce engineered antigens with improved stability and solubility, to enable efficient protein production (Campeotto et al., 2017). Further protein engineering efforts to enhance recombinant vaccine efficacy include fusion of antigens to adjuvants (Gregory et al., 2013; Moyle, 2017), antigen-chemokine fusion (Biragyn et al., 1999) and protein conjugates targeting the antigen to dendritic cells (Liu et al., 2014).

Despite genomic and structural optimisation of antigens, subunit vaccines still often show limited efficacy. Major improvement can be achieved by antigen multimerisation, which is discussed below.

1.3.2. Antigen Multimerisation

One reason for the stronger immune response often elicited by whole pathogens compared to recombinant antigen is the repeat architecture of antigens. B-cells react efficiently against those highly repetitive, rigidly ordered arrays of antigens (Bachmann et al., 1997; Zinkernagel, 2003). To improve the efficacy of protein vaccines, this concept was harnessed by fusing antigens to viral coat proteins. Virus-like particles (VLPs) are derived from and mimic the structural organisation of viruses but lack the pathogenic genetic material (Bárcena and Blanco, 2013; Jeong and Seong, 2017). Besides serving as multimerisation platform, VLPs are transported highly effectively to the lymphatic system, which enables them to stimulate a strong immune response (Manolova et al., 2008). Another advantage of VLPs is that they may encapsulate nucleic acids, which stimulate a Toll-like receptor (TLR) based response and act as an adjuvant (Schmitz et al., 2012).

Antigens have been successfully displayed on VLPs derived from a variety of viruses, such as Hepatitis B Virus (Rutgers et al., 1988; Valenzuela et al., 1985), Cowpea Mosaic Virus (Cañizares et al., 2005), Alfalfa Mosaic Virus (Modelska et al., 1998) or Tobacco Mosaic Virus (McComb et al., 2015). Artificial protein cages have also been developed by computational means (Bale et al., 2016; Hsia et al., 2016). The increased antibody titre caused by antigen multimerisation has been demonstrated in numerous cases (Baschong et al., 2003; Kushnir et al., 2012).

However, generation of fusion constructs is not always successful (Jain et al., 2015; Stockley and Mastico, 2000). This has led to the development of modular systems, in which the VLP and the antigen are produced separately and subsequently conjugated. This guarantees the correct folding of both the nanoparticle and the antigen. To generate reproducible results, high site-specificity of the conjugation method is important (Sletten and Bertozzi, 2009). For this purpose, Karl Brune in the Howarth laboratory developed SpyCatcher-VLPs which could be used to display SpyTag-fused antigens (Brune et al., 2016). A similar approach was taken by the Sander laboratory but with inverse SpyCatcher-SpyTag placement (Thrane et al., 2016). Sortase has been employed for a comparable system (Tang et al., 2016).

Apart from virus-derived structures, there are a large number of multimerisation domains to enhance the efficacy of recombinant vaccines (Engel and Kammerer, 2000; De Filette et al., 2008; Negahdaripour et al., 2017). One of the most important, the heptameric oligomerisation domain IMX313, is described below.

1.3.3. IMX313

IMX313 is an engineered oligomeric protein derived from the oligomerisation domain of the *Gallus gallus* complement inhibitor C4-binding protein (C4bp) (Ogun et al., 2008). Like the human C4bp, *Gallus gallus* C4bp mainly adopts a heptameric conformation and monomers are linked by inter-chain disulfide bonds (Figure 1.8) (Hofmeyer et al., 2013). Engineering of *Gallus gallus* C4bp aimed at removing self-antigen structural features for human and murine administration, giving rise to IMX313.

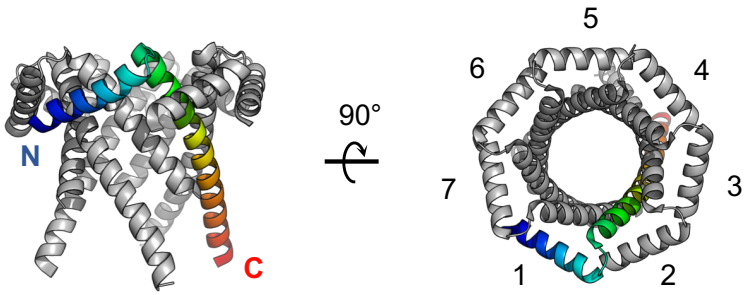


Figure 1.8 Structure of IMX313. Crystal structure adapted from *Gallus gallus* C4bp, Protein Data Bank 2YF2. One monomer shown in rainbow colour code from blue (N-terminus) to red (C-terminus). Individual monomers annotated on the right.

As the oligomer consists mainly of seven monomers, only a low number of antigens can be displayed on IMX313 compared to the dozens to hundreds of subunits that VLPs are made of (Michel and Tiollais, 2010; Qu et al., 2004). However, a dense array of only 2 – 3 antigens can be sufficient to elicit a strong B-cell response (Klaus and Humphrey, 1976). It has been shown that genetic fusions of antigens to IMX313 enhance the immune response (Li et al., 2016; Ogun et al., 2008). Based on antibody titre, antigen multimerisation on IMX313 can be nearly as effective as on VLPs (Leneghan et al., 2017). An IMX313-based tuberculosis vaccine has recently shown promising results in a phase I clinical trial (Minhinnick et al., 2016).

Attempts by Karl Brune in the Howarth laboratory to produce dually-addressable VLPs with both SpyCatcher and SnoopCatcher were unsuccessful due to failed or faulty assembly of fusion constructs or compromised reactivity of Catcher proteins (Brune, 2016). In a project parallel to this thesis, I have worked together with others from the Howarth and Biswas laboratories to develop and test an IMX313-based modular vaccination platform particle. Each IMX313 monomer was genetically fused to SpyCatcher and SnoopCatcher, allowing conjugation of the particle with two distinct antigens. Simultaneous antigen multimerisation increased

the antibody titres against both antigens (Brune et al., 2017). The platform particle developed in the Howarth lab has proven to be highly tolerant to freezing and elevated temperatures and could be lyophilised – aspects which can be problematic for VLPs (Brune, 2016; Brune et al., 2017). Since protein stability is an important aspect in vaccine development, IMX313 can be considered a reasonable alternative to VLPs (Chen and Kristensen, 2009; Jezek et al., 2009).

1.4. Thesis Objectives

The primary objective of this thesis was to develop a functional peptide-peptide ligation system named SnoopLigase as a general tool for bioconjugation. Desired features of the reaction included i) high conjugation yield ii) at acceptable reaction rate iii) under mild reaction conditions. After tripartite splitting of the isopeptide bond-containing parent domain, the focus was on characterising the reaction and engineering the modules to improve performance.

The second goal was to develop protocols for simple and economical purification of SnoopLigase reaction products to enable the production of highly uniform conjugate samples. To allow wide applicability of the approach, the aim was to identify mild purification conditions that are not disruptive to most proteins.

The third objective was to develop a highly stable modular immunisation platform, based on the mainly heptameric protein IMX313 and the newly designed SnoopLigase system. The focus was on establishing high conjugation yields of the platform particle with different model proteins and antigens to demonstrate that SnoopLigase-mediated bioconjugation can enable the development of potentially useful therapeutic products.

2. Methods

2.1. Molecular Cloning

2.1.1. Polymerase Chain Reaction (PCR)

Desiccated DNA oligonucleotides were purchased from Integrated DNA Technologies (IDT), re-constituted to a final concentration of 100 µM in 10 mM Tris–HCl pH 8.5 and stored at -20 °C until usage. Polymerase chain reactions (PCR) were prepared at 25 µL final volume, containing 0.2 ng plasmid DNA template, 0.5 µM of each forward and reverse primers, and 12.5 µL 2× Q5 high-fidelity DNA polymerase master mix (New England Biolabs). Thermocycling was performed using a C1000 Thermal Cycler (Bio-Rad). The standard thermocycling protocol consisted of an initial denaturation step for 30 s at 95 °C, 32 cycles of denaturation for 20 s at 95 °C, primer annealing for 15 s at 60 °C, and elongation for 40 s per kb at 72 °C, and one final elongation step for 5 min at 72°C. For amplicons shorter than 200 bp the number of cycles was increased to 35.

2.1.2. Agarose Gel Electrophoresis and PCR Product Purification

PCR reaction samples of 25 µL volume were mixed with 5 µL of 6× loading dye (1.2 M sucrose, 0.7 mM bromophenol blue, 0.9 mM xylene cyanol FF) and loaded on 0.5 – 2% (w/v) agarose gels (depending on the expected DNA fragment size) containing 0.01% (v/v) SybrSafe (Thermo Fisher Scientific). Samples were run for 45 min at 130 V in TAE buffer (40 mM Tris base, 20 mM acetic acid, 1 mM EDTA, pH 8.0). Agarose gels were placed on an XcitaBlue Conversion Screen and Filter (Bio-Rad) and imaged using a Gel Imager Universal Hood II (Bio-Rad). Images were analysed using Image Lab Version 5.2.1 (Bio-Rad). For PCR product purification,

PCR product bands of interest were excised and purified using the Wizard SV Gel and PCR Clean-Up Kit (Promega). DNA concentration was determined using an ND-1000 Nanodrop (NanoDrop).

2.1.3. Gibson Assembly

Gibson assembly of PCR fragments was performed using the standard protocol (Gibson et al., 2009), but with two of the enzymes cloned, expressed and purified by Karl Brune in the Howarth laboratory. The enzymes were a Phusion-like DNA polymerase based on a patent (Wang et al., 2003b) and Taq DNA ligase. T5 exonuclease was obtained from New England Biolabs. 2× Gibson assembly master mix was stored as 10 µL aliquots at -20 °C.

Agarose gel-purified PCR fragments were mixed on ice at equimolar ratio at the highest concentration possible to a final volume of 10 µL. 10 µL ice-cold 2× Gibson assembly master mix was added and samples were incubated for 1 – 8 h at 50 °C in a pre-heated C1000 Thermal Cycler (Bio-Rad).

2.1.4. *E. coli* Transformation

8 µL of Gibson assembly samples or 1 µL of supercoiled plasmid were transformed into 50 – 100 µL aliquots of heat competent *E. coli* NEB 5α, NEB Turbo or XL1-Blue cells. Heat shock was performed for 45 s in a 42 °C water bath, followed by 2 min incubation on ice. Alternatively, up to 2 µL of Gibson assembly samples were transformed into 50 µL aliquots of electrocompetent NEB 5α cells.

2.1.5. Plasmid Preparation

For plasmid preparation the GeneJET Plasmid Miniprep Kit (Thermo Fisher Scientific) or the QIAprep Spin Miniprep Kit (Qiagen) was used with double the amount of recommended buffer volumes. 5 – 10 mL of *E. coli* overnight cultures were employed.

2.1.6. DNA Sequence Verification

For all constructs made, the genes of interest were sequenced with Sanger sequencing performed by either Source BioScience (Nottingham, UK) or Eurofins Genomics (Ebersberg, Germany).

2.1.7. Protein Construct Sequences

Amino acid sequences of proteins used in this thesis are shown in the Appendix.

2.2. Protein Expression and Purification

2.2.1. Media and *E. coli* Growth Conditions

Media were obtained from the Media Glassware Sterilization Laboratory of the University of Oxford Department of Biochemistry. LB-Miller (high salt) consisted of 10 g tryptone, 5 g yeast extract and 10 g NaCl dissolved in Milli Q water to a final volume of 1 L. 2× YT was composed of 16 g tryptone, 10 g yeast extract and 5 g NaCl dissolved in Milli Q water to a final volume of 1 L, adjusted to pH 7.0. All media were sterilised by autoclaving for 20 minutes at 121 °C prior to storage at room

temperature. Bacterial cultures were grown in an Infors Multitron shaker (Infors UK Ltd) with 2.5 cm throw at 200 rpm.

2.2.2. Protein Expression

Proteins were expressed in *E. coli* BL21 (DE3) RIPL (Agilent), except for IMX313-DogTag, which was expressed in BL21 (DE3) RIPL +PGL. PGL is phosphogluconolactonase, an enzyme of the pentose phosphate pathway. With endogenous levels of PGL, 6-phosphogluconolactone accumulates in the cells and causes protein gluconylation. Overexpression of the PGL gene reduces gluconylation of proteins (Aon et al., 2008). BL21 (DE3) RIPL containing the PGL gene was made by Karl Brune in the Howarth laboratory.

Single colonies were picked from LB-agar plates and grown overnight at 37 °C in 2× YT containing 0.5 mg/mL kanamycin or 0.1 mg/mL ampicillin, depending on the plasmid's resistance gene. 10 mL of overnight cultures were added to 1 L 2× YT containing 0.8% (w/v) glucose and the appropriate antibiotic. IMX313-DogTag was expressed without glucose. Cells were grown at 37 °C to OD600 of 0.4 – 0.8 and expression induced with 0.42 mM IPTG for 4 h at 30 °C. IMX313-DogTag was expressed for 4 h at 25 °C. HaloTag7-SS-SnoopLigase was expressed for 6 h at 18 °C. Cultures were centrifuged for 20 min at 4,000 g at 4 °C, and pellets processed immediately or stored at -80 °C.

2.2.3. Non-Denaturing Purification of His_6-Tagged Proteins

E. coli culture pellets derived from 1 L cultures were resuspended in 10 mL ice-cold Ni-NTA binding buffer (50 mM Tris base, 300 mM NaCl, pH 7.8) containing

10 mM imidazole pH 7.8, 10 mM phenylmethylsulfonylfluoride (PMSF) (Sigma) and EDTA-free mixed protease inhibitors (Roche). Samples were sonicated on ice six times for 30 s with 1 min breaks between sonication cycles using a Sonics Vibra-Cell sonicator (Ultrasonic Processor). Cell lysates were centrifuged for 30 min at 17,000 g, 4 °C. The supernatant was incubated with 0.5 mL Ni-NTA resin (Qiagen) per L of bacterial culture. For high-expressing proteins such as MBP- or AffiHER2-containing constructs the amount of Ni-NTA resin was doubled. Incubation with Ni-NTA resin was performed for 30 min at 4 °C on a rolling incubator. Samples were applied to polyprep columns (Bio-Rad) at 4 °C and the flow-through collected for later analysis. The resin was washed five times with ten resin volumes of Ni-NTA binding buffer containing 10 mM imidazole, and once with two resin volumes of Ni-NTA binding buffer containing 30 mM imidazole. For elution, the polyprep columns were capped and the resin incubated with one resin volume of Ni-NTA buffer containing 200 mM imidazole for 5 min. The sample was collected and the elution repeated twice more, but without incubation time. For high-expressing proteins such as MBP- or AffiHER2-containing constructs, one additional elution step was performed. All washing and elution steps were performed at 4 °C. Purified proteins were dialysed three times against 1000-fold volume excess of suitable buffer using a Spectra/Por 3 3,500 Da MWCO membrane (Spectrum Labs). Dialysis buffers were 25 mM sodium borate pH 10 – 10.5 for RrgALigase, variants thereof and constructs containing AffiHER2, 25 mM Tris borate pH 8.0 for constructs containing MBP or SUMO and 25 mM Tris borate pH 7.4 for HaloTag7-SnoopLigase or HaloTag7SS-SnoopLigase.

2.2.4. Denaturing Purification of His_6-Tagged IMX313-DogTag

The purification of His_6-Tagged IMX313-DogTag was based on a previously reported protocol (Chae et al., 2004). *E. coli* culture pellets derived from 1 L cultures were resuspended in 10 mL ice-cold Ni-NTA binding buffer (50 mM Tris base, 300 mM NaCl, pH 7.8) containing 10 mM phenylmethylsulfonylfluoride (PMSF) (Sigma) and EDTA-free mixed protease inhibitors (Roche). Samples were sonicated on ice six times for 30 s with 1 min breaks between sonication cycles using a Sonics Vibra-Cell sonicator (Ultrasonic Processor). To 12 mL of cell lysate 24 g of Guanidinium-HCl was added, yielding a final volume of approximately 30 mL. 20 mM 2-Mercaptoethanol was added and the sample incubated for 1 h at 25 °C on an end-over-end tube rotor. Cell lysates were centrifuged for 30 min at 35,000 g, 4 °C using a JA-20 rotor (Beckman Coulter). The supernatant was centrifuged once more and subsequently incubated with 1 mL Ni-NTA resin (Qiagen) per L of bacterial. Incubation with Ni-NTA resin was performed for 1 h at 4 °C on a rolling incubator. Samples were applied to polyprep columns (BioRad) at 4 °C and the flow-through collected for later analysis. The resin was washed twice with ten resin volumes of wash buffer (50 mM Tris base, 300 mM NaCl + 10 mM imidazole pH 7.8 + 20 mM 2-Mercaptoethanol) containing 8 M urea, twice with ten resin volumes of wash buffer containing 4 M urea, twice with ten resin volumes of wash buffer containing 2 M urea, twice with ten resin volumes of wash buffer containing 1 M urea and no 2-Mercaptoethanol, twice with ten resin volumes of wash buffer containing 0.5 M urea and no 2-Mercaptoethanol and twice with ten resin volumes of Ni-NTA binding buffer containing 100 mM imidazole. For elution, the polyprep column was capped and the resin incubated with one resin volume of Ni-NTA buffer containing 2 M

imidazole for 10 min. The sample was collected and the elution repeated twice more, but without extended incubation. All washing and elution steps were performed at 4 °C. Purified His_6-IMX313-DogTag was dialysed three times against 1000-fold volume excess of 50 mM Tris borate pH 8.0 using a regenerated cellulose 6-8 kDa MWCO membrane (Fisher Scientific).

2.2.5. Purification of His_6-Tag-Free IMX313-DogTag

E. coli culture pellets derived from 1 L cultures were resuspended in 10 mL ice-cold lysis buffer (25 mM Tris phosphate, 25 mM citric acid, pH 5.0) containing 10 mM phenylmethylsulfonylfluoride (PMSF) (Sigma) and EDTA-free mixed protease inhibitors (Roche). Samples were sonicated on ice six times for 1 min with 1 min breaks between sonication cycles using a Sonics Vibra-Cell sonicator (Ultrasonic Processor). The cell lysate was subjected to a series of purification steps described below.

2.2.5.1. Heat Purification

The cell lysate was divided into 1.5 mL aliquots in 2 mL tubes (Anachem) and incubated for 15 min at 80 °C, 400 rpm in a ThermoMixer comfort (Eppendorf). Samples were cooled on ice for 5 min, 100 mM 2-Mercaptoethanol added, and samples were centrifuged for 30 min at 35,000 g, 4 °C using a JA-20 rotor (Beckman Coulter). The supernatant was collected for further processing.

2.2.5.2. Acid Precipitation

Per 10 mL of supernatant, 3 mL of 500 mM glycine pH 2.0 was added and samples centrifuged for 30 min at 35,000 g, 4 °C using a JA-20 rotor (Beckman Coulter). The supernatant was discarded and the tubes left standing inverted on a tissue for 5 min to remove as much supernatant as possible. The pellet was dissolved by thorough vortexing at 25 °C in 2.5 mL of 50 mM Tris–HCl pH 8.0 + 2 mM DTT per 10 mL of supernatant.

2.2.5.3. PEI Precipitation

0.1% polyethyleneimine (PEI) (v/v) was added and samples were centrifuged for 30 min at 35,000 g, 4 °C using a JA-20 rotor (Beckman Coulter). The supernatant was collected. PEI-treated supernatant showed an approximately 6× lower A_{280} value compared to non-treated supernatant. Furthermore, the sample showed lower viscosity, indicating the removal of DNA from the protein sample.

2.2.5.4. Ammonium Sulfate Precipitation

5 mL of 3.375 M ammonium sulfate was added to the PEI-treated supernatant (section 2.2.5.3.), giving a final concentration of 2.5 M. Samples were centrifuged for 30 min at 35,000 g, 4 °C using a JA-20 rotor (Beckman Coulter). The supernatant was discarded and the pellet dissolved in 2.5 mL of 50 mM Tris–HCl pH 8.0. The ammonium sulfate precipitation was repeated once more. The sample was dialysed for 2 h at 25 °C into 2.5 L of 10 mM Tri-HCl pH 8.0 using a Spectra/Por 3 3,500 Da MWCO membrane (Spectrum Labs).

2.2.5.5. Anion Exchange Chromatography

Anion exchange chromatography of IMX313-DogTag was performed on an ÄKTA pure 25 fast protein liquid chromatography (FPLC) system (GE Healthcare). The sample was loaded onto a 1 mL quaternary high performance (Q-HP) column (GE Healthcare) at 0.5 mL/min flow rate. The column was washed with 10 column volumes of 10 mM Tris-HCl pH 8.0, followed by elution with 10 column volumes of buffer containing 180 mM NaCl. Protein purity of the 0.5 mL fractions was determined using SDS-PAGE. Purified IMX313-DogTag was dialysed three times against 1000-fold volume excess of 50 mM Tris–HCl pH 8.0 using a regenerated cellulose 6-8 kDa MWCO membrane (Fisher Scientific).

2.2.5.6. Endotoxin Depletion and Quantification

Endotoxin depletion of the purified IMX313-DogTag sample was performed by Triton X-114 phase separation as described previously (Aida and Pabst, 1990). Endotoxin concentration was quantified using the Pierce LAL Chromogenic Endotoxin Quantitation Kit (Thermo Fisher Scientific) according to the manufacturer's protocol.

2.2.6. Purified Proteins Obtained from Others

The following purified proteins were obtained from others in the Howarth laboratory:

- SnoopTag-4E6-SpyTag, Gianluca Veggiani (Veggiani et al., 2016)
- SnoopTagJr-CIDR, Anne-Marie Andersson
- CyRPA-SnoopTagJr, Anne-Marie Andersson

2.3. Quantification of Protein Concentration

Protein concentrations were measured using Bicinchoninic Acid (BCA) assay according to the kit manufacturer's protocol (Thermo Fisher Scientific). Alternatively, A_{280} was measured using an ND-1000 Nanodrop (NanoDrop), the extinction coefficient predicted by ExPASy ProtParam (Gasteiger et al., 2005).

2.4. Protein Biotinylation

AP-SnoopLigase was biotinylated as described (Fairhead and Howarth, 2015). 220 µM AP-SnoopLigase was incubated with 14.7 µM GST-BirA, 0.5 mM $MgCl_2$, 3.3 mM D-biotin and 1 mM ATP in Tris borate pH 8.0 for 1 h at 25 °C, before addition of the same amount of GST-BirA and D-biotin and incubation for 1 h at 25 °C. To remove GST-BirA, the sample was incubated with 100 µL of glutathione-HiCap resin (Qiagen) for 30 min at 25 °C on a sample rotor and centrifuged for 30 s at 17,000 g. The supernatant was dialysed three times 1:1000 into Tris borate pH 8.0 using a Spectra/Por 3 3.5 kDa MWCO regenerated cellulose membrane (Spectrum Labs).

2.5. SDS-PAGE

Tris-glycine sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) was performed using Tris-glycine gels and XCell SureLock Mini-cell gel chambers (Thermo Fisher Scientific). 6× SDS loading buffer (0.23 M Tris–HCl pH 6.8, 24% v/v glycerol, 0.23 M SDS, 120 µM bromophenol blue) was added to samples at 1:6 ratio, heated to 99 °C for 3 min using a C1000 thermal cycler (Bio-

Rad) and cooled to room temperature for 5 min. 100 mM 2-Mercaptoethanol or 2 mM DTT was included in the SDS loading buffer for samples that required reduction of disulfide bonds. Tris-glycine gels were run in Tris-glycine running buffer [25 mM Tris–HCl, 192 mM glycine, 0.1% (w/v) SDS] for approximately 1 h at 200 V, room temperature. Pre-stained PageRuler (Thermo Fisher Scientific) was used as molecular weight marker. To visualise proteins, gels were stained using InstantBlue Coomassie stain (Expedeon) and destained with MilliQ water. Images were acquired using a ChemiDoc XRS+ Imager (Bio-Rad). Quantification of protein bands (densitometry) was performed using the Image Lab Software, version 5.2.1 (Bio-Rad).

2.6. Protein Characterization

2.6.1. Mass Spectrometry

Proteins to be analysed by mass spectrometry were dialysed into 10 mM ammonium acetate pH 7.0. Electrospray ionisation mass spectrometry (ESI-MS) was performed by Dr David Staunton from the Department of Biochemistry Biophysical Instrument Facility. A m/z (mass/charge) spectrum was obtained for each sample using a Micromass LCT time-of-flight (TOF) ESI-MS (Micromass UK) instrument. The m/z spectrum was converted to a molecular mass spectrum using MassLynx V 4.00.00 (Water Corporation).

2.6.2. Size Exclusion Chromatography

Size exclusion chromatography (SEC) was performed at 4 °C on an ÄKTA pure 25 fast protein liquid chromatography (FPLC) system (GE Healthcare). Protein

samples were loaded onto a HiLoad 16/600 Superdex 200 pg column (GE Healthcare) at 1 mL/min flow rate using 50 mM Tris borate pH 7.25 as buffer. Proteins in the 1 mL fractions were analysed using SDS-PAGE. Fractions containing the desired protein were concentrated to a final volume of approximately 100 µL using a Vivaspin centrifugal concentrator with 10 kDa MWCO (GE Healthcare). The Gel Filtration Standard (Bio-Rad, catalog number 151-1901) run was performed by Anne-Marie Andersson in the Howarth Laboratory.

2.6.3. Dynamic Light Scattering

Concentrated protein samples from SEC peaks were centrifuged for 30 min at 17,000 g, 4 °C to remove aggregates and loaded into a reusable cuvette. The size distribution of each protein sample was measured at 1:1, 1:2 and 1:4 dilution in buffer at 20 °C using an Omnisizer (Viscotek). Samples were measured by 10 scans each at 10 s accumulation time and the intensity distribution was plotted.

2.7. *In Vitro* Protein Conjugation Assays

2.7.1. General Set-Up

Conjugation assay set-ups varied strongly and are described in the results section for each experiment. Protein conjugations were generally performed at 10 µM of each protein component. To push the reaction to completion, for example to efficiently conjugate proteins to IMX-DogTag, the concentration of SnoopTagJr-linked protein and SnoopLigase was doubled. Short reaction times were chosen to compare the reactivity of different proteins or reaction conditions and longer reaction times to push conjugation reactions to completion. The final standard SnoopLigase

reaction conditions were 50 mM Tris borate pH 7.25 + 15% (v/v) glycerol as reaction buffer and 4 °C reaction temperature.

2.7.2. Reaction quantification

Assays required for quantification were performed in triplicate and analysed by densitometry (section 2.5.). The following formulas were used for calculations:

$$\%\ Product\ formed = \frac{[Product]}{[Product] + [Substrates]} \times 100\%$$

$$Relative\ product\ formed = \frac{\%\ Product\ formed\ (sample)}{\%\ Product\ formed\ (reference)}$$

$$\%\ Change\ in\ protein\ amount\ over\ time = \frac{[Sample]\ at\ time\ X}{[Sample]\ at\ time\ 0} \times 100\%$$

2.8. Protein Thermal Resilience Analysis

SnoopLigase at 12.5 μM in 50 mM Tris borate pH 7.25 + 15% (v/v) glycerol was incubated at 25 – 99 °C for 15 min using a C1000 thermal cycler (Bio-Rad) and cooled to 4 °C for 5 min. Heat-treated SnoopLigase was used for *in vitro* protein conjugation assays with SUMO-DogTag and SnoopTagJr-AffiHER2 as substrates in 50 mM Tris borate pH 7.25 + 15% (v/v) glycerol, 4 °C for 2 h, followed by SDS-PAGE and densitometry (section 2.5).

IMX313-DogTag at 50 μM in 50 mM Tris borate pH 7.25 was incubated at 25 – 99 °C for 15 min using a C1000 thermal cycler (Bio-Rad) and cooled to 12 °C for

30 min. Heat-treated IMX313-DogTag was used for *in vitro* protein conjugation assays with SnoopLigase and SnoopTagJr-AffiHER2 in 50 mM Tris borate pH 7.25 + 15% (v/v) glycerol, 4 °C for 2 h, followed by SDS-PAGE and densitometry (section 2.5.).

2.9. Protein Lyophilisation

30 µL aliquots of SnoopLigase or IMX313-DogTag at 10 µM in 50 mM Tris borate pH 7.25 were prepared in 100 µL thin-wall PCR tubes (StarLab). Special care was taken that the liquid was at the bottom of the tubes. The lids of the tubes were closed and pierced with a 1.2 × 38 mm syringe needle (Terumo). Up to five tubes were placed upright at the bottom of a 50 mL Falcon tube containing 4 mL MilliQ water, so that the bottoms of the PCR tubes were immersed in water. The Falcon tube was placed into a dry ice-ethanol bath for 10 min for sample shock-freezing. The purpose of placing the PCR tubes into a water-containing Falcon tube was to enable the transfer of the samples into the lyophiliser without thawing of the samples. The Falcon tube was placed without cap into a BenchTop 2K freeze-dryer (VirTis) and lyophilisation was performed for 48 h under vacuum at 0.14 mbar and −72.5 °C. Samples were stored in a glass scintillation vial sealed with Parafilm (Sigma) on a bed of Drierite (Sigma) to minimise sample hydration. The scintillation vial was stored at 37 °C to simulate suboptimal storage conditions and aliquots were removed at different time points to perform conjugation assays. Samples were briefly centrifuged to collect the lyophilised powder at the bottom of the tube, and reconstituted in reaction buffer to give a final concentration of 10 µM of all protein

components and 50 mM Tris borate pH 7.25 + 15% (v/v) glycerol. Reactions were performed for 2 h at 4 °C, followed by SDS-PAGE and densitometry (section 2.5.).

2.10. Purification of SnoopLigase Reaction Product

2.10.1. General Protein Conjugation and Immobilisation Set-Up

Protein conjugations were generally performed at 10 µM of each protein component. Proteins employed were a DogTag- and a SnoopTagJr-linked substrate protein, as well as either biotinylated SnoopLigase or HaloTag7-SnoopLigase. To push the reaction to completion, for example to efficiently conjugate proteins to IMX-DogTag, the concentration of SnoopTagJr-linked protein and functionalised SnoopLigase was doubled. After reaction, samples were incubated with HiCap streptavidin agarose (Thermo Fisher Scientific) or HaloLink resin (Promega) on a tube rotor. Typical resin volumes were 5 µL streptavidin agarose or 10 µL of HaloLink resin per 100 µL of reaction mix containing 10 µM functionalised SnoopLigase. For HaloLink coupling, 0.01% v/v Tween-20 was added. Furthermore, 500 µM imidazole pH 7.25 could improve solubility of reaction samples that showed protein aggregation. Incubation with streptavidin agarose was performed for 30 min at 25 °C, and for 30 min – 12 h with HaloLink resin (conjugated IMX313-DogTag required long incubation times). Samples were transferred to 1 mL polyprep columns (Bio-Rad) and centrifuged for 1 min at 300 g. All subsequent spin steps were also performed at 300 g.

2.10.2. Purification by Glycine

All steps were performed at 4 °C. The resin with immobilised protein was washed twice with five resin volumes of ice-cold 50 mM glycine pH 3.0 containing 300 mM NaCl and three times with five resin volumes of ice-cold 50 mM glycine pH 3.0. One extra spin ensured thorough removal of remaining buffer from the resin. The resin was incubated for 1 min with one resin volume of 50 mM glycine pH 2.0 at 4 °C and centrifuged into an Eppendorf tube containing 0.3 resin volumes of 1 M Tris–HCl pH 9.5. The elution was repeated twice more with centrifugation into the same Eppendorf tube and without addition of fresh Tris–HCl.

2.10.3. SnoopLigase Reaction Cycling

To analyse SnoopLigase re-usability after reaction product stripping by low pH buffer, 50 µL of biotinylated SnoopLigase at 50 µM in Tris borate pH 8.0 was coupled to 10 µL washed and equilibrated HiCap Streptavidin Agarose (Thermo Fisher Scientific) for 30 min at 25 °C on a rotating tube incubator. The sample was transferred to a 1 mL polyprep column (Bio-Rad) and centrifuged for 1 min at 300 g, followed by five washes with 100 µL of 50 mM Tris borate pH 8.0. The polyprep column was capped and the reaction started by addition of 50 µL reaction mix, containing 50 µM SUMO-DogTag and 50 µM SnoopTagJr-AffiHER2 in Tris borate pH 7.25 with 15% (v/v) glycerol. The sample was incubated for 1 h at 4 °C on a thermomixer at 800 rpm. The reaction product was purified by the glycine method (section 2.10.2). After reaction product elution, the resin was washed twice with 100 µL glycine pH 2.0 to ensure complete removal of residual reaction product and twice with 100 µL Tris borate pH 7.25. To start the next reaction cycle, fresh reaction mix

was added, and the reaction and purification procedure repeated. Overall eight reaction cycles were performed. Samples were analysed by SDS-PAGE and densitometry (section 2.5.).

2.10.4. Generation of Competitor

To generate competing SnoopLigase reaction product that could aid in the elution of other SnoopLigase reaction product, biotinylated SnoopLigase (75 µM), SUMO-DogTag (75 µM) and SnoopTag peptide (GKLGDIEFIKVNKGY, 150 µM, obtained from Tomohiko Nakamura in the Howarth laboratory, synthesised by Insight Biotechnology with 95% purity as determined by mass spectrometry) were incubated in 50 mM Tris borate pH 7.25 + 15% (v/v) glycerol in a total volume of 200 µL for 16 h at 4 °C. Subsequently, glycine purification was performed as described in section 2.10.2.

2.10.5. Purification by High Temperature and Competitor

The resin with immobilised protein was washed five times with five resin volumes of ice-cold phosphate buffered saline (PBS) (10 mM $Na_2HPO_4•H_2O$, 2 mM KH_2PO_4, 137 mM NaCl, 2.7 mM KCl, pH 7.4). One extra spin ensured thorough removal of remaining buffer from the resin. The resin was incubated for 5 min with one column volume of PBS with or without 35 µM of purified competitor (section 2.10.4.) at 25, 37, 45 or 55 °C on a ThermoMixer comfort (Eppendorf) at 500 rpm. Samples were centrifuged for 1 min at 300 g, 25 °C. The elution was repeated once more.

2.10.6. Purification by Imidazole

All steps were performed at 25 °C. The resin with immobilised protein was washed twice with five resin volumes of Tris phosphate pH 7.0 (25 mM phosphoric acid, pH adjusted with Tris base) containing 300 mM NaCl and 500 mM imidazole pH 7.0 and three times with five resin volumes of Tris phosphate pH 7.0 containing 500 mM imidazole pH 7.0. Samples containing multimeric IMX313-DogTag conjugates were washed with 1 M instead of 0.5 M imidazole wash buffers. One extra spin ensured thorough removal of remaining buffer from the resin. The resin was incubated for 5 min with one column volume of Tris phosphate pH 7.0 (25 mM phosphoric acid, pH adjusted with Tris base) containing 2 M imidazole pH 7.0 on a ThermoMixer comfort (Eppendorf) at 500 rpm. Multimeric IMX313-DogTag conjugates were eluted with 3.5 M instead of 2 M imidazole in the elution buffer. Samples were centrifuged for 1 min at 300 g, 25 °C. The elution was repeated twice more.

2.11. Analysis of Functionalised SnoopLigase Binding to Solid-Phase Support

2.11.1. Streptavidin Resin Binding Capacity for Biotin-SnoopLigase

To test the biotin-SnoopLigase binding capacity of HiCap streptavidin agarose, 10 µL of 10 µM biotinylated SnoopLigase in 50 mM Tris borate pH 8.0 was incubated with 1 µL washed and equilibrated HiCap streptavidin agarose (Thermo Fisher Scientific) or dilutions of the resin. After incubation for 1 h at 25 °C on a tube rotor, the samples were centrifuged for 1 min at 17,000 g to pellet the resin.

Supernatants of the samples were analysed by SDS-PAGE and densitometry (section 2.5.).

2.11.2. HaloLink Resin Binding of HaloTag7-SnoopLigase Containing Samples

180 µL reaction samples were prepared, containing 10 µM ligase protein in 50 mM Tris borate pH 7.25 + 15% (v/v) glycerol and / or 10 µM SUMO-DogTag + 10 µM SnoopTagJr-AffiHER2, 10 µM IMX313-DogTag + 20 µM SnoopTagJr-AffiHER2, 10 µM IMX313-DogTag + 20 µM SnoopTagJr-MBP or 100 mM 2-Mercaptoethanol. Samples were incubated for 12 h at 4 °C. Subsequently, 0.01% Tween-20 and 0.5 M imidazole were added and aliquots taken as 0 h time points. 18 µL of washed and equilibrated HaloLink resin (Promega) was added and samples were incubated on a tube rotor at 25 °C. After 1, 2, 4 and 8 h, 30 µL aliquots were collected. The aliquots were stored at -80 °C until all samples had been collected. The samples were thawed and centrifuged for 1 min at 17,000 g to pellet the resin. Supernatants of the samples were analysed by SDS-PAGE and densitometry (section 2.5.).

2.12. Computational Analyses

2.12.1. Ramachandran Analysis for Proline Substitution

To identify sites suitable for Pro residue substitutions in RrgALigase, the crystal structure of the parent domain of RrgA (Protein Data Bank 2WW8 residues 734 – 860 (Izoré et al., 2010)) was analysed using MolProbity (Davis et al., 2007). Amino acid residues and the respective Φ angles were listed in Excel (Microsoft). Entries were sorted by Φ angle and residues outside the suitable range for Pro

substitution (-75° to -45°) were removed. Furthermore, Pro residues and amino acids involved in isopeptide bond formation or immediately next to isopeptide bond-forming residues were excluded. The remaining entries were examined in the RrgA crystal structure using PyMOL (Schrödinger). Residues in the middle of β-strands, pointing into the hydrophobic protein core or involved in intramolecular hydrogen bonding were excluded. The remaining amino acid residues were chosen for Pro substitution.

2.12.2. Protein Repair One Stop Shop (PROSS)

As the PROSS sequence alignment yielded < 50 homologous sequences, a separate multiple sequence alignment (MSA) was generated. The C-terminal domain of RrgA (residues 734 – 860) was subjected to Position-Specific Iterated Basic Local Alignment Search Tool (PSI-BLAST) (Altschul et al., 1997; Bhagwat and Aravind, 2007) using a statistical significance threshold of 0.005. Five PSI-BLAST iterations were performed, giving 28,602 homologous sequences. Cluster Database at High Identity with Tolerance (CD-HIT) was used to eliminate sequences shorter than 90 residues (2,785 sequences remained) and to remove redundant sequences, using a clustering threshold of 0.63 (606 sequences remained) (Fu et al., 2012; Li and Godzik, 2006). Multiple Sequence Comparison by Log-Expectation (MUSCLE) (Edgar, 2004) was used to generate a multiple sequence alignment. The MSA was submitted along with the amino acid sequence and crystal structure to the PROSS (Goldenzweig et al., 2016). The PROSS-suggested amino acid mutations were examined in the RrgA crystal structure using PyMOL (Schrödinger). Excluded were mutations of the reactive triad residues

(K742, E803, N854), residues immediately next to the reactive triad, mutations potentially disrupting important hydrogen bonds, exposing hydrophobic amino acids on the protein surface or potentially disrupting secondary structure elements. The remaining amino acid mutations were chosen for testing.

3. RrgALigase Development and Optimisation

Covalent conjugation of proteins is an important tool in research and the development of novel therapeutics. For example, protein conjugation has allowed the development of fluorescently labelled affinity probes (Zola et al., 1992), antibody-protein drug conjugates (Panowksi et al., 2014), bispecific antibodies (Nisonoff and Rivers, 1961), half-life improvement of protein drugs in blood (Larsen et al., 2016), adjuvant-protein vaccine conjugates (Kreutz et al., 2012), and enhanced enzyme cascades (Pröschel et al., 2015).

Existing methods for the conjugation of proteins include a wide selection of chemical (Glazer, 1970; Kalia and Raines, 2010), and a narrower selection of enzymatic approaches (Walper et al., 2017). While chemical conjugation approaches often struggle with site-specificity and unwanted side reactions, limitations of existing enzymatic bioconjugation methods include low specificity (Griffin et al., 2002), limited solubility and dependence on flanking sequences (Aranko et al., 2014b), low affinity and large amount of substrate required (Guimaraes et al., 2013), laborious enzyme production and incomplete reaction (Yang et al., 2017).

The Howarth laboratory has developed Catcher-Tag systems for bioconjugation, which covalently link one peptide tag and one large protein part, the “catcher” (Veggiani et al., 2016; Zakeri et al., 2012). Catcher-Tag systems are fully genetically encoded and are functional over wide range of conditions, making them widely applicable tools *in vitro* and *in vivo* (Reddington and Howarth, 2015). However, Catcher-Tag systems comprise the necessity of fusing one of the conjugation partners to a protein domain, which can disturb its function and increase

immunogenicity. Therefore, it would be better to be able to exclusively use peptide tags for conjugation.

SpyLigase is a tripartite system based on the isopeptide bond-containing CnaB2 domain, developed by Jacob O. Fierer in the Howarth laboratory (Fierer et al., 2014). The enzyme covalently links the two peptide tags SpyTag and KTag, overcoming the problem of heterologous protein insertion. SpyLigase has proven to be a valuable tool for covalent protein conjugation (Fierer et al., 2014; Siegmund et al., 2016). However, the reaction requires high concentration of the chemical chaperone trimethylamine N-oxide (TMAO) (1.5 M) and low temperature. Furthermore, maximum substrate conversion achieved was around 60%. These properties limit the utility of the SpyLigase system to a narrow range of *in vitro* applications.

Hence, my aim was to develop a novel peptide-peptide ligase that had a more moderate buffer requirement and higher maximum substrate conversion, as well as being an orthogonal system to SpyLigase. Orthogonality can be useful to create protein architectures with defined subunit arrangements, such as performed previously with Catcher-Tag systems (Veggiani et al., 2016).

In this chapter, I describe the development of RrgALigase, a tripartite split protein system based on the isopeptide bond-containing domain of the protein RrgA. I show that the design concept, based on lessons from previous two-way split projects of the same domain, was feasible and the expected activity of the split protein system was detectable under specific reaction conditions. Optimisation of reaction conditions and the split points of the modules enhanced performance further.

3.1. Tripartite Splitting of RrgA Domain 4 as Basis for RrgALigase

RrgA is a protein from *Streptococcus pneumoniae* TIGR4 (GenBank genome accession number and version AE005672.3, locus tag SP_RS02280, protein accession number and version WP_000757844.1) (Tettelin et al., 2001) and is part of a multi-subunit pilus (LeMieux et al., 2006). The C-terminal domain of *Streptococcus pneumoniae* protein RrgA contains an intramolecular isopeptide bond between lysine 742 and asparagine 854 and the formation is catalysed by the apposed glutamate 803 (Figure 3.1a) (Izoré et al., 2010). In the transamidation reaction ammonia gets eliminated (Figure 3.1b).

The isopeptide-containing domain of RrgA had been engineered previously in two ways in the Howarth lab. First, Raphael Gayet, Gianluca Veggiani and Tomohiko Nakamura had split the domain into a peptide tag and a catcher protein (SnoopTag and SnoopCatcher), with SnoopTag (residues 734 – 745) containing the reactive lysine, and SnoopCatcher (residues 749 – 860) containing the reactive asparagine and the catalytic glutamate. Upon mixing, the two modules formed a covalent isopeptide bond (Veggiani et al., 2016). Several rational point mutations had been tested by Raphael Gayet, Gianluca Veggiani and Tomohiko Nakamura in the Howarth laboratory with the aim of enhancing SnoopCatcher and SnoopTag reactivity. The main focus was on replacement of hydrophobic residues exposed on the protein surface, and stabilising secondary structure elements according to relative amino acid propensities (Chou and Fasman, 1974). These screenings yielded two point mutations that improve SnoopCatcher reaction with SnoopTag: G842T and D848G (Veggiani et al., 2016).

In a second project based on RrgA (unpublished), Jin Huang in the Howarth laboratory had created RrgACatcher2 (residues 734 – 847) containing the reactive lysine and the catalytic glutamate, and SnoopTag2 (residues 848 – 856) containing the reactive asparagine. The two modules were reactive and formed an isopeptide bond upon mixing.

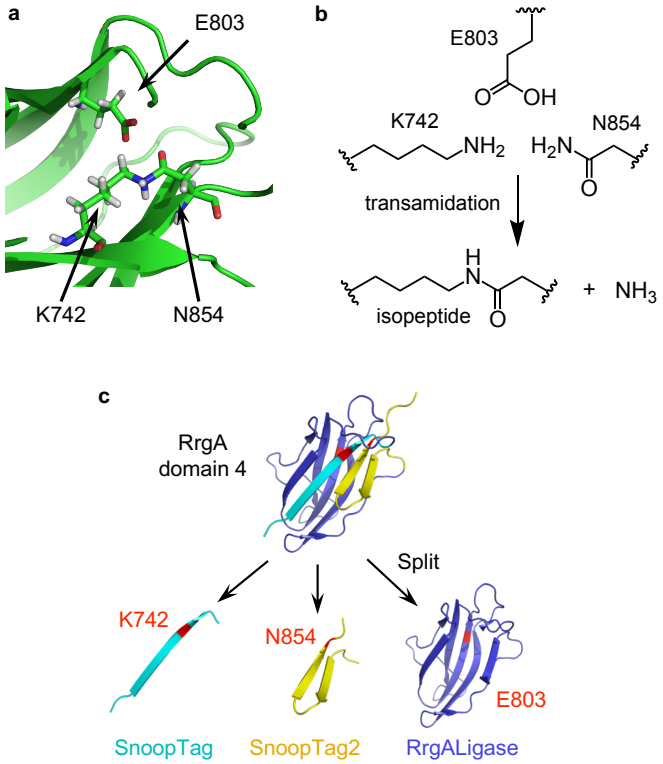


Figure 3.1 Tripartite splitting of RrgA domain 4. a) The C-terminal domain of RrgA (Protein Data Bank 2WW8) contains an isopeptide bond, formed between Lys 742 and Asn 854, catalysed by Glu 803. b) Molecular basis of isopeptide bond formation in RrgA. The transamidation reaction between Lys 742 and Asn 854 is catalysed by Glu 803, eliminating ammonia. c) Cartoon of RrgA splitting. The C-terminal domain of RrgA was split into three modules, with RrgALigase (blue) containing the catalytic Glu 803, SnoopTag (cyan) the reactive Lys 742, and SnoopTag2 (yellow) the reactive Asn 854. Key residues for isopeptide bond formation are highlighted in red.

The two projects described above laid the foundation for the splitting of RrgALigase: SnoopTag (residues 734 – 745) contained the reactive lysine, RrgALigase (residues 749 – 847) contained the catalytic glutamate, and SnoopTag2 (residues 848 – 856) contained the reactive asparagine (Figure 3.1c). In theory, upon mixing the three modules should reconstitute the protein domain, and an isopeptide bond should be formed between SnoopTag and SnoopTag2.

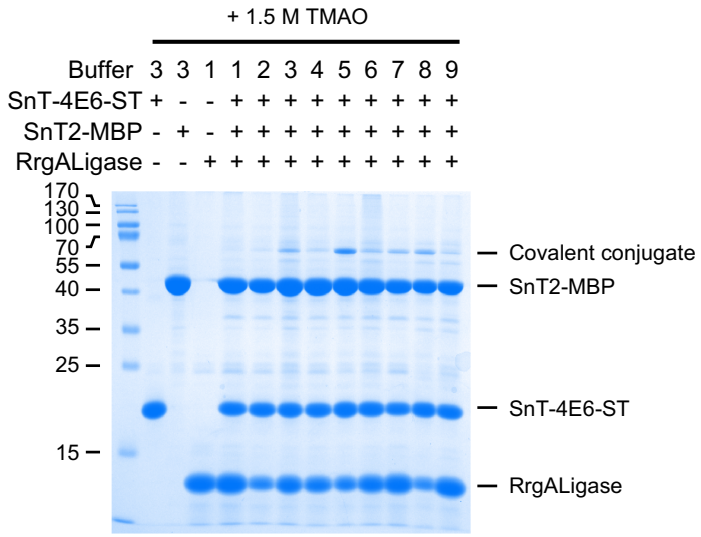
3.2. Crude Reaction Condition Screen for RrgALigase Reactivity

To test whether the designed modules were functional, I produced RrgALigase with an N-terminal His_6-tag and the model proteins SnoopTag-4E6-SpyTag (cloned by Gianluca Veggiani) and SnoopTag2-MBP (cloned by Jin Huang). 4E6 is a nanobody against human death receptor 5 (Huet et al., 2014) and MBP is maltose-binding protein (Duplay et al., 1984). Mixing of the two model proteins with RrgALigase should result in covalent isopeptide linkage between SnoopTag and SnoopTag2. The covalent linkage could be visualised by SDS-PAGE, by appearance of a new higher molecular weight band after reaction.

After mixing the proteins at 10 µM each and incubating them at 25 °C for 48 h in PBS, I was not able to detect reactivity of RrgALigase. Hence, I tested a range of different reaction conditions.

Reactivity of RrgALigase was first observed in buffers containing 1.5 M trimethylamine N-oxide (TMAO) (Figure 3.2). TMAO is an osmolyte shown to stabilise proteins and promote protein folding (Baskakov and Bolen, 1998; Qu et al., 1998). I had tested TMAO since it had shown a positive effect on the split protein systems of SpyLigase and SnoopCatcher (Fierer et al., 2014; Veggiani et al., 2016).

From the reaction condition screen, it also became apparent that NaCl inhibits the ligation reaction. The phosphate-based buffer worked best in the absence of added NaCl, intermediate with 137 mM NaCl, and worst with 300 mM NaCl (Figure 3.2).



Buffers

1	50 mM Tris + 300 mM NaCl + 200 mM imidazole pH 7.8
2	50 mM Tris + 300 mM NaCl pH 7.8
3	10 mM Na_2HPO_4 + 137 mM NaCl pH 7.4 (PBS)
4	10 mM Na_2HPO_4 + 300 mM NaCl pH 7.4
5	10 mM Na_2HPO_4 pH 7.4
6	PBS + 0.1 % Tween 20
7	PBS + 10% glycerol
8	50 mM Tris pH 8.0
9	40 mM Na_2HPO_4 + 20 mM citric acid pH 5.0

Figure 3.2 RrgALigase is reactive. SnoopTag-4E6-SpyTag and SnoopTag2-MBP formed a covalent conjugate upon mixing with RrgALigase in different reaction buffers. Reaction of 10 µM of each protein in indicated buffers for 16 h at 25 °C. All buffers contained 1.5 M TMAO. Samples were boiled, run on SDS-PAGE and visualised by Coomassie staining.

3.3. RrgALigase MBP-Fusion for Improved Solubility

Although His_6-tagged RrgALigase expressed well in *E. coli*, the solubility of the protein was low. Using a dialysis buffer screen, I found RrgALigase to be most soluble in 50 mM boric acid pH 10.5. However, even in boric acid buffer, RrgALigase gradually precipitated during storage at 4 °C and freeze-thawing also led to partial precipitation. Furthermore, RrgALigase reaction mixtures showed precipitation, especially in buffers other than boric acid pH 10.5.

I reasoned that improving RrgALigase solubility could increase its reactivity, as more protein would be available to perform reactions. Moreover, higher solubility could minimise fluctuations in RrgALigase concentrations when setting up reaction mixtures and allow more reliable testing of reaction conditions. It has been shown that solubility of an aggregation-prone protein can be improved by producing the protein as fusion construct with a highly soluble protein (Butt et al., 1989). Maltose-binding protein (MBP) is a commonly used solubilisation tag (Pryor and Leiting, 1997) and can sometimes promote correct folding of the protein it is fused to (Kapust and Waugh, 1999).

I could produce MBP-RrgALigase at higher yield and concentration (≥ 500 μM) than free RrgALigase (≤ 200 μM), and it was highly soluble in a wide range of buffers. No precipitation was observed upon freeze-thawing or prolonged storage at 4°C (tested up to 7 d). To compare RrgALigase and MBP-RrgALigase reactivity, I incubated 10 μM of the constructs with SnoopTag-AffiHER2 and SUMO-SnoopTag2 at 10 μM each for 16 h, followed by SDS-PAGE analysis. AffiHER2 is the Z(HER2:342) affibody protein targeting HER2 (Orlova et al., 2006) and SUMO is the small ubiquitin-like modifier protein Smt3p from *Saccharomyces cerevisiae* (Johnson et al., 1997). Reactivity of MBP-RrgALigase was similar to that of RrgALigase (Figure 3.3).

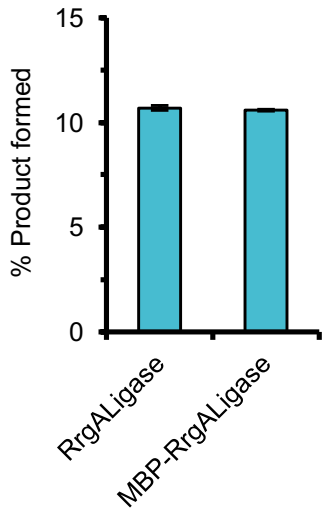


Figure 3.3 MBP fusion to RrgALigase does not alter RrgALigase activity. RrgALigase or MBP-RrgALigase was incubated with SUMO-SnoopTag2 and SnoopTag-AffiHER2 at 10 μM each in 50 mM boric acid pH 10.0 + 30% (v/v) glycerol + 1.5 M TMAO for 16 h at 4 °C. Samples were boiled, run on SDS-PAGE and visualised by Coomassie staining. Quantification by densitometry (mean ± 1 S.D., n=3).

3.4. Reaction Condition Optimisation

Initial reactivity of RrgALigase had been achieved in low salt buffer with TMAO (Chapter 3.2). To better characterise and optimise reaction conditions, I tested the effects of temperature, protein stabilising agents, reaction buffers and ion concentration on RrgALigase reaction in more detail.

3.4.1. Temperature

To test temperature dependence of RrgALigase, I incubated SnoopTag-AffiHER2, SUMO-SnoopTag2 and MBP-RrgALigase for 24 h at different temperatures and analysed formation of the covalent conjugate (Figure 3.4). As it had been reported that some chemical reactions are accelerated by freezing (Agten et al., 2016), I included -80 °C and -20 °C in the assay.

MBP-RrgALigase reacted best at low temperatures (4 – 8 °C), and reactivity was detectable up to 25 °C. A small amount of covalent conjugate was also formed at -20 °C (Figure 3.4).

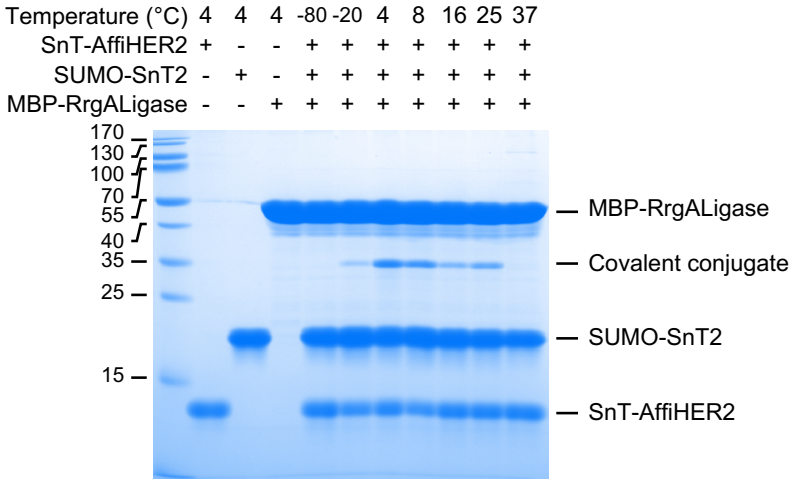


Figure 3.4 RrgALigase works best at low temperature. MBP-RrgALigase was incubated with SnoopTag-AffiHER2 and SUMO-SnoopTag2 at 10 μM each in 50 mM boric acid pH 10.5 + 1.5 M TMAO + 40% (v/v) glycerol for 16 h at indicated temperatures. Samples were boiled, run on SDS-PAGE and visualised by Coomassie staining.

3.4.2. Protein Stabilising Agents

RrgALigase was designed by removal of three β-strands from the parent domain (Chapter 3.1.), which represents a major structural modification. Considering that the whole domain contains only eight β-strands, it is likely that the modification could reduce the stability of the protein. Hence, I tested the effect of various protein stabilising agents on RrgALigase reaction.

There is a wide range of compounds known to stabilise proteins (Arakawa and Timasheff, 1985; Leibly et al., 2012). Effects of different agents vary from protein to protein and are difficult to predict (Barnett et al., 2016; Huang et al., 1995). Therefore, I tested a range of common protein stabilising agents, with the aim to enhance reactivity of RrgALigase (Table 1).

Table 1: Qualitative effect of protein stabilising agents on RrgALigase reactivity

Compound	Concentration	Effect	References
Glycerol	5 - 50% (v/v)	Positive	(Bradbury and Jakoby, 1972; Myers and Jakoby, 1973)
Ethylene glycol	10 - 40% (v/v)	Negative	(Huang et al., 1995; Zaroog et al., 2013)
Polyethylene glycol 400	10 - 40% (v/v)	Negative	(Ye et al., 1988; Yu et al., 2014)
Ficoll 400	10 - 40 mg/ml	None	(Alfano et al., 2017; Fritz et al., 2015)
Trimethylamine N-oxide (TMAO)	0.5 - 2.5 M	Positive	(Baskakov and Bolen, 1998; Mashino and Fridovich, 1987)
Betaine (trimethyl glycine)	1 - 2.5 M	Positive	(Arakawa and Timasheff, 1985; Papageorgiou and Murata, 1995)
NDSB-201	38 - 600 mM	Negative	(Heiss et al., 2003; Vuillard et al., 1995)
Sodium 4-phenylbutyrate	0.5 - 8 mM	Negative	(Burrows et al., 2000; Ray et al., 2011)
Trehalose	0.5 - 2 M	Positive	(Carninci et al., 1998; Sola-Penna and Meyer-Fernandes, 1996)
L-arginine	0.6 - 10 mg/ml	Negative	(Arakawa and Tsumoto, 2003; Brinkmann et al., 1992)
D_2O	25 - 50% (v/v)	None	(Kresheck et al., 1965; Makhatadze et al., 1995)

Out of the compounds tested, only glycerol, TMAO, betaine and trehalose enhanced RrgALigase reaction. Glycerol and TMAO had the strongest effect and therefore I focused further tests on these two additives. To determine the optimum concentration of the additives, I incubated MBP-RrgALigase with SUMO-SnoopTag2 and SnoopTag-4E6-SpyTag or SnoopTag-AffiHER2 with varying amounts of glycerol and TMAO (Figure 3.5).

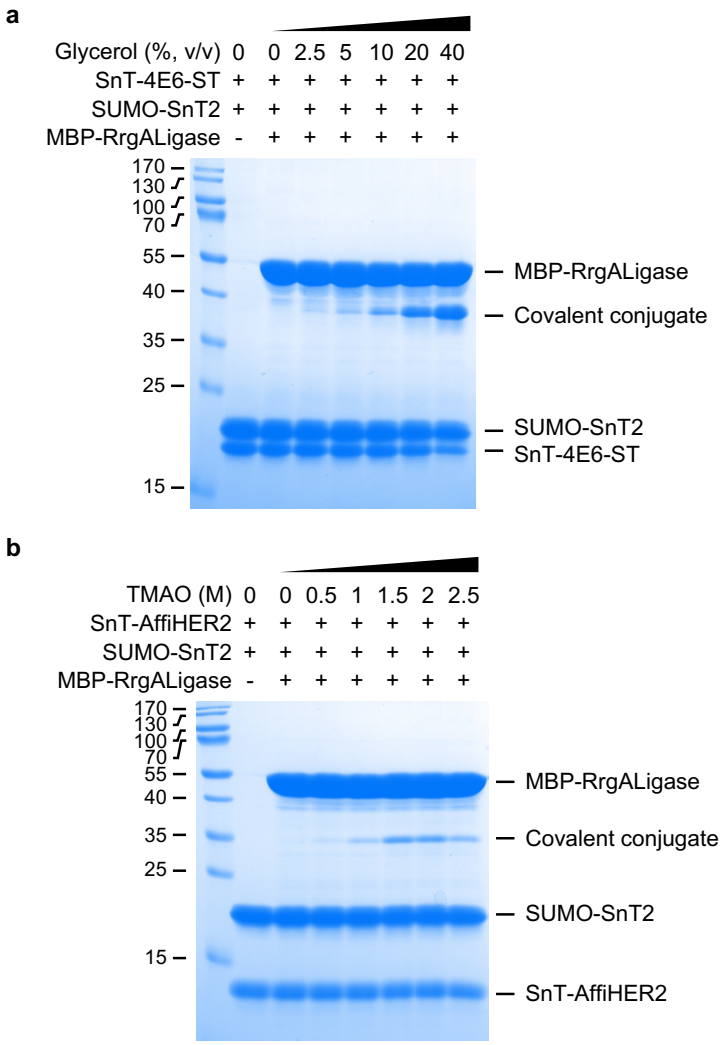


Figure 3.5 Glycerol and TMAO enhance MBP-RrgALigase reactivity. MBP-RrgALigase was incubated with SUMO-SnoopTag2 and SnoopTag-4E6-SpyTag or SnoopTag-AffiHER2 at 10 µM each in 50 mM boric acid pH 10.5 for 16 h at 4°C in presence of a) 1.5 M TMAO + indicated concentrations of glycerol or b) 20% (v/v) glycerol + indicated concentrations of TMAO. Samples were boiled, run on SDS-PAGE and visualised by Coomassie staining.

Addition of glycerol to the reaction enhanced MBP-RrgALigase reactivity. High concentrations of glycerol had the greatest effect (Figure 3.5a). Presence of TMAO enhanced MBP-RrgALigase reactivity, with an optimum concentration of ca. 1.5 – 2 M (Figure 3.5b).

3.4.3. Ion Compatibility

Presence of NaCl in the reaction buffer seemed to have a negative effect on RrgALigase reactivity (Figure 3.2). However, some proteins require salts for the protein to remain soluble (Fágáin, 1995). For RrgALigase to be a widely applicable tool it would be advantageous for the protein to be active in the presence of salt. Therefore, I tested whether the inhibitory effect observed previously was specific to NaCl or also caused by other salts. I incubated MBP-RrgALigase with SUMO-SnoopTag2 and SnoopTag-AffiHER2 at 10 μM each for 16 h with Na_2SO_4, NaCl, KCl, $CaCl_2$ or $MgCl_2$ (Figure 3.6). Salt concentrations were adjusted to the total charge of 100 mM NaCl.

MBP-RrgALigase reaction worked best at low ionic strength. Addition of any of the salts tested led to a reduction in product formation (Figure 3.6). Therefore, the addition of salts to RrgALigase reactions was avoided where possible in subsequent assays.

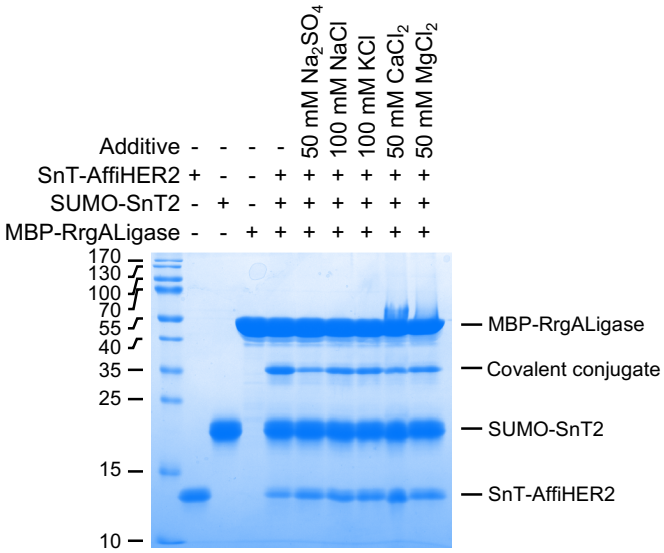


Figure 3.6 Effect of different salts on MBP-RrgALigase reactivity. MBP-RrgALigase was incubated with SUMO-SnoopTag2 and SnoopTag-AffiHER2 at 10 μM each in 50 mM boric acid pH 10.5 + 1.5 M TMAO + 50% (v/v) glycerol for 16 h at 4 °C in presence of indicated salt additives. Samples were boiled, run on SDS-PAGE and visualised by Coomassie staining.

3.5. Domain Splitting Optimisation

The splitting of the C-terminal RrgA domain into three parts was guided by the previous successful splitting endeavours, as described in Chapter 3.1. However, possibly other ways of splitting the domain could allow better reactivity of the system. As the length and split points of SnoopTag had been extensively studied previously by Gianluca Veggiani and Tomohiko Nakamura, I focused my splitting optimisation on RrgALigase and SnoopTag2.

The primary aim of the splitting optimisation was to accelerate reaction speed of the split protein system. Two secondary aims guided the splitting optimisation endeavour: minimising the length of SnoopTag2, and improving the solubility of RrgALigase. Short peptide tags are generally desirable, as they are likely to be less disruptive when fused to proteins of interest, and are cheaper to be synthesised as

nucleotide or amino acid sequence. The initial design of RrgALigase yielded a protein lacking three β-strands compared to its parent domain. This protein showed limited solubility. I reasoned that elongation of RrgALigase could improve solubility, by increasing resemblance of the split protein to the parent domain and reducing exposure of the hydrophobic protein core.

3.5.1. SnoopTag2 Termini Optimisation and DogTag Development

The initial SnoopTag2, as developed by Jin Huang, consisted of 19 amino acid residues, spanning the amino acid residues 838 – 856 of the RrgA gene. The sequence space explored for the SnoopTag2 length alteration study is illustrated in the context of SnoopTag and RrgALigase in Figure 3.7a and as alignment with the amino acid sequence in Figure 3.7b. The individual sequences I designed, N- and/or C-terminally shortened or elongated versions of SnoopTag2, are shown in Figure 3.7c. To test reactivity of the designed SnoopTag2 variants, I produced them as C-terminal fusions to SUMO and tested them as substrates for RrgALigase-mediated conjugation to SnoopTag-AffiHER2.

Shortening SnoopTag2 on the N-terminus by three, nine, ten or eleven residues abolished reactivity, while a four-residue elongation showed an approximately three-fold reduction in product formed compared to the wildtype (Figure 3.7d). C-terminal removal of one or two residues yielded non-reactive tags. Increasing the length of SnoopTag2 on the C-terminus by the four residues IPPK (SnT2+4) improved reactivity, with about 80% more product formed compared to the wildtype (Figure 3.7e).

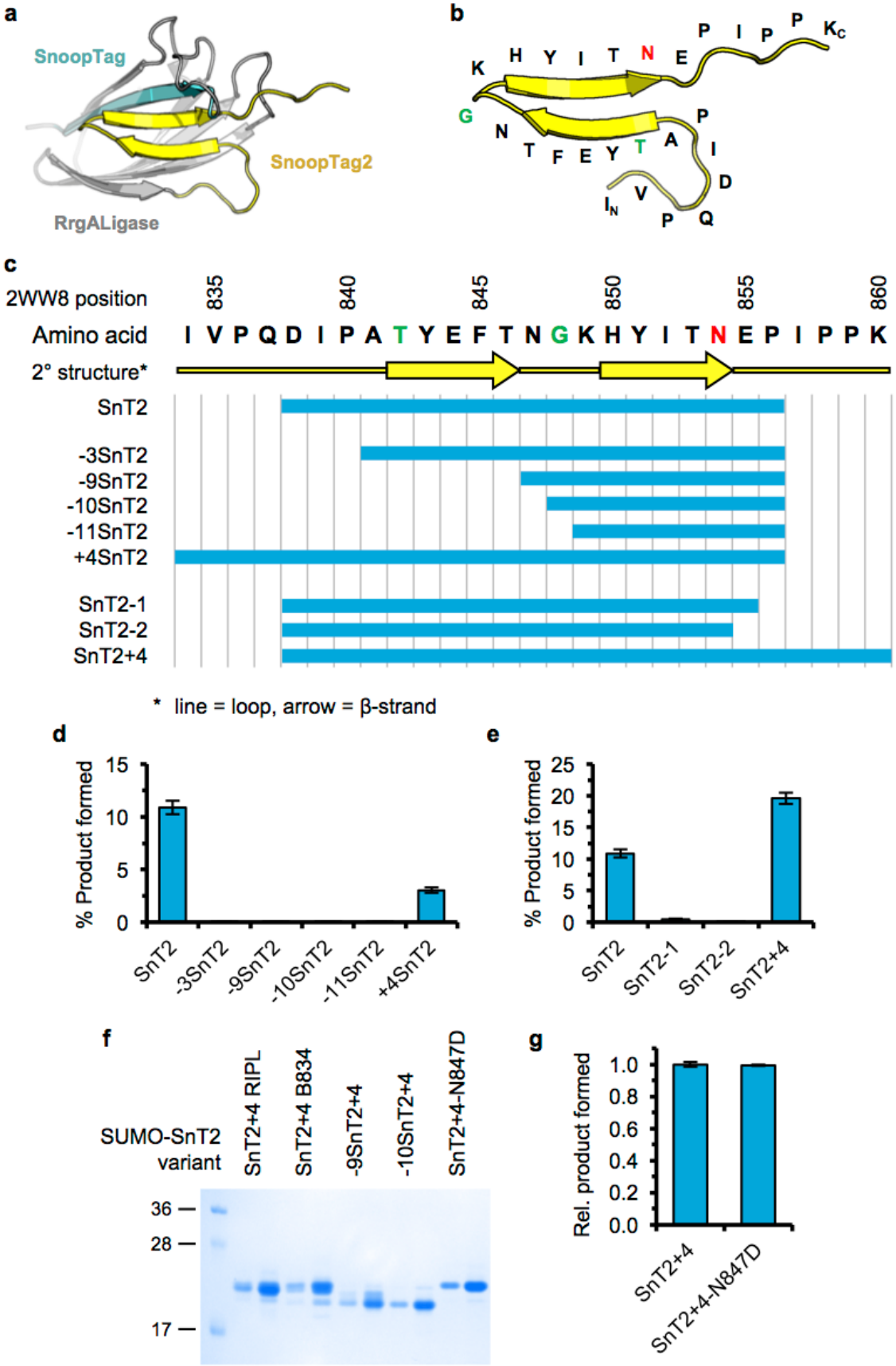

a
SnoopTag
SnoopTag2
RrgALigase
b
K H Y I T N E P I P P Kc
G N T F E Y T A P I D V P Q
IN
c
2WW8 position
835
840
845
850
855
860
Amino acid
I V P Q D I P A T Y E F T N G K H Y I T N E P I P P K
2° structure*
SnT2
-3SnT2
-9SnT2
-10SnT2
-11SnT2
+4SnT2
SnT2-1
SnT2-2
SnT2+4
* line = loop, arrow = β-strand
d
% Product formed
0
5
10
15
SnT2
-3SnT2
-9SnT2
-10SnT2
-11SnT2
+4SnT2
e
% Product formed
0
5
10
15
20
25
SnT2
SnT2-1
SnT2-2
SnT2+4
f
SUMO-SnT2 variant
SnT2+4 RIPL
SnT2+4 B834
-9SnT2+4
-10SnT2+4
SnT2+4-N847D
36
28
17
g
Rel. product formed
0.0
0.2
0.4
0.6
0.8
1.0
SnT2+4
SnT2+4-N847D

Figure 3.7 SnoopTag2 Termini optimisation. a) Cartoon of SnoopTag2 (yellow) sequence space explored in the context of RrgALigase (grey) and SnoopTag (cyan). b) Cartoon of SnoopTag2 (yellow) sequence space explored with annotated amino acid residues. Green residues represent G842T and D848G point mutations made by Gianluca Veggiani and Tomohiko Nakamura during SnoopCatcher development. Reactive residue N854 highlighted in red. c) Illustration of SnoopTag2 elongation and truncation variants tested. Amino acid sequence aligned to secondary structure as determined by PyMOL. d) and e) MBP-RrgALigase was incubated with SnTg-AffiHER2 and d) N-terminal or e) C-terminal variants of SUMO-SnT2 at 10 μM each in 50 mM boric acid pH 10.0 + 1.5 M TMAO + 30% (v/v) glycerol for 16 h at 4 °C. f) SDS-PAGE analysis of double band appearance of SUMO-SnT2+4 produced in BL21(DE3) RIPL or B834(DE3), or variants of SUMO-SnT2+4. Each protein was loaded at 2.5 μM (left lane) and 10 μM (right lane). g) N847D mutation of SnT2+4 does not alter reactivity. MBP-RrgALigase was incubated with SnoopTag-AffiHER2 and SUMO-SnT2+4 or SUMO-SnT2+4-N847D at 10 μM each in 50 mM boric acid pH 10.5 + 1.5 M TMAO + 50% (v/v) glycerol for 16 h at 4 °C. Samples of d) – g) were boiled, run on SDS-PAGE and visualised by Coomassie staining. Quantification of d), e) and g) by densitometry (mean ± 1 S.D., n=3).

SUMO-SnoopTag2+4 appeared as a double band on SDS-PAGE gels. Both bands seemed to be reactive, suggesting that a fraction of the protein might undergo a modification during expression or purification that does not affect the ability of the protein to act as substrate for RrgALigase. While most variants of SUMO-SnoopTag2+4 also appeared as double band, I noticed a difference between SUMO-9SnoopTag2+4 (double band) and SUMO-10SnoopTag2+4 (single band) (Figure 3.7f). The only difference between these two constructs was the absence of N847 in SUMO-10SnoopTag2+4. Mutation of this asparagine residue to aspartate in SUMO-SnoopTag2+4 caused the protein to appear as single band on SDS-PAGE gels (Figure 3.7f). While the double-band appearance is not necessarily a fundamental problem when studying RrgALigase-mediated conjugations, SDS-

PAGE analysis and densitometry of protein reactions are easier with single-band proteins. Reactivity of SUMO-SnnoopTag2+4-N847D was comparable to that of SUMO-SnoopTag2+4 (Figure 3.7g). Therefore, I decided to proceed with SnoopTag2+4-N847D in my downstream work. SnoopTag2+4-N847D was named “DogTag”.

3.5.2. RrgALigase Termini Optimisation

The original RrgALigase lacked three β-strands compared to the parent domain (Figure 3.1). This modification potentially causes improper folding of the protein by absence of stabilising structure elements and solvent exposure of parts of the hydrophobic core. Indeed, RrgALigase showed limited solubility (Chapter 3.3.). I reasoned that elongating the termini of RrgALigase might stabilise the protein and increase solubility. Additionally, a higher fraction of properly folded protein could translate into enhanced reactivity. Hence my RrgALigase termini optimisation efforts were focused on elongation rather than truncation. The sequence space explored for the RrgALigase termini elongation study is illustrated in Figure 3.8a. The elongations caused partial sequence overlaps with SnoopTag and DogTag (Figure 3.8b). The individual RrgALigase versions I designed are shown as alignment to SnoopTag and DogTag in Figure 3.8c.

To test reactivity of the designed RrgALigase variants, I produced them as MBP-fusion constructs and tested their ligation activity with SnoopTag-AffiHER2 and SUMO-SnoopTag2+4 as substrates.

Elongating RrgALigase on the C-terminus by nine, ten or eleven residues increased reactivity (Figure 3.8d). Nine and ten residue elongation had the best

effect with a ca. 18 – 20% increase in covalent product formed after 16 h reaction (Figure 3.8d). N-terminal elongation variants were designed based on RrgALigase+9. Four and six residue elongations had a positive effect on ligase reactivity, whereas a nine or ten residue elongation (with mutation of the reactive lysine 742 to alanine) had a negative effect (Figure 3.8e). Four and six residue elongation showed ca. 30 – 35% increase in covalent product formed after 16 h reaction (Figure 3.8e). I decided to proceed with +6RrgALigase+9 in my downstream work.

The amino acid sequence of elongated RrgALigase (+6RrgALigase+9) resembled the parent domain more than the initial split. Hence, I reasoned that the protein might be able to fold better and be less prone to aggregation. When expressed without MBP fusion, +6RrgALigase+9 was highly soluble and could be produced at high yield in *E. coli*. Reactivity of MBP+6RrgALigase+9 and +6RrgALigase+9 was similar when using substrates of rather low molecular weight, such as SnoopTag-AffiHER2 (12.7 kDa) and SUMO-SnoopTag2+4 (15.1 kDa) (Figure 3.8f). For ligations of bulkier or multimeric proteins, +6RrgALigase+9 performed better without MBP fusion, possibly by avoiding steric clashes of the bulky MBP protein (40.2 kDa) with the substrate proteins.

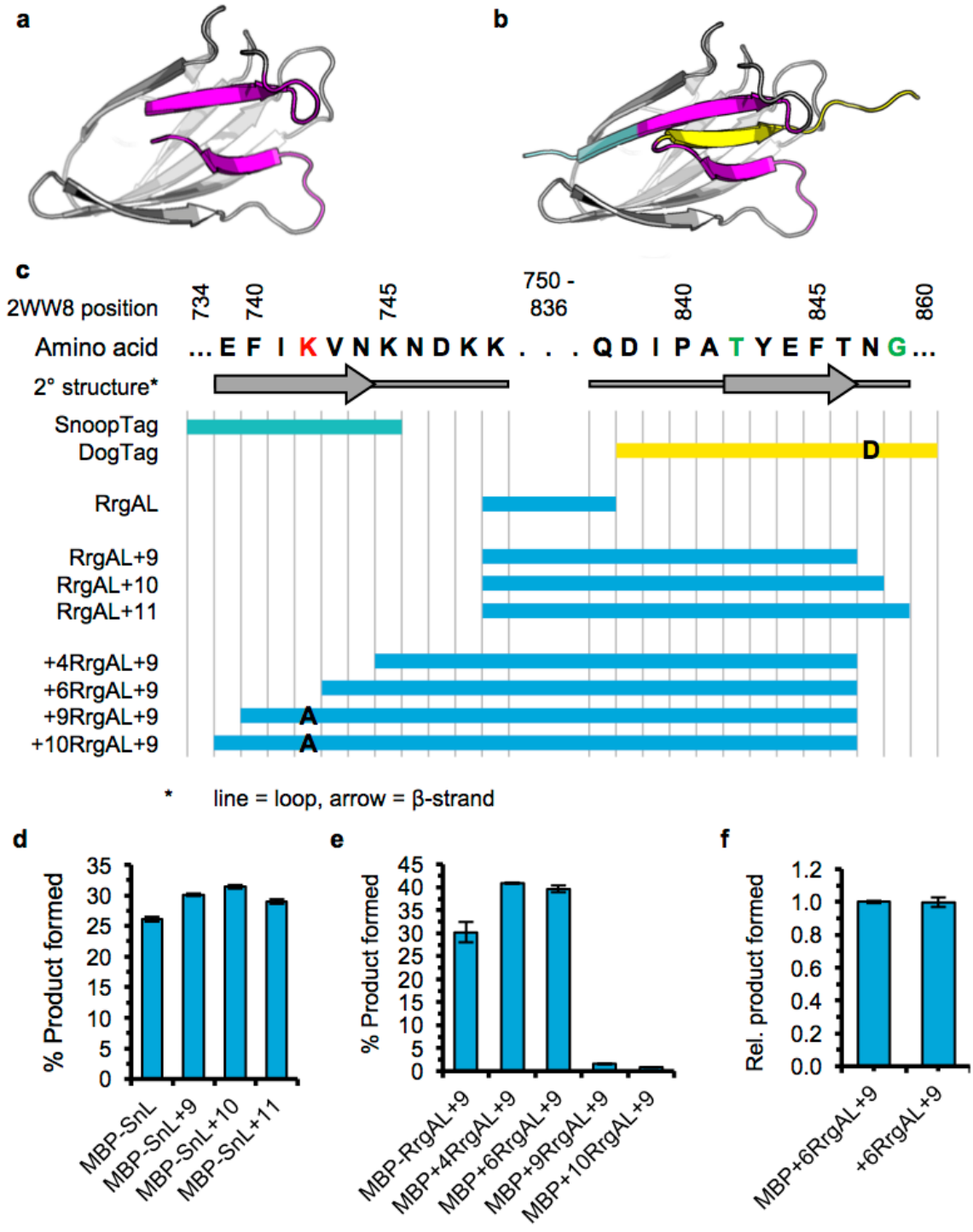


Figure 3.8 RrgALigase elongation improves RrgALigase reactivity. a) Cartoon of RrgALigase sequence space explored. RrgALigase in grey, sequence space explored by elongation study in magenta. b) Cartoon of RrgALigase (grey), SnoopTag (cyan) and DogTag (yellow). Sequence overlaps between ligase and tags explored by the ligase elongation study shown in magenta. c) Alignment of RrgALigase elongation variants with SnoopTag and DogTag. Green residues represent G842T and D848G point mutations

made by Gianluca Veggiani and Tomohiko Nakamura in the Howarth laboratory during SnoopCatcher development. Reactive residue K742 highlighted in red. Amino acid changes in individual constructs annotated. Amino acid sequence aligned to secondary structure as determined by PyMOL. d) SnT-AffiHER2 and SUMO-SnT2+4 was incubated with C-terminal elongation variants of MBP-RrgALigase at 10 μM each in 50 mM boric acid pH 10.5 + 1.5 M TMAO + 40 % (v/v) glycerol for 16 h at 4°C. e) SnT-AffiHER2 and SUMO-SnT2+4 was incubated with N-terminal elongation variants of MBP-RrgALigase at 10 μM each in 50 mM boric acid + 1.5 M TMAO pH 9 + 50 % (v/v) glycerol for 16 h at 4°C. f) MBP+6RrgALigase+9 and +6RrgALigase+9 are equally reactive. Proteins were incubated with SnoopTag-AffiHER2 and SUMO-DogTag at 10 μM each in 50 mM boric acid pH 10 + 1.5 M TMAO + 30% (v/v) glycerol for 24 h at 4°C. Samples of d) – f) were boiled, run on SDS-PAGE and visualised by Coomassie staining. Quantification by densitometry (mean ± 1 S.D., n=3).

3.5.3. Elongated RrgALigase Reacts with N-terminally Shortened DogTag

One of the aims of the RrgALigase and SnoopTag2 elongation study was to reduce the size of SnoopTag2. Previously, reaction of N-terminally shortened SnoopTag2 (e.g. -10SnoopTag2) with RrgALigase had failed (Figure 3.7d). The same applied to the N-terminally shortened version of DogTag (-10DogTag). The lack of reactivity was possibly caused by the absence of an entire β-strand in the reconstituted domain (residues 838 – 847) when using RrgALigase with SnoopTagJr and -10DogTag (Figure 3.9a). However, after elongating RrgALigase by one β-strand on the C-terminus (RrgALigase+9), the missing β-strand was introduced into the ligase, and therefore, N-terminally shortened DogTag might show reactivity.

To test whether +6RrgALigase+9 can react with N-terminally shortened DogTag, I incubated MBP+6RrgALigase+9 with SnoopTag-AffiHER2 and SUMO-

DogTag or SUMO-10DogTag at 10 μM each for 24 h. -10DogTag showed reactivity, although lower than that of full DogTag (Figure 3.9b).

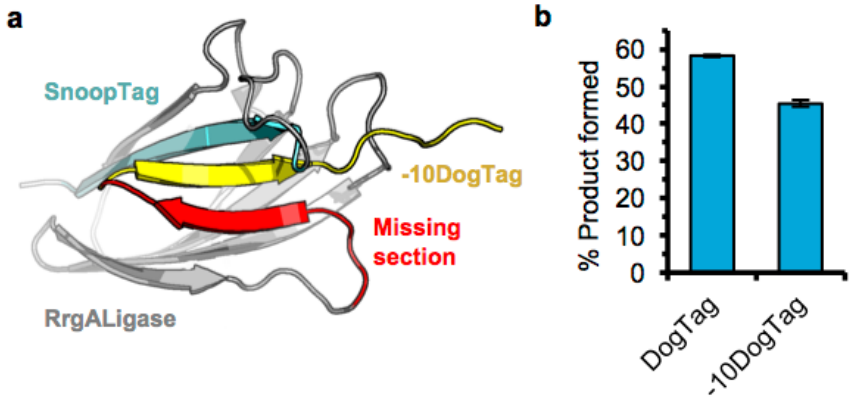


Figure 3.9 Elongated RrgALigase reacts with N-terminally shortened DogTag. a) Cartoon of RrgALigase (grey), SnoopTag (cyan) and -10DogTag (yellow), missing a 10-amino acid section (red) compared to the parent domain. b) MBP+6RrgALigase+9 was incubated with SnoopTag-AffiHER2 and SUMO-DogTag or SUMO-10DogTag at 10 μM each in 50 mM boric acid pH 10 + 1.5 M TMAO + 30% (v/v) glycerol for 24 h at 4°C. Samples were boiled, run on SDS-PAGE and visualised by Coomassie staining. Quantification by densitometry (mean ± 1 S.D., n=3).

3.6. Discussion

Genetically encoded protein conjugation methods are a valuable part of the molecular biology toolkit. They have wide functionality and are used in diverse applications including cancer cell capture, metabolic engineering, vaccine enhancement and protein stabilisation. However, limitations of existing approaches can be low specificity (Griffin et al., 2002), low affinity (Guimaraes et al., 2013), or restrictive reaction conditions (Fierer et al., 2014), among others. Therefore, there was need to design a new tool for protein conjugation.

The C-terminal, isopeptide bond-containing domain of RrgA was split into three modules: the two peptides SnoopTag and SnoopTag2, and the protein

RrgALigase. Upon mixing of the three modules, a covalent isopeptide bond should be formed between SnoopTag and SnoopTag2.

Initially, no activity of the split system was detectable. Adjustment of the reaction conditions (1.5 M TMAO, 40 – 50% glycerol, low temperature) allowed detection and improvement of reactivity. TMAO and glycerol are protein stabilising agents, and low temperature limits thermal motion of the protein. Together with the necessity of fusing RrgALigase to the protein chaperone MBP to improve protein solubility, the reaction conditions indicate limited structural stability of RrgALigase.

RrgALigase showed tolerance of N-terminal MBP fusion, which suggests that RrgALigase could possibly be fused to other proteins as well, such as fluorescent reporter proteins, alternative affinity tags, or even a third protein of interest when attempting protein heterotrimerisation.

Elongation of SnoopTag2 and RrgALigase purely based on the RrgA gene sequence enhanced reaction rate and yielded a more soluble ligase protein. Elongation of SnoopTag2 by four residues (SnT2+4) improved reactivity of the tag. The IPPK sequence had been excluded in the original design of SnoopTag2, as the crystal structure of RrgA did not suggest extensive interaction of the IPPK region with the rest of the protein domain (Izoré et al., 2010). The strong effect of the elongation on SnoopTag2 reactivity, however, indicates that folding of the C-terminal region of the domain could be different than suggested by the crystal structure. It would be interesting to further investigate whether all four of the added residues (IPPK) are required to enhance SnoopTag2 reactivity, or whether a shorter addition is sufficient.

The original RrgALigase lacked three β-strands compared to the parent domain, which was likely a major source of improper folding and structural

instability. Elongating the termini of RrgALigase (+6RrgALigase+9) enhanced reactivity and solubility of the ligase, possibly due to better resemblance of the parent domain, and therefore decreased exposure of hydrophobic core residues and more stabilising amino acid residue interactions.

After optimising the split points of RrgALigase and SnoopTag2, the split protein system showed terminal overlap between the modules: three amino acid residues between SnoopTag and +6RrgALigase+9 and nine amino acid residues between +6RrgALigase+9 and DogTag. To reconstitute the parent domain, only one set of sequences can occupy the defined space. Possibly, the sequence overlaps created by RrgALigase elongation improve folding and stability of RrgALigase in the absence of substrate, but could move away upon substrate binding.

The split-improved system allowed covalent conjugation of two proteins of interest by fusing SnoopTag (12 amino acids) to one of them and DogTag (23 amino acids) or -10DogTag (13 amino acids) to the other one, followed by incubation of the tagged proteins with +6RrgALigase+9. Using DogTag allowed maximum reactivity, while -10DogTag might be useful in some occasions to reduce structural perturbation of the protein of interest.

However, the established split protein system had several limitations. First, the ligation reaction rate was slow, with only ca. 60% substrate conversion after 24 h. Second, the ligation reaction conditions (high pH, 1.5 M TMAO, 40 – 50% glycerol, low temperature, low salt) limited the utility of the RrgALigase system *in vivo* and *in vitro*. It is unlikely that cells could survive treatment with the reaction buffer (Burrows et al., 2000; Ceccarini and Eagle, 1971; Waymouth, 1970; Wiebe and Dinsdale, 1991). Furthermore, some proteins are not soluble under the

RrgALigase reaction conditions (Zayas, 1997). Third, a large portion of the final reaction mixture volume is occupied by the buffer components (especially TMAO and glycerol), such that substrate proteins would need to be produced at high concentration (≥ 100 μM), to achieve a final concentration of 10 μM. Not all proteins are soluble to such high degree (Trevino et al., 2008).

The work presented in this chapter enabled the development of a new functional peptide-peptide ligation system. It represented a good starting point but left room for further enhancement.

4. Rational and Bioinformatic Protein Engineering of Peptide Ligase

Methods for enhancing protein properties include rational design, computational studies, directed protein evolution and combinations thereof (Lutz, 2010). Rational design involves manual examination and modification of protein structure based on the knowledge of a scientist (Pokkuluri et al., 2002; Thakkar et al., 2014). This approach can yield strong and quick improvements, but relies on the availability of accurate structural data and suffers from low throughput (Nixon and Firestine, 2000). Computational approaches can take vast amounts of relevant sequence and structure data into account, but the accuracy of predictions is often limited by imperfect algorithms, insufficient computing power of modern machines and the overwhelming complexity of molecular interactions (Chica, 2015; Markov, 2014; Rosenfeld et al., 2016). Directed evolution does not require extensive functional, structural and mechanistic information and allows the screening of millions of protein variants simultaneously (Cobb et al., 2013). However, directed evolution assays can be difficult to set up, require extensive optimisation and are prone to yield mutants that have dodged the selection pressure in an unexpected way (Leemhuis et al., 2009; Wong et al., 2006).

In the previous chapter I described the development of +6RrgALigase+9, an enzyme covalently linking the two peptide tags SnoopTag and DogTag. Although the system was functional, low reactivity and stringent reaction conditions limited its utility as a universal bioconjugation tool. To improve the properties of the split protein system, I decided to engineer the split protein domain. A range of rational point mutations had been tested previously by Raphael Gayet, Gianluca Veggiani and Tomohiko Nakamura in the Howarth laboratory during SnoopCatcher/

SnoopTag development. The focus was on stabilisation of secondary structure elements according to relative amino acid propensities (Chou and Fasman, 1974). These screenings yielded two reaction rate-enhancing point mutations, G842T and D848G (Veggiani et al., 2016), which had been already included in the RrgALigase system.

In this chapter I describe the successful use of two rational and bioinformatic protein engineering approaches to improve RrgALigase reactivity. The enhanced version was termed “SnoopLigase” which, apart from strongly improved reaction rate, showed more desirable optimal reaction conditions.

4.1. Loop Stabilisation Through Proline Substitutions

Naturally occurring proteins are composed of 22 canonical amino acids (Ambrogelly et al., 2007). Using their α-amino and α-carboxyl groups, individual amino acid residues are joint together into a polypeptide chain. Rotation of individual amino acid residues along their Φ and Ψ dihedral angles (Figure 4.1) contributes to polypeptide chain backbone folding in three-dimensional space, giving rise to secondary structure elements and protein tertiary structure (Ramachandran et al., 1963).

Proline is the only amino acid containing a secondary amine as α-amino group (tertiary amine in polypeptide chains). The direct attachment of the α-amino group to the amino acid side chain locks the rotational freedom of proline’s Φ angle to approximately -65° (Figure 4.1a) (Morris et al., 1992). The conformational restriction makes proline a loop-stabilising residue in naturally occurring proteins, and has been harnessed to engineer more stable protein variants, by insertion of

prolines at suitable positions (Fu et al., 2009; Trevino et al., 2007). Therefore, I decided to test whether proline substitutions could positively impact RrgALigase.

I screened the crystal structure of RrgA domain 4 for possible proline substitution locations, by analysing the Φ angles of the protein's amino acid residues. The screen was based on the known preferred proline Φ angle of ca -65°. I limited the search to residues with Φ angles of -45° to -75°.

I then manually examined hits to exclude residues in the middle of β-strands, close to the catalytic or reactive residues, or involved in intramolecular hydrogen bonding. Four +6RrgALigase+9 proline substitution mutants were tested for their ligation reactivity on SnoopTag-AffiHER2 and SUMO-DogTag: A771P, E785P, A808P and Q837P. Under previously optimised standard reaction conditions (section 3.4), A771P and E785P mutants were less reactive than +6RrgALigase+9, while A808P and Q837P mutations had no effect. Only at elevated reaction temperature (25 °C instead of 4 °C) reaction-enhancing effects of A808P and Q837P mutations became apparent. As an aim of RrgALigase engineering was to create a protein conjugation system with broad applicability *in vitro* and *in vivo*, improved reactivity at elevated temperature was considered beneficial. A771P and E785P mutants performed worse relative to +6RrgALigase+9 at elevated temperature (Figure 4.1b).

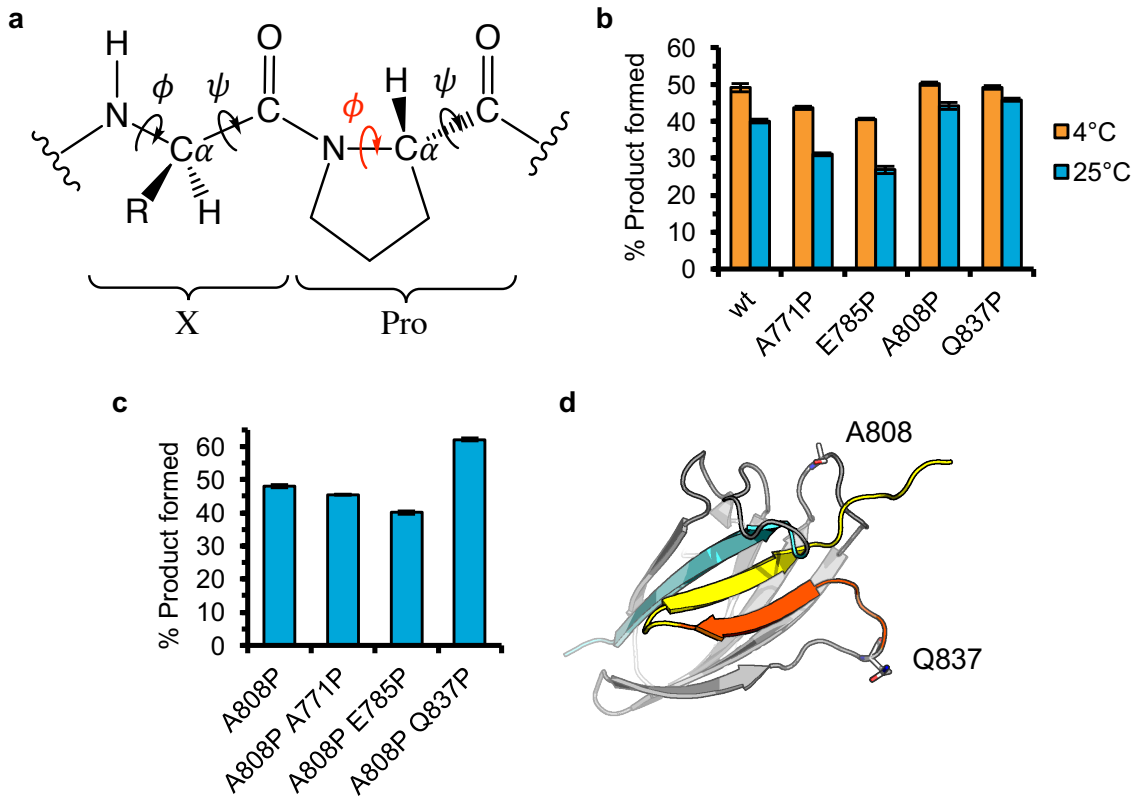


Figure 4.1 Proline substitutions improved RrgALigase reactivity. a) Proline is conformationally restricted in polypeptides. Structure of the dipeptide X–Pro inside a polypeptide chain. X is an amino acid other than Pro with side-chain R. Dihedral angles *Φ* and *Ψ* indicated for both residues. Conformations that can be adopted by the polypeptide backbone are restricted at Pro sites by limited *Φ* angle (red) possibilities. b) Reactivity of +6RrgALigase+9 single proline mutants. SnT-AffiHER2 and SUMO-SnT2+4 was incubated with mutants of +6RrgALigase+9 at 10 µM each in 50 mM boric acid pH 10.5 + 1.5 M TMAO + 45% (v/v) glycerol for 4 h at indicated temperature. c) Reactivity of +6RrgALigase+9 double proline mutants. SnT-AffiHER2 and SUMO-SnT2+4 were incubated with mutants of +6RrgALigase+9 at 10 µM each in 50 mM boric acid pH 9 + 1 M TMAO + 30% (v/v) glycerol for 4 h at 4 °C. Samples of b) and c) were boiled, run on SDS-PAGE and visualised by Coomassie staining. Quantification by densitometry (mean ± 1 S.D., n=3). d) Cartoon of RrgALigase (grey), SnoopTag (cyan) and DogTag (yellow). The 9-amino acid overlap between +6RrgALigase+9 and DogTag shown in orange. Successful proline mutations at locations A808 and Q837 indicated as stick residues.

Building on +6RrgALigase+9 A808P, I tested the effect of double proline mutations under standard reaction conditions. Again, A771P and E785P had negative effects on substrate conversion. Q837P increased product formation

(Figure 4.1c). A808P and Q837P lie in loops involved in DogTag interaction (Figure 4.1d). As A771P and E785P had shown decreased RrgALigase reactivity in both single and double mutants, no proline triple or quadruple mutants were tested. Further RrgALigase engineering was performed on +6RrgALigase+9 A808P Q837P.

4.2. Protein Repair One-Stop Shop and SnoopLigase Development

The Protein Repair One-Stop Shop (PROSS) is an automated algorithm aimed at identifying stabilising point mutations for proteins of interest. It is based on homology analysis and atomistic Rosetta modelling (Goldenzweig et al., 2016). Apart from the five proteins engineered in the original publication, the PROSS algorithm has been used to enhance solubility of RH5, a malaria protein that has gained interest as promising vaccination target (Campeotto et al., 2017). Enhancements of split protein systems have not been reported yet. However, we suspected that enhanced stability of the reconstituted domain might improve reactivity of the modules.

When submitting the RrgA domain 4 to the PROSS, the algorithm was only able to find a low number of homologous sequences, limiting the reliability of the output. The low number of hits was due to the parameters of the algorithm's sequence alignment method. Therefore, I generated a separate sequence alignment for the PROSS to use in the Rosetta algorithm. I then manually examined the hits to exclude residues in the middle of β-strands, close to the catalytic or reactive residues, or involved in intramolecular hydrogen bonding. Four PROSS mutations were tested: D737S in SnoopTag, and D838G, A820E, D830N and I839V in RrgALigase.

To test the effect of D737S on SnoopTag reactivity, I produced SnoopTag-D737S-AffiHER2 and tested it as substrate for +6RrgALigase+9 A808P-mediated conjugation to SUMO-DogTag. Substrate conversion with SnoopTag D737S was approximately 60 % higher compared to wildtype SnoopTag (Figure 4.2a). The enhanced SnoopTag D737S was named “SnoopTagJr”.

The four point mutations suggested by the PROSS for RrgALigase were designed as single mutants of +6RrgALigase+9 A808P, since at that time the second proline substitution Q837P had not been confirmed yet. All four mutants, especially D838G and I839V, performed better at ligating SnoopTagJr-AffiHER2 and SUMO-DogTag than the regular +6RrgALigase+9 (Figure 4.2c). To test whether combining mutations would enhance reactivity further, A820E, D830N and I839V mutants of +6RrgALigase+9 A808P Q837P D838G were tested for ligation efficiency. Only the addition of the I839V mutation enhanced product formation (Figure 4.2c). +6RrgALigase+9 A808P Q837P D838G I839V was considered the final enhanced version of RrgALigase, and was named “SnoopLigase”.

To confirm that this engineered conjugation system still uses the reactive triad described in Figure 3.1b for isopeptide bond formation, I mutated the key residues of each component: SnoopTagJr-K742A-AffiHER2, SnoopLigase E803Q and SUMO-DogTag-N854A. Using any of these mutants for conjugation reactions prevented formation of the covalent conjugate, indicating that the molecular mechanism of covalent conjugation is not altered by the point mutations made in this chapter.

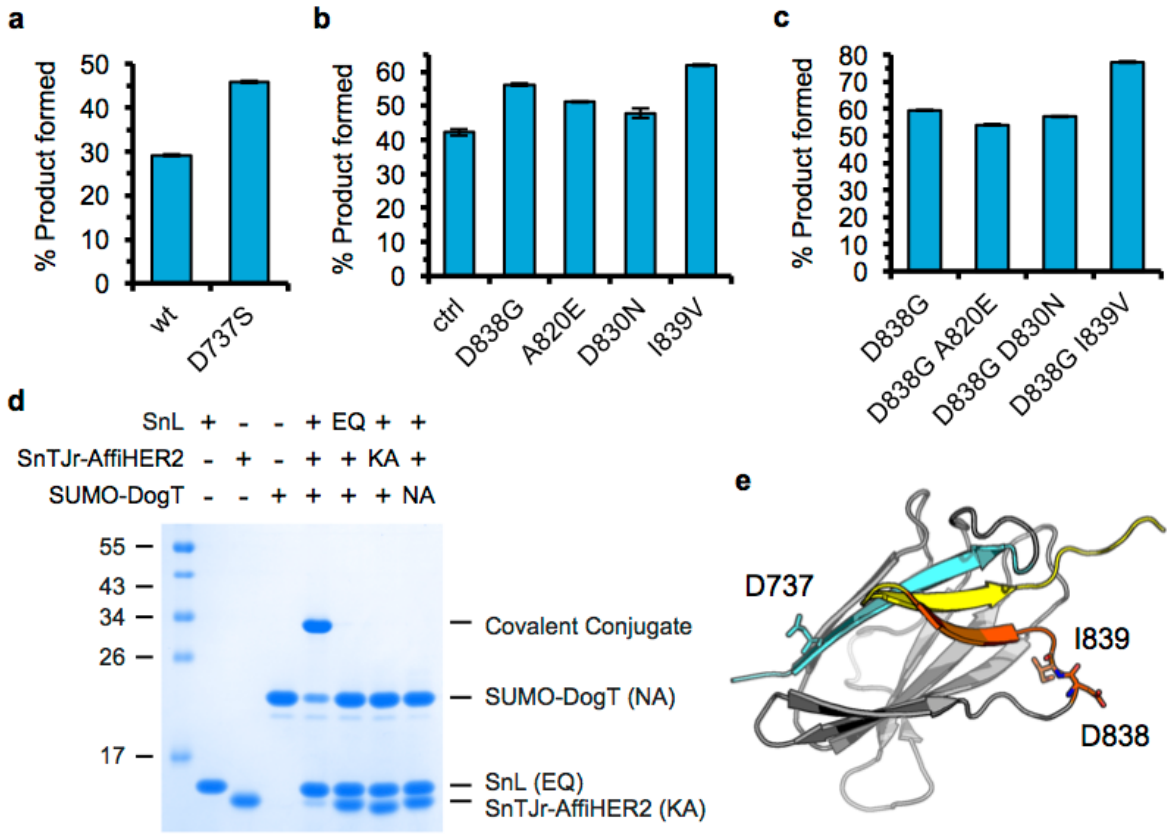


Figure 4.2 PROSS mutations improved RrgALigase reactivity. a) Reactivity of wildtype (wt) SnoopTag and SnoopTag D737S. Reaction with +6RrgALigase+9 A808P and SUMO-DogTag at 10 µM each in 50 mM boric acid pH 9.0 + 1 M TMAO + 30% (v/v) glycerol for 3 h at 4 °C. b) Reactivity of +6RrgALigase+9 A808P (ctrl) and single PROSS mutants thereof. Reaction with SnoopTagJr-AffiHER2 and SUMO-DogTag at 10 µM each in 50 mM Tris borate + 0.5 M TMAO pH 8.0 + 30% (v/v) glycerol for 4 h at 4 °C. c) Reactivity of +6RrgALigase+9 A808P Q837P D838G (D838G) and double PROSS mutants thereof. Reaction with SnoopTagJr-AffiHER2 and SUMO-DogTag at 10 µM each in 50 mM Tris borate + 0.75 M TMAO pH 8.0 + 7.5% (v/v) glycerol for 1.5 h at 4 °C. d) The K E N catalytic triad mediates covalent bond formation of the SnoopLigase system. Reaction of SnoopLigase, SnoopTagJr-AffiHER2 and SUMO-DogTag at 10 µM each in 50 mM Tris borate pH 7.25 + 15% (v/v) glycerol for 24 h at 4 °C enabled covalent product formation, while E803Q (EQ), K742A (KA) or N854A (NA) mutations of the components did not. Samples of a), b), c) and d) were boiled, run on SDS-PAGE and visualised by Coomassie staining. Quantification of a), b) and c) by densitometry (mean ± 1 S.D., n=3). e) Cartoon of RrgALigase (grey), SnoopTag (cyan) and DogTag (yellow). The 9-amino acid overlap between +6RrgALigase+9 and DogTag shown in orange. Successful PROSS mutations at locations D737, D838 and I839 indicated as stick residues.

The locations of the PROSS mutations are illustrated in Figure 4.2e. The two mutations D838G and I839V are positioned directly next to each other, and mark the start of the sequence overlap between SnoopLigase and DogTag.

4.3. Optimisation of SnoopLigase Reaction Conditions

The aim of my thesis project was to develop a novel peptide-peptide ligase system with good kinetic properties and moderate (close to physiological) reaction condition requirements, to make it a widely applicable tool for research. The initial RrgALigase version, however, was reacting slowly and needed reaction conditions far from physiological (section 3.4.). Although not specifically selecting for it, during proline substitution and PROSS-based RrgALigase engineering it became apparent that improved ligase variants had altered reaction condition requirements. Therefore, I characterised pH and temperature optima of SnoopLigase, requirements for the protein stabilising agents glycerol and TMAO, and sensitivity to salts.

4.3.1. pH Dependence

Previously, RrgALigase reactions were performed in boric acid pH 10.5, mainly a remnant from the initial RrgALigase solubility screen (section 3.2.). To determine the optimal pH for SnoopLigase reaction, I tested SnoopLigase-mediated SnoopTagJr-AffiHER2 and SUMO-DogTag substrate conversion in two different buffers (Tris-phosphate-citrate and Tris-borate) covering pH values of 3 – 8.75. In Tris-phosphate-citrate, at low pH up to 4.5 no ligation activity was detected. Substrate conversion increased between pH 5.0 and 6.0, reaching maximum at pH

6.5 and 7.0 (Figure 4.3a). In Tris-borate, product formation was highest at pH 7.25 – 8.25, decreasing at higher values (Figure 4.3b).

Tris-borate pH 7.25 was chosen as standard buffer, as it is a) close to physiological pH, b) within the pH optimum range of the reaction, and c) does not require NaOH or HCl for pH adjustment, which would be potentially inhibitory to the ligase reaction, as seen in section 3.4.3.

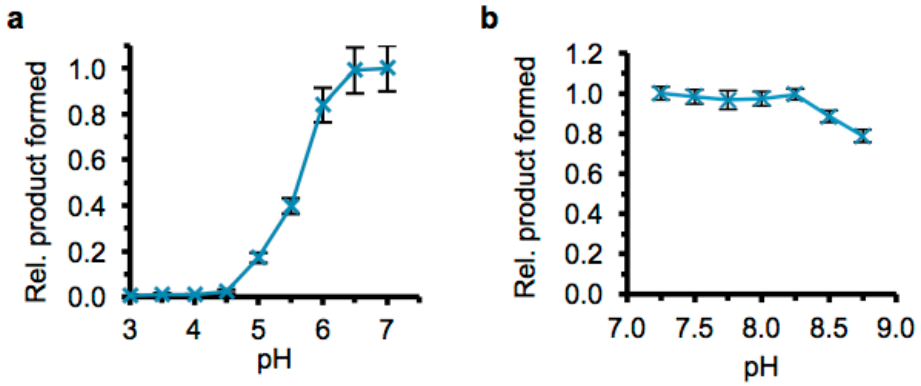


Figure 4.3 SnoopLigase pH dependence. SnoopLigase reaction with SnoopTagJr-AffiHER2 and SUMO-DogTag at 10 μM each at 4°C a) for 2 h in 25 mM Tris phosphate citrate pH 3.0 – 7.0 or b) for 1.5 h in 50 mM Tris borate pH 7.25 – 8.75. Samples were boiled, run on SDS-PAGE and visualised by Coomassie staining. Quantification by densitometry (mean ± 1 S.D., n=3).

4.3.2. Temperature Dependence

RrgALigase reaction worked best at low temperature (section 3.4.1.), which limited the applicability of the system *in vivo.* During SnoopLigase development it appeared that enhanced ligase versions performed better at higher temperature (Figure 4.1b). To test the temperature profile of SnoopLigase reaction, I analysed SnoopLigase-mediated SnoopTagJr-AffiHER2 and SUMO-DogTag substrate conversion at 4 – 37 °C. The reaction performed best at 4 – 20 °C, gradually declining to approximately 70 % relative reaction efficiency at 37 °C (Figure 4.4).

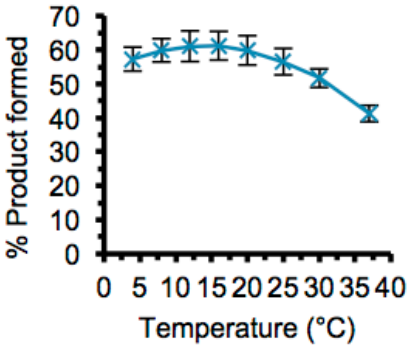


Figure 4.4 SnoopLigase temperature dependence. SnoopLigase reaction with SnoopTagJr-AffiHER2 and SUMO-DogTag at 10 μM each in 50 mM Tris borate pH 7.25 + 15% (v/v) glycerol for 2 h at indicated temperatures. Samples were boiled, run on SDS-PAGE and visualised by Coomassie staining. Quantification by densitometry (mean ± 1 S.D., n=3).

4.3.3. Protein Stabilising Agents

RrgALigase reaction required 40 – 50 % glycerol and 1.5 M TMAO (section 3.4.2.). limiting the system's applicability. Modification of the protein system by optimising split points (section 3.5.), introduction of prolines (section 4.1.) and usage of the PROSS algorithm (section 4.2.) yielded a more reactive ligation system, and possibly also altered the stability of the ligase module. To test whether these modifications caused changes in protein stabilising agent requirements, I tested SnoopLigase-mediated SnoopTagJr-AffiHER2 and SUMO-DogTag substrate conversion in the presence of different glycerol and TMAO concentrations.

Addition of 15 – 30 % (v/v) glycerol enhanced SnoopLigase reaction by approximately 70 % relative to the no glycerol condition. Lower or higher concentrations had a weaker effect (Figure 4.5a). TMAO led to decreased SnoopLigase-mediated substrate conversion, with the highest concentration of 1.5 M tested causing an approximately 25 % decrease (Figure 4.5b). Hence, 15 % (v/v) glycerol and no TMAO was chosen as standard SnoopLigase reaction condition.

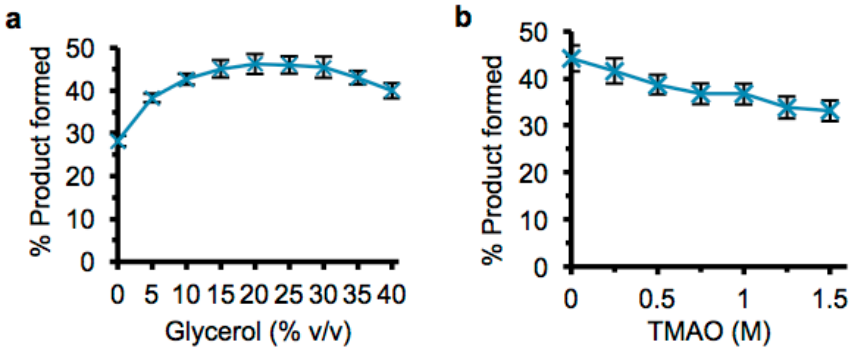


Figure 4.5 SnoopLigase glycerol and TMAO dependence. SnoopLigase reaction with SnoopTagJr-AffiHER2 and SUMO-DogTag at 10 μM each in 50 mM Tris borate pH 7.25 for 1.5 h at 4 °C in presence of a) indicated concentrations of glycerol and b) 15% (v/v) glycerol + indicated concentrations of TMAO pH 7.5. Samples were boiled, run on SDS-PAGE and visualised by Coomassie staining. Quantification by densitometry (mean ± 1 S.D., n=3).

4.3.4. Ion Compatibility

RrgALigase reactivity was impaired by the presence of NaCl (section 3.2.). However, some proteins require salts to remain soluble (Fágáin, 1995), and cells require salt to survive (Waymouth, 1970). To be a broadly applicable tool, it would be advantageous for SnoopLigase to be active in the presence of salt. Therefore, I tested the effect of different salts on SnoopLigase reaction. I incubated reactions samples of SnoopLigase, SUMO-DogTag and SnoopTagJr-AffiHER2 with a range of sodium- or chloride-based salts. Cation concentration was adjusted to match Na^+ and anion concentration to match Cl^-.

Generally, SnoopLigase reaction worked best at low ionic strength. All salts tested impaired SnoopLigase reactivity. The anion study showed a weaker inhibitory effect of sodium fluoride and sodium acetate compared to sodium chloride, while sodium iodide and sodium sulfate were more inhibitory (Figure 4.6a). In the cation study the effects clustered according to valency – divalent cations had a stronger inhibitory effect than monovalent ones (Figure 4.6b). Therefore, the addition of salts

to SnoopLigase reactions was avoided where possible in subsequent assays. Where salt is required, it would be recommended to use fluoride- or acetate-based salts and specifically avoid iodide- or sulfate-based salts and high concentrations of divalent cations. As fluoride is toxic to cells, it should not be used for *in vivo* applications (Yan et al., 2017).

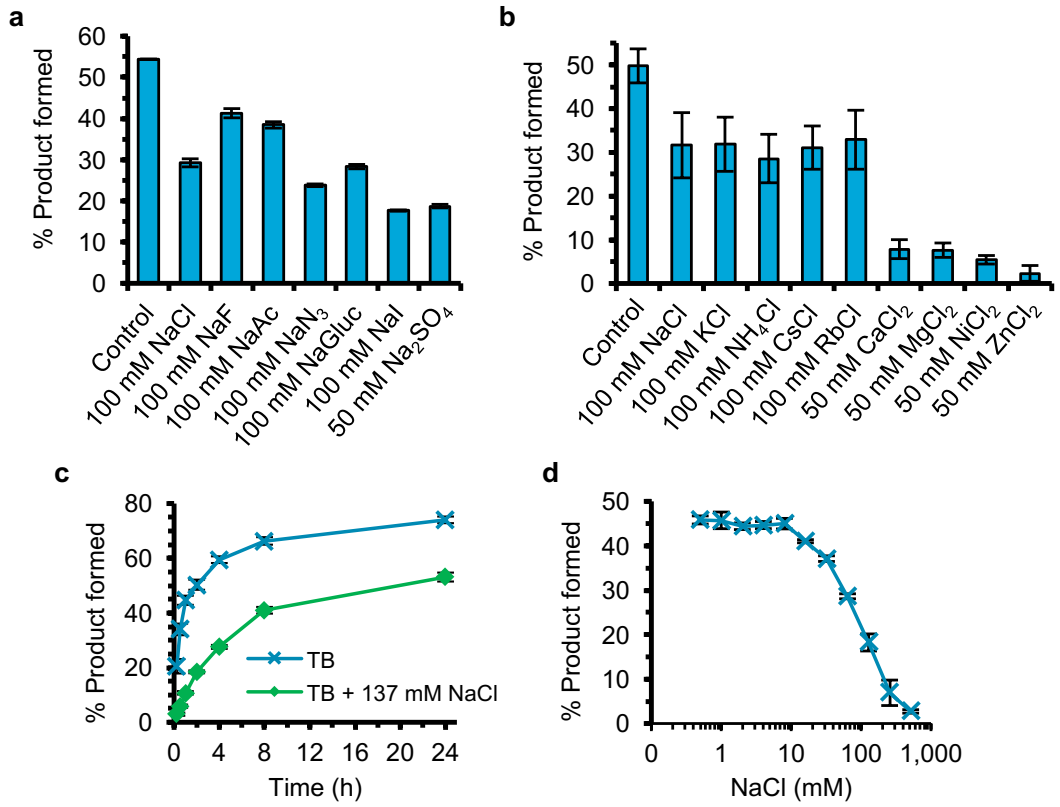


Figure 4.6 Effect of salts on SnoopLigase reactivity. SnoopLigase reaction with SnoopTagJr-AffiHER2 and SUMO-DogTag at 10 µM each in 50 mM Tris borate pH 7.25 + 15% (v/v) glycerol for 1.5 h at 4 °C in presence of indicated concentrations of a) sodium-based and b) chloride-based salts. c) Time course of SnoopLigase reaction with SnoopTagJr-AffiHER2 and SUMO-DogTag at 10 µM each in 50 mM Tris borate (TB) pH 7.25 + 15% (v/v) glycerol or 50 mM TB pH 7.25 + 15% (v/v) glycerol + 137 mM NaCl at 4 °C. d) SnoopLigase reaction with SnoopTagJr-AffiHER2 and SUMO-DogTag at 10 µM each in 50 mM Tris borate pH 7.25 + 15% (v/v) glycerol for 2 h at 4 °C in presence of different NaCl concentrations. Samples were boiled, run on SDS-PAGE and visualised by Coomassie staining. Quantification by densitometry (mean ± 1 S.D., n=3).

Since sodium chloride is present in many protein and cell culture buffer formulations, I further tested the effect of NaCl on SnoopLigase reaction. PBS is a commonly buffer and contains 137 mM NaCl (Dulbecco and Vogt, 1954). Addition of 137 mM NaCl to SnoopLigase reaction shows impaired reaction rate in a time course (Figure 4.6c). Titration of NaCl on 2 h SnoopLigase reactions showed a half maximal inhibitory concentration of approximately 100 mM (Figure 4.6d). In summary, SnoopLigase reactions are possible yet substantially slower under physiological sodium chloride concentration.

4.4. Comparison of RrgALigase Mutant Reaction Rates

Engineering of RrgALigase has not only yielded a faster enzyme, but also the optimum reaction conditions of the engineered SnoopLigase are closer to physiological conditions than those of the initial RrgALigase.

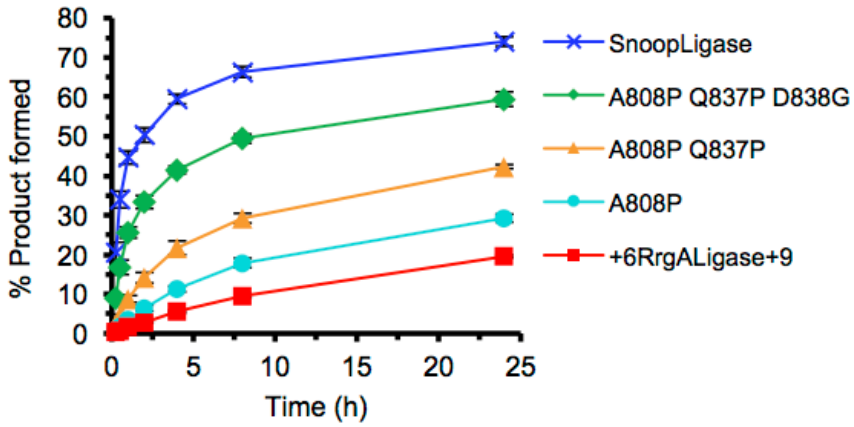


Figure 4.7 RrgALigase variants reaction time course. +6RrgALigase+9 and variants thereof were used for reaction with SnoopTagJr-AffiHER2 and SUMO-DogTag at 10 μM each in 50 mM Tris borate pH 7.25 + 15% (v/v) glycerol for indicated times at 4 °C. Samples were boiled, run on SDS-PAGE and visualised by Coomassie staining. Quantification by densitometry (mean ± 1 S.D., n=3).

To assess whether each of the point mutations were helpful in improving SnoopLigase performance under the optimised reaction conditions, I determined reaction rates of individual variants. I used five different RrgALigase variants generated during SnoopLigase development for ligation time-course reactions of SnoopTagJr-AffiHER2 and SUMO-DogTag at 10 μM each in the newly optimised reaction conditions. As expected, each of the mutations improved RrgALigase reactivity (Figure 4.7).

From the assay, the initial reaction rates were calculated. Starting from 3.5×10^{-11} M s^{-1} by +6RrgALigase+9, SnoopLigase achieved the highest initial reaction rate of 2.3×10^{-9} M s^{-1}, representing an approximately 65-fold increase (Table 2).

Table 2: Reaction rates of RrgALigase variants. +6RrgALigase+9 and variants thereof were used for reaction with SnoopTagJr-AffiHER2 and SUMO-DogTag at 10 μM each in 50 mM Tris borate pH 7.25 + 15% (v/v) glycerol for 15 min at 4 °C. Samples were boiled, run on SDS-PAGE and visualised by Coomassie staining. Quantification by densitometry (mean ± 1 S.D., n=3) and calculation of R. "Fold of previous" represents the improvement of R by individual point mutations compared to the parent construct.

RrgALigase variant	Initial reaction rate R (M s^{-1})	Fold of previous	Fold of +6RrgALigase+9
+6RrgALigase+9	3.5E-11	N/A	1.0
+6RrgALigase+9 A808P	1.1E-10	3.3	3.3
+6RrgALigase+9 A808P Q837P	2.4E-10	2.1	6.8
+6RrgALigase+9 A808P Q837P D838G	9.9E-10	4.2	28.5
SnoopLigase	2.3E-09	2.3	65.3

4.5. Characterisation of SnoopLigase Reaction and Storage Characteristics

To assess SnoopLigase's capacity as versatile conjugation tool with broad applicability, I tested the reaction's compatibility with commonly used detergents

and the possibility of reaction quenching by SDS and urea. Furthermore, I tested protein thermoresilience and effects of lyophilisation.

4.5.1. Detergent Compatibility

Nonionic surfactants are common additives to improve protein solubility (Allison et al., 1995; Hjertén and Johansson, 1972; Leibly et al., 2012), prevent aggregation (Bam et al., 1998; Machida et al., 2000), and as blocking agents prevent non-specific binding of proteins to surfaces (Batteiger et al., 1982; Lauer and Nolan, 2002). To be a broadly applicable tool, it would be advantageous for SnoopLigase to be active in the presence of nonionic detergents. This would be especially useful for SnoopLigase utility on solubilised membrane proteins (Lin and Guidotti, 2009), for example for NMR-based structural studies by segmental isotopic labelling (Freiburger et al., 2015; Kainosho et al., 2006; Tatulian et al., 2005) or defined arrangement of membrane proteins in nanodiscs (Raschle et al., 2015).

Sodium dodecyl sulfate (SDS) is an anionic surfactant used to denature proteins (Yonath et al., 1977). For some applications it could be useful to specifically terminate SnoopLigase reaction, for example in hydrogel preparation (Kim et al., 2016). Therefore, I analysed SnoopLigase reaction-quenching capacity of SDS.

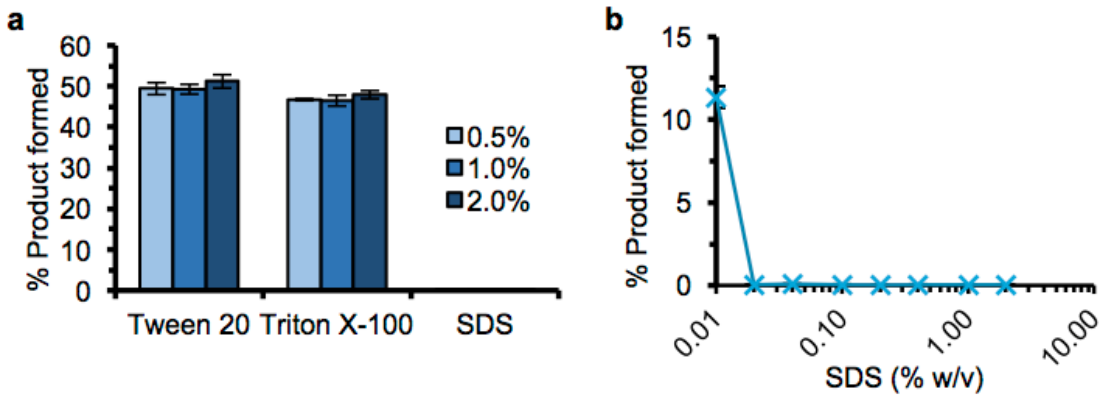


Figure 4.8 SnoopLigase detergent-tolerance. a) SnoopLigase reaction with SnoopTagJr-AffiHER2 and SUMO-DogTag at 10 µM each in 50 mM Tris borate pH 7.25 + 15% (v/v) glycerol for 2 h at 4 °C in presence of indicated concentrations of Tween 20, Triton X-100 or SDS. b) SDS titration on SnoopLigase reaction with SnoopTagJr-AffiHER2 and SUMO-DogTag at 10 µM each in 50 mM Tris borate pH 7.25 + 15% (v/v) glycerol for 3 h at 4 °C. Samples were boiled, run on SDS-PAGE and visualised by Coomassie staining. Quantification by densitometry (mean ± 1 S.D., n=3).

I tested the effects of SDS and the widely used non-ionic detergents Tween 20 and Triton X-100 on SnoopLigase-mediated conjugation of SnoopTagJr-AffiHER2 and SUMO-DogTag. Tween 20 and Triton X-100 at concentrations up to 2 % (v/v) did not alter SnoopLigase reactivity, while SDS completely inhibited SnoopLigase reactivity already at the lowest concentration (0.5 %) tested (Figure 4.7a). A test with lower SDS concentrations revealed total SnoopLigase reaction inhibition at concentrations of ≥ 0.02 % (w/v). At 0.01 % (w/v) SDS, SnoopLigase reactivity was strongly impaired, with approximately 23 % relative ligation efficiency (Figure 4.7b).

4.5.2. Effect of Urea on SnoopLigase Reaction

Urea is a commonly used protein denaturing agent (Pace, 1986; Zou et al., 1998). I reasoned that it could be a useful tool for quenching SnoopLigase reactions. Furthermore, urea sensitivity can be used to assess protein stability (Grant et al., 1992; Xie et al., 2014)

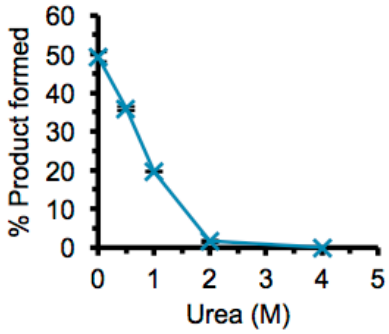


Figure 4.9 SnoopLigase urea tolerance. Urea titration on SnoopLigase reaction with SnoopTagJr-AffiHER2 and SUMO-DogTag at 10 μM each in 50 mM Tris borate pH 7.25 + 15% (v/v) glycerol for 3 h at 4 °C. Samples were boiled, run on SDS-PAGE and visualised by Coomassie staining. Quantification by densitometry (mean ± 1 S.D., n=3).

I tested the effects of 0.5 – 4 M urea on SnoopLigase-mediated conjugation of SnoopTagJr-AffiHER2 and SUMO-DogTag. All urea concentrations tested impaired SnoopLigase reactivity, with a half-maximal inhibitory concentration of approximately 0.8 M. At 4 M urea, no product formation could be detected after 3 h (Figure 4.8).

4.5.3. SnoopLigase Thermal Resistance

Thermal resistance can be an important characteristic in the development of biochemical assays. For example, while polymerase chain reaction (PCR) requires thermoresistant DNA polymerase that can survive the high nucleic acid melting temperature (Saiki et al., 1988), other methods rely on enzyme heat inactivation

(Dahl et al., 2004; Moran et al., 2001). To assess thermoresilience of SnoopLigase, the protein was incubated for 15 min at elevated temperatures, cooled down and used for ligation reactions of SnoopTagJr-AffiHER2 and SUMO-DogTag.

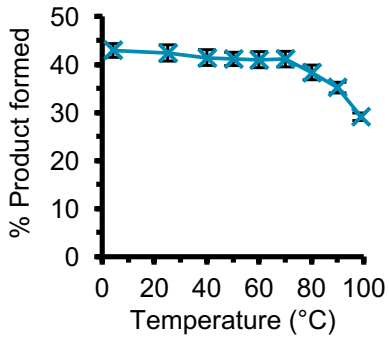


Figure 4.10 SnoopLigase thermoresilience. SnoopLigase was incubated for 15 min at indicated temperatures. Heat-treated SnoopLigase was used for reaction of SnoopTagJr-AffiHER2 and SUMO-DogTag at 10 µM each in 50 mM Tris borate pH 7.25 + 15% (v/v) glycerol for 2 h at 4 °C. Samples were boiled, run on SDS-PAGE and visualised by Coomassie staining. Quantification by densitometry (mean ± 1 S.D., n=3).

SnoopLigase treatment up to 70 °C did not lead to loss in activity, while higher temperatures caused partial activity loss. 99 °C heat treatment led to an approximately 30 % loss in relative substrate conversion (Figure 4.9).

4.5.4. SnoopLigase lyophilisation

As the aim of my thesis project was to make SnoopLigase a widely applicable tool, it was important at this stage to think about long-term storage possibilities for the protein. Often proteins are stored as frozen solutions at -20 or -80 °C (Carpenter et al., 2002). However, protein storage as lyophilised powder has several advantages, including i) reduced storage space required, ii) simplification of transport by avoiding the necessity of a cold chain, iii) reduced energy cost of storage, and iv) possibility to increase protein concentration by redissolution in lower

volume (Matejtschuk, 2007). However, not all proteins can be lyophilised as the dehydration process can lead to protein aggregation (Jain and Roy, 2010).

To test whether SnoopLigase could be lyophilised and stored as powder outside the cold chain, the protein was lyophilised in aliquot form and stored at 37 °C for up to 120 days. The storage temperature of 37 °C was chosen to mimic transient or permanent exposure to elevated temperatures in countries with hot climate. Substrate conversion activity of regular or lyophilised SnoopLigase was compared. Lyophilisation and storage for up to 120 days had little or no impact on SnoopLigase reactivity (Figure 4.10).

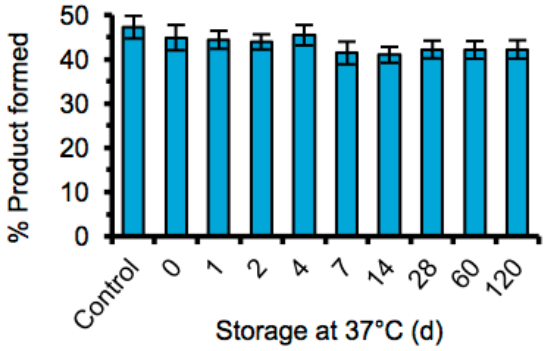


Figure 4.11 SnoopLigase can be stored in lyophilised form. SnoopLigase was lyophilised and stored for indicated periods at 37 °C. Lyophilised SnoopLigase was re-constituted in buffer and used for reaction of SnoopTagJr-AffiHER2 and SUMO-DogTag at 10 µM each in 50 mM Tris borate pH 7.25 + 15% (v/v) glycerol for 2 h at 4 °C. Samples were boiled, run on SDS-PAGE and visualised by Coomassie staining. Quantification by densitometry (mean ± 1 S.D., n=3).

4.6. Discussion

In the previous chapter I described the development of +6RrgALigase+9, which was able to covalently link two peptide tags. However, this prototype system was far from optimal, with slow reaction rate and reaction conditions strongly differing from the aspired physiological ones. Many engineering approaches exist

to enhance protein reactions, including rational mutations, computational or evolution approaches (Steiner and Schwab, 2012).

In this chapter I used two approaches to engineer the ligation system – introduction of proline residues at suitable sites, and computational analysis of the ligase parent domain by the PROSS algorithm. These engineering efforts gave rise to the final versions of the split protein modules, SnoopTagJr and SnoopLigase. Overall, five mutations were taken forward: D737S in SnoopTagJr, and A808P, Q837P, D838G and I839V in SnoopLigase. Given that only nine mutations had been tested overall, this engineering endeavour could be considered successful and efficient.

With Q837P, D838G and I839V, three of the mutations clustered directly next to each other. This mutated region is located at the beginning of the sequence overlap section between SnoopLigase and DogTag, which might be a flexible structure element in SnoopLigase and move away upon DogTag binding. Potentially, the dynamics of this movement could have strong influence on DogTag binding, which would explain the strong effect of the three sequential Q837P, D838G and I839V mutations. The hypothesis is supported by the fact that positions 838 and 839 are also the first two residues of DogTag, but mutation of these residues had no effect on DogTag reactivity. The mutations modulate reactivity only in the context of the SnoopLigase loop.

The A808P mutation is positioned in a loop that interacts with DogTag. Stabilisation of this loop might be important to bind DogTag and thereby enhance reaction rate.

This study is the first to my knowledge using the PROSS system to enhance a split protein system. It would be interesting to see whether PROSS-suggested

mutations could also improve the performance of other split protein systems, like SpyLigase (Fierer et al., 2014) or split GFP (Cabantous et al., 2005). Given the strong dependence of the PROSS outcome on the multiple sequence alignment, it would also be worth testing how recent advances in multiple sequence alignment alter the PROSS suggestions (Oda et al., 2017). Furthermore, newly released protein engineering portals might be useful to identify further stabilising and rate-enhancing mutations in SnoopLigase (Musil et al., 2017; Schwarte et al., 2017).

Before engineering, optimal reaction conditions of RrgALigase were far from physiological (high pH, 4 °C, 1.5 M TMAO, 40 % glycerol, no salt). With the development of SnoopLigase, reaction conditions have moved closer towards physiological conditions: around physiological pH, 4 – 37 °C no TMAO, 0 – 20% glycerol, salt acceptable. Together with the enhanced reaction rate, the new reaction conditions should make SnoopLigase a useful tool for a wide range of applications. The sensitivity of the reaction to the addition of ions could indicate either the importance of electrostatic interactions between the modules for domain re-constitution, or an altered conformation of SnoopLigase in the presence of salt, which is less suitable for tag interaction.

SnoopLigase is reactive in high concentrations of Tween 20 and Triton X-100, which can be useful when ligating proteins that require these detergents for solubility, or in assays that use these detergents to block protein sticking to plastic surfaces.

Quenching of SnoopLigase reaction can be performed by addition of ≥ 0.02 % (w/v) SDS or > 2 M urea. SDS binds strongly to proteins and denatures them already at millimolar concentrations (Reynolds et al., 1967). Instead of protein

denaturation, the negative charge imposed on the split protein components by SDS could also prevent the interaction of the three modules.

With an IC50 of approximately 0.8 M, SnoopLigase reaction is inhibited by relatively low concentration of urea. Many other proteins are only affected by higher urea concentrations (Kleppe et al., 1999; Xie et al., 2014; Yano and Irie, 1975). The low IC50 could either indicate limited stability of the SnoopLigase protein, or that the interaction of urea with specific parts of the polypeptide prevents domain reconstitution (Lim et al., 2009).

SnoopLigase was thermostable up to 70 °C and could be lyophilised with subsequent storage at 37 °C for at least 120 days without major loss of activity. These characteristics enable convenient storage and transport of the SnoopLigase protein outside the cold-chain, which is an important consideration for wide application.

5. Development of a Plug-and-Display Nanoparticle as Vaccination Platform

Vaccines are important medical interventions, preventing infections and saving millions of lives each year (Matthiessen et al., 2017; Orenstein and Ahmed, 2017). Although highly successful products are available against many diseases including measles, polio and hepatitis B (Hajj Hussein et al., 2015), other conditions like malaria and AIDS still lack effective vaccines (Ouattara et al., 2015).

Recombinantly produced proteins are a cost-effective and safe form of vaccines (Moyle and Toth, 2013). However, recombinant antigens are often poorly immunogenic (Moyle, 2017). Protein multimerisation can strongly enhance the immune response against the antigen and has been achieved with a wide range of multimerisation domains (Baschong et al., 2003; Kushnir et al., 2012; Ogun et al., 2008). However, the generation of multimeric fusion constructs is not always successful (Jain et al., 2015; Stockley and Mastico, 2000). This problem can be overcome by modular systems, in which the multimerisation platform and the antigen are produced separately to yield correctly folded proteins and subsequently coupled together.

In the previous chapter I improved the reaction characteristics of the prototype peptide-peptide ligase by protein engineering. SnoopLigase could efficiently conjugate DogTag- and SnoopTagJr-fused proteins. In this chapter I describe the development of a robust modular multimeric vaccine platform. SnoopLigase-mediated conjugation of antigens to the heptameric protein IMX313 may have the potential to enhance the efficacy of a wide range of recombinant protein vaccines. The generation of immunisation-grade conjugate vaccines also

required the development of a general method to separate SnoopLigase-conjugated proteins from unreacted substrate and the ligase.

5.1. Nanoparticle Development

IMX313 is an engineered, mainly heptameric protein that has been used for the development of multimerised protein vaccines (see section 1.3.3.) (Li et al., 2016; Ogun et al., 2008). Previously, I had worked together with Karl Brune, Yuanyuan Li, Iona Taylor and Florian Brod from the Howarth and Biswas laboratories to develop a Catcher-based modular IMX313 vaccination platform. SpyCatcher-IMX313-SnoopCatcher allowed simultaneous immunisation with two different antigens and enhanced antibody titres of both antigens compared to the free proteins (Brune et al., 2017).

Despite its immunogenicity-enhancing effect, the Catcher-based IMX313 platform system had several limitations, such as moderate production yield, limited thermostability and the tendency to self-react when stored at ≥ 4 °C. Moreover, the system required the injection of large protein domains (SpyCatcher and SnoopCatcher) that can be a source of off-target antigenicity. It would be desirable to exclude large protein domains like Catchers from platform particles, as irrelevant proteins can reduce the immune response against the antigen (Cohn et al., 1949; Ringe et al., 2017; Sutherland et al., 2001) Therefore, I generated IMX313-DogTag which could allow multimerisation of SnoopTagged antigen by SnoopLigase-mediated conjugation, with the aim to overcome the disadvantages of the Catcher-based vaccine platform.

5.1.1. Purification of IMX313-DogTag

The purification of IMX313-DogTag was optimised for two constructs – one with and one without N-terminal His_6-tag. Both constructs expressed well in *E. coli*. The His_6-tag carrying construct could be purified using a denaturing IMAC procedure. As poly-His-tags are not desirable for immunisation grade preparations (Gaberc-Porekar and Menart, 2001), I established a protocol to purify His_6-tag-free IMX313-DogTag, which would be more suitable for vaccinations.

Heat treatment of the *E. coli* cell lysate caused most proteins to precipitate, while IMX313-DogTag remained soluble. Treatment of the lysate with low pH glycine buffer caused IMX313-DogTag to precipitate, but the major contaminating protein remained soluble. Therefore, resuspension of the pellet in fresh buffer yielded a cleaner preparation. Polyethyleneimine (PEI) addition to the sample was required to precipitate the remaining DNA contamination. Ammonium sulfate precipitation of IMX313-DogTag allowed removal of PEI and low molecular weight impurities from the preparation. Finally, ion exchange chromatography yielded a highly pure IMX313-DogTag preparation. After oxidation and without the addition of reducing agent, IMX313-DogTag showed the expected disulfide-linked multimeric organisation (Figure 5.1).

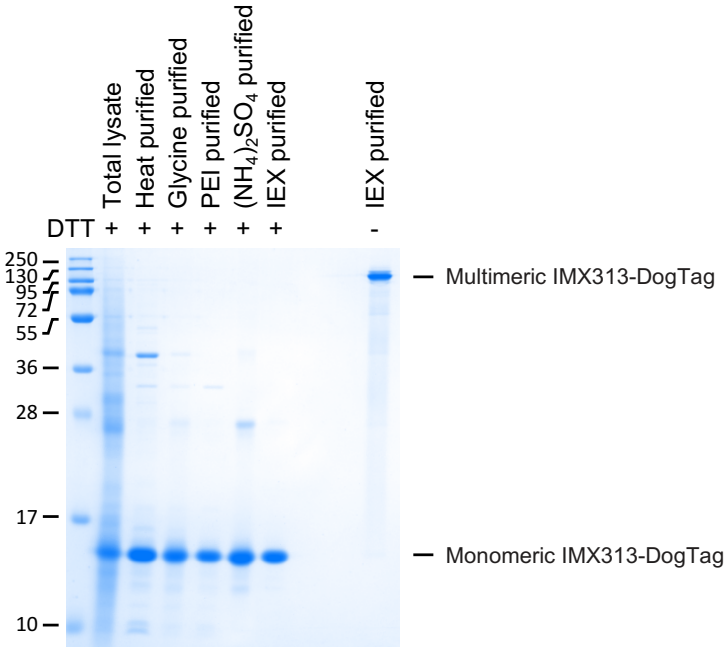


Figure 5.1 Purification of IMX313-DogTag without His-Tag. Analysis of samples at different stages of the IMX313-DogTag purification procedure. "IEX" refers to ion exchange chromatography. DTT was added where indicated to reduce inter-chain disulfide bonds. Samples were boiled, run on SDS-PAGE and visualised by Coomassie staining.

5.1.2. IMX313-DogTag Conjugation with Model Proteins

To test whether the purified IMX313-DogTag was reactive, I conjugated the model proteins SnoopTagJr-AffiHER2, SnoopTagJr-MBP and SnoopTagJr-CIDR onto IMX313-DogTag using SnoopLigase. Approximately 42 kDa in size, MBP is a relatively large protein and was chosen to indicate potential steric clashes when conjugating large proteins on IMX313-DogTag, which could prevent complete conjugation. Cysteine-rich interdomain region (CIDR) is a highly variable domain type from the *Plasmodium falciparum* erythrocyte membrane protein-1 family (PfEMP1). The specific variant used here was IT4var07 (Lau et al., 2015). CIDR was chosen as model protein for conjugation on IMX313-DogTag due to its structural complexity, containing several disulfide bonds (Lau et al., 2015).

SnoopTagJr-CIDR was cloned by me, but expressed and purified by Anne-Marie Andersson in the Howarth laboratory.

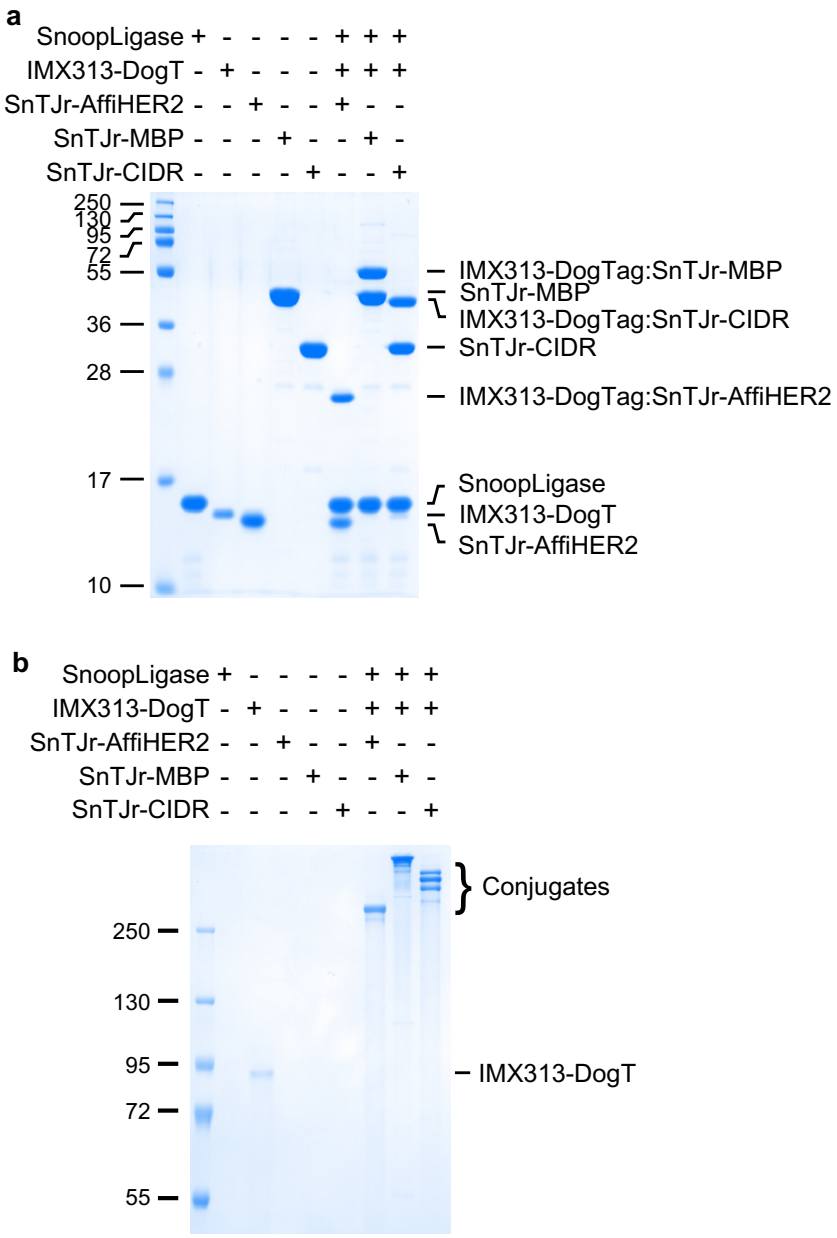


Figure 5.2 IMX313-DogTag conjugation with model proteins. SnoopTagJr-AffiHER2, SnoopTagJr-MBP or SnoopTagJr-CIDR at 20 μM were conjugated to 10 μM IMX313-DogTag using 20 μM SnoopLigase in 50 mM Tris borate pH 7.25 + 15% (v/v) glycerol for 48 h at 4 °C. Samples were boiled in the a) presence or b) absence of DTT, run on SDS-PAGE and visualised by Coomassie staining. Unconjugated antigens ran off the gel in b).

Conjugation of all three model proteins on IMX313-DogTag was successful. SnoopTagJr-AffiHER2 and SnoopTagJr-MBP conjugations nearly completely depleted unconjugated IMX313-DogTag, while some unconjugated monomer was still visible in the SnoopTagJr-CIDR reaction sample (Figure 5.2a). Conjugated IMX313-DogTag retained its multimeric inter-chain disulfide-linked state (Figure 5.2b).

5.1.3. Characterization of IMX313-DogTag

5.1.3.1. IMX313-DogTag Thermal Resilience

Thermostability is an important aspect in the development of therapeutic protein products. High protein thermostability allows room for temperature variation during manufacturing, handling and storage, which is especially relevant in situations without an efficient cold-chain (Asial et al., 2013; Chirino and Mire-Sluis, 2004). Thermal denaturation of proteins is mostly caused by disruption of the protein's tertiary structure (Tanford, 1968). Consisting of only two α-helices (IMX313 structure (Hofmeyer et al., 2013; Ogun et al., 2008)), a GS-rich flexible linker and two β-strands (DogTag) per monomer, as well as disulfide bonds between individual monomers, I hypothesised that IMX313-DogTag would be highly thermoresilient.

Incubation of IMX313-DogTag at temperatures up to 99 °C did not lead to loss in reactivity of the protein (Figure 5.3). This endorses IMX313-DogTag as robust immunisation platform protein.

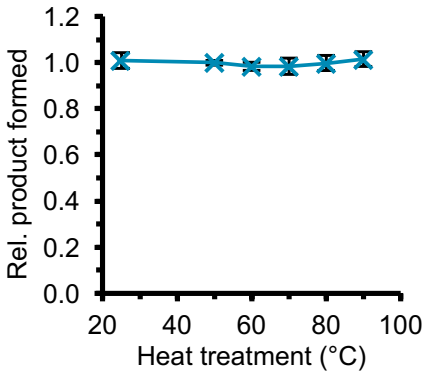


Figure 5.3 IMX313-DogTag thermoresilience. IMX313-DogTag was incubated in 50 mM Tris borate pH 7.25 for 2 h at indicated temperatures. Heat-treated IMX313-DogTag was used for reaction with SnoopTagJr-AffiHER2 mediated by SnoopLigase at 10 µM each in 50 mM Tris borate pH 7.25 + 15% (v/v) glycerol for 2 h at 4 °C. Samples were boiled in the presence of DTT, run on SDS-PAGE and visualised by Coomassie staining. Quantification by densitometry (mean ± 1 S.D., n=3).

5.1.3.2. IMX313-DogTag Lyophilisation

As described in section 4.5.4. lyophilisation of proteins allows more convenient storage and transport. This is specifically important considering that IMX313-DogTag is intended to be used as vaccine platform protein. Failure of the cold-chain is a frequent event during vaccine handling (Bell et al., 2001) and vaccine transport (Guthridge and Miller, 1996), especially in developing countries due to limited access to resources and electricity (Kartoglu and Milstien, 2014; Murhekar et al., 2013). Hence, high protein stability is an important aspect in vaccine development (Chen and Kristensen, 2009; Jezek et al., 2009).

To test whether IMX313-DogTag could be lyophilised and stored as powder outside the cold chain, the protein was lyophilised in aliquot form and stored at 37 °C for up to 7 days. The storage temperature of 37 °C was chosen to mimic transient or permanent exposure to elevated temperatures in countries with hot climate.

SnoopLigase-mediated conjugation of SnoopTagJr-AffiHER2 with regular or lyophilised IMX313-DogTag was compared. Lyophilisation and storage for up to 7 days had no to minor negative impact on SnoopLigase reactivity (Figure 5.4).

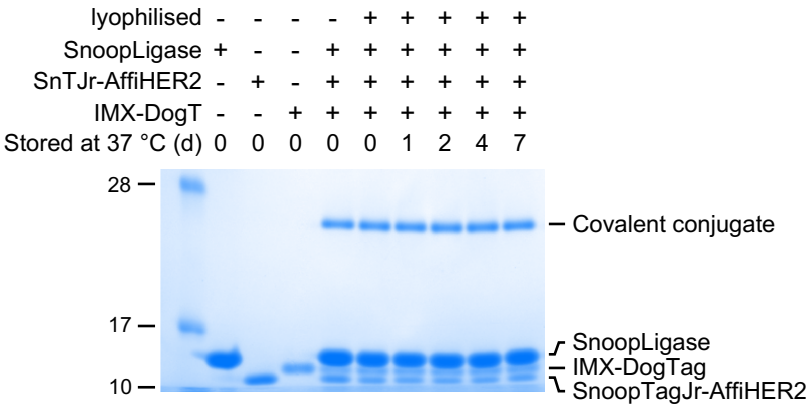


Figure 5.4 IMX313-DogTag can be stored in lyophilised form. IMX313-DogTag was lyophilised and stored for indicated periods at 37 °C. Lyophilised IMX313-DogTag was re-constituted in buffer to a final concentration of 10 µM and used for reaction with 10 µM SnoopTagJr-AffiHER2 mediated by 20 µM SnoopLigase in 50 mM Tris borate pH 7.25 + 15% (v/v) glycerol for 2 h at 4 °C. Samples were boiled in the presence of DTT, run on SDS-PAGE and visualised by Coomassie staining.

5.2. Development of SnoopLigase Removal Procedures

An aim of IMX313-DogTag development as a SnoopLigase-based vaccine platform was to reduce off-target antigenicity caused by the presence of Catcher proteins in the platform protein. However, during the development of SnoopLigase it became apparent that the enzyme remains bound to the product after ligation of the SnoopTagJr and DogTag substrates. Therefore, ligation of antigens onto IMX313-DogTag with SnoopLigase would possibly have the same off-target antigenicity effect as Catcher-fused IMX313.

To achieve the aim of reducing off-target antigenicity, I investigated methods to remove SnoopLigase from the ligated product after reaction.

5.2.1. Biotin-SnoopLigase Solid-Phase Immobilisation

The non-covalent interaction between biotin and streptavidin is a widely used tool for protein immobilisation due to the femtomolar affinity, high on-rate and resilience to a wide range of temperatures and pH values (Laitinen et al., 2006; Stayton et al., 1999). The biotin-streptavidin-based concept for SnoopLigase reaction product purification is illustrated in Figure 5.5a. After covalent conjugation of SnoopTagJr- and DogTag-linked proteins of interest using biotinylated SnoopLigase, the complex of ligase and reaction product are immobilised on streptavidin resin, while unreacted substrate is washed away. Subsequently, disruption of the interaction between SnoopLigase and the reaction product allows elution of the reaction product while biotinylated SnoopLigase remains bound to the streptavidin resin. Hence, besides SnoopLigase removal from the reaction product, the procedure also incorporates the removal of unreacted substrate. In this regard, SnoopLigase's trait of remaining bound to the reaction product is advantageous, as it abolishes the need of laborious product purification steps, such as size exclusion chromatography. Biotinylated SnoopLigase was efficiently captured by streptavidin agarose (Figure 5.5b).

5.2.2. SnoopLigase Elution Conditions

After being able to immobilise biotinylated SnoopLigase on streptavidin resin, the challenge was to identify conditions that allow the release of the reaction product from SnoopLigase without disrupting the biotin-streptavidin interaction.

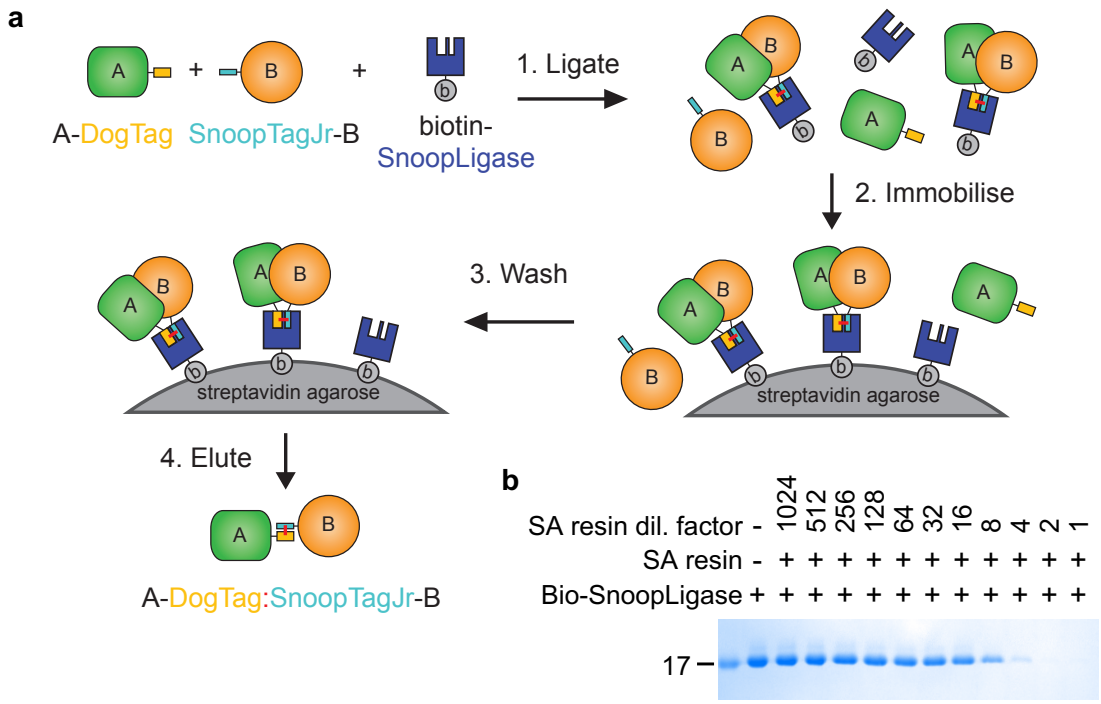


Figure 5.5 SnoopLigase Immobilisation for Reaction Product Purification. a) Cartoon of the SnoopLigase reaction product purification methodology principle. SnoopTagJr- and DogTag-fused proteins A and B are covalently conjugated using biotinylated SnoopLigase. SnoopLigase is immobilised on streptavidin agarose resin, washing allows removal of unreacted substrate, and ligated reaction product is eluted. b) Analysis of streptavidin resin binding capacity for biotin-SnoopLigase. 10 µM biotin-SnoopLigase was incubated with 1 µl packed HiCap streptavidin agarose or dilutions thereof in 50 mM Tris borate pH 8.0 for 1 h at 25 °C. To determine the amount of biotin-SnoopLigase remaining unbound to the streptavidin resin, samples were centrifuged, the resin-free supernatant boiled, run on SDS-PAGE and visualised by Coomassie staining.

5.2.2.1. Initial Elution Condition Screen

To test potential conditions for SnoopLigase reaction product elution, I ligated model proteins using biotinylated SnoopLigase, added streptavidin agarose resin and after washing I incubated the samples with different buffers containing a range of additives which can disrupt protein-protein interaction. These included high salt (Sawyer and Puckridge, 1973), detergents (Kloet et al., 2015; Sihag and Deutscher, 1983) and low or high pH buffers (Prajapati et al., 1998; Roberts et al., 2015;

Watanabe et al., 2009). Salts can dissociate charge-based interactions while detergents may disrupt hydrophobic contacts (Masters, 2004). Furthermore, elevated temperature (Stromer et al., 2004) as well as competition by ligated product (Pollard, 2010) were investigated for SnoopLigase reaction product elution.

The screens identified glycine pH 2.0 (Figure 5.6a), high temperature and product competition (data shown in section 5.2.2.3.), and imidazole (Figure 5.6b) as promising elution approaches. The conditions were analysed in more detail to optimise the SnoopLigase reaction product purification protocol.

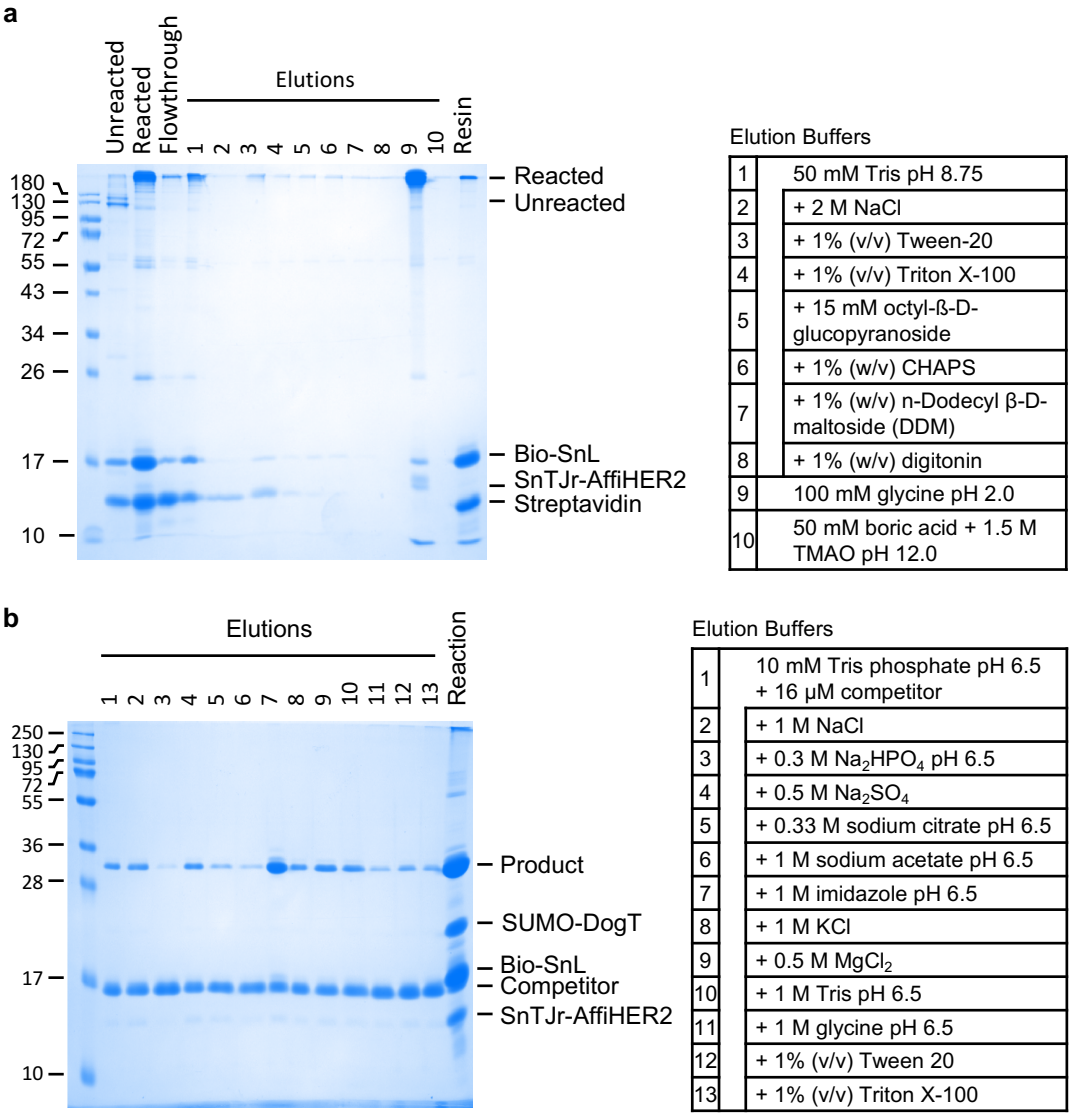


Figure 5.6 Low pH and imidazole allow reaction product elution from immobilised SnoopLigase. a) SnoopTagJr-AffiHER2 at 20 µM and IMX313-DogTag at 10 µM were conjugated using biotinylated SnoopLigase at 10 µM in 50 mM Tris borate pH 8.75 + 7.5% (v/v) + 0.75 M TMAO for 4 h at 4°C. SnoopLigase was immobilised on streptavidin agarose, unreacted substrate washed away, and reaction product eluted sequentially with indicated buffers and washes of 50 mM Tris borate pH 8.75 in between each elution. The resin was resuspended in 50 mM Tris borate pH 8.0. b) SnoopTagJr-AffiHER2 and SUMO-DogTag were conjugated using biotinylated SnoopLigase at 50 µM each in 50 mM Tris borate pH 7.25 + 15% (v/v) for 24 h at 4°C. SnoopLigase was immobilised on streptavidin agarose, unreacted substrate washed away, and reaction product eluted with indicated buffers. Competitor refers to AffiHER2-DogTag conjugated to SnoopTag peptide using SnoopLigase and purified using SnoopLigase pulldown. Samples of a) and b) were boiled, run on SDS-PAGE and visualised by Coomassie staining.

5.2.2.2. Low pH Elution

Conjugation of SnoopTagJr-AffiHER2 to SUMO-DogTag using biotinylated SnoopLigase, followed by immobilisation on streptavidin agarose and elution with glycine pH 2.0 yielded the purified reaction product without detectable levels of substrate or ligase. The streptavidin agarose retained biotinylated SnoopLigase, as well as minor amounts of reaction product (Figure 5.7a). Mass spectrometry analysis of purified SUMO-DogTag conjugated to SnoopTag peptide matched the expected molecular weight (Figure 5.7b).

Since, after product elution, biotinylated SnoopLigase was retained on the streptavidin agarose resin, I hypothesised that the immobilised enzyme could be re-used to conjugate new substrate (Figure 5.7c). SnoopLigase reaction cycling could allow for more economical use of the purified and biotinylated SnoopLigase, as well as the streptavidin resin. To test whether immobilised SnoopLigase could be recycled, I incubated biotinylated SnoopLigase with streptavidin agarose, incubated the sample with SnoopTagJr-AffiHER2 and SUMO-DogTag and purified the product using glycine pH 2.0. Subsequently I started a new reaction by addition of fresh substrate. After up to eight reaction cycles, no loss in product formation was detected, showing that immobilised SnoopLigase can be re-used.

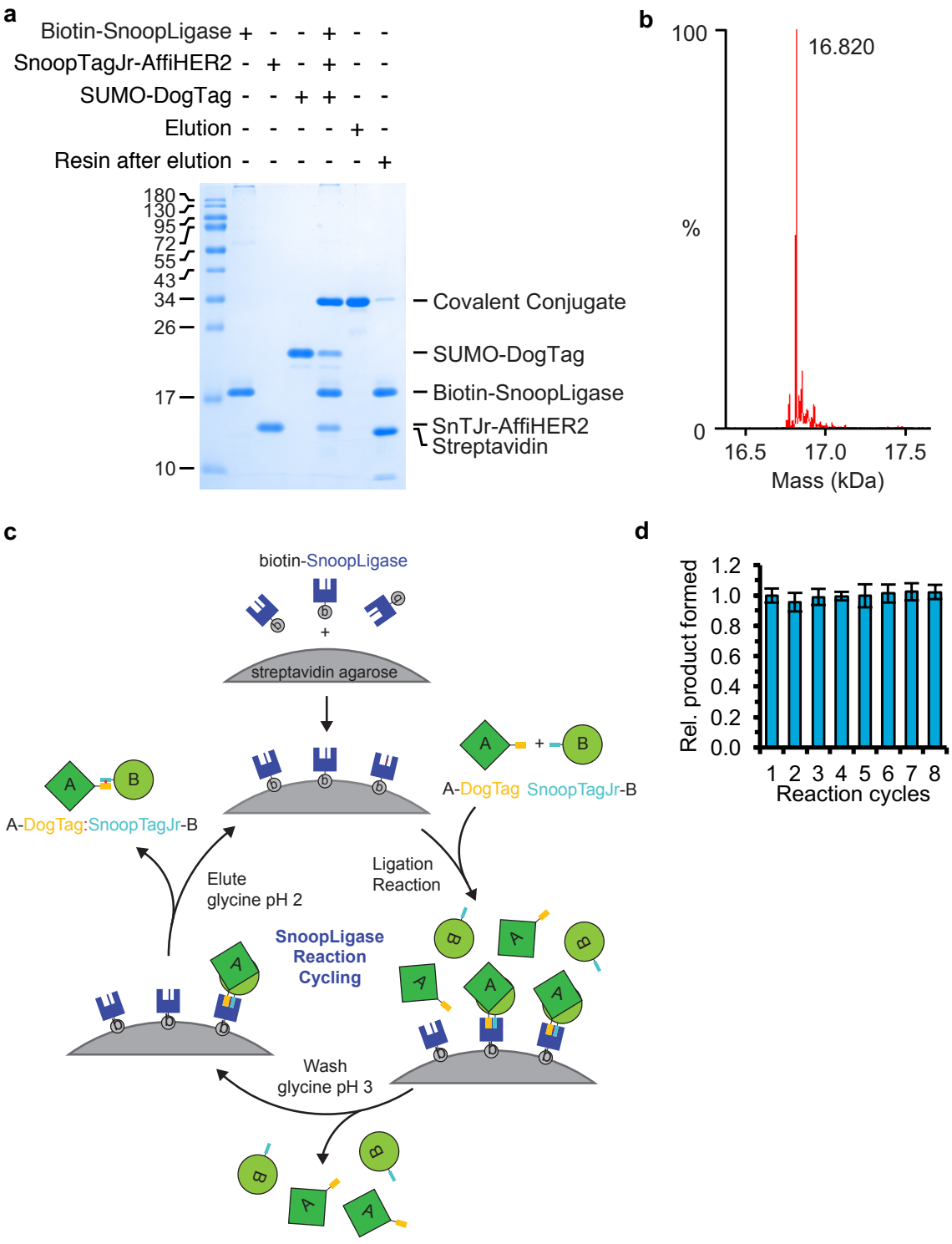


Figure 5.7 Low pH elution allows repeated purification of SnoopLigase reaction product. a) SnoopLigase immobilisation allows purification of reaction product. SnoopTagJr-AffiHER2 and SUMO-DogTag were conjugated using biotinylated SnoopLigase at 10 µM each for 20 h at 4°C. SnoopLigase was immobilised on streptavidin agarose, unreacted substrate was washed away, and reaction product eluted with 50 mM glycine pH 2.0. The resin was resuspended in 50 mM Tris borate pH 8.0. Samples were boiled, run on SDS-PAGE and visualised by Coomassie staining. b) Electrospray ionization

mass spectrometry analysis of the purified reaction product from a). c) Cartoon illustrating the principle for immobilised SnoopLigase reaction cycling. Biotinylated SnoopLigase is immobilised on streptavidin agarose resin. SnoopTagJr- and DogTag-fused proteins A and B are covalently conjugated using immobilised SnoopLigase. Washing allows removal of unreacted substrate, and ligated reaction product is eluted. Immobilised SnoopLigase can then be used for further cycles of ligation reactions. d) Immobilised SnoopLigase reaction cycling does not decrease SnoopLigase reactivity. SnoopTagJr-AffiHER2 and SUMO-DogTag at 50 µM each were conjugated using immobilised biotin-SnoopLigase in 50 mM Tris borate pH 7.25 + 15% (v/v) glycerol for 1 h at 4°C. Unreacted substrate was washed away and reaction product was eluted with 50 mM glycine pH 2.0. The resin was used for further reaction rounds. Elution samples of the different reaction cycles were boiled, run on SDS-PAGE and visualised by Coomassie staining. Quantification of b) by densitometry (mean ± 1 S.D., n=9).

5.2.2.3. Elution at High Temperature

Elevated temperatures can disrupt protein-protein interactions (Schwesinger et al., 2000). To test whether high temperatures could dissociate SnoopLigase from the reaction product, I conjugated SnoopTagJr-AffiHER2 to SUMO-DogTag using biotinylated SnoopLigase, followed by immobilisation on streptavidin agarose and incubation at 25 – 55 °C. Reaction product elution was inefficient at 25, 37 and 45 °C but efficient at 55 °C (Figure 5.8).

Solubility and folding of proteins is temperature-dependent and for many proteins incubation at 55 °C would not be advisable (Ku et al., 2009; Stojanoski et al., 2015). To test whether addition of competing SnoopLigase product could decrease the required elution temperature, I repeated the thermal elution experiment with addition of SUMO-DogTag coupled to SnoopTag peptide as competitor during the elution. The presence of competitor improved SnoopLigase reaction product elution at all temperatures, decreasing the required temperature for efficient elution to 45 °C (Figure 5.8).

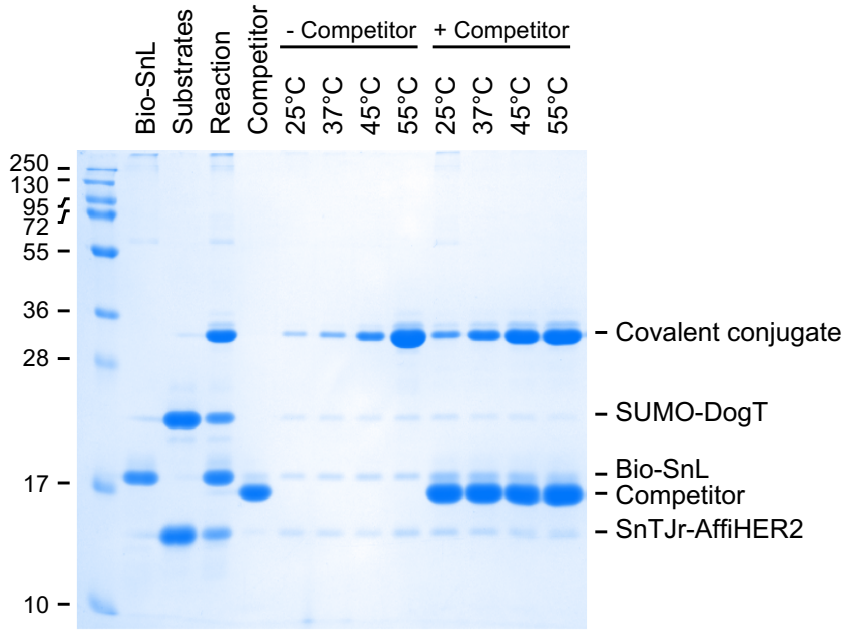


Figure 5.8 Elevated temperature and competitor efficiently elute SnoopLigase reaction product. SnoopTagJr-AffiHER2 and SUMO-DogTag were conjugated using biotinylated SnoopLigase at 50 µM each in 50 mM Tris borate pH 7.25 + 15% (v/v) for 5 h at 4°C. SnoopLigase was immobilised on streptavidin agarose, unreacted substrate washed away, and reaction product eluted in PBS with or without 35 µM competitor at indicated temperatures. Competitor refers to AffiHER2-DogTag conjugated to SnoopTag peptide using SnoopLigase and purified using SnoopLigase pulldown. Samples were boiled, run on SDS-PAGE and visualised by Coomassie staining.

5.2.2.4. Imidazole Elution

Previously, a buffer screen had identified imidazole as promising candidate for SnoopLigase product elution (Figure 5.6b). To test what concentrations of imidazole could efficiently dissociate SnoopLigase from the reaction product, I conjugated SnoopTagJr-AffiHER2 to SUMO-DogTag using biotinylated SnoopLigase, followed by immobilisation on streptavidin agarose and incubation with elution buffer containing 0.5 – 4 M imidazole. Efficient elution occurred at ≥ 2 M imidazole (Figure 5.9).

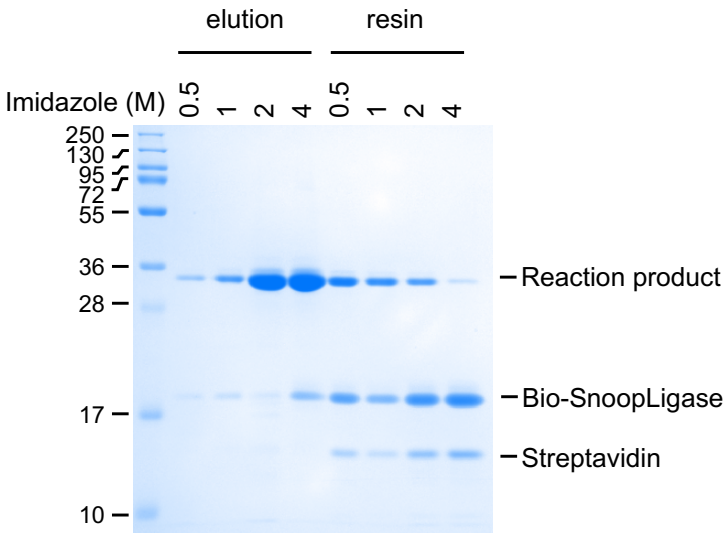


Figure 5.9 Imidazole efficiently elutes reacted product from SnoopLigase. SnoopTagJr-AffiHER2 and SUMO-DogTag were conjugated using biotinylated SnoopLigase at 50 µM each in 50 mM Tris borate pH 7.25 + 15% (v/v) for 24 h at 4°C. SnoopLigase was immobilised on streptavidin agarose, unreacted substrate washed away, and reaction product eluted in 25 mM Tris phosphate pH 7.0 with indicated concentrations of imidazole pH 7.0 at 25 °C. Resins after elution were resuspended in 25 mM Tris phosphate pH 7.0. Samples were boiled, run on SDS-PAGE and visualised by Coomassie staining.

5.2.3. HaloTag7-SnoopLigase Solid-Phase Immobilisation

Biotin-streptavidin-based immobilisation of SnoopLigase allowed purification of SnoopLigase reaction product. However, detectable amounts of biotin-SnoopLigase were occasionally present in the elution samples (Figure 5.8 and 5.9). The level of SnoopLigase contamination was substrate-dependent, but also varied between replicates of identical sample set-ups. Even after extensive troubleshooting, these fluctuations could not be eliminated and influenced experiment reproducibility. Despite careful experimentation, it occurred that the fluctuations were likely caused by minute differences in sample handling, which can hardly be avoided by human experimenters in laboratory environments. To make

the SnoopLigase reaction product purification approach more robust for general use, I sought to develop a more reliable method.

The interaction between biotin and streptavidin is strong, yet non-covalent and reversible (Stayton et al., 1999). While the dissociation rate is low under mild conditions, it can increase with pH, temperature, shear forces (Pierres et al., 2002) and vary depending on the conjugated molecules (Swift et al., 2006). To improve SnoopLigase removal, an option would have been to use traptavidin rather than streptavidin agarose for the immobilisation of SnoopLigase. Traptavidin is a streptavidin mutant with an approximately 10-fold lower biotin-dissociation rate (Chivers et al., 2010). However, traptavidin resin is not available commercially, and would need to be produced manually. This would represent an inconvenient extra step and potentially hinder wide-spread use of the SnoopLigase system. Furthermore, I reasoned that a covalent attachment method could prevent unwanted SnoopLigase to detach from the resin and thereby improve reproducibility and purity of sample preparations.

A commonly used system for covalent protein immobilisation is the HaloTag system (England et al., 2015; Hata and Nakayama, 2007; Ohana et al., 2009). HaloTag7-SnoopLigase showed equal ligation activity as biotinylated SnoopLigase with SUMO-DogTag, IMX313-DogTag, and SnoopTagJr-AffiHER2 as substrates (Figure 5.10a). To test whether HaloTag7-mediated immobilisation could reduce SnoopLigase contamination in the elution, I conjugated SUMO-DogTag and SnoopTagJr-AffiHER2 with HaloTag7-SnoopLigase, followed by immobilisation on HaloLink resin and incubation with elution buffer containing 2 M imidazole. Conjugated product elution was efficient, without detectable levels of substrate or ligase impurities (Figure 5.10b).

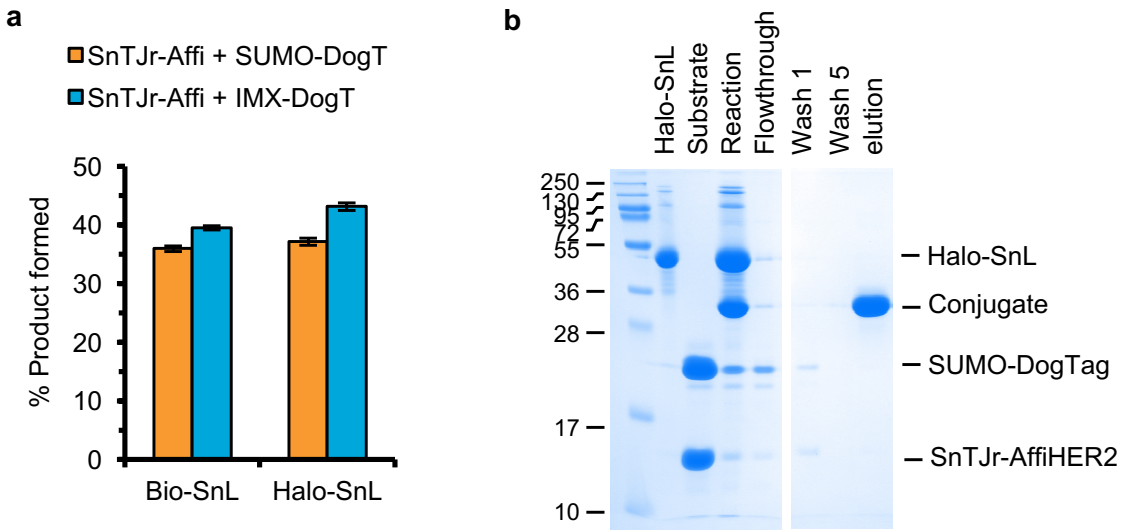


Figure 5.10 HaloTag7-SnoopLigase Immobilisation for Reaction Product Purification. a) HaloTag7-SnoopLigase is as reactive as biotinylated SnoopLigase. Biotinylated SnoopLigase or HaloTag7-SnoopLigase were incubated with SUMO-DogTag and SnoopTagJr-AffiHER2 or IMX313-DogTag and SnoopTagJr-AffiHER2 at 10 µM each in 50 mM Tris borate pH 7.25 + 15% (v/v) glycerol for 2 h at 4 °C. b) SnoopTagJr-AffiHER2 and SUMO-DogTag were conjugated using HaloTag7-SnoopLigase at 10 µM each in 50 mM Tris borate pH 7.25 + 15% (v/v) for 12 h at 4°C. SnoopLigase was immobilised on HaloLink resin for 1 h at 25 °C, unreacted substrate washed away, and reaction product eluted using 25 mM Tris phosphate + 2 M imidazole pH 7.0. Samples were boiled, run on SDS-PAGE and visualised by Coomassie staining. Quantification of a) by densitometry (mean ± 1 S.D., n=3).

5.3. Generation and Purification of Protein-Decorated IMX313-DogTag

To test whether HaloTag7-SnoopLigase could efficiently ligate proteins to IMX313-DogTag, I performed conjugations with SnoopTagJr-AffiHER2, SnoopTagJr-MBP, or SnoopTagJr-CIDR. After 48 h reaction, SnoopTagJr-AffiHER2 and SnoopTagJr-MBP conjugations to IMX313-DogTag reached > 95% and approximately 85% for SnoopTagJr-CIDR conjugation (Figure 5.11a). A similar result was previously observed using standard SnoopLigase, indicating that the fusion of HaloTag7 to SnoopLigase was not the cause of the incomplete reaction (Figure 5.2).

Purification of SnoopLigase reaction product using the HaloTag7-mediated immobilisation approach had proven successful with SUMO-DogTag and SnoopTagJr-AffiHER2 as substrates (section 5.2.3.). I used the established procedure to conjugate SnoopTagJr-MBP to IMX313-DogTag, immobilise the HaloTag7-SnoopLigase on HaloLink resin and elute the reaction product with imidazole. The eluted conjugate retained its multimeric structure as shown by non-reducing SDS-PAGE, but also showed a detectable amount of HaloTag7-SnoopLigase contamination (Figure 5.11b).

To further validate the conjugation of SnoopTagJr-MBP to IMX313-DogTag, I analysed the eluted reaction product using size exclusion chromatography (SEC) and compared the trace to unconjugated IMX313-DogTag. The conjugated protein eluted at lower volume than the unconjugated protein, verifying the expected increase in particle size (Figure 5.11c). SDS-PAGE analysis of the respective peaks confirmed the multimeric nature of the proteins (Figure 5.11d). Dynamic light scattering (DLS) of SEC-purified proteins revealed an increase in mean hydrodynamic radius from 5.2 to 14.5 nm upon conjugation (Figure 5.11e).

SnoopTagJr-MBP was useful as model protein, but is not relevant as vaccine target. Cysteine-rich protective antigen (CyRPA) is a protein involved in *Plasmodium falciparum* merozoite invasion into erythrocytes and represents a promising malaria vaccine candidate (Dreyer et al., 2012). CyRPA-SnoopTagJr was expressed and purified by Anne-Marie Andersson in the Howarth laboratory. To test whether the established method could be used to purify IMX313-DogTag nanoparticles conjugated to CyRPA-SnoopTagJr, I conjugated the two proteins using HaloTag7-SnoopLigase. After 48 h reaction, most of the IMX313-DogTag

monomer had reacted and the product could be purified, albeit with detectable level of HaloTag7-SnoopLigase contamination (Figure 5.11a).

An aim of this thesis was to be able to generate antigen-conjugated platform particles free of contaminant substrate or enzyme. Despite promising results with the HaloTag7-based approach on monomeric proteins (section 5.2.3.), application to the multimeric IMX313 particle caused HaloTag7-SnoopLigase presence in the product elution samples. To analyse and troubleshoot the problem, I tested the binding of HaloTag7-SnoopLigase to HaloLink resin over time. Nearly complete coupling of the protein to the resin occurred within one hour, but binding did not reach completion even after eight hours of incubation (Figure 5.11g). Furthermore, I found that coupling of substrate-bound HaloTag7-SnoopLigase showed strongly decreased binding rate (Figure 5.11g). Hence the standard resin incubation time was increased to six hours, which usually did not allow the complete capture of conjugated reaction product.

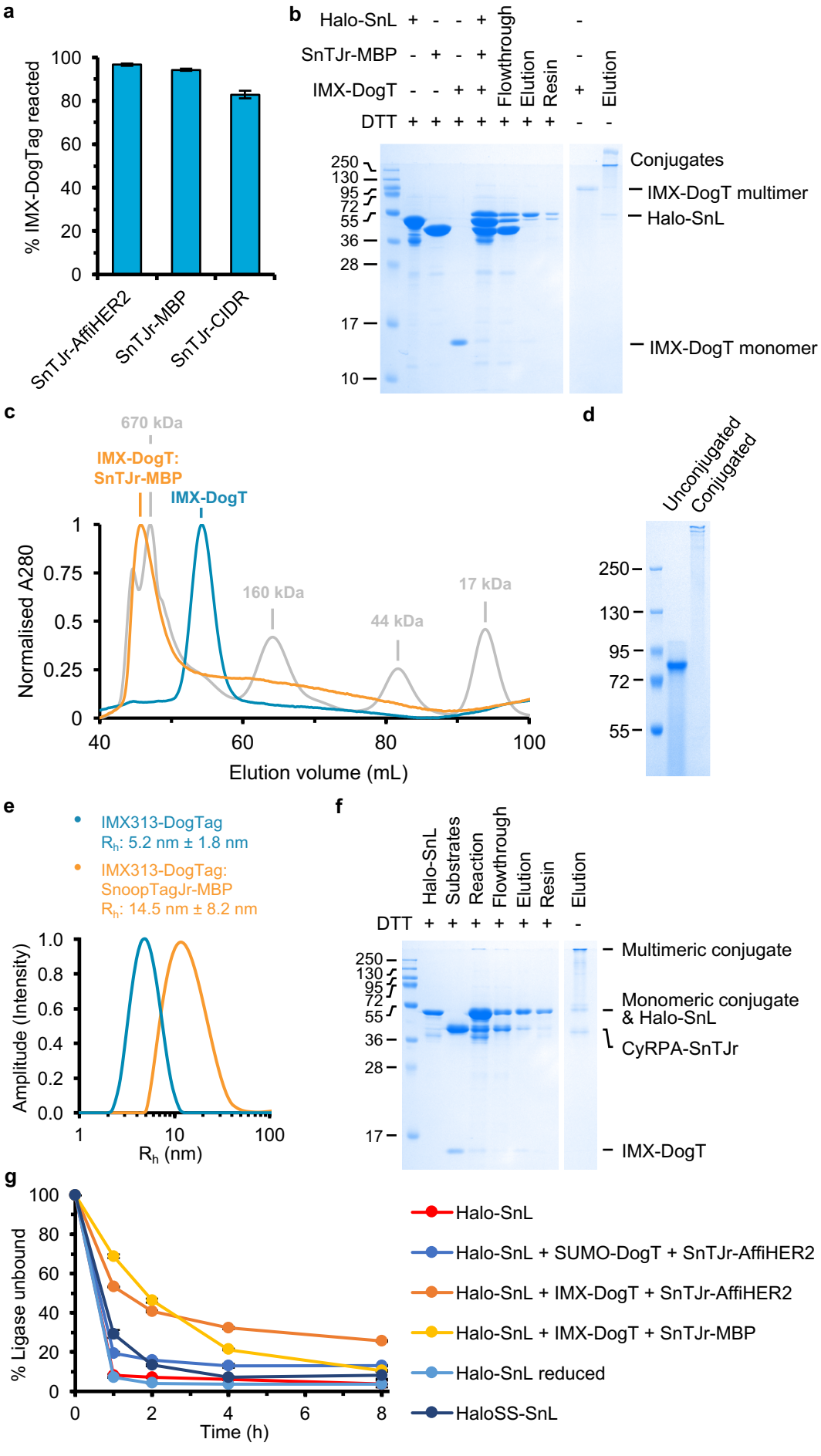

a
% IMX-DogTag reacted
100
80
60
40
20
0
SnTJr-AffiHER2
SnTJr-MBP
SnTJr-CIDR
b
Halo-SnL + - - + -
SnTJr-MBP - + - + -
IMX-DogT - - + + +
DTT + + + + + + + - -
Flowthrough
Elution
Resin
Elution
250
130
95
72
55
36
28
17
10
Conjugates
IMX-DogT multimer
Halo-SnL
IMX-DogT monomer
c
670 kDa
IMX-DogT:
SnTJr-MBP
IMX-DogT
160 kDa
44 kDa
17 kDa
Normalised A280
1
0.75
0.5
0.25
0
40
60
80
100
Elution volume (mL)
d
Unconjugated
Conjugated
250
130
95
72
55
e
IMX313-DogTag
Rh: 5.2 nm ± 1.8 nm
IMX313-DogTag:
SnoopTagJr-MBP
Rh: 14.5 nm ± 8.2 nm
Amplitude (Intensity)
1.0
0.8
0.6
0.4
0.2
0.0
1
10
100
Rh (nm)
f
Halo-SnL
Substrates
Reaction
Flowthrough
Elution
Resin
Elution
DTT + + + + + + -
250
130
95
72
55
36
28
17
Multimeric conjugate
Monomeric conjugate
& Halo-SnL
CyRPA-SnTJr
IMX-DogT
g
% Ligase unbound
100
80
60
40
20
0
0
2
4
6
8
Time (h)
Halo-SnL
Halo-SnL + SUMO-DogT + SnTJr-AffiHER2
Halo-SnL + IMX-DogT + SnTJr-AffiHER2
Halo-SnL + IMX-DogT + SnTJr-MBP
Halo-SnL reduced
HaloSS-SnL

Figure 5.11 Generation of protein-decorated IMX313 nanoparticles. a) SnoopTagJr-AffiHER2, SnoopTagJr-MBP or SnoopTagJr-CIDR (20 μM) were conjugated to IMX313-DogTag (10 μM) using HaloTag7-SnoopLigase (20 μM) in 50 mM Tris borate pH 7.25 + 15% (v/v) glycerol for 48 h at 4 °C. b) IMX313-DogTag (10 μM) and SnoopTagJr-MBP (20 μM) were conjugated using HaloTag7-SnoopLigase (20 μM) in 50 mM Tris borate pH 7.25 + 15% (v/v) glycerol for 48 h at 4 °C. SnoopLigase was immobilised on HaloLink resin for 12 h at 25 °C, unreacted substrate washed away, and reaction product eluted using 25 mM Tris phosphate + 3.5 M imidazole pH 7. c) Size exclusion chromatography (Superdex 200 16/600pg) of the eluted reaction product from b) or IMX313-DogTag compared to Mw standards. Mw standards were analysed by Anne-Marie Andersson in the Howarth laboratory. d) Peak fractions from the size exclusion chromatography in c) of unconjugated or SnoopTagJr-MBP-conjugated IMX313-DogTag analysed by SDS-PAGE. e) Hydrodynamic radius distribution of unconjugated or SnoopTagJr-MBP-conjugated IMX313-DogTag determined by dynamic light scattering. f) CyRPA-SnoopTagJr (produced by Anne-Marie Andersson in the Howarth laboratory) at 15 μM was conjugated to IMX313-DogTag (10 μM) using HaloTag7-SnoopLigase (20 μM) in 50 mM Tris borate pH 7.25 + 15% (v/v) glycerol for 48 h at 4 °C. g) HaloTag7-SnoopLigase, reaction samples with substrate proteins or the dual C61S C262S mutant “HaloTag7SS-SnoopLigase” were incubated for 12 h at 4 °C to allow substrate conversion where applicable and subsequently incubated with HaloLink resin. The “Halo-SnL reduced” sample contained 100 mM 2-Mercaptoethanol during resin binding. Supernatant samples were obtained at different time points to analyse the amount of Halo-SnoopLigase protein not bound to HaloLink resin. Samples of a), b), d), f) and g) were boiled, run on SDS-PAGE and visualised by Coomassie staining. Quantification of a) and g) by densitometry (mean ± 1 S.D., n=3).

The incomplete capture of HaloTag7-SnoopLigase is not a problem for purifying monomeric proteins, as unbound enzyme is simply removed during the resin washing steps. However, for use on multimeric proteins like IMX313-DogTag, the presence of minor amounts of unreactive HaloTag7-SnoopLigase is problematic. A single reactive enzyme could link the entire multimeric protein to the resin, but unreactive HaloTag7-SnoopLigase bound to the same particle would not be removed during the washing steps, resulting in co-elution with the IMX-particle.

HaloTag7 contains two Cys residues (Ke et al., 2016; Newman et al., 1999), which can form intermolecular disulfide bonds (Figure 5.10b). Although the Cys are not positioned directly at the active site, I hypothesised that dimerised or multimerised HaloTag7 could show compromised binding and prevent complete coupling of the purified protein to HaloLink resin. However, even in the presence of reducing agent detectable amounts of HaloTag7-SnoopLigase remained unbound after eight hours incubation. Similarly, mutation of Cys 61 and Cys 262 to Ser residues did not allow complete coupling (Figure 5.11g).

5.4. Discussion

In the previous chapter I enhanced the prototype peptide-peptide ligase using protein engineering, which yielded a widely applicable bioconjugation tool. In this chapter I used the newly developed SnoopLigase to generate a multimeric modular vaccine platform system with the potential to enhance the immune response against a wide range of antigens. The usage of merely peptide tags as opposed to proteins for conjugation minimizes off-target antigenicity. This is a desirable feature for vaccines, as irrelevant proteins might reduce the immune response against the antigen (Cohn et al., 1949; Ringe et al., 2017; Sutherland et al., 2001).

IMX313-DogTag could be produced at high yield, high purity and proved to be a highly robust protein, resistant to prolonged exposure to high temperatures and lyophilisation. SnoopLigase-mediated conjugation of proteins to IMX313-DogTag allowed nearly complete coupling, making it a promising approach to generate highly uniform vaccine products. Less efficient coupling (ca 85%) was observed with SnoopTagJr-CIDR which could be caused by structure-related steric

clashes. To stimulate a strong immune response it might not be necessary to conjugate all sites in the platform particle, as a dense array of only 2 – 3 antigens can already be sufficient to elicit a strong B-cell response (Klaus and Humphrey, 1976).

SnoopLigase strongly binds to its reaction product. To generate pure protein vaccines without contaminating enzyme, it was necessary to establish a procedure to remove the ligase from the product. The installation of an affinity handle on SnoopLigase and immobilisation on a solid-phase support promised convenient elimination of unreacted substrate and the enzyme in a single step. Enzymatic biotinylation of SnoopLigase and capture with streptavidin agarose resin was successful and allowed screening conditions to elute the reaction product. Most conditions tested did not disrupt the interaction between SnoopLigase and the reaction product. This shows the great stability of the re-constituted domain and could explain why SnoopLigase is product-inhibited.

Low pH (2.0), elevated temperature (55 °C), competition by SnoopLigase reaction product and 2 M imidazole allowed product elution. Low pH elution is commonly performed in immunoprecipitation to disrupt the interaction between antibody and target protein (Sousa et al., 2011). However, not all proteins can withstand the low pH and denature (Kishore et al., 2012). The same applies to high temperature (Volkin and Klibanov, 1992). Addition of competing SnoopLigase reaction product to the elution buffer decreased the required elution temperature. It would be interesting to study whether competitor based on SnoopTagJr and DogTag peptides could be more potent than the SUMO-DogTag and SnoopTag peptide conjugate employed here to even further decrease the required elution temperature. Higher competitor concentration could also have a desirable effect.

Although potentially the costliest option for SnoopLigase product elution, peptide competitor would be the least disruptive to the structure of the conjugated substrates.

Imidazole was chosen as main purification approach, as it is cheaper than peptides and can increase protein solubility (Hamilton et al., 2003; Huyet et al., 2011; Shi et al., 2007). The concentration of 2 M imidazole employed here is relatively high but is occasionally used to elute multimeric proteins from Ni-NTA resin (Leneghan et al., 2017).

The imidazole structure represents the side chain of His and can form interactions with aromatic or charged amino acids (Liao et al., 2013), potentially disrupting intra- or intermolecular interactions of such amino acids and thereby altering protein structure. Imidazole could also compete with interactions of His residues. His 850 in DogTag is positioned at the interface between DogTag, SnoopTagJr and SnoopLigase (Izoré et al., 2010), potentially sealing the hydrophobic core of the assembled protein (Figure 5.12). It would be interesting to see whether mutation of His850 would prevent the elution of reaction product by imidazole, which would indicate a key role of this amino acid in reaction product binding to SnoopLigase.

With low pH, high temperature, competition and imidazole, there are four possible ways to elute SnoopLigase reaction products; it might be necessary to choose an approach according to the substrate proteins employed. Glycine pH 2.0 followed by neutralisation, and elution by elevated temperature would not require dialysis after purification. Imidazole could potentially be better tolerated by many proteins, but could disrupt protein complexes and requires dialysis after elution.

Peptide competitor-based elution would be costly and requires dialysis upon elution, but would be the least disruptive option to substrate proteins.

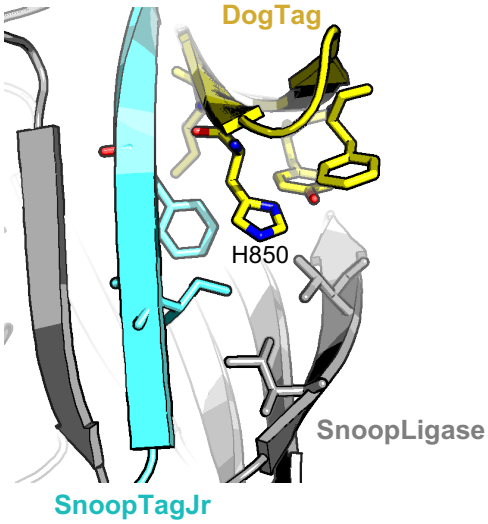


Figure 5.12 His850 is at the interface of DogTag, SnoopTagJr and SnoopLigase. Cartoon of DogTag (yellow), SnoopLigase (grey) and SnoopTagJr (cyan), Protein Data Bank 2WW8. His850 and surrounding residues displayed as sticks.

The initially used biotin-streptavidin interaction for SnoopLigase reaction product purification was replaced by the HaloTag system, with the aim to reduce SnoopLigase contaminations in elution samples. HaloTag was preferred over biotin-streptavidin, as it forms a covalent bond to its substrate and could be produced as genetic fusion, without the need of enzymatic biotin labelling. HaloTag-SnoopLigase mediated efficient conjugation of model proteins and the malaria antigens CIDR and CyRPA to IMX313-DogTag. Purification and analysis of IMX313-DogTag conjugates confirmed the multimeric nature of the particle after conjugation and the expected increase in particle size.

The work presented in this chapter enabled the development of a method for the generation of purified SnoopLigase reaction products, in particular IMX313-DogTag-conjugated multimeric vaccines. However, there is room for improvement.

HaloTag-SnoopLigase contaminations were detectable in purified IMX313-DogTag conjugates. Analysis of HaloTag-SnoopLigase binding to the HaloLink resin revealed that binding of the purified protein did not reach completion after up to 8 h. This could be due to misfolded HaloTag protein, unable to react with its substrate. As the binding problem does not seem to be related to intermolecular disulfide bonds between the Cys residues of HaloTag, it might be possible to overcome the problem by expression at lower temperature to prevent potential misfolding of the protein. Another possibility to reduce SnoopLigase contaminations in elution sample would be to test an alternative covalent interaction system, such as SpyTag-SpyCatcher (Zakeri et al., 2012).

6. Summary and General Discussion

Protein conjugation is an important tool for research and drug development (Jeger et al., 2010; McCluskey and Collier, 2013; Moyle, 2017; Schellenberger et al., 2009). However, limitations of existing coupling methods can include low specificity (Griffin et al., 2002), low affinity (Guimaraes et al., 2013), or restrictive reaction conditions (Fierer et al., 2014). Therefore, there was need to design a new tool for protein conjugation.

In this thesis, I have developed a novel peptide-peptide ligase termed “SnoopLigase”, which can covalently link the two peptide tags SnoopTagJr and DogTag. The system is functional under mild reaction conditions, allowing it to be used for a wide range of applications. SnoopLigase enabled the development of a highly stable modular multimeric vaccine platform with the potential to enhance the immunogenicity of recombinant protein antigens.

In Chapter 3 I generated a prototype ligase (RrgALigase) based on an isopeptide bond-containing domain of the *Streptococcus pneumoniae* protein RrgA. The initial version was only functional at low temperature, high pH and in the presence of high concentrations of the molecular chaperones TMAO and glycerol, and was not able to achieve efficient substrate conversion. Poor solubility of RrgALigase was overcome by fusion to the protein chaperone MBP. These findings indicated limited stability of the protein, potentially caused by the absence of three β-strands compared to the parent domain. As the characteristics of the prototype system were not desirable for most potential applications, there was need for improvement.

Optimising the termini of the modules allowed enhancement of the system’s reactivity. The improved RrgALigase version was elongated by six residues on the

N-terminus and nine residues on the C-terminus, potentially allowing the protein to fold to a tertiary structure resembling the parent domain more closely. +6RrgALigase+9 exhibited improved solubility and did not require fusion to MBP anymore. While MBP-fusion for solubilisation might not be a problem in many situations, the large extra protein could cause steric clashes when ligating bulky substrates or complicate the functionalisation of SnoopLigase with other proteins, such as HaloTag7 for covalent immobilisation (Ohana et al., 2009).

The results of the termini optimisation study could be relevant to the development of other split proteins, indicating that alteration of the split points can have major effects on performance and should be the first step of enhancing split proteins. Such studies can be carried out within less than a month and could strongly enhance downstream engineering efforts.

Despite exhibiting reactivity, the reaction characteristics of +6RrgALigase+9 developed in Chapter 3 prohibited broad utility of the tool. In Chapter 4 I enhanced the performance of the split protein system by protein engineering. Pro insertion into protein loops has yielded two reaction-enhancing point mutations and the homology- and Rosetta-based tool Protein Repair One Stop Shop (PROSS) identified three successful point mutations. Several PROSS-suggested mutations were not tested as they were suspected to have a strong and potentially disruptive effect to SnoopLigase structure and function. It would be interesting to analyse the effects of the excluded mutations and thereby assess the reliability of the tool. The PROSS has been developed to enhance protein solubility and stability (Goldenzweig et al., 2016). Similarly, Pro insertions into protein loops tend to improve protein stability (Fu et al., 2009), suggesting that improving parent domain

stability might improve the performance of other self-assembling split protein systems (Cabantous et al., 2005; Fierer et al., 2014; Veggiani et al., 2016).

The final effective reaction conditions were 4 – 25 °C, pH 6.5 – 8.5 and 0 – 15% (v/v) glycerol. Compared to the preferred reaction conditions of the prototype developed in Chapter 3, this represents a major improvement and broadens the applicability of the system. However, remaining disadvantages of the SnoopLigase system are the relatively slow reaction, requiring 24 – 48 h to plateau at 10 µM protein concentrations, and the considerable sensitivity to salts such as NaCl, which would influence the utility of SnoopLigase in extracellular environments.

In Chapter 5 I used SnoopLigase to generate a multimeric modular vaccine platform system with the potential to enhance the immune response against a wide range of antigens. Rigidly ordered arrays of antigens have proven to enhance the immune response compared to monomeric antigen (Bachmann et al., 1997; Zinkernagel, 2003). To improve the efficacy of protein vaccines, this concept was harnessed by fusing antigens to multimeric proteins (Cañizares et al., 2005; McComb et al., 2015; Ogun et al., 2008; Rutgers et al., 1988). However, generation of fusion constructs is often unsuccessful (Jain et al., 2015; Stockley and Mastico, 2000) which has led to the development of modular systems, in which the multimerisation platform and the antigen are produced separately and subsequently conjugated (Brune et al., 2016, 2017; Tang et al., 2016; Thrane et al., 2016). The usage of merely peptide tags as opposed to proteins for conjugation minimises off-target antigenicity. This is a desirable feature for vaccines, as irrelevant proteins might reduce the immune response against the antigen (Cohn et al., 1949; Ringe et al., 2017; Sutherland et al., 2001) and represents a major advantage to comparable Catcher/Tag-based modular systems (Brune et al., 2016, 2017; Thrane

et al., 2016). The mainly heptameric platform particle IMX313-DogTag could be produced at high yield and purity, showed high thermostability and could be lyophilised. Conjugation of model proteins and the promising malaria vaccine candidate CyRPA (Dreyer et al., 2012) to IMX313-DogTag were efficient, reaching 85 – 98% coupling efficiency.

In order to remove SnoopLigase from the reaction product, ligase immobilisation and product elution strategies were explored. Biotinylated SnoopLigase or HaloTag7-SnoopLigase could be immobilised on streptavidin agarose or HaloLink resin, respectively. Screening of conditions identified low pH, high temperature, product competition or imidazole as efficient elution strategies. SnoopLigase immobilisation and product elution allowed efficient purification of reaction product. However, SnoopLigase contaminations were detectable in purified IMX313-DogTag conjugates. This problem could probably be overcome by further adjustment of the protocol, for example by optimising HaloTag7-SnoopLigase expression and purification to remove non-functional protein, longer incubation of the resin with elution buffer to allow ejected SnoopLigase to bind to the resin, or testing an alternative immobilisation handle such as the SpyTag-SpyCatcher system (Zakeri et al., 2012).

Purification of model protein-conjugated IMX313-DogTag was achieved in Chapter 5. After adjusting the SnoopLigase product purification method to prevent enzyme contamination, it would be interesting to test the ability of antigen-decorated IMX313-DogTag at enhancing antibody generation in a mouse immunisation study. Given that a previous endeavour with C-terminally conjugated antigen on IMX313 had shown good results (Brune et al., 2017), it is likely that the SnoopLigase-conjugated vaccine is also able to increase the antibody titre. It would be interesting

to test the system with a range of different antigens to confirm its general utility as vaccine-enhancing platform system. Furthermore, an idea would be to test the reactivity of IMX313-(-10DogTag) to further decrease the size of the bioconjugation scar.

Apart from enhancing vaccination, multimeric proteins are important in signal transduction mediated by receptor cross-linking (Kiessling et al., 2006). IMX313-DogTag could serve as platform to multimerise receptor-stimulating molecules, for example targeting death receptor 5 for cancer cell killing (Allen et al., 2012; Chaudhary et al., 1997).

SnoopLigase is a useful alternative to existing site-specific conjugation methods. The previously developed SpyLigase is a peptide-peptide ligation system which used a similar development principle as SnoopLigase (Fierer et al., 2014). However, SpyLigase showed low reaction rate, low maximum substrate conversion and stringent reaction conditions. A hypothesis was that tripartite domain splitting could not generate broadly applicable bioconjugation tools. Conversely, the work presented in this thesis shows that tripartite split isopeptide bond-containing protein domains can generate powerful bioconjugation tools but may require extensive optimisation and engineering.

Sortase is one of the most commonly used bioconjugation systems, which exhibits high site-specificity, but requires high substrate concentration (up to millimolar) and still rarely achieves more than 90% substrate conversion (David Row et al., 2015; Guimaraes et al., 2013; Levary et al., 2011). In contrast, SnoopLigase has been tested at standard concentrations of only 10 μM of each protein component and addition of as low as 1.5-fold molar excess of one substrate allows

>95% reaction of the other substrate. Hence SnoopLigase can be useful to efficiently conjugate poorly soluble proteins which cannot be produced at concentrations required for sortase-mediated ligation. More recently, butelase and its relative OaAEP1 have been described which require lower substrate concentration (as low as 50 μM) and achieve higher substrate conversion than sortase (Nguyen et al., 2014, 2016a; Yang et al., 2017). The reactions of butelase and engineered sortase require only minutes (McCluskey and Collier, 2013), while SnoopLigase requires 24 – 48 h.

Sortase has first been used for bioconjugation in 2004 (Mao et al., 2004) and since then underwent extensive optimisation and engineering (Antos et al., 2016; Chen et al., 2011; Dorr et al., 2014; Schmohl et al., 2017). It is likely that further engineering of SnoopLigase through rational design or protein evolution could strongly enhance its reaction characteristics.

Compared to other bioconjugation strategies SnoopLigase can show higher site-specificity (transglutaminase) (Gundersen et al., 2014) or simpler reaction set-up (unnatural amino acids) (Xie and Schultz, 2005), making the tool easily accessible to non-specialised laboratories.

SnoopLigase is functional, yet improvement of various characteristics could make it a more broadly applicable system. These include the reaction rate, substrate turnover or reactivity in NaCl. Potential engineering approaches for further SnoopLigase enhancement include computational and directed protein evolution methods. Similar to the successfully employed PROSS, two similar tools have been developed recently (Musil et al., 2017; Schwarte et al., 2017). As the NewProt and FireProt servers use different computational approaches to identify point mutations,

it would be interesting to see whether these tools could further enhance SnoopLigase performance.

Directed evolution could be performed either on the peptide tags or the ligase protein. Phage display is an established and widely used evolution method that could be employed to identify improved peptide tags or ligase enzymes with high product affinity (Pande et al., 2010). Approaches that are more suitable for the identification of SnoopLigase mutants with substrate turnover include the dihydrofolate reductase (DHFR) protein-fragment complementation assay (Remy et al., 2007) or yeast endoplasmic reticulum sequestration screening (YESS) (Yi et al., 2013).

The SnoopLigase reaction product elution studies have given an insight into the product binding capability of the enzyme. Only relatively harsh conditions such as pH 2.0, 55 °C or 2 M imidazole enabled the disruption of the SnoopLigase–product complex, suggesting a high affinity. It would be interesting to measure the affinity, for example by isothermal titration calorimetry (Pierce et al., 1999) or surface plasmon resonance (Douzi, 2017). The high affinity for the reaction product probably explains why SnoopLigase does not turn over substrate (product inhibition). Introducing substrate turnover into the system would likely require a weaker interaction between SnoopLigase and the reaction product.

While not desirable for efficient bioconjugation measured by enzyme input, the high affinity between SnoopLigase and its reaction product has proven to be a useful feature for purification of reacted protein conjugates. *In vitro* bioconjugation is usually performed with molar excess of one of the reaction partners in order to maximise conjugation of the other reaction partner (Stephanopoulos and Francis, 2011). This requires purification of free substrate from product, which could be

technically challenging depending on the substrates. Because of its high affinity for the reaction product, SnoopLigase-mediated bioconjugation allows to use the same purification approach regardless of the substrates employed. While there is a wide range of site-specific bioconjugation tools available, SnoopLigase is the only system with an integrated reaction product purification feature.

Despite being useful as a general purification approach, the identified elution conditions limit the general applicability of the system, both *in vivo* and *in vitro*. Cells would not be able to survive the harsh conditions, and some proteins might denature. To allow SnoopLigase elution at milder conditions, switchable domains could be placed into SnoopLigase that allow elution upon change in temperature or pH (Choi et al., 2015; Megeed et al., 2006; Naik et al., 2001), addition of molecules such as rapamycin or Ca^{2+} (Baird et al., 1999; Dagliyan et al., 2013), or exposure to light (Wang et al., 2016). Homology studies of SnoopLigase have identified a region on the opposite face of the Tag-binding site with low sequence conservation. This region could be suitable for insertion of a switchable domain without disrupting the functionality of SnoopLigase.

Another approach to tune SnoopLigase elution to milder conditions could be strategically introducing histidine residues in the ligase protein, such that disruption of the SnoopLigase-product complex could be achieved at higher pH. A similar pH switch has been achieved by introducing histidine residues into *Staphylococcus aureus* protein A, to allow elution under milder condition (Tsukamoto et al., 2014; Watanabe et al., 2013).

Bibliography

van Aarle, P. (2006). Making avian influenza vaccines available, an industry point of view (IFAH). Dev Biol (Basel) *124*, 151–155.

Agten, S.M., Suylen, D.P.L., and Hackeng, T.M. (2016). Oxime catalysis by freezing. Bioconjug Chem *27*, 42–46.

Aida, Y., and Pabst, M.J. (1990). Removal of endotoxin from protein solutions by phase separation using Triton X-114. J Immunol Methods *132*, 191–195.

Alegre-Cebollada, J., Badilla, C.L., and Fernández, J.M. (2010). Isopeptide bonds block the mechanical extension of pili in pathogenic Streptococcus pyogenes. J Biol Chem *285*, 11235–11242.

Alfano, C., Sanfelice, D., Martin, S.R., Pastore, A., and Temussi, P.A. (2017). An optimized strategy to measure protein stability highlights differences between cold and hot unfolded states. Nat Commun *8*, 15428.

Allen, J.E., Ferrini, R., Dicker, D.T., Batzer, G., Chen, E., Oltean, D.I., Lin, B., Renshaw, M.W., Kretz-Rommel, A., and El-Deiry, W.S. (2012). Targeting TRAIL death receptor 4 with trivalent DR4 Atrimer complexes. Mol Cancer Ther *11*, 2087–2095.

Allison, S.L., Schalich, J., Stiasny, K., Mandl, C.W., Kunz, C., and Heinz, F.X. (1995). Oligomeric rearrangement of tick-borne encephalitis virus envelope proteins induced by an acidic pH. J Virol *69*, 695–700.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res *25*, 3389–3402.

Ambrogelly, A., Palioura, S., and Söll, D. (2007). Natural expansion of the genetic code. Nat Chem Biol *3*, 29–35.

Amer, B.R., Macdonald, R., Jacobitz, A.W., Liauw, B., and Clubb, R.T. (2016). Rapid addition of unlabeled silent solubility tags to proteins using a new substrate-fused sortase reagent. J Biomol NMR *64*, 197–205.

Antos, J.M., and Francis, M.B. (2004). Selective tryptophan modification with rhodium carbenoids in aqueous solution. J Am Chem Soc *126*, 10256–10257.

Antos, J.M., Truttmann, M.C., and Ploegh, H.L. (2016). Recent advances in sortase-catalyzed ligation methodology. Curr Opin Struct Biol *38*, 111–118.

Aon, J.C., Caimi, R.J., Taylor, A.H., Lu, Q., Oluboyede, F., Dally, J., Kessler, M.D., Kerrigan, J.J., Lewis, T.S., Wysocki, L.A., et al. (2008). Suppressing posttranslational gluconoylation of heterologous proteins by metabolic engineering of Escherichia coli. Appl Environ Microbiol *74*, 950–958.

Arakawa, T., and Timasheff, S.N. (1985). The stabilization of proteins by osmolytes. Biophys J *47*, 411–414.

Arakawa, T., and Tsumoto, K. (2003). The effects of arginine on refolding of aggregated proteins: not facilitate refolding, but suppress aggregation. Biochem Biophys Res Commun *304*, 148–152.

Aranko, A.S., Wlodawer, A., and Iwaï, H. (2014a). Nature's recipe for splitting inteins. Protein Eng Des Sel *27*, 263–271.

Aranko, A.S., Oeemig, J.S., Zhou, D., Kajander, T., Wlodawer, A., and Iwaï, H. (2014b). Structure-based engineering and comparison of novel split inteins for protein ligation. Mol Biosyst *10*, 1023–1034.

Asial, I., Cheng, Y.X., Engman, H., Dollhopf, M., Wu, B., Nordlund, P., and Cornvik, T. (2013). Engineering protein thermostability using a generic activity-independent biophysical screen inside the cell. Nat Commun *4*, 2901.

Asquith, R.S., Otterburn, M.S., and Sinclair, W.J. (1974). Isopeptide crosslinks--their occurrence and importance in protein structure. Angew Chem Int Ed Engl *13*, 514–520.

Bachmann, M.F., Kalinke, U., Althage, A., Freer, G., Burkhart, C., Roost, H., Aguet, M., Hengartner, H., and Zinkernagel, R.M. (1997). The role of antibody concentration and avidity in antiviral protection. Science *276*, 2024–2027.

Baird, G.S., Zacharias, D.A., and Tsien, R.Y. (1999). Circular permutation and receptor insertion within green fluorescent proteins. Proc Natl Acad Sci U S A *96*, 11241–11246.

Bale, J.B., Gonen, S., Liu, Y., Sheffler, W., Ellis, D., Thomas, C., Cascio, D., Yeates, T.O., Gonen, T., King, N.P., et al. (2016). Accurate design of megadalton-scale two-component icosahedral protein complexes. Science *353*, 389–394.

Bam, N.B., Cleland, J.L., Yang, J., Manning, M.C., Carpenter, J.F., Kelley, R.F., and Randolph‖, T.W. (1998). Tween protects recombinant human growth hormone against agitation-induced damage via hydrophobic interactions. J Pharm Sci *87*, 1554–1559.

Bárcena, J., and Blanco, E. (2013). Design of novel vaccines based on virus-like particles or chimeric virions. Subcell Biochem *68*, 631–665.

Barnett, G.V., Razinkov, V.I., Kerwin, B.A., Blake, S., Qi, W., Curtis, R.A., and Roberts, C.J. (2016). Osmolyte Effects on Monoclonal Antibody Stability and Concentration-Dependent Protein Interactions with Water and Common Osmolytes. J Phys Chem B *120*, 3318–3330.

Baschong, W., Hasler, L., Häner, M., Kistler, J., and Aebi, U. (2003). Repetitive versus monomeric antigen presentation: direct visualization of antibody affinity and specificity. J Struct Biol *143*, 258–262.

Baskakov, I., and Bolen, D.W. (1998). Forcing thermodynamically unfolded proteins to fold. J Biol Chem *273*, 4831–4834.

Batteiger, B., Newhall V, W.J., and Jones, R.B. (1982). The use of tween 20 as a blocking agent in the immunological detection of proteins transferred to nitrocellulose membranes. J Immunol Methods *55*, 297–307.

Behbehani, A.M. (1983). The smallpox story: life and death of an old disease. Microbiol Rev *47*, 455–509.

Bell, K.N., Hogue, C.J., Manning, C., and Kendal, A.P. (2001). Risk factors for improper vaccine storage and handling in private provider offices. Pediatrics *107*, E100.

Bhagwat, M., and Aravind, L. (2007). PSI-BLAST tutorial. Methods Mol Biol *395*, 177–186.

Bi, X., Yin, J., Nguyen, G.K.T., Rao, C., Halim, N.B.A., Hemu, X., Tam, J.P., and Liu, C.-F. (2017). Enzymatic engineering of live bacterial cell surfaces using butelase 1. Angew Chem Int Ed Engl *56*, 7822–7825.

Biragyn, A., Tani, K., Grimm, M.C., Weeks, S., and Kwak, L.W. (1999). Genetic fusion of chemokines to a self tumor antigen induces protective, T-cell dependent antitumor immunity. Nat Biotechnol *17*, 253–258.

Bourzac, K. (2014). Infectious disease: Beating the big three. Nature *507*, S4–S7.

Bradbury, S.L., and Jakoby, W.B. (1972). Glycerol as an enzyme-stabilizing agent: effects on aldehyde dehydrogenase. Proc Natl Acad Sci U S A *69*, 2373–2376.

Brinkmann, U., Buchner, J., and Pastan, I. (1992). Independent domain folding of Pseudomonas exotoxin and single-chain immunotoxins: influence of interdomain connections. Proc Natl Acad Sci U S A *89*, 3075–3079.

Brofelth, M., Städe, L.W., Ekstrand, A.I., Edfeldt, L.P., Kovačič, R., Nielsen, T.T., Larsen, K.L., Duroux, L., and Wingren, C. (2017). Site-specific photocoupling of pBpa mutated scFv antibodies for use in affinity proteomics. Biochim Biophys Acta *1865*, 985–996.

Brune, K.D. (2016). Engineering modular platforms for rapid vaccine development. Doctoral dissertation. University of Oxford.

Brune, K.D., Leneghan, D.B., Brian, I.J., Ishizuka, A.S., Bachmann, M.F., Draper, S.J., Biswas, S., and Howarth, M. (2016). Plug-and-Display: decoration of Virus-Like Particles via isopeptide bonds for modular immunization. Sci Rep *6*, 19234.

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. Bioconjug Chem *28*, 1544–1551.

Burrows, J.A., Willis, L.K., and Perlmutter, D.H. (2000). Chemical chaperones mediate increased secretion of mutant alpha 1-antitrypsin (alpha 1-AT) Z: A potential pharmacological strategy for prevention of liver injury and emphysema in alpha 1-AT deficiency. Proc Natl Acad Sci U S A *97*, 1796–1801.

Butt, T.R., Jonnalagadda, S., Monia, B.P., Sternberg, E.J., Marsh, J.A., Stadel, J.M., Ecker, D.J., and Crooke, S.T. (1989). Ubiquitin fusion augments the yield of cloned gene products in Escherichia coli. Proc Natl Acad Sci U S A *86*, 2540–2544.

Cabantous, S., Terwilliger, T.C., and Waldo, G.S. (2005). Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. Nat Biotechnol *23*, 102–107.

Campeotto, I., Goldenzweig, A., Davey, J., Barfod, L., Marshall, J.M., Silk, S.E., Wright, K.E., Draper, S.J., Higgins, M.K., and Fleishman, S.J. (2017). One-step design of a stable variant of the malaria invasion protein RH5 for use as a vaccine immunogen. Proc Natl Acad Sci U S A *114*, 998–1002.

Cañizares, M.C., Lomonossoff, G.P., and Nicholson, L. (2005). Development of cowpea mosaic virus-based vectors for the production of vaccines in plants. Expert Rev Vaccines *4*, 687–697.

Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki, N., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y. (1998). Thermostabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc Natl Acad Sci U S A *95*, 520–524.

Carpenter, J.F., Manning, M.C., and Randolph, T.W. (2002). Long-term storage of proteins. Curr Protoc Protein Sci *Chapter 4*, Unit 4.6.

Ceccarini, C., and Eagle, H. (1971). pH as a determinant of cellular growth and contact inhibition. Proc Natl Acad Sci U S A *68*, 229–233.

Chae, Y.K., Im, H., Zhao, Q., Doelling, J.H., Vierstra, R.D., and Markley, J.L. (2004). Prevention of aggregation after refolding by balanced stabilization-destabilization: production of the Arabidopsis thaliana protein APG8a (At4g21980) for NMR structure determination. Protein Expr Purif *34*, 280–283.

Chalker, J.M., Bernardes, G.J.L., Lin, Y.A., and Davis, B.G. (2009). Chemical modification of proteins at cysteine: opportunities in chemistry and biology. Chem Asian J *4*, 630–640.

Chaudhary, P.M., Eby, M., Jasmin, A., Bookwalter, A., Murray, J., and Hood, L. (1997). Death receptor 5, a new member of the TNFR family, and DR4 induce FADD-dependent apoptosis and activate the NF-kappaB pathway. Immunity *7*, 821–830.

Chen, D., and Kristensen, D. (2009). Opportunities and challenges of developing thermostable vaccines. Expert Rev Vaccines *8*, 547–557.

Chen, I., Dorr, B.M., and Liu, D.R. (2011). A general strategy for the evolution of bond-forming enzymes using yeast display. Proc Natl Acad Sci U S A *108*, 11399–11404.

Chen, X., Muthoosamy, K., Pfisterer, A., Neumann, B., and Weil, T. (2012). Site-selective lysine modification of native proteins and peptides via kinetically controlled labeling. Bioconjug Chem *23*, 500–508.

Chica, R.A. (2015). Protein engineering in the 21st century. Protein Sci *24*, 431–433.

Chirino, A.J., and Mire-Sluis, A. (2004). Characterizing biological products and assessing comparability following manufacturing changes. Nat Biotechnol *22*, 1383–1391.

Chivers, C.E., Crozat, E., Chu, C., Moy, V.T., Sherratt, D.J., and Howarth, M. (2010). A streptavidin variant with slower biotin dissociation and increased mechanostability. Nat Methods *7*, 391–393.

Choi, J.H., Laurent, A.H., Hilser, V.J., and Ostermeier, M. (2015). Design of protein switches based on an ensemble model of allostery. Nat Commun *6*, 6968.

Chong, S., Mersha, F.B., Comb, D.G., Scott, M.E., Landry, D., Vence, L.M., Perler, F.B., Benner, J., Kucera, R.B., Hirvonen, C.A., et al. (1997). Single-column purification of free recombinant proteins using a self-cleavable affinity tag derived from a protein splicing element. Gene *192*, 271–281.

Chou, P.Y., and Fasman, G.D. (1974). Conformational parameters for amino acids in helical, beta-sheet, and random coil regions calculated from proteins. Biochemistry *13*, 211–222.

Cobb, R.E., Sun, N., and Zhao, H. (2013). Directed evolution as a powerful synthetic biology tool. Methods *60*, 81–90.

Cohn, M., Wetter, L.R., and Deutsch, H.F. (1949). Immunological studies on egg white proteins; precipitation of chicken-ovalbumin and conalbumin by rabbit- and horse-antisera. J Immunol *61*, 283–296.

Coolbaugh, M.J., Shakalli Tang, M.J., and Wood, D.W. (2017). High-throughput purification of recombinant proteins using self-cleaving intein tags. Anal Biochem *516*, 65–74.

Cortez, J., Bonner, P.L.., and Griffin, M. (2004). Application of transglutaminases in the modification of wool textiles. Enzyme Microb Technol *34*, 64–72.

Dagliyan, O., Shirvanyants, D., Karginov, A.V., Ding, F., Fee, L., Chandrasekaran, S.N., Freisinger, C.M., Smolen, G.A., Huttenlocher, A., Hahn, K.M., et al. (2013). Rational design of a ligand-controlled protein conformational switch. Proc Natl Acad Sci U S A *110*, 6800–6804.

Dahl, F., Banér, J., Gullberg, M., Mendel-Hartvig, M., Landegren, U., and Nilsson, M. (2004). Circle-to-circle amplification for precise and sensitive DNA analysis. Proc Natl Acad Sci U S A *101*, 4548–4553.

David Row, R., Roark, T.J., Philip, M.C., Perkins, L.L., and Antos, J.M. (2015). Enhancing the efficiency of sortase-mediated ligations through nickel-peptide complex formation. Chem Commun (Camb) *51*, 12548–12551.

Davis, I.W., Leaver-Fay, A., Chen, V.B., Block, J.N., Kapral, G.J., Wang, X., Murray, L.W., Arendall, W.B., Snoeyink, J., Richardson, J.S., et al. (2007). MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res *35*, W375–W383.

Delany, I., Rappuoli, R., and De Gregorio, E. (2014). Vaccines for the 21st century. EMBO Mol Med *6*, 708–720.

Dorr, B.M., Ham, H.O., An, C., Chaikof, E.L., and Liu, D.R. (2014). Reprogramming the specificity of sortase enzymes. Proc Natl Acad Sci U S A *111*, 13343–13348.

Douzi, B. (2017). Protein-Protein Interactions: Surface Plasmon Resonance. Methods Mol Biol *1615*, 257–275.

Dreyer, A.M., Matile, H., Papastogiannidis, P., Kamber, J., Favuzza, P., Voss, T.S., Wittlin, S., and Pluschke, G. (2012). Passive immunoprotection of Plasmodium falciparum-infected mice designates the CyRPA as candidate malaria vaccine antigen. J Immunol *188*, 6225–6237.

Dulbecco, R., and Vogt, M. (1954). Plaque formation and isolation of pure lines with poliomyelitis viruses. J Exp Med *99*, 167–182.

Duplay, P., Bedouelle, H., Fowler, A., Zabin, I., Saurin, W., and Hofnung, M. (1984). Sequences of the malE gene and of its product, the maltose-binding protein of Escherichia coli K12. J Biol Chem *259*, 10606–10613.

Edgar, R.C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics *5*, 113.

Egan, J.E., Weber, J.L., Ballou, W.R., Hollingdale, M.R., Majarian, W.R., Gordon, D.M., Maloy, W.L., Hoffman, S.L., Wirtz, R.A., and Schneider, I. (1987). Efficacy of murine malaria sporozoite vaccines: implications for human vaccine development. Science *236*, 453–456.

Engel, J., and Kammerer, R.A. (2000). What are oligomerization domains good for? Matrix Biol *19*, 283–288.

England, C.G., Luo, H., and Cai, W. (2015). HaloTag technology: a versatile platform for biomedical applications. Bioconjug Chem *26*, 975–986.

Fágáin, C.Ó. (1995). Understanding and increasing protein stability. Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology *1252*, 1–14.

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

Farias, S.E., Strop, P., Delaria, K., Galindo Casas, M., Dorywalska, M., Shelton, D.L., Pons, J., and Rajpal, A. (2014). Mass spectrometric characterization of transglutaminase based site-specific antibody-drug conjugates. Bioconjug Chem *25*, 240–250.

Feikin, D.R., Flannery, B., Hamel, M.J., Stack, M., and Hansen, P.M. (2016). Vaccines for Children in Low- and Middle-Income Countries. In Reproductive, Maternal, Newborn, and Child Health: Disease Control Priorities, Third Edition (volume 2), R.E. Black, R. Laxminarayan, M. Temmerman, and N. Walker, eds. (Washington (DC): The International Bank for Reconstruction and Development / The World Bank),.

Fierer, J.O. (2013). Development of spontaneous isopeptide bond formation for ligation of peptide tags. Doctoral dissertation. University of Oxford.

Fierer, J.O., Veggiani, G., and Howarth, M. (2014). SpyLigase peptide-peptide ligation polymerizes affibodies to enhance magnetic cancer cell capture. Proc Natl Acad Sci U S A *111*, E1176–E1181.

De Filette, M., Martens, W., Roose, K., Deroo, T., Vervalle, F., Bentahir, M., Vandekerckhove, J., Fiers, W., and Saelens, X. (2008). An influenza A vaccine based on tetrameric ectodomain of matrix protein 2. J Biol Chem *283*, 11382–11387.

Folk, J.E., Park, M.H., Chung, S.I., Schrode, J., Lester, E.P., and Cooper, H.L. (1980). Polyamines as physiological substrates for transglutaminases. J Biol Chem *255*, 3695–3700.

Fontana, A., Spolaore, B., Mero, A., and Veronese, F.M. (2008). Site-specific modification and PEGylation of pharmaceutical proteins mediated by transglutaminase. Adv Drug Deliv Rev *60*, 13–28.

Forsgren, N., Lamont, R.J., and Persson, K. (2010). Two intramolecular isopeptide bonds are identified in the crystal structure of the Streptococcus gordonii SspB C-terminal domain. J Mol Biol *397*, 740–751.

Freiburger, L., Sonntag, M., Hennig, J., Li, J., Zou, P., and Sattler, M. (2015). Efficient segmental isotope labeling of multi-domain proteins using Sortase A. J Biomol NMR *63*, 1–8.

Fritz, B.R., Jamil, O.K., and Jewett, M.C. (2015). Implications of macromolecular crowding and reducing conditions for in vitro ribosome construction. Nucleic Acids Res *43*, 4774–4784.

Fu, H., Grimsley, G.R., Razvi, A., Scholtz, J.M., and Pace, C.N. (2009). Increasing protein stability by improving beta-turns. Proteins *77*, 491–498.

Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics *28*, 3150–3152.

Gaberc-Porekar, V., and Menart, V. (2001). Perspectives of immobilized-metal affinity chromatography. J Biochem Biophys Methods *49*, 335–360.

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. In The Proteomics Protocols Handbook, J.M. Walker, ed. (Totowa, NJ: Humana Press), pp. 571–607.

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

Giepmans, B.N.G., Adams, S.R., Ellisman, M.H., and Tsien, R.Y. (2006). The fluorescent toolbox for assessing protein location and function. Science *312*, 217–224.

Gilbert, C., Howarth, M., Harwood, C.R., and Ellis, T. (2017). Extracellular Self-Assembly of Functional and Tunable Protein Conjugates from Bacillus subtilis. ACS Synth Biol *6*, 957–967.

Giuliani, M.M., Adu-Bobie, J., Comanducci, M., Aricò, B., Savino, S., Santini, L., Brunelli, B., Bambini, S., Biolchi, A., Capecchi, B., et al. (2006). A universal vaccine for serogroup B meningococcus. Proc Natl Acad Sci U S A *103*, 10834–10839.

Glazer, A.N. (1970). Specific chemical modification of proteins. Annu Rev Biochem *39*, 101–130.

Goldenzweig, A., Goldsmith, M., Hill, S.E., Gertman, O., Laurino, P., Ashani, Y., Dym, O., Unger, T., Albeck, S., Prilusky, J., et al. (2016). Automated Structure- and Sequence-

Based Design of Proteins for High Bacterial Expression and Stability. Mol Cell *63*, 337–346.

Grant, S.K., Deckman, I.C., Culp, J.S., Minnich, M.D., Brooks, I.S., Hensley, P., Debouck, C., and Meek, T.D. (1992). Use of protein unfolding studies to determine the conformational and dimeric stabilities of HIV-1 and SIV proteases. Biochemistry *31*, 9491–9501.

Greenwood, B. (2014). The contribution of vaccination to global health: past, present and future. Philos Trans R Soc Lond, B, Biol Sci *369*, 20130433.

Gregory, J.A., Topol, A.B., Doerner, D.Z., and Mayfield, S. (2013). Alga-produced cholera toxin-Pfs25 fusion proteins as oral vaccines. Appl Environ Microbiol *79*, 3917–3925.

Griffin, M., Casadio, R., and Bergamini, C.M. (2002). Transglutaminases: nature's biological glues. Biochem J *368*, 377–396.

Groenen, P.J., Smulders, R.H., Peters, R.F., Grootjans, J.J., van den Ijssel, P.R., Bloemendal, H., and de Jong, W.W. (1994). The amine-donor substrate specificity of tissue-type transglutaminase. Influence of amino acid residues flanking the amine-donor lysine residue. Eur J Biochem *220*, 795–799.

Guimaraes, C.P., Witte, M.D., Theile, C.S., Bozkurt, G., Kundrat, L., Blom, A.E.M., and Ploegh, H.L. (2013). Site-specific C-terminal and internal loop labeling of proteins using sortase-mediated reactions. Nat Protoc *8*, 1787–1799.

Gundersen, M.T., Keillor, J.W., and Pelletier, J.N. (2014). Microbial transglutaminase displays broad acyl-acceptor substrate specificity. Appl Microbiol Biotechnol *98*, 219–230.

Guthridge, S.L., and Miller, N.C. (1996). Cold chain in a hot climate. Aust N Z J Public Health *20*, 657–660.

Hajj Hussein, I., Chams, N., Chams, S., Sayegh, S. El, Badran, R., Raad, M., Gerges-Geagea, A., Leone, A., and Jurjus, A. (2015). Vaccines through centuries: major cornerstones of global health. Front Public Health *3*, 269.

Hallam, T.J., Wold, E., Wahl, A., and Smider, V.V. (2015). Antibody conjugates with unnatural amino acids. Mol Pharm *12*, 1848–1862.

Hamilton, S., Odili, J., Pacifico, M.D., Wilson, G.D., and Kupsch, J.-M. (2003). Effect of imidazole on the solubility of a his-tagged antibody fragment. Hybrid Hybridomics *22*, 347–355.

Hanley, K.A. (2011). The double-edged sword: How evolution can make or break a live-attenuated virus vaccine. Evolution *4*, 635–643.

Harris, K.S., Durek, T., Kaas, Q., Poth, A.G., Gilding, E.K., Conlan, B.F., Saska, I., Daly, N.L., van der Weerden, N.L., Craik, D.J., et al. (2015). Efficient backbone cyclization of linear peptides by a recombinant asparaginyl endopeptidase. Nat Commun *6*, 10199.

Hata, T., and Nakayama, M. (2007). Rapid single-tube method for small-scale affinity purification of polyclonal antibodies using HaloTag Technology. J Biochem Biophys Methods *70*, 679–682.

Heck, T., Pham, P.-H., Yerlikaya, A., Th?ny-Meyer, L., and Richter, M. (2014). Sortase A catalyzed reaction pathways: a comparative study with six SrtA variants. Catal. Sci. Technol. *4*, 2946–2956.

Heinz, F.X., Allison, S.L., Stiasny, K., Schalich, J., Holzmann, H., Mandl, C.W., and Kunz, C. (1995). Recombinant and virion-derived soluble and particulate immunogens for vaccination against tick-borne encephalitis. Vaccine *13*, 1636–1642.

Heiss, A., DuChesne, A., Denecke, B., Grötzinger, J., Yamamoto, K., Renné, T., and Jahnen-Dechent, W. (2003). Structural basis of calcification inhibition by alpha 2-HS glycoprotein/fetuin-A. Formation of colloidal calciprotein particles. J Biol Chem *278*, 13333–13341.

Hemu, X., Qiu, Y., Nguyen, G.K.T., and Tam, J.P. (2016). Total synthesis of circular bacteriocins by butelase 1. J Am Chem Soc *138*, 6968–6971.

Hirakawa, H., Ishikawa, S., and Nagamune, T. (2012). Design of Ca2+-independent Staphylococcus aureus sortase A mutants. Biotechnol Bioeng *109*, 2955–2961.

Hjertén, S., and Johansson, K.-E. (1972). Selective solubilization with tween 20 of membrane proteins from Acholeplasma laidlawii. Biochimica et Biophysica Acta (BBA) - Biomembranes *288*, 312–325.

Hofmeyer, T., Schmelz, S., Degiacomi, M.T., Dal Peraro, M., Daneschdar, M., Scrima, A., van den Heuvel, J., Heinz, D.W., and Kolmar, H. (2013). Arranged sevenfold: structural insights into the C-terminal oligomerization domain of human C4b-binding protein. J Mol Biol *425*, 1302–1317.

Hsia, Y., Bale, J.B., Gonen, S., Shi, D., Sheffler, W., Fong, K.K., Nattermann, U., Xu, C., Huang, P.-S., Ravichandran, R., et al. (2016). Design of a hyperstable 60-subunit protein dodecahedron. [corrected]. Nature *535*, 136–139.

Hu, Q.-Y., Berti, F., and Adamo, R. (2016). Towards the next generation of biomedicines by site-selective conjugation. Chem Soc Rev *45*, 1691–1719.

Hu, X., Hu, H., Melvin, J.A., Clancy, K.W., McCafferty, D.G., and Yang, W. (2011). Autocatalytic intramolecular isopeptide bond formation in gram-positive bacterial pili: a QM/MM simulation. J Am Chem Soc *133*, 478–485.

Huang, P., Caughey, W.S., and Dong, A. (1995). Effects of dimethyl sulfoxide, glycerol, and ethylene glycol on secondary structures of cytochrome c and lysozyme as observed by infrared spectroscopy. J Pharm Sci *84*, 387–392.

Huet, H.A., Growney, J.D., Johnson, J.A., Li, J., Bilic, S., Ostrom, L., Zafari, M., Kowal, C., Yang, G., Royo, A., et al. (2014). Multivalent nanobodies targeting death receptor 5 elicit superior tumor cell killing through efficient caspase induction. MAbs *6*, 1560–1570.

Huyet, J., Gilbert, M., Popoff, M.R., and Basak, A. (2011). Crystallization and preliminary X-ray diffraction studies of delta-toxin from Clostridium perfringens. Acta Crystallogr Sect F Struct Biol Cryst Commun *67*, 369–371.

Ito, A., Mase, A., Takizawa, Y., Shinkai, M., Honda, H., Hata, K.-I., Ueda, M., and Kobayashi, T. (2003). Transglutaminase-mediated gelatin matrices incorporating cell adhesion factors as a biomaterial for tissue engineering. J Biosci Bioeng *95*, 196–199.

Izoré, T., Contreras-Martel, C., Mortaji, L. El, Manzano, C., Terrasse, R., Vernet, T., Di Guilmi, A.M., and Dessen, A. (2010). Structural basis of host cell recognition by the pilus adhesin from Streptococcus pneumoniae. Structure *18*, 106–115.

Jacobson, R.M. (2003). Vaccine safety. Immunol Allergy Clin North Am *23*, 589–603.

Jain, N.K., and Roy, I. (2010). Trehalose and protein stability. Curr Protoc Protein Sci *Chapter 4*, Unit 4.9.

Jain, N.K., Sahni, N., Kumru, O.S., Joshi, S.B., Volkin, D.B., and Russell Middaugh, C. (2015). Formulation and stabilization of recombinant protein based virus-like particle vaccines. Adv Drug Deliv Rev *93*, 42–55.

Jeger, S., Zimmermann, K., Blanc, A., Grünberg, J., Honer, M., Hunziker, P., Struthers, H., and Schibli, R. (2010). Site-specific and stoichiometric modification of antibodies by bacterial transglutaminase. Angew Chem Int Ed Engl *49*, 9995–9997.

Jeong, H., and Seong, B.L. (2017). Exploiting virus-like particles as innovative vaccines against emerging viral infections. J Microbiol *55*, 220–230.

Jezek, J., Chen, D., Watson, L., Crawford, J., Perkins, S., Tyagi, A., and Jones-Braun, L. (2009). A heat-stable hepatitis B vaccine formulation. Hum Vaccin *5*, 529–535.

Johnson, E.S., Schwienhorst, I., Dohmen, R.J., and Blobel, G. (1997). The ubiquitin-like protein Smt3p is activated for conjugation to other proteins by an Aos1p/Uba2p heterodimer. EMBO J *16*, 5509–5519.

Johnsson, N., and Varshavsky, A. (1994). Split ubiquitin as a sensor of protein interactions in vivo. Proc Natl Acad Sci U S A *91*, 10340–10344.

Joshi, N.S., Whitaker, L.R., and Francis, M.B. (2004). A three-component Mannich-type reaction for selective tyrosine bioconjugation. J Am Chem Soc *126*, 15942–15943.

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

Kaether, C., and Gerdes, H.H. (1995). Visualization of protein transport along the secretory pathway using green fluorescent protein. FEBS Lett *369*, 267–271.

Kainosho, M., Torizawa, T., Iwashita, Y., Terauchi, T., Mei Ono, A., and Güntert, P. (2006). Optimal isotope labelling for NMR protein structure determinations. Nature *440*, 52–57.

Kalia, J., and Raines, R.T. (2010). Advances in Bioconjugation. Current Organic Chemistry *14*, 138–147.

Kane, P.M., Yamashiro, C.T., Wolczyk, D.F., Neff, N., Goebl, M., and Stevens, T.H. (1990). Protein splicing converts the yeast TFP1 gene product to the 69-kD subunit of the vacuolar H(+)-adenosine triphosphatase. Science *250*, 651–657.

Kang, H.J., and Baker, E.N. (2009). Intramolecular isopeptide bonds give thermodynamic and proteolytic stability to the major pilin protein of Streptococcus pyogenes. J Biol Chem *284*, 20729–20737.

Kang, H.J., and Baker, E.N. (2011). Intramolecular isopeptide bonds: protein crosslinks built for stress? Trends Biochem Sci *36*, 229–237.

Kang, H.J., Coulibaly, F., Clow, F., Proft, T., and Baker, E.N. (2007). Stabilizing isopeptide bonds revealed in gram-positive bacterial pilus structure. Science *318*, 1625–1628.

Kanno, A., Yamanaka, Y., Hirano, H., Umezawa, Y., and Ozawa, T. (2007). Cyclic luciferase for real-time sensing of caspase-3 activities in living mammals. Angew Chem Int Ed Engl *46*, 7595–7599.

Kapust, R.B., and Waugh, D.S. (1999). Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. Protein Sci *8*, 1668–1674.

Kartoglu, U., and Milstien, J. (2014). Tools and approaches to ensure quality of vaccines throughout the cold chain. Expert Rev Vaccines *13*, 843–854.

Kay, L.E., Ikura, M., Tschudin, R., and Bax, A. (1990). Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. Journal of Magnetic Resonance (1969) *89*, 496–514.

Ke, N., Landgraf, D., Paulsson, J., and Berkmen, M. (2016). Visualization of Periplasmic and Cytoplasmic Proteins with a Self-Labeling Protein Tag. J Bacteriol *198*, 1035–1043.

Kennedy, P.J., Oliveira, C., Granja, P.L., and Sarmento, B. (2017). Antibodies and associates: Partners in targeted drug delivery. Pharmacol Ther.

Kieliszek, M., and Misiewicz, A. (2014). Microbial transglutaminase and its application in the food industry. A review. Folia Microbiol (Praha) *59*, 241–250.

Kiessling, L.L., Gestwicki, J.E., and Strong, L.E. (2006). Synthetic multivalent ligands as probes of signal transduction. Angew Chem Int Ed Engl *45*, 2348–2368.

Kim, J., Kong, Y.P., Niedzielski, S.M., Singh, R.K., Putnam, A.J., and Shikanov, A. (2016). Characterization of the crosslinking kinetics of multi-arm poly(ethylene glycol) hydrogels formed via Michael-type addition. Soft Matter *12*, 2076–2085.

Kishore, D., Kundu, S., and Kayastha, A.M. (2012). Thermal, chemical and pH induced denaturation of a multimeric β-galactosidase reveals multiple unfolding pathways. PLoS ONE *7*, e50380.

Klaus, G.G., and Humphrey, J.H. (1976). B cell tolerance induced by polymeric antigens. I. Comparison of the dose and epitope density requirements for inactivation of primed and unprimed B cells in vivo. Eur J Immunol *5*, 361–365.

Kleppe, R., Uhlemann, K., Knappskog, P.M., and Haavik, J. (1999). Urea-induced denaturation of human phenylalanine hydroxylase. J Biol Chem *274*, 33251–33258.

Kloet, S.L., Baymaz, H.I., Makowski, M., Groenewold, V., Jansen, P.W.T.C., Berendsen, M., Niazi, H., Kops, G.J., and Vermeulen, M. (2015). Towards elucidating the stability, dynamics and architecture of the nucleosome remodeling and deacetylase complex by using quantitative interaction proteomics. FEBS J *282*, 1774–1785.

Kresheck, G.C., Schneider, H., and Scheraga, H.A. (1965). The Effect of D_2 O on the Thermal Stability of Proteins. Thermodynamic Parameters for the Transfer of Model Compounds from H_2 O to D_2 $O^{1,2}$. J Phys Chem *69*, 3132–3144.

Kreutz, M., Giquel, B., Hu, Q., Abuknesha, R., Uematsu, S., Akira, S., Nestle, F.O., and Diebold, S.S. (2012). Antibody-antigen-adjuvant conjugates enable co-delivery of antigen and adjuvant to dendritic cells in cis but only have partial targeting specificity. PLoS ONE *7*, e40208.

Ku, T., Lu, P., Chan, C., Wang, T., Lai, S., Lyu, P., and Hsiao, N. (2009). Predicting melting temperature directly from protein sequences. Comput Biol Chem *33*, 445–450.

Kuramoto, K., Yamasaki, R., Shimizu, Y., Tatsukawa, H., and Hitomi, K. (2013). Phage-displayed peptide library screening for preferred human substrate peptide sequences for transglutaminase 7. Arch Biochem Biophys *537*, 138–143.

Kushnir, N., Streatfield, S.J., and Yusibov, V. (2012). Virus-like particles as a highly efficient vaccine platform: diversity of targets and production systems and advances in clinical development. Vaccine *31*, 58–83.

Kwon, H., Young, P.G., Squire, C.J., and Baker, E.N. (2017). Engineering a Lys-Asn isopeptide bond into an immunoglobulin-like protein domain enhances its stability. Sci Rep *7*, 42753.

Laitinen, O.H., Hytönen, V.P., Nordlund, H.R., and Kulomaa, M.S. (2006). Genetically engineered avidins and streptavidins. Cell Mol Life Sci *63*, 2992–3017.

Lambert, S.B., Chuk, L.R., Nissen, M.D., Nolan, T.M., McVernon, J., Booy, R., Heron, L., Richmond, P.C., Walls, T., Marshall, H.S., et al. (2013). Safety and tolerability of a 2009

trivalent inactivated split-virion influenza vaccine in infants, children and adolescents. Influenza Other Respi Viruses *7*, 676–685.

Larsen, M.T., Kuhlmann, M., Hvam, M.L., and Howard, K.A. (2016). Albumin-based drug delivery: harnessing nature to cure disease. Molecular and Cellular Therapies *4*, 3.

Lau, C.K.Y., Turner, L., Jespersen, J.S., Lowe, E.D., Petersen, B., Wang, C.W., Petersen, J.E.V., Lusingu, J., Theander, T.G., Lavstsen, T., et al. (2015). Structural conservation despite huge sequence diversity allows EPCR binding by the PfEMP1 family implicated in severe childhood malaria. Cell Host Microbe *17*, 118–129.

Lauer, S.A., and Nolan, J.P. (2002). Development and characterization of Ni-NTA-bearing microspheres. Cytometry *48*, 136–145.

Leemhuis, H., Kelly, R.M., and Dijkhuizen, L. (2009). Directed evolution of enzymes: Library screening strategies. IUBMB Life *61*, 222–228.

Leibly, D.J., Nguyen, T.N., Kao, L.T., Hewitt, S.N., Barrett, L.K., and Van Voorhis, W.C. (2012). Stabilizing additives added during cell lysis aid in the solubilization of recombinant proteins. PLoS ONE *7*, e52482.

LeMieux, J., Hava, D.L., Basset, A., and Camilli, A. (2006). RrgA and RrgB are components of a multisubunit pilus encoded by the Streptococcus pneumoniae rlrA pathogenicity islet. Infect Immun *74*, 2453–2456.

Leneghan, D.B., Miura, K., Taylor, I.J., Li, Y., Jin, J., Brune, K.D., Bachmann, M.F., Howarth, M., Long, C.A., and Biswas, S. (2017). Nanoassembly routes stimulate conflicting antibody quantity and quality for transmission-blocking malaria vaccines. Sci Rep *7*, 3811.

Lennard, K.R., and Tavassoli, A. (2014). Peptides come round: using SICLOPPS libraries for early stage drug discovery. Chemistry (Easton) *20*, 10608–10614.

Leroux-Roels, G. (2015). Old and new adjuvants for hepatitis B vaccines. Med Microbiol Immunol *204*, 69–78.

Lesaicherre, M.-L., Lue, R.Y.P., Chen, G.Y.J., Zhu, Q., and Yao, S.Q. (2002). Intein-mediated biotinylation of proteins and its application in a protein microarray. J Am Chem Soc *124*, 8768–8769.

Levary, D.A., Parthasarathy, R., Boder, E.T., and Ackerman, M.E. (2011). Protein-protein fusion catalyzed by sortase A. PLoS ONE *6*, e18342.

Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics *22*, 1658–1659.

Li, Y., Leneghan, D.B., Miura, K., Nikolaeva, D., Brian, I.J., Dicks, M.D.J., Fyfe, A.J., Zakutansky, S.E., de Cassan, S., Long, C.A., et al. (2016). Enhancing immunogenicity and transmission-blocking activity of malaria vaccines by fusing Pfs25 to IMX313 multimerization technology. Sci Rep *6*, 18848.

Liao, S.-M., Du, Q.-S., Meng, J.-Z., Pang, Z.-W., and Huang, R.-B. (2013). The multiple roles of histidine in protein interactions. Chem Cent J *7*, 44.

Liljeroos, L., Malito, E., Ferlenghi, I., and Bottomley, M.J. (2015). Structural and computational biology in the design of immunogenic vaccine antigens. J Immunol Res *2015*, 156241.

Lim, W.K., Rösgen, J., and Englander, S.W. (2009). Urea, but not guanidinium, destabilizes proteins by forming hydrogen bonds to the peptide group. Proc Natl Acad Sci U S A *106*, 2595–2600.

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. J Am Chem Soc *128*, 4542–4543.

Lin, S.-H., and Guidotti, G. (2009). Chapter 35 Purification of Membrane Proteins. In Guide to Protein Purification, 2nd Edition, (Elsevier), pp. 619–629.

Lin, Y., Li, M., Song, H., Xu, L., Meng, Q., and Liu, X.-Q. (2013). Protein trans-splicing of multiple atypical split inteins engineered from natural inteins. PLoS ONE *8*, e59516.

Liu, Z., Zhou, H., Wang, W., Tan, W., Fu, Y.-X., and Zhu, M. (2014). A novel method for synthetic vaccine construction based on protein assembly. Sci Rep *4*, 7266.

Lutz, S. (2010). Beyond directed evolution--semi-rational protein engineering and design. Curr Opin Biotechnol *21*, 734–743.

Machida, S., Ogawa, S., Xiaohua, S., Takaha, T., Fujii, K., and Hayashi, K. (2000). Cycloamylose as an efficient artificial chaperone for protein refolding. FEBS Lett *486*, 131–135.

Makhatadze, G.I., Clore, G.M., and Gronenborn, A.M. (1995). Solvent isotope effect and protein stability. Nat Struct Biol *2*, 852–855.

Manolova, V., Flace, A., Bauer, M., Schwarz, K., Saudan, P., and Bachmann, M.F. (2008). Nanoparticles target distinct dendritic cell populations according to their size. Eur J Immunol *38*, 1404–1413.

Mao, H., Hart, S.A., Schink, A., and Pollok, B.A. (2004). Sortase-mediated protein ligation: a new method for protein engineering. J Am Chem Soc *126*, 2670–2671.

Marion, D. (2013). An introduction to biological NMR spectroscopy. Mol Cell Proteomics *12*, 3006–3025.

Markov, I.L. (2014). Limits on fundamental limits to computation. Nature *512*, 147–154.

Marmion, B.P. (1980). Prospects for new viral vaccines. Philos Trans R Soc Lond, B, Biol Sci *290*, 395–407.

Mashino, T., and Fridovich, I. (1987). Effects of urea and trimethylamine-N-oxide on enzyme activity and stability. Arch Biochem Biophys *258*, 356–360.

Masters, S.C. (2004). Co-Immunoprecipitation from Transfected Cells. In Protein-Protein Interactions: Methods and Applications, H. Fu, ed. (Springer Science & Business Media), pp. 337–350.

Matejtschuk, P. (2007). Lyophilization of proteins. Methods Mol Biol *368*, 59–72.

Matsumoto, T., Furuta, K., Tanaka, T., and Kondo, A. (2016). Sortase A-Mediated Metabolic Enzyme Ligation in Escherichia coli. ACS Synth Biol *5*, 1284–1289.

Matthiessen, L., Lång, H., Klimathianaki, M., Hanrahan, F., Kerstiëns, B., Martini, A., and Draghia-Akli, R. (2017). European strategy for vaccine development against infectious diseases. Vaccine *35 Suppl 1*, A20–A23.

McCluskey, A.J., and Collier, R.J. (2013). Receptor-directed chimeric toxins created by sortase-mediated protein fusion. Mol Cancer Ther *12*, 2273–2281.

McComb, R.C., and Martchenko, M. (2016). Neutralizing antibody and functional mapping of Bacillus anthracis protective antigen-The first step toward a rationally designed anthrax vaccine. Vaccine *34*, 13–19.

McComb, R.C., Ho, C.-L., Bradley, K.A., Grill, L.K., and Martchenko, M. (2015). Presentation of peptides from Bacillus anthracis protective antigen on Tobacco Mosaic Virus as an epitope targeted anthrax vaccine. Vaccine *33*, 6745–6751.

Meeusen, E.N.T., Walker, J., Peters, A., Pastoret, P.-P., and Jungersen, G. (2007). Current status of veterinary vaccines. Clin Microbiol Rev *20*, 489–510, table of contents.

Megeed, Z., Winters, R.M., and Yarmush, M.L. (2006). Modulation of single-chain antibody affinity with temperature-responsive elastin-like polypeptide linkers. Biomacromolecules *7*, 999–1004.

Meldal, M., and Schoffelen, S. (2016). Recent advances in covalent, site-specific protein immobilization. [version 1; referees: 3 approved]. F1000Res *5*.

Mero, A., Spolaore, B., Veronese, F.M., and Fontana, A. (2009). Transglutaminase-mediated PEGylation of proteins: direct identification of the sites of protein modification by mass spectrometry using a novel monodisperse PEG. Bioconjug Chem *20*, 384–389.

Michel, M.L., and Tiollais, P. (2010). Hepatitis B vaccines: protective efficacy and therapeutic potential. Pathol Biol *58*, 288–295.

Minhinnick, A., Satti, I., Harris, S., Wilkie, M., Sheehan, S., Stockdale, L., Manjaly Thomas, Z.-R., Lopez-Ramon, R., Poulton, I., Lawrie, A., et al. (2016). A first-in-human phase 1 trial to evaluate the safety and immunogenicity of the candidate tuberculosis vaccine MVA85A-IMX313, administered to BCG-vaccinated adults. Vaccine *34*, 1412–1421.

Modelska, A., Dietzschold, B., Sleysh, N., Fu, Z.F., Steplewski, K., Hooper, D.C., Koprowski, H., and Yusibov, V. (1998). Immunization against rabies with plant-derived antigen. Proc Natl Acad Sci U S A *95*, 2481–2485.

Moran, A.J., Hills, M., Gunton, J., and Nano, F.E. (2001). Heat-labile proteases in molecular biology applications. FEMS Microbiol Lett *197*, 59–63.

Morris, A.L., MacArthur, M.W., Hutchinson, E.G., and Thornton, J.M. (1992). Stereochemical quality of protein structure coordinates. Proteins *12*, 345–364.

Moyle, P.M. (2017). Biotechnology approaches to produce potent, self-adjuvanting antigen-adjuvant fusion protein subunit vaccines. Biotechnol Adv *35*, 375–389.

Moyle, P.M., and Toth, I. (2013). Modern subunit vaccines: development, components, and research opportunities. ChemMedChem *8*, 360–376.

Murhekar, M.V., Dutta, S., Kapoor, A.N., Bitragunta, S., Dodum, R., Ghosh, P., Swamy, K.K., Mukhopadhyay, K., Ningombam, S., Parmar, K., et al. (2013). Frequent exposure to

suboptimal temperatures in vaccine cold-chain system in India: results of temperature monitoring in 10 states. Bull World Health Organ *91*, 906–913.

Musil, M., Stourac, J., Bendl, J., Brezovsky, J., Prokop, Z., Zendulka, J., Martinek, T., Bednar, D., and Damborsky, J. (2017). FireProt: web server for automated design of thermostable proteins. Nucleic Acids Res.

Myers, J.S., and Jakoby, W.B. (1973). Effect of polyhydric alcohols on kinetic parameters of enzymes. Biochem Biophys Res Commun *51*, 631–636.

Naik, R.R., Kirkpatrick, S.M., and Stone, M.O. (2001). The thermostability of an α-helical coiled-coil protein and its potential use in sensor applications. Biosensors and Bioelectronics *16*, 1051–1057.

Negahdaripour, M., Golkar, N., Hajighahramani, N., Kianpour, S., Nezafat, N., and Ghasemi, Y. (2017). Harnessing self-assembled peptide nanoparticles in epitope vaccine design. Biotechnol Adv.

Newman, J., Peat, T.S., Richard, R., Kan, L., Swanson, P.E., Affholter, J.A., Holmes, I.H., Schindler, J.F., Unkefer, C.J., and Terwilliger, T.C. (1999). Haloalkane dehalogenases: structure of a Rhodococcus enzyme. Biochemistry *38*, 16105–16114.

Nguyen, G.K.T., Wang, S., Qiu, Y., Hemu, X., Lian, Y., and Tam, J.P. (2014). Butelase 1 is an Asx-specific ligase enabling peptide macrocyclization and synthesis. Nat Chem Biol *10*, 732–738.

Nguyen, G.K.T., Kam, A., Loo, S., Jansson, A.E., Pan, L.X., and Tam, J.P. (2015). Butelase 1: A versatile ligase for peptide and protein macrocyclization. J Am Chem Soc *137*, 15398–15401.

Nguyen, G.K.T., Qiu, Y., Cao, Y., Hemu, X., Liu, C.-F., and Tam, J.P. (2016a). Butelase-mediated cyclization and ligation of peptides and proteins. Nat Protoc *11*, 1977–1988.

Nguyen, G.K.T., Hemu, X., Quek, J.-P., and Tam, J.P. (2016b). Butelase-Mediated Macrocyclization of d-Amino-Acid-Containing Peptides. Angew Chem Int Ed Engl *55*, 12802–12806.

Nisonoff, A., and Rivers, M.M. (1961). Recombination of a mixture of univalent antibody fragments of different specificity. Arch Biochem Biophys *93*, 460–462.

Nixon, A.E., and Firestine, S.M. (2000). Rational and “irrational” design of proteins and their use in biotechnology. IUBMB Life *49*, 181–187.

Novikova, O., Jayachandran, P., Kelley, D.S., Morton, Z., Merwin, S., Topilina, N.I., and Belfort, M. (2016). Intein clustering suggests functional importance in different domains of life. Mol Biol Evol *33*, 783–799.

Oda, T., Lim, K., and Tomii, K. (2017). Simple adjustment of the sequence weight algorithm remarkably enhances PSI-BLAST performance. BMC Bioinformatics *18*, 288.

Odii, B.O., and Coussons, P. (2014). Biological functionalities of transglutaminase 2 and the possibility of its compensation by other members of the transglutaminase family. ScientificWorldJournal *2014*, 714561.

Oeemig, J.S., Aranko, A.S., Djupsjöbacka, J., Heinämäki, K., and Iwaï, H. (2009). Solution structure of DnaE intein from Nostoc punctiforme: structural basis for the design of a new split intein suitable for site-specific chemical modification. FEBS Lett *583*, 1451–1456.

Ogun, S.A., Dumon-Seignovert, L., Marchand, J.-B., Holder, A.A., and Hill, F. (2008). The oligomerization domain of C4-binding protein (C4bp) acts as an adjuvant, and the fusion protein comprised of the 19-kilodalton merozoite surface protein 1 fused with the murine C4bp domain protects mice against malaria. Infect Immun *76*, 3817–3823.

Ohana, R.F., Encell, L.P., Zhao, K., Simpson, D., Slater, M.R., Urh, M., and Wood, K.V. (2009). HaloTag7: a genetically engineered tag that enhances bacterial expression of soluble proteins and improves protein purification. Protein Expr Purif *68*, 110–120.

Orenstein, W.A., and Ahmed, R. (2017). Simply put: Vaccination saves lives. Proc Natl Acad Sci U S A *114*, 4031–4033.

Orlova, A., Magnusson, M., Eriksson, T.L.J., Nilsson, M., Larsson, B., Höidén-Guthenberg, I., Widström, C., Carlsson, J., Tolmachev, V., Ståhl, S., et al. (2006). Tumor imaging using a picomolar affinity HER2 binding affibody molecule. Cancer Res *66*, 4339–4348.

Oscherwitz, J., Yu, F., Jacobs, J.L., and Cease, K.B. (2013). Recombinant vaccine displaying the loop-neutralizing determinant from protective antigen completely protects rabbits from experimental inhalation anthrax. Clin Vaccine Immunol *20*, 341–349.

Ouattara, A., Barry, A.E., Dutta, S., Remarque, E.J., Beeson, J.G., and Plowe, C.V. (2015). Designing malaria vaccines to circumvent antigen variability. Vaccine *33*, 7506–7512.

Pace, C.N. (1986). [14]Determination and analysis of urea and guanidine hydrochloride denaturation curves. In Enzyme Structure Part L, (Elsevier), pp. 266–280.

Pande, J., Szewczyk, M.M., and Grover, A.K. (2010). Phage display: concept, innovations, applications and future. Biotechnol Adv *28*, 849–858.

Panowksi, S., Bhakta, S., Raab, H., Polakis, P., and Junutula, J.R. (2014). Site-specific antibody drug conjugates for cancer therapy. MAbs *6*, 34–45.

Papageorgiou, G.C., and Murata, N. (1995). The unusually strong stabilizing effects of glycine betaine on the structure and function of the oxygen-evolving Photosystem II complex. Photosyn Res *44*, 243–252.

Paterson, N.G., and Baker, E.N. (2011). Structure of the full-length major pilin from Streptococcus pneumoniae: implications for isopeptide bond formation in gram-positive bacterial pili. PLoS ONE *6*, e22095.

Peschke, T., Rabe, K.S., and Niemeyer, C.M. (2017). Orthogonal Surface Tags for Whole-Cell Biocatalysis. Angew Chem Int Ed Engl *56*, 2183–2186.

Pickart, C.M. (2001). Mechanisms underlying ubiquitination. Annu Rev Biochem *70*, 503–533.

Pierce, M.M., Raman, C.S., and Nall, B.T. (1999). Isothermal titration calorimetry of protein-protein interactions. Methods *19*, 213–221.

Pierres, A., Touchard, D., Benoliel, A.-M., and Bongrand, P. (2002). Dissecting streptavidin-biotin interaction with a laminar flow chamber. Biophys J *82*, 3214–3223.

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

Pokkuluri, P.R., Raffen, R., Dieckman, L., Boogaard, C., Stevens, F.J., and Schiffer, M. (2002). Increasing protein stability by polar surface residues: domain-wide consequences of interactions within a loop. Biophys J *82*, 391–398.

Pollard, T.D. (2010). A guide to simple and informative binding assays. Mol Biol Cell *21*, 4061–4067.

Powers, D.C., Smith, G.E., Anderson, E.L., Kennedy, D.J., Hackett, C.S., Wilkinson, B.E., Volvovitz, F., Belshe, R.B., and Treanor, J.J. (1995). Influenza A virus vaccines containing purified recombinant H3 hemagglutinin are well tolerated and induce protective immune responses in healthy adults. J Infect Dis *171*, 1595–1599.

Prajapati, S., Bhakuni, V., Babu, K.R., and Jain, S.K. (1998). Alkaline unfolding and salt-induced folding of bovine liver catalase at high pH. Eur J Biochem *255*, 178–184.

Pritz, S., Wolf, Y., Kraetke, O., Klose, J., Bienert, M., and Beyermann, M. (2007). Synthesis of biologically active peptide nucleic acid-peptide conjugates by sortase-mediated ligation. J Org Chem *72*, 3909–3912.

Pröschel, M., Detsch, R., Boccaccini, A.R., and Sonnewald, U. (2015). Engineering of metabolic pathways by artificial enzyme channels. Frontiers in Bioengineering and Biotechnology *3*, 168.

Pryor, K.D., and Leiting, B. (1997). High-level expression of soluble protein in Escherichia coli using a His6-tag and maltose-binding-protein double-affinity fusion system. Protein Expr Purif *10*, 309–319.

Qu, Q., Sawa, H., Suzuki, T., Semba, S., Henmi, C., Okada, Y., Tsuda, M., Tanaka, S., Atwood, W.J., and Nagashima, K. (2004). Nuclear entry mechanism of the human polyomavirus JC virus-like particle: role of importins and the nuclear pore complex. J Biol Chem *279*, 27735–27742.

Qu, Y., Bolen, C.L., and Bolen, D.W. (1998). Osmolyte-driven contraction of a random coil protein. Proc Natl Acad Sci U S A *95*, 9268–9273.

Rachel, N.M., and Pelletier, J.N. (2013). Biotechnological applications of transglutaminases. Biomolecules *3*, 870–888.

Raeeszadeh-Sarmazdeh, M., Parthasarathy, R., and Boder, E.T. (2015). Site-specific immobilization of protein layers on gold surfaces via orthogonal sortases. Colloids Surf B Biointerfaces *128*, 457–463.

Ramachandran, G.N., Ramakrishnan, C., and Sasisekharan, V. (1963). Stereochemistry of polypeptide chain configurations. J Mol Biol *7*, 95–99.

Ramirez, M.A., and Chen, Z. (2017). Split-Intein Triggered Protein Hydrogels. Methods Mol Biol *1495*, 161–171.

Rappuoli, R., and Aderem, A. (2011). A 2020 vision for vaccines against HIV, tuberculosis and malaria. Nature *473*, 463–469.

Raschle, T., Lin, C., Jungmann, R., Shih, W.M., and Wagner, G. (2015). Controlled Co-reconstitution of Multiple Membrane Proteins in Lipid Bilayer Nanodiscs Using DNA as a Scaffold. ACS Chem Biol *10*, 2448–2454.

Ray, S., Taylor, M., Burlingame, M., Tatulian, S.A., and Teter, K. (2011). Modulation of toxin stability by 4-phenylbutyric acid and negatively charged phospholipids. PLoS ONE *6*, e23692.

Reason, D.C., Ullal, A., Liberato, J., Sun, J., Keitel, W., and Zhou, J. (2008). Domain specificity of the human antibody response to Bacillus anthracis protective antigen. Vaccine *26*, 4041–4047.

Reddington, S.C., and Howarth, M. (2015). Secrets of a covalent interaction for biomaterials and biotechnology: SpyTag and SpyCatcher. Curr Opin Chem Biol *29*, 94–99.

Reed, S.G., Orr, M.T., and Fox, C.B. (2013). Key roles of adjuvants in modern vaccines. Nat Med *19*, 1597–1608.

Remy, I., Campbell-Valois, F.X., and Michnick, S.W. (2007). Detection of protein-protein interactions using a simple survival protein-fragment complementation assay based on the enzyme dihydrofolate reductase. Nat Protoc *2*, 2120–2125.

Reynolds, J.A., Herbert, S., Polet, H., and Steinhardt, J. (1967). The binding of divers detergent anions to bovine serum albumin*. Biochemistry *6*, 937–947.

Rich, K.M., and Perry, B.D. (2011). The economic and poverty impacts of animal diseases in developing countries: new roles, new demands for economics and epidemiology. Prev Vet Med *101*, 133–147.

Ringe, R.P., Ozorowski, G., Rantalainen, K., Struwe, W.B., Matthews, K., Torres, J.L., Yasmeen, A., Cottrell, C.A., Ketas, T.J., LaBranche, C.C., et al. (2017). Reducing V3 Antigenicity and Immunogenicity on Soluble, Native-Like HIV-1 Env SOSIP Trimers. J Virol *91*.

Roberts, D., Keeling, R., Tracka, M., van der Walle, C.F., Uddin, S., Warwicker, J., and Curtis, R. (2015). Specific ion and buffer effects on protein-protein interactions of a monoclonal antibody. Mol Pharm *12*, 179–193.

Rosen, C.B., and Francis, M.B. (2017). Targeting the N terminus for site-selective protein modification. Nat Chem Biol *13*, 697–705.

Rosenfeld, L., Heyne, M., Shifman, J.M., and Papo, N. (2016). Protein engineering by combined computational and in vitro evolution approaches. Trends Biochem Sci *41*, 421–433.

Rota, P.A., De, B.K., Shaw, M.W., Black, R.A., Gamble, W.C., and Kendal, A.P. (1990). Comparison of inactivated, live and recombinant DNA vaccines against influenza virus in a mouse model. Virus Res *16*, 83–93.

Rutgers, T., Gordon, D., Gathoye, A.M., Hollingdale, M., Hockmeyer, W., Rosenberg, M., and Wilde, M.D. (1988). Hepatitis B surface antigen as carrier matrix for the repetitive epitope of the circumsporozoite protein of plasmodium falciparum. Nat Biotechnol *6*, 1065–1070.

Sahdev, P., Ochyl, L.J., and Moon, J.J. (2014). Biomaterials for nanoparticle vaccine delivery systems. Pharm Res *31*, 2563–2582.

Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B., and Erlich, H.A. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science *239*, 487–491.

Saleh, L., and Perler, F.B. (2006). Protein splicing in cis and in trans. Chem Rec *6*, 183–193.

Salk, J., and Salk, D. (1977). Control of influenza and poliomyelitis with killed virus vaccines. Science *195*, 834–847.

Sawyer, W.H., and Puckridge, J. (1973). The dissociation of proteins by chaotropic salts. J Biol Chem *248*, 8429–8433.

Schellenberger, V., Wang, C.-W., Geething, N.C., Spink, B.J., Campbell, A., To, W., Scholle, M.D., Yin, Y., Yao, Y., Bogin, O., et al. (2009). A recombinant polypeptide extends the in vivo half-life of peptides and proteins in a tunable manner. Nat Biotechnol *27*, 1186–1190.

Schmitz, N., Beerli, R.R., Bauer, M., Jegerlehner, A., Dietmeier, K., Maudrich, M., Pumpens, P., Saudan, P., and Bachmann, M.F. (2012). Universal vaccine against influenza virus: linking TLR signaling to anti-viral protection. Eur J Immunol *42*, 863–869.

Schmohl, L., Bierlmeier, J., Gerth, F., Freund, C., and Schwarzer, D. (2017). Engineering sortase A by screening a second-generation library using phage display. J Pept Sci.

Schoene, C., Fierer, J.O., Bennett, S.P., and Howarth, M. (2014). SpyTag/SpyCatcher cyclization confers resilience to boiling on a mesophilic enzyme. Angew Chem Int Ed Engl *53*, 6101–6104.

Schwarte, A., Genz, M., Skalden, L., Nobili, A., Vickers, C., Melse, O., Kuipers, R., Joosten, H.-J., Stourac, J., Bendl, J., et al. (2017). NewProt - a protein engineering portal. Protein Eng Des Sel 1–7.

Schwesinger, F., Ros, R., Strunz, T., Anselmetti, D., Güntherodt, H.J., Honegger, A., Jermutus, L., Tiefenauer, L., and Pluckthun, A. (2000). Unbinding forces of single antibody-antigen complexes correlate with their thermal dissociation rates. Proc Natl Acad Sci U S A *97*, 9972–9977.

Shan, X., Gardner, K.H., Muhandiram, D.R., Rao, N.S., Arrowsmith, C.H., and Kay, L.E. (1996). Assignment of ^{15}N, $^{13}C^{\alpha}$, $^{13}C^{\beta}$, and HN Resonances in an ^{15}N, ^{13}C, ^{2}H Labeled 64 kDa Trp Repressor−Operator Complex Using Triple-Resonance NMR Spectroscopy and ^{2}H-Decoupling. J Am Chem Soc *118*, 6570–6579.

Shekhawat, S.S., and Ghosh, I. (2011). Split-protein systems: beyond binary protein-protein interactions. Curr Opin Chem Biol *15*, 789–797.

Shen, M., Rusling, J.F., and Dixit, C.K. (2017). Site-selective orientated immobilization of antibodies and conjugates for immunodiagnostics development. Methods *116*, 95–111.

Shi, R., Pan, Q., Guan, Y., Hua, Z., Huang, Y., Zhao, M., and Li, Y. (2007). Imidazole as a catalyst for in vitro refolding of enhanced green fluorescent protein. Arch Biochem Biophys *459*, 122–128.

Siegmund, V., Piater, B., Zakeri, B., Eichhorn, T., Fischer, F., Deutsch, C., Becker, S., Toleikis, L., Hock, B., Betz, U.A.K., et al. (2016). Spontaneous Isopeptide Bond Formation as a Powerful Tool for Engineering Site-Specific Antibody-Drug Conjugates. Sci Rep *6*, 39291.

Siegrist, C.-A. (2013). Vaccine immunology. In Vaccines, (Elsevier), pp. 14–32.

Sihag, R.K., and Deutscher, M.P. (1983). Perturbation of the aminoacyl-tRNA synthetase complex by salts and detergents. Importance of hydrophobic interactions and possible involvement of lipids. J Biol Chem *258*, 11846–11850.

Singh, Y., Khanna, H., Chopra, A.P., and Mehra, V. (2001). A dominant negative mutant of Bacillus anthracis protective antigen inhibits anthrax toxin action in vivo. J Biol Chem *276*, 22090–22094.

Sletten, E.M., and Bertozzi, C.R. (2009). Bioorthogonal chemistry: fishing for selectivity in a sea of functionality. Angew Chem Int Ed Engl *48*, 6974–6998.

Slomovic, S., and Collins, J.J. (2015). DNA sense-and-respond protein modules for mammalian cells. Nat Methods *12*, 1085–1090.

Sochaj, A.M., Świderska, K.W., and Otlewski, J. (2015). Current methods for the synthesis of homogeneous antibody-drug conjugates. Biotechnol Adv *33*, 775–784.

Sola-Penna, M., and Meyer-Fernandes, J.R. (1996). Trehalose protects yeast pyrophosphatase against structural and functional damage induced by guanidinium chloride. Z Naturforsch, C, J Biosci *51*, 160–164.

Sousa, M.M., Steen, K.W., Hagen, L., and Slupphaug, G. (2011). Antibody cross-linking and target elution protocols used for immunoprecipitation significantly modulate signal-to noise ratio in downstream 2D-PAGE analysis. Proteome Sci *9*, 45.

Southworth, M.W., Amaya, K., Evans, T.C., Xu, M.Q., and Perler, F.B. (1999). Purification of proteins fused to either the amino or carboxy terminus of the Mycobacterium xenopi gyrase A intein. BioTechniques *27*, 110–114, 116, 118.

Spolaore, B., Raboni, S., Ramos Molina, A., Satwekar, A., Damiano, N., and Fontana, A. (2012). Local unfolding is required for the site-specific protein modification by transglutaminase. Biochemistry *51*, 8679–8689.

Stayton, P.S., Freitag, S., Klumb, L.A., Chilkoti, A., Chu, V., Penzotti, J.E., To, R., Hyre, D., Le Trong, I., Lybrand, T.P., et al. (1999). Streptavidin-biotin binding energetics. Biomol Eng *16*, 39–44.

Steiner, K., and Schwab, H. (2012). Recent advances in rational approaches for enzyme engineering. Comput Struct Biotechnol J *2*, e201209010.

Stephanopoulos, N., and Francis, M.B. (2011). Choosing an effective protein bioconjugation strategy. Nat Chem Biol *7*, 876–884.

Stevens, A.J., Brown, Z.Z., Shah, N.H., Sekar, G., Cowburn, D., and Muir, T.W. (2016). Design of a Split Intein with Exceptional Protein Splicing Activity. J Am Chem Soc *138*, 2162–2165.

Stockley, P.G., and Mastico, R.A. (2000). Use of fusions to viral coat proteins as antigenic carriers for vaccine development. Meth Enzymol *326*, 551–569.

Stojanoski, V., Chow, D.-C., Hu, L., Sankaran, B., Gilbert, H.F., Prasad, B.V.V., and Palzkill, T. (2015). A triple mutant in the Ω-loop of TEM-1 β-lactamase changes the substrate profile via a large conformational change and an altered general base for catalysis. J Biol Chem *290*, 10382–10394.

Stranges, P.B., Palla, M., Kalachikov, S., Nivala, J., Dorwart, M., Trans, A., Kumar, S., Porel, M., Chien, M., Tao, C., et al. (2016). Design and characterization of a nanopore-coupled polymerase for single-molecule DNA sequencing by synthesis on an electrode array. Proc Natl Acad Sci U S A *113*, E6749–E6756.

Stratton, K.R., Howe, C.J., and Johnston, R.B. (1994). Adverse events associated with childhood vaccines other than pertussis and rubella. Summary of a report from the Institute of Medicine. JAMA *271*, 1602–1605.

Stromer, T., Fischer, E., Richter, K., Haslbeck, M., and Buchner, J. (2004). Analysis of the regulation of the molecular chaperone Hsp26 by temperature-induced dissociation: the N-terminal domail is important for oligomer assembly and the binding of unfolding proteins. J Biol Chem *279*, 11222–11228.

Strop, P., Dorywalska, M.G., Rajpal, A., Shelton, D., Liu, S., Pons, J., and Dushin, R. (2012). Engineered polypeptide conjugates and methods for making thereof using transglutaminase (United States of America).

Sugimura, Y., Hosono, M., Wada, F., Yoshimura, T., Maki, M., and Hitomi, K. (2006). Screening for the preferred substrate sequence of transglutaminase using a phage-displayed peptide library: identification of peptide substrates for TGASE 2 and Factor XIIIA. J Biol Chem *281*, 17699–17706.

Sugimura, Y., Yokoyama, K., Nio, N., Maki, M., and Hitomi, K. (2008). Identification of preferred substrate sequences of microbial transglutaminase from Streptomyces mobaraensis using a phage-displayed peptide library. Arch Biochem Biophys *477*, 379–383.

Sutherland, R.M., McKenzie, B.S., Corbett, A.J., Brady, J.L., and Lew, A.M. (2001). Overcoming the poor immunogenicity of a protein by DNA immunization as a fusion construct. Immunol Cell Biol *79*, 49–53.

Swee, L.K., Lourido, S., Bell, G.W., Ingram, J.R., and Ploegh, H.L. (2015). One-step enzymatic modification of the cell surface redirects cellular cytotoxicity and parasite tropism. ACS Chem Biol *10*, 460–465.

Swift, J.L., Heuff, R., and Cramb, D.T. (2006). A two-photon excitation fluorescence cross-correlation assay for a model ligand-receptor binding system using quantum dots. Biophys J *90*, 1396–1410.

Sydor, J.R., Mariano, M., Sideris, S., and Nock, S. (2002). Establishment of intein-mediated protein ligation under denaturing conditions: C-terminal labeling of a single-chain antibody for biochip screening. Bioconjug Chem *13*, 707–712.

Tanford, C. (1968). Protein denaturation. Adv Protein Chem *23*, 121–282.

Tang, S., Xuan, B., Ye, X., Huang, Z., and Qian, Z. (2016). A Modular Vaccine Development Platform Based on Sortase-Mediated Site-Specific Tagging of Antigens onto Virus-Like Particles. Sci Rep *6*, 25741.

Tatulian, S.A., Qin, S., Pande, A.H., and He, X. (2005). Positioning membrane proteins by novel protein engineering and biophysical approaches. J Mol Biol *351*, 939–947.

Teixeira, L.S.M., Feijen, J., van Blitterswijk, C.A., Dijkstra, P.J., and Karperien, M. (2012). Enzyme-catalyzed crosslinkable hydrogels: emerging strategies for tissue engineering. Biomaterials *33*, 1281–1290.

Tesfaw, A., Assefa, F., Tesfaw, A., and Assefa, F. (2014). Applications of Transglutaminase in Textile, Wool, and Leather Processing. International Journal of Textile Science *3*, 64–69.

Tettelin, H., Nelson, K.E., Paulsen, I.T., Eisen, J.A., Read, T.D., Peterson, S., Heidelberg, J., DeBoy, R.T., Haft, D.H., Dodson, R.J., et al. (2001). Complete genome sequence of a virulent isolate of Streptococcus pneumoniae. Science *293*, 498–506.

Thakkar, S., Nanaware-Kharade, N., Celikel, R., Peterson, E.C., and Varughese, K.I. (2014). Affinity improvement of a therapeutic antibody to methamphetamine and amphetamine through structure-based antibody engineering. Sci Rep *4*, 3673.

Thomassen, Y.E., van ’t Oever, A.G., van Oijen, M.G.C.T., Wijffels, R.H., van der Pol, L.A., and Bakker, W.A.M. (2013). Next generation inactivated polio vaccine manufacturing to support post polio-eradication biosafety goals. PLoS ONE *8*, e83374.

Thrane, S., Janitzek, C.M., Matondo, S., Resende, M., Gustavsson, T., de Jongh, W.A., Clemmensen, S., Roeffen, W., van de Vegte-Bolmer, M., van Gemert, G.J., et al. (2016). Bacterial superglue enables easy development of efficient virus-like particle based vaccines. J Nanobiotechnology *14*, 30.

Trevino, S.R., Schaefer, S., Scholtz, J.M., and Pace, C.N. (2007). Increasing protein conformational stability by optimizing beta-turn sequence. J Mol Biol *373*, 211–218.

Trevino, S.R., Scholtz, J.M., and Pace, C.N. (2008). Measuring and increasing protein solubility. J Pharm Sci *97*, 4155–4166.

Tsukamoto, M., Watanabe, H., Ooishi, A., and Honda, S. (2014). Engineered protein A ligands, derived from a histidine-scanning library, facilitate the affinity purification of IgG under mild acidic conditions. J Biol Eng *8*, 15.

Ulmer, J.B., Valley, U., and Rappuoli, R. (2006). Vaccine manufacturing: challenges and solutions. Nat Biotechnol *24*, 1377–1383.

Valdez, J., Cook, C.D., Ahrens, C.C., Wang, A.J., Brown, A., Kumar, M., Stockdale, L., Rothenberg, D., Renggli, K., Gordon, E., et al. (2017). On-demand dissolution of modular, synthetic extracellular matrix reveals local epithelial-stromal communication networks. Biomaterials *130*, 90–103.

Valenzuela, P., Medina, A., Rutter, W.J., Ammerer, G., and Hall, B.D. (1982). Synthesis and assembly of hepatitis B virus surface antigen particles in yeast. Nature *298*, 347–350.

Valenzuela, P., Coit, D., Medina-Selby, M.A., Kuo, C.H., Van Nest, G., Lyn Burke, R., Bull, P., Urdea, M.S., and Graves, P.V. (1985). Antigen Engineering in Yeast: Synthesis and Assembly of Hybrid Hepatitis B Surface Antigen-Herpes Simplex 1 gD Particles. Nat Biotechnol *3*, 323–326.

Veggiani, G., Nakamura, T., Brenner, M.D., Gayet, R.V., Yan, J., Robinson, C.V., and Howarth, M. (2016). Programmable polyproteams built using twin peptide superglues. Proc Natl Acad Sci U S A *113*, 1202–1207.

Vengadesan, K., Ma, X., Dwivedi, P., Ton-That, H., and Narayana, S.V.L. (2011). A model for group B Streptococcus pilus type 1: the structure of a 35-kDa C-terminal fragment of the major pilin GBS80. J Mol Biol *407*, 731–743.

Volkin, D.B., and Klibanov, A.M. (1992). Alterations in the structure of proteins that cause their irreversible inactivation. Dev Biol Stand *74*, 73–80; discussion 80.

Volkmann, G., and Iwaï, H. (2010). Protein trans-splicing and its use in structural biology: opportunities and limitations. Mol Biosyst *6*, 2110–2121.

Vuillard, L., Madern, D., Franzetti, B., and Rabilloud, T. (1995). Halophilic protein stabilization by the mild solubilizing agents nondetergent sulfobetaines. Anal Biochem *230*, 290–294.

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 *82*, 517–527.

Wallace, S., and Chin, J.W. (2014). Strain-promoted sydnone bicyclo-[6.1.0]-nonyne cycloaddition†Electronic supplementary information (ESI) available: Full experimental details, (1)H/(13)C NMR spectral data, protein synthesis and purification. See DOI: 10.1039/c3sc53332h. Chem Sci *5*, 1742–1744.

Walper, S.A., Turner, K.B., and Medintz, I.L. (2017). Bioorthogonal labeling of cellular proteins by enzymatic and related mechanisms. In Chemoselective and Bioorthogonal Ligation Reactions: Concepts and Applications, W.R. Algar, P.E. Dawson, and I.L. Medintz, eds. (Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA), pp. 165–230.

Wang, H., Vilela, M., Winkler, A., Tarnawski, M., Schlichting, I., Yumerefendi, H., Kuhlman, B., Liu, R., Danuser, G., and Hahn, K.M. (2016). LOVTRAP: an optogenetic system for photoinduced protein dissociation. Nat Methods *13*, 755–758.

Wang, H.H., Altun, B., Nwe, K., and Tsourkas, A. (2017). Proximity-Based Sortase-Mediated Ligation. Angew Chem Int Ed Engl *56*, 5349–5352.

Wang, L., Zhang, Z., Brock, A., and Schultz, P.G. (2003a). Addition of the keto functional group to the genetic code of Escherichia coli. Proc Natl Acad Sci U S A *100*, 56–61.

Wang, Y., Vander, H.P., and Xi, L. (2003b). Methods of using improved polymerases (USA).

Watanabe, H., Matsumaru, H., Ooishi, A., Feng, Y., Odahara, T., Suto, K., and Honda, S. (2009). Optimizing pH response of affinity between protein G and IgG Fc: how electrostatic modulations affect protein-protein interactions. J Biol Chem *284*, 12373–12383.

Watanabe, H., Matsumaru, H., Ooishi, A., and Honda, S. (2013). Structure-based histidine substitution for optimizing pH-sensitive Staphylococcus protein A. J Chromatogr B Analyt Technol Biomed Life Sci *929*, 155–160.

Waymouth, C. (1970). Osmolality of mammalian blood and of media for culture of mammalian cells. In Vitro *6*, 109–127.

Wehr, M.C., Laage, R., Bolz, U., Fischer, T.M., Grünewald, S., Scheek, S., Bach, A., Nave, K.-A., and Rossner, M.J. (2006). Monitoring regulated protein-protein interactions using split TEV. Nat Methods *3*, 985–993.

Wiebe, J.P., and Dinsdale, C.J. (1991). Inhibition of cell proliferation by glycerol. Life Sci *48*, 1511–1517.

Wikoff, W.R., Liljas, L., Duda, R.L., Tsuruta, H., Hendrix, R.W., and Johnson, J.E. (2000). Topologically linked protein rings in the bacteriophage HK97 capsid. Science *289*, 2129–2133.

Williamson, D.J., Fascione, M.A., Webb, M.E., and Turnbull, W.B. (2012). Efficient N-terminal labeling of proteins by use of sortase. Angew Chem Int Ed Engl *51*, 9377–9380.

Wong, T.S., Zhurina, D., and Schwaneberg, U. (2006). The diversity challenge in directed protein evolution. Comb Chem High Throughput Screen *9*, 271–288.

Wu, Z., and Guo, Z. (2012). Sortase-Mediated Transpeptidation for Site-Specific Modification of Peptides, Glycopeptides, and Proteins. J Carbohydr Chem *31*, 48–66.

Wu, H., Hu, Z., and Liu, X.Q. (1998). Protein trans-splicing by a split intein encoded in a split DnaE gene of Synechocystis sp. PCC6803. Proc Natl Acad Sci U S A *95*, 9226–9231.

Xie, J., and Schultz, P.G. (2005). An expanding genetic code. Methods *36*, 227–238.

Xie, Y., An, J., Yang, G., Wu, G., Zhang, Y., Cui, L., and Feng, Y. (2014). Enhanced enzyme kinetic stability by increasing rigidity within the active site. J Biol Chem *289*, 7994–8006.

Xue, J., Burz, D.S., and Shekhtman, A. (2012). Segmental labeling to study multidomain proteins. Adv Exp Med Biol *992*, 17–33.

Yan, X., Wang, L., Yang, X., Qiu, Y., Tian, X., Lv, Y., Tian, F., Song, G., and Wang, T. (2017). Fluoride induces apoptosis in H9c2 cardiomyocytes via the mitochondrial pathway. Chemosphere *182*, 159–165.

Yang, J., Meng, Q., and Liu, X.-Q. (2004). Intein harbouring large tandem repeats in replicative DNA helicase of Trichodesmium erythraeum. Mol Microbiol *51*, 1185–1192.

Yang, R., Wong, Y.H., Nguyen, G.K.T., Tam, J.P., Lescar, J., and Wu, B. (2017). Engineering a catalytically efficient recombinant protein ligase. J Am Chem Soc *139*, 5351–5358.

Yano, Y., and Irie, M. (1975). Renaturation of Yeast Inorganic Pyrophosphatase Denatured in Urea and Guanidine Hydrochloride. The Journal of Biochemistry.

Ye, W.N., Combes, D., and Monsan, P. (1988). Influence of additives on the thermostability of glucose oxidase. Enzyme Microb Technol *10*, 498–502.

Yi, L., Gebhard, M.C., Li, Q., Taft, J.M., Georgiou, G., and Iverson, B.L. (2013). Engineering of TEV protease variants by yeast ER sequestration screening (YESS) of combinatorial libraries. Proc Natl Acad Sci U S A *110*, 7229–7234.

Yonath, A., Podjarny, A., Honig, B., Sielecki, A., and Traub, W. (1977). Crystallographic studies of protein denaturation and renaturation. 2. Sodium dodecyl sulfate induced structural changes in triclinic lysozyme. Biochemistry *16*, 1418–1424.

Youde, J. (2013). Cattle scourge no more. The eradication of rinderpest and its lessons for global health campaigns. Politics Life Sci *32*, 43–57.

Yu, J., Wei, X., Zhang, L., Fang, X., Yang, T., Huang, F., and Liang, W. (2014). Poly(ethylene glycol)-mediated conformational alteration of α-chymotrypsin prevents inactivation of insulin by stabilizing active intermediates. Mol Pharm *11*, 3361–3370.

Yuen, C.M., Rodda, S.J., Vokes, S.A., McMahon, A.P., and Liu, D.R. (2006). Control of transcription factor activity and osteoblast differentiation in mammalian cells using an evolved small-molecule-dependent intein. J Am Chem Soc *128*, 8939–8946.

Yumura, K., Akiba, H., Nagatoishi, S., Kusano-Arai, O., Iwanari, H., Hamakubo, T., and Tsumoto, K. (2017). Use of SpyTag/SpyCatcher to construct bispecific antibodies that target two epitopes of a single antigen. J Biochem.

Zähner, D., Gandhi, A.R., Stuchlik, O., Reed, M., Pohl, J., and Stephens, D.S. (2011). Pilus backbone protein PitB of Streptococcus pneumoniae contains stabilizing intramolecular isopeptide bonds. Biochem Biophys Res Commun *409*, 526–531.

Zakeri, B., and Howarth, M. (2010). Spontaneous intermolecular amide bond formation between side chains for irreversible peptide targeting. J Am Chem Soc *132*, 4526–4527.

Zakeri, B., Fierer, J.O., Celik, E., Chittock, E.C., Schwarz-Linek, U., Moy, V.T., and Howarth, M. (2012). Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. Proc Natl Acad Sci U S A *109*, E690–E697.

Zaroog, M.S., Abdul Kadir, H., and Tayyab, S. (2013). Stabilizing effect of various polyols on the native and the denatured states of glucoamylase. ScientificWorldJournal *2013*, 570859.

Zayas, J.F. (1997). Solubility of Proteins. In Functionality of Proteins in Food, (Berlin, Heidelberg: Springer Berlin Heidelberg), pp. 6–75.

Zettler, J., Schütz, V., and Mootz, H.D. (2009). The naturally split Npu DnaE intein exhibits an extraordinarily high rate in the protein trans-splicing reaction. FEBS Lett *583*, 909–914.

Zhang, S., Ma, C., and Chalfie, M. (2004). Combinatorial marking of cells and organelles with reconstituted fluorescent proteins. Cell *119*, 137–144.

Zhao, X., Shaw, A.C., Wang, J., Chang, C.-C., Deng, J., and Su, J. (2010). A novel high-throughput screening method for microbial transglutaminases with high specificity toward Gln141 of human growth hormone. J Biomol Screen *15*, 206–212.

Zheng, Y., Wu, Q., Wang, C., Xu, M.-Q., and Liu, Y. (2012). Mutual synergistic protein folding in split intein. Biosci Rep *32*, 433–442.

Zinkernagel, R.M. (2003). On natural and artificial vaccinations. Annu Rev Immunol *21*, 515–546.

Zinkernagel, R.M., LaMarre, A., Ciurea, A., Hunziker, L., Ochsenbein, A.F., McCoy, K.D., Fehr, T., Bachmann, M.F., Kalinke, U., and Hengartner, H. (2001). Neutralizing antiviral antibody responses. Adv Immunol *79*, 1–53.

Zola, H., Flego, L., and Sheldon, A. (1992). Detection of cytokine receptors by high-sensitivity immunofluorescence/flow cytometry. Immunobiology *185*, 350–365.

Zou, Q., Habermann-Rottinghaus, S.M., and Murphy, K.P. (1998). Urea effects on protein stability: hydrogen bonding and the hydrophobic effect. Proteins *31*, 107–115.

Zuckerman, A.J. (1985). Subunit, recombinant and synthetic hepatitis B vaccines. Scand J Gastroenterol Suppl *117*, 27–38.

(2016). A little less conjugation, a little more accuracy. Nat Chem *8*, 91.

Appendix

Amino acid sequences of proteins employed in this thesis:
Note: Ligase and Tag mutants described in Chapters 3 and 4 are not listed separately.

+6RrgALigase+9 (PCB199)
MGSWSHHHHHHSSGGSGVNKNDKKPLRGAVFSLQKQHPDYPDIYGAIDQNGT
YQNVRTGEDGKLTFKNLSDGKYRLFENSEPAGYKPVQNKPIVAFQIVNGEVRDV
TSIVPQDIPATYEFT*

AP-SnoopLigase (PCB269)
MGSSHHHHHHSSGGLNDIFEAQKIEWHEGSGGSGSGVNKNDKKPLRGAVFSL
QKQHPDYPDIYGAIDQNGTYQNVRTGEDGKLTFKNLSDGKYRLFENSEPPGYKP
VQNKPIVAFQIVNGEVRDVTSIVPPGVPATYEFT*

HaloTag7-SnoopLigase (PCB391)
MGSSHHHHHHSSGEIGTGFPFDPHYVEVLGERMHYVDVGPRDGTPVLFLHGNP
TSSYVWRNIIPHVAPTHRCIAPDLIGMGKSDKPDLGYFFDDHVRFMDAFIEALGLE
EVVLVIHDWGSALGFHWAKRNPERVKGIAFMEFIRPIPTWDEWPEFARETFQAF
RTTDVGRKLIIDQNVFIEGTLPMGVVRPLTEVEMDHYREPFLNPVDREPLWRFPN
ELPIAGEPANIVALVEEYMDWLHQSPVPKLLFWGTPGVLIPPAEAARLAKSLPNC
KAVDIGPGLNLLQEDNPDLIGSEIARWLSTLEISGGSPANLKALEAQKQKEQRQA
AEELANAKKLKEQLEKGSHMVNKNDKKPLRGAVFSLQKQHPDYPDIYGAIDQNG
TYQNVRTGEDGKLTFKNLSDGKYRLFENSEPPGYKPVQNKPIVAFQIVNGEVRD
VTSIVPPGVPATYEFT*

HaloTag7SS-SnoopLigase (PCB402)
MGSSHHHHHHSSGEIGTGFPFDPHYVEVLGERMHYVDVGPRDGTPVLFLHGNP
TSSYVWRNIIPHVAPTHRSIAPDLIGMGKSDKPDLGYFFDDHVRFMDAFIEALGLE
EVVLVIHDWGSALGFHWAKRNPERVKGIAFMEFIRPIPTWDEWPEFARETFQAF
RTTDVGRKLIIDQNVFIEGTLPMGVVRPLTEVEMDHYREPFLNPVDREPLWRFPN
ELPIAGEPANIVALVEEYMDWLHQSPVPKLLFWGTPGVLIPPAEAARLAKSLPNS
KAVDIGPGLNLLQEDNPDLIGSEIARWLSTLEISGGSPANLKALEAQKQKEQRQA
AEELANAKKLKEQLEKGSHMVNKNDKKPLRGAVFSLQKQHPDYPDIYGAIDQNG
TYQNVRTGEDGKLTFKNLSDGKYRLFENSEPPGYKPVQNKPIVAFQIVNGEVRD
VTSIVPPGVPATYEFT*

His_6-IMX313-DogTag (PCB230)
MGSSHHHHHHSSGLVPRGSHMSGKKQGDADVCGEVAYIQSVVSDCHVPTAEL
RTLLEIRKLFLEIQKLKVELQGLSKEGSGSGEGSGDIPATYEFTDGKHYITNEPIPP
K*

IMX313-DogTag (PCB388)
MGSSKKQGDADVCGEVAYIQSVVSDCHVPTAELRTLLEIRKLFLEIQKLKVELQG
LSKEGSGSGEGSGDIPATYEFTDGKHYITNEPIPPK*

MBP-RrgALigase (PCB068)
MKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATG
DGPDIIFWAHDRFGGYAQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIA
VEALSLIYNKDLLPNPPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADG
GYAFKYENGKYDIKDVGVDNAGAKAGLTFLVDLIKNKHMNADTDYSIAEAAFNKG
ETAMTINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKE
LAKEFLENYLLTDEGLEAVNKDKPLGAVALKSYEEELAKDPRIAATMENAQKGEI
MPNIPQMSAFWYAVRTAVINAASGRQTVDEALKDAQTNSSSGSGGSGKPLRGA
VFSLQKQHPDYPDIYGAIDQNGTYQNVRTGEDGKLTFKNLSDGKYRLFENSEPA
GYKPVQNKPIVAFQIVNGEVRDVTSIVPQ*

RrgALigase (PCB040)
MGSSHHHHHHSSGLVPRGSHMKPLRGAVFSLQKQHPDYPDIYGAIDQNGTYQN
VRTGEDGKLTFKNLSDGKYRLFENSEPAGYKPVQNKPIVAFQIVNGEVRDVTSIV
PQ*

SnoopLigase (PCB259)
MGSWSHHHHHHSSGGSGVNKNDKKPLRGAVFSLQKQHPDYPDIYGAIDQNGT
YQNVRTGEDGKLTFKNLSDGKYRLFENSEPPGYKPVQNKPIVAFQIVNGEVRDV
TSIVPPGVPATYEFT*

SnoopTag peptide
GKLGDIEFIKVNKGY

SnoopTag-4E6-SpyTag (PCB049, obtained from Gianluca Veggiani)
MKKTAIAIAVALAGLATVAQASGKLGDIEFIKVNKGGGGSGGGGSEVQLVESGG
GSVQAGDSLRLSCAASGRTFGSIRVGWFRQTPGKEREFVAAINRNDGTTYYAD
SVKGRFTISRDNAKNTVYMQMASLKPEDTAVYYCAAGLQYNRSADRVPVGAVY
WGQGTQVTVSSAAAGGGGSGGGGSAHIVMVDAYKPTKGAAHHHHHH*

SnoopTag-AffiHER2 (PCB031)
MGSSHHHHHHSSGLVPRGSHMGKLGDIEFIKVNKGSGESGSGASMTGGQQMG
RDPGVDNKFNKEMRNAYWEIALLPNLNNQQKRAFIRSLYDDPSQSANLLAEAKK
LNDAQAPKGLE*

SnoopTag2-MBP (PCB043, obtained from Jin Huang)
MGSSHHHHHHSSGLVPRGSHMGDIPATYEFTNGKHYITNEPGSGESGKIEEGKL
VIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFW
AHDRFGGYAQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSLIYN
KDLLPNPPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYEN
GKYDIKDVGVDNAGAKAGLTFLVDLIKNKHMNADTDYSIAEAAFNKGETAMTING
PWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFLEN
YLLTDEGLEAVNKDKPLGAVALKSYEEELAKDPRIAATMENAQKGEIMPNIPQMS
AFWYAVRTAVINAASGRQTVDEALKDAQTNSSS*

SnoopTagJr-AffiHER2 (PCB325)
MGSSHHHHHHSSGGKLGSIEFIKVNKGSGESGSGASMTGGQQMGRDPGVDNK
FNKEMRNAYWEIALLPNLNNQQKRAFIRSLYDDPSQSANLLAEAKKLNDAQAPK
GLE*

SnoopTagJr-CIDR (PCB332)
MGHHHHHHENLYFQGGSKLGSIEFIKVNKGGSEPAPDVKTTEINVIVSGDEQGIIT
KKLQDFCMNPNNENGTNNQIWKCYYKDEKENKCKVETKSGNSTYKEKITSFDEF
FDFWVRKLLIDTIKWETELTYCINNTTNADCNNECNKNCVCFDKWVKQKEKEWK
NIMDLFTNKHDIPKKYYLNINDLFNSFFFQVIYKFNEGEAKWNKLKENLKKKIESSK
KNKGTKDSEAAIKVLFDHLKETATICKDNNTNEAC*

CyRPA-SnoopTagJr (obtained from Anne-Marie Andersson)
MEFQTQVLMSLLLCMSGAAADSRHVFIRTELSFIKNNVPCIRDMFFIYKRELYNIC
LDDLKGEEDETHIYVQKKVKDSWITLNDLFKETDLTGRPHIFAYVDVEEIIILLCED
EEFSNRKKDMTCHRFYSNDGKEYNNAEITISDYILKDKLLSSYVSLPLKIENREYF
LICGVSPYKFKDDNKKDDILCMASHDKGETWGTKIVIKYDNYKLGVQYFFLRPYIS
KNDLSFHFYVGDNINNVKNVNFIECTHEKDLEFVCSNRDFLKDNKVLQDVSTLND
EYIVSYGNDNNFAECYIFFNNENSILIKPEKYGNTAAGCYGGTFVKIDENRALFIYS
SSQGIYNIHTIYYANYEGSGGSGGSGKLGSIEFIKVNKGSGEPEA*

SUMO-DogTag (PCB218)
MGSWSHHHHHHGSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTT
PLRRLMEAFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHREQIGG
GDIPATYEFTDGKHYITNEPIPPK*

SUMO-SnoopTag2 (PCB067)
MGSSHHHHHHGSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPL
RRLMEAFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHREQIGGG
DIPATYEFTNGKHYITNEP*